Міністерство освіти і науки України
Прикарпатський національний університет
імені Василя Стефаника

М.М. Осипчук, Р.В. Шевчук

ТЕОРІЯ ЙМОВІРНОСТЕЙ
І МАТЕМАТИЧНА СТАТИСТИКА

Конспект лекцій
для студентів напряму підготовки
«Прикладна математика»
вищих навчальних закладів

Івано-Франківськ
2019

УДК 519.2
ББК 22.17
О 74

Друкується за рішенням Вченої ради факультету математики та інформатики Прикарпатського національного університету імені Василя Стефаника (протокол №7 від 28 березня 2019 р.)

Рецензенти:
Копитко Б.І., доктор фізико-математичних наук, професор (Ченстоховський політехнічний університет)
Нитребич З.М., доктор фізико-математичних наук, професор (Національний університет «Львівська політехніка»)

Осипчук М.М., Шевчук Р.В.

Теорія ймовірностей і математична статистика: Конспект лекцій. – Івано-Франківськ: Голіней, 2019. – 170с.

У вигляді конспекту лекцій викладено основи теорії ймовірностей та математичної статистики. Наведено приклади і задачі, які ілюструють основні поняття, а також пояснюють можливе застосування доведених теорем.

Конспект лекцій призначений для студентів напряму підготовки «Прикладна математика» вищих навчальних закладів.

УДК 519.2
ББК 22.17

Розділ 1

Випадкові події та ймовірності

1.1 Стохастичний експеримент та ймовірнісний простір

1.1.1 Стохастичний експеримент, елементарні події, події

Вихідним поняттям теорії ймовірностей є поняття стохастичного експерименту, випадкової події та ймовірності випадкової події. Стохастичним (випадковим) експериментом будемо називати експеримент (деяку дію), результат якого неможливо точно передбачити наперед. При цьому будемо розглядати тільки такі експерименти, які можна повторювати при незмінних умовах будь-яку кількість раз. Будь-який результат стохастичного експерименту, який може бути зафіксованим (спостерігатися), будемо називати випадковою подією[1](подією). Серед випадкових подій виділимо ті з них, які не можуть відбуватися одночасно і не можуть бути розкладені на простіші випадкові події. Такі події будемо називати елементарними подіями.

Таким чином, з кожним стохастичним експериментом можна пов'язати деяку множину Ω елементарних подій (простір елементарних подій). Кожну подію, яка може настати в результаті цього експерименту, можна тому розглядати як деяку підмножину простору елементарних подій.

Елементарні події будемо позначати буквами ω або ω_i ($\omega \in \Omega$ або $\omega_i \in \Omega$), випадкові події будемо позначати великими буквами латинського алфавіту: A, B, і т. д. Виняток становлять дві події: Ω — вірогідна подія та $\emptyset$ — неможлива подія.

Будемо говорити, що подія A настала (відбулася), якщо настала яка-небуть елементарна подія $\omega \in A$, при цьому вважатимемо, що елементарна подія ω сприяє події A.

Приклади.

1. Експеримент полягає в трикратному підкиданні монети і фіксації сторони монети, якою вона випаде кожного разу. В такому експерименті є всього 8 найпростіших результатів (елементарних подій). Тобто

$$\Omega = \{\text{ггг, ггц, гцг, цгг, гцц, цгц, ццг, ццц}\}.$$

Нехай подія A полягає в тому, що випаде принаймні один герб, а подія B — в тому, що другим випаде герб. Тому

$$A = \{\text{ггг, ггц, гцг, цгг, гцц, цгц, ццг}\},$$

$$B = \{\text{ггг, ггц, цгг, цгц}\}.$$

[1]Пізніше дамо більш точне означення випадкової події.

2. Експеримент полягає в киданні точки в деякий квадрат. Елементарною подією слід в цьому випадку вважати попадання в деяку фіксовану точку квадрата. Тому простір елементарних подій можна задати як множину точок квадрата.

Нехай подія A полягає в попаданні точки у вписаний в даний квадрат круг. Отже, A задається кругом, вписаним в даний квадрат.

Зауваження. Потрібно відзначити, що в першому прикладі простір елементарних подій складається із скінченної кількості елементів, а в другому — із нескінченної (незліченної).

1.1.2 Дії над подіями

Розглянемо тепер можливі дії над подіями та відношення, в яких ці події можуть бути. Дамо такі означення.

Означення. Будемо говорити, що подія A є частковим випадком події B (або B є наслідком A), якщо множина елементарних подій, що сприяють події A є підмножиною елементарних подій, що сприяють події B. Позначається це так: $A \subset B$ або $B \supset A$

Якщо $A \subset B$ і $B \subset A$, то кажуть, що події A і B співпадають, тобто $A = B$.

Означення. Сумою (або об'єднанням) двох подій A і B називається подія, яка полягає в тому, що відбудеться або подія A, або подія B. Позначається сума подій так: $A + B$ або $A \cup B$.

Тому сумі двох подій A і B сприяють ті і тільки ті елементарні події, які сприяють хоча б одній з подій A чи B.

Означення. Добутком (або перерізом) двох подій A і

B називається подія, яка полягає в тому, що відбудеться і подія A, і подія B. Позначається добуток подій так: $A \cdot B$ або $A \cap B$.

Отже, добутку двох подій A і B сприяють ті і тільки ті елементарні події, які сприяють обидвом подіям A і B.

Означення. Протилежною подією до події A називається подія, яка полягає в тому, що подія A не відбудеться. Позначається протилежна подія так: $\overline{A}$.

Тобто, події $\overline{A}$ сприяють ті і тільки ті елементарні події, які не сприяють події A.

Крім розглянутих дій над подіями можна виділити ще й такі:

1. різниця подій $A \setminus B = A \cdot \overline{B}$; тому $\overline{A} = \Omega \setminus A$;

2. симетрична різниця подій $A \nabla B = (A \setminus B) + (B \setminus A)$.

Наведемо найважливіші властивості дій над подіями:

1. $A + B = B + A$, $A \cdot B = B \cdot A$ (комутативність додавання та множення);

2. $A + (B + C) = (A + B) + C$, $A \cdot (B \cdot C) = (A \cdot B) \cdot C$ (асоціативність додавання та множення);

3. $(A+B) \cdot C = A \cdot C + B \cdot C$ (дистрибутивність множення відносно додавання);

4. $A \cdot B + C = (A + C) \cdot (B + C)$ (дистрибутивність додавання відносно множення);

5. $\overline{A + B} = \overline{A} \cdot \overline{B}$, $\overline{A \cdot B} = \overline{A} + \overline{B}$ (закони де Моргана);

6. $A + \overline{A} = \Omega$, $A \cdot \overline{A} = \emptyset$, $A = \overline{\overline{A}}$.

Вправа. Довести властивості 1-6.

Означення. Події A і B називаються несумісними, якщо $A \cdot B = \emptyset$. В іншому випадку ці події сумісні.

Очевидно, що події A і $\overline{A}$ несумісні.

Нехай $\mathcal{A}$ — деяка множина подій (система підмножин Ω).

Означення. Множина подій $\mathcal{A}$ називається алгеброю, якщо:

1. $\Omega \in \mathcal{A}$;
2. з того, що A, $B \in \mathcal{A}$, випливає, що $A + B \in \mathcal{A}$;
3. з того, що $A \in \mathcal{A}$, випливає, що $\overline{A} \in \mathcal{A}$.

З цього означення легко слідує, що $\emptyset \in \mathcal{A}$, бо $\emptyset = \overline{\Omega}$, і якщо A, $B \in \mathcal{A}$, то $A \cdot B \in \mathcal{A}$, бо $A \cdot B = \overline{\overline{A} + \overline{B}}$.

Означення. Множина подій $\mathcal{F}$ називається σ-алгеброю (сигма-алгеброю), якщо вона є алгеброю і для будь-якої послідовності подій $\{A_n,\ n \geq 1\}$ з $\mathcal{F}$ їх сума $\bigcup\limits_{n=1}^{\infty} A_n \in \mathcal{F}$.

Очевидно також, що [2]

$$\bigcap_{n=1}^{\infty} A_n = \overline{\bigcup_{n=1}^{\infty} \overline{A_n}} \in \mathcal{F},$$

якщо $\{A_n,\ n \geq 1\} \subset \mathcal{F}$.

Надалі для кожного стохастичного експеримента будемо визначати разом з простором елементарних подій Ω також і σ-алгебру подій $\mathcal{F}$, і тільки елементи σ-алгебри $\mathcal{F}$ будемо називати подіями.

---

[2] Значками $\bigcup\limits_{n=1}^{\infty}$ та $\bigcap\limits_{n=1}^{\infty}$ позначені, відповідно, сума і добуток подій, записаних після них.

1.1.3 Ймовірність

В теорії ймовірностей та її застосуваннях майже завжди виникає потреба оцінювати події (що відносяться до одного чи різних стохастичних експериментів) за ступенем їх «достовірності» як, наприклад, подія Ω відбувається завжди (вірогідна подія), а подія $\emptyset$ не відбувається ніколи (неможлива подія). Для цього вводиться поняття ймовірності події.

Означення. Ймовірністю (ймовірнісною мірою) називається числова функція $\mathrm{P}(A)$, яка визначена на σ-алгебрі подій $\mathcal{F}$ та має такі властивості:

1. $\mathrm{P}(A) \geq 0$ для всіх $A \in \mathcal{F}$;

2. $\mathrm{P}(\Omega) = 1$;

3. якщо події A_1, A_2, ... попарно несумісні (тобто $A_i \cdot A_j = \emptyset$ при всіх $i \neq j$), то

$$\mathrm{P}\left(\bigcup_{n=1}^{\infty} A_n\right) = \sum_{n=1}^{\infty} \mathrm{P}(A_n).$$

Означення. Трійку елементів $(\Omega, \mathcal{F}, \mathrm{P})$, де Ω — простір елементарних подій, $\mathcal{F}$ — σ-алгебра подій (підмножин Ω), P — ймовірнісна міра на $\mathcal{F}$, будемо називати ймовірнісним простором.

Зауваження. Властивості з означень σ-алгебри $\mathcal{F}$, та ймовірнісної міри P ще називають аксіомами теорії ймовірностей. Аксіоматичний підхід до побудови теорії ймовірностей запропонував в 1929 р. А.М.Колмогоров.

1.2 Властивості та приклади ймовірностей

1.2.1 Властивості ймовірності

Розглянемо деякі найпростіші властивості ймовірності, що випливають з означення.

Теорема (про ймовірність протилежної події). Для будь-якої події A

$$\mathrm{P}(\overline{A}) = 1 - \mathrm{P}(A).$$

◀ Розглянемо події A і $\overline{A}$. Оскільки $A+\overline{A} = \Omega$ і $A\cdot\overline{A} = \emptyset$, то $1 = \mathrm{P}(\Omega) = \mathrm{P}(A + \overline{A}) = \mathrm{P}(A) + \mathrm{P}(\overline{A})$. Звідси $\mathrm{P}(\overline{A}) = 1 - \mathrm{P}(A)$. ▶

Наслідок. $\mathrm{P}(\emptyset) = 0$.

◀ Оскільки $\emptyset = \overline{\Omega}$, то за теоремою про ймовірність протилежної події $\mathrm{P}(\emptyset) = 1 - \mathrm{P}(\Omega) = 1 - 1 = 0$. ▶

Теорема (про монотонність ймовірності). Якщо $A \subset B$, то $\mathrm{P}(A) \le \mathrm{P}(B)$

◀ Подамо подію B у вигляді $B = A + \overline{A} \cdot B$. Оскільки $A \cdot (\overline{A} \cdot B) = \emptyset$, то $\mathrm{P}(B) = \mathrm{P}(A) + \mathrm{P}(\overline{A} \cdot B) \ge \mathrm{P}(A)$, бо $\mathrm{P}(\overline{A} \cdot B) \ge 0$. ▶

Теорема (про додавання ймовірностей). Якщо A і B довільні події, то $\mathrm{P}(A + B) = \mathrm{P}(A) + \mathrm{P}(B) - \mathrm{P}(A \cdot B)$.

◀ Оскільки $A + B = A + \overline{A} \cdot B$ і $B = \overline{A} \cdot B + A \cdot B$ та $A \cdot (\overline{A} \cdot B) = \emptyset$ і $(\overline{A} \cdot B) \cdot (A \cdot B) = \emptyset$, то

$$\mathrm{P}(A + B) = \mathrm{P}(A) + \mathrm{P}(\overline{A} \cdot B) \text{ і}$$

$$\mathrm{P}(B) = \mathrm{P}(\overline{A} \cdot B) + \mathrm{P}(A \cdot B).$$

Звідси $\mathrm{P}(\overline{A} \cdot B) = \mathrm{P}(B) - \mathrm{P}(A \cdot B)$ і

$$\mathrm{P}(A + B) = \mathrm{P}(A) + \mathrm{P}(B) - \mathrm{P}(A \cdot B). \quad ▶$$

Теорема. Якщо $A \subset B$, то $\mathrm{P}(B \setminus A) = \mathrm{P}(B) - \mathrm{P}(A)$.

◀ Подамо подію B у вигляді $B = B \setminus A + A$. Очевидно, що $(B \setminus A) \cdot A = \emptyset$ і тому

$$\mathrm{P}(B) = \mathrm{P}(B \setminus A) + \mathrm{P}(A),$$

звідси і випливає твердження теореми. ▶

Теорема (про неперервність ймовірності). Нехай дано послідовність подій $A_1, A_2, \ldots$ таку, що $A_1 \supset A_2 \supset \ldots$ і $A = \bigcap\limits_{n=1}^{\infty} A_n$. Тоді $\mathrm{P}(A) = \lim\limits_{n\to\infty} \mathrm{P}(A_n)$.

◀ Розкладемо кожну подію A_n на суму попарно несумісних подій $A_n = \bigcup\limits_{k=n}^{\infty} A_k \cdot \overline{A_{k+1}} \cup A$. Тому $\mathrm{P}(A_n) = \sum\limits_{k=n}^{\infty} \mathrm{P}(A_k \cdot \overline{A_{k+1}} + \mathrm{P}(A)$, звідки

$$\mathrm{P}(A_n) - \mathrm{P}(A) = \sum_{k=n}^{\infty} \mathrm{P}(A_k \cdot \overline{A_{k+1})}. \tag{1.1}$$

При $n = 1$ з цієї рівності випливає, що ряд $\sum\limits_{k=1}^{\infty} \mathrm{P}(A_k \cdot \overline{A_{k+1})}$ збігається. Оскільки за формулою (1.1) $\mathrm{P}(A_n) - \mathrm{P}(A)$ є залишком цього ряду, то

$$\lim_{n\to\infty} \mathrm{P}(A_n) - \mathrm{P}(A) = 0. \quad ▶$$

Перейдемо тепер до розгляду деяких конкретних прикладів ймовірнісних просторів.

1.2.2 Класичне означення ймовірності

Припустимо, що деякому стохастичному експериментові відповідає ймовірнісний простір $(\Omega,\ \mathcal{F},\ \mathrm{P})$. Нехай простір елементарних подій Ω є скінченною множиною і $n =$

$|\Omega|$. Нехай $\mathcal{F}$ — σ-алгебра всіх підмножин Ω. Крім того, нехай ймовірнісна міра P така, що всі елементарні події мають однакові ймовірності (рівноможливі). Тому $\mathrm{P}(\omega_i) = \frac{1}{n}$.

Нехай далі події A сприяють $m = |A|$ елементарних подій тобто $A = \bigcup\limits_{k=1}^{m} \omega_{i_k}$.

Оскільки $\mathrm{P}(\bigcup\limits_{k=1}^{m} \omega_{i_k}) = \sum\limits_{k=1}^{m} \mathrm{P}(\omega_{i_k}) = \frac{m}{n}$, то

$$\mathrm{P}(A) = \frac{m}{n}. \tag{1.2}$$

Рівність (1.2) і є класичним означенням ймовірності.

Приклад. Підкидається гральний кубик. Знайти ймовірність того, що число очок, яке випаде, кратне трьом.

Найпростіші результати описаного експерименту (елементарні події) $\omega_i = \{\text{випало число } i\}$, $i = \overline{1,6}$.

Очевидно (якщо кубик правильний — виготовленний з однорідного матеріалу), що ймовірнісну міру потрібно визначити так, щоб $\mathrm{P}(\omega_i) = \frac{1}{6}$.

Описаній події (позначимо її A) сприяють дві елементарні події $A = \{\omega_3, \omega_6\}$.

Тому $|A| = 2$ і $\mathrm{P}(A) = \frac{2}{6} = \frac{1}{3}$.

Зауважимо, що функція задана на $\mathcal{F}$ рівністю (1.2) задовольняє всі вимоги, які повинні виконуватись для ймовірнісної міри. Дійсно

1. $\mathrm{P}(A) = \frac{m}{n} \geq 0$;

2. $\mathrm{P}(\Omega) = \frac{n}{n} = 1$;

3. якщо $A \cdot B = \emptyset$ ($|A| = m$ і $|B| = k$), то

$$\mathrm{P}(A + B) = \frac{m + k}{n} = \frac{m}{n} + \frac{k}{n} = \mathrm{P}(A) + \mathrm{P}(B).$$

Зліченної адитивності від функції P вимагати не потрібно, бо $\mathcal{F}$ складається із скінченної кількості (2^n) підмножин.

1.2.3 Геометричне означення ймовірності

Розглянемо ситуацію коли стохастичному експерименту відповідає ймовірнісний простір $(\Omega,\ \mathcal{F},\ \mathrm{P})$ в якому простір елементарних подій Ω є нескінченним, навіть незліченним.

Нехай Ω є деякою множиною евклідового простору R^n $(n = 1, 2, 3)$, σ-алгебра подій $\mathcal{F}$ — найменша σ-алгебра, що містить всі відкриті підмножини Ω. Як і в попередньому випадку, будемо вважати, що всі елементарні події рівноможливі. Проте ніякої ненульової ймовірності елементарним подіям приписати неможна, бо їх нескінченне число. Відштовхуючись від рівноможливості елементарних подій природно визначити ймовірність кожної події пропорційну її n-вимирній мірі (довжині при $n = 1$, площі при $n = 2$, об'єму при $n = 3$) $mes(A)$, як множини з R^n.

Тобто $\mathrm{P}(A) = C \cdot mes(A)$, а оскільки $1 = \mathrm{P}(\Omega) = C \times mes(\Omega)$, то $C = \frac{1}{mes(\Omega)}$. Тому

$$\mathrm{P}(A) = \frac{mes(A)}{mes(\Omega)} \tag{1.3}$$

Рівність (1.3) і є геометричним означенням ймовірності.

Очевидно, що всі характеристичні властивості ймовірнісної міри виконуються, бо властивості 1) і 3) має сама міра mes, а $\mathrm{P}(\Omega) = \frac{mes(\Omega)}{mes(\Omega)} = 1$.

Приклад (задача про зустріч). Два студенти призначили зустріч у певному місці між третьою та четвертою годинами дня. Той, хто прийде перший, чекає іншого впродовж 15 хвилин, після чого йде з місця зустрічі. Знайти ймовірність того, що зустріч відбудеться.

Позначимо через x час (у годинах) приходу на місце зустрічі першого студента, а через y — другого. Очевидно, що година, під час якої має відбутися зустріч, несуттєва, тобто студенти мають зустрітися протягом однієї години. Тоді для x і y виконуються умови

$$0 \leq x \leq 1, \quad 0 \leq y \leq 1,$$

і

$$\Omega = \{(x; y) \in R^2 : 0 \leq x \leq 1,\ 0 \leq y \leq 1\}.$$

Геометричним зображенням простору Ω є одиничний квадрат $OMBN$ (рис. 1.1).

Нехай подія A полягає в тому, що зустріч відбулася. Це можливо лише тоді, коли різниця між часом приходу на місце зустрічі першого та другого студентів не більша за 15 хв, або $\frac{1}{4}$ год, тобто

$$|x - y| \leq \frac{1}{4}. \tag{1.4}$$

Звідси отримуємо систему нерівностей

$$y \leq x + \frac{1}{4}, \quad y \geq x - \frac{1}{4}.$$

Множина точок, координати яких задовольняють нерівності (1.4), утворює фігуру $ABCDOE$ (рис. 1.1).

Оскільки

$$S_{ABCDOE} = S_{MBNO} - 2S_{EMA},$$

причому

$$S_{MBNO} = 1,\ S_{EMA} = \frac{1}{2} \cdot \frac{3}{4} \cdot \frac{3}{4} = \frac{9}{32},$$

то

$$S_{ABCDOE} = 1 - \frac{9}{16} = \frac{7}{16}.$$

Отже, ймовірність того, що зустріч відбудеться,

$$\mathrm{P}(A) = \frac{7}{16}$$

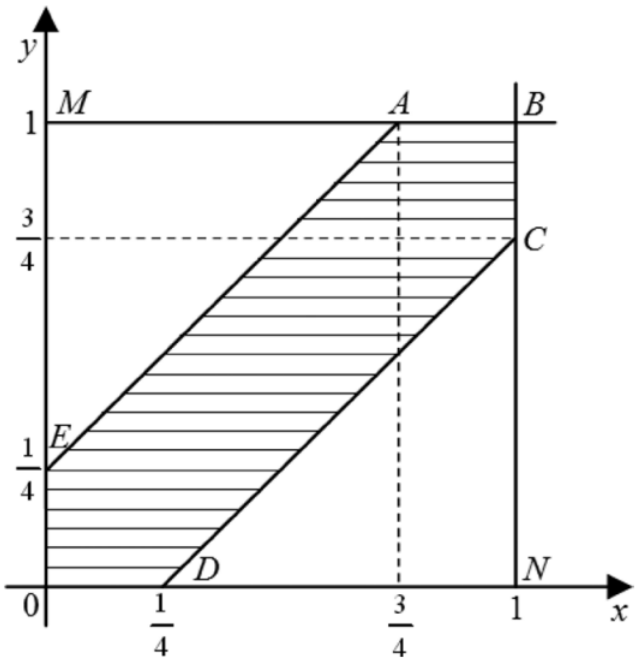

Рис. 1.1

1.3 Умовні ймовірності. Незалежність випадкових подій

1.3.1 Умовна ймовірність

Нехай, деякому стохастичному експерименту відповідає ймовірнісний простір $(\Omega,\ \mathcal{F},\ \mathrm{P})$. Розглянемо дві події A і B, причому $\mathrm{P}(B) \neq 0$. Задамось питанням, чи впливає на ймовірність («достовірність») події A інформація про те, що подія B відбулась, та як знайти цю ймовірність. Очевидно, що відповідь на першу частину питання повинна бути, взагалі кажучі, позитивною. Абсолютно зрозумілим є те, що якщо ми знаємо, що подія B відбулася, то простір елементарних подій необхідно звузити до множини, яка описує подію B. Ясно також, що замість σ-алгебри подій $\mathcal{F}$ слід взяти σ-алгебру $\mathcal{F}_B$, що є слідом $\mathcal{F}$ на події B. Тобто $\mathcal{F}_B = \{C \cap B; C \in \mathcal{F}\}$.

Легко довести, що множина подій $\mathcal{F}_B$ утворює σ-алгебру підмножин B.

Використовуючи ідеї побудови ймовірнісних мір, які ми розглядали вище, та поняття ймовірності як деякої міри («ваги») події, дамо таке означення умовної ймовірності:

Означення. Умовною ймовірністю події A при умові, що відбулася подія B $(\mathrm{P}(B) \neq 0)$, називається число $\mathrm{P}(A/B) = \frac{\mathrm{P}(A \cdot B)}{\mathrm{P}(B)}$.

Умовна ймовірність $\mathrm{P}(\cdot/B)$ має такі прості властивості:

1. $\mathrm{P}(A/B) \geq 0$ для всіх $A \in \mathcal{F}$;

2. $\mathrm{P}(\Omega/B) = \mathrm{P}(B/B) = 1$;

3. якщо події A_1, $A_2, \ldots$ попарно несумісні, то

$$\mathrm{P}\left(\bigcup_{n=1}^{\infty} A_n/B\right) = \frac{\mathrm{P}\left(\bigcup\limits_{n=1}^{\infty} A_n \cdot B\right)}{\mathrm{P}(B)} = \frac{\sum\limits_{n=1}^{\infty} \mathrm{P}(A_n \cdot B)}{\mathrm{P}(B)} =$$

$$= \sum_{n=1}^{\infty} \mathrm{P}(A_n/B),$$

бо події $\{A_n \cdot B : n \geq 1\}$ попарно несумісні.

Отже, $\mathrm{P}(\cdot/B)$ є ймовірнісною мірою як на $\mathcal{F}$ так і на $\mathcal{F}_B$. Оскільки для всіх $A \in \mathcal{F}$ таких, що $A \cdot B = \emptyset$, $\mathrm{P}(A/B) = 0$, то якщо відомо, що подія B відбулася, можна обмежитись умовним стохастичним експериментом, якому відповідає ймовірнісний простір $(B,\ \mathcal{F}_B,\ \mathrm{P}(\cdot/B))$.

Приклад. Двічі витягують без повернення по одній кульці з урни, яка містить n білих і m чорних кульок. Розглянемо дві події: A=\{друга кулька чорна\}, B=\{перша кулька біла\}.

Якщо подія B відбулась то в урні залишилось $n-1$ білих і m чорних кульок. Тому ймовірність події A при умові, що відбулася подія B дорівнює $\mathrm{P}(A/B) = \frac{m}{n-1+m}$. З іншого боку оскільки $\mathrm{P}(A \cdot B) = \frac{n \cdot m}{A^2_{n+m}} = \frac{n \cdot m}{(n+m)\cdot(n+m-1)}$, $\mathrm{P}(B) = \frac{n}{n+m}$, то $\mathrm{P}(A/B) = \frac{\mathrm{P}(A \cdot B)}{\mathrm{P}(B)} = \frac{m}{n+m-1}$.

Одержали два однакові результати, що є підтвердженням сказаного вище.

1.3.2 Теорема множення ймовірностей

Розглянемо деякі прості, але важливі властивості умовної ймовірності. Безпосередньо з означення випливає, що якщо $\mathrm{P}(B) \neq 0$, то

$$\mathrm{P}(A \cdot B) = \mathrm{P}(A/B) \cdot \mathrm{P}(B). \tag{1.5}$$

Очевидно, що якщо $P(A) \neq 0$, $P(B) \neq 0$, то

$$P(A \cdot B) = P(A/B) \cdot P(B) = P(B/A) \cdot P(A).$$

Формулу (1.5) можна легко розповсюдити на випадок будь-якої скінченної кількісті подій.

Теорема (множення ймовірностей). *Якщо події A_1, $A_2, \ldots, A_n$ такі, що $P(A_1 \cdot A_2 \ldots A_{n-1}) \neq 0$, то*

$$P(A_1 \cdot A_2 \ldots A_n) = \\ = P(A_1) \cdot P(A_2/A_1) \ldots P(A_n/A_1 \cdot A_2 \ldots A_{n-1}). \quad (1.6)$$

◀ Очевидно, що умовні ймовірності в формулі (1.6) визначені, бо $P(A_1 \cdot A_2 \ldots A_k) \neq 0$ при всіх
$k = 1, 2, \ldots, n-1$.

Дійсно оскільки

$$A_1 \supset A_1 \cdot A_2 \supset \cdots \supset A_1 \cdot A_2 \ldots A_{n-1}$$

і ймовірнісна міра має властивість монотонності, то

$$P(A_1 \cdot A_2 \ldots A_k) \geq P(A_1 \cdot A_2 \ldots A_{n-1}) > 0$$

Формула (1.6) при $n = 2$ правильна (див. 1.5).

Нехай (1.6) правильна при деякому $n \in N$, тобто

$$P(A_1 \cdot A_2 \ldots A_n) = \\ = P(A_1) \cdot P(A_2/A_1) \ldots P(A_n/A_1 \cdot A_2 \ldots A_{n-1}).$$

Розглянемо

$$P(A_1 \cdot A_2 \ldots A_{n+1}) = \\ = P(A_1 \cdot A_2 \ldots A_n) \cdot P(A_{n+1}/A_1 \cdot A_2 \ldots A_n) = \\ = P(A_1) \cdot P(A_2/A_1) \ldots P(A_n/A_1 \cdot A_2 \ldots A_{n-1}) \times \\ \times P(A_{n+1}/A_1 \cdot A_2 \ldots A_n).$$

Тому за принципом математичної індукції рівність (1.6) має місце при кожному натуральному n. ▶

1.3.3 Незалежні випадкові події

Незалежними, природно, назвати випадкові події, ймовірності кожної з яких не залежать від того, відбулась чи ні інша подія. Як ми побачимо дальше, для незалежних подій умовні та безумовні ймовірності співпадають. Тому, виходячи з теореми множення, дамо таке означення незалежності двох подій:

Означення. Випадкові події A і B називаються незалежними, якщо

$$\mathrm{P}(A \cdot B) = \mathrm{P}(A) \cdot \mathrm{P}(B). \tag{1.7}$$

Звідси легко одержимо сказане вище.

Теорема. *Якщо випадкові події A і B незалежні і $\mathrm{P}(A) \neq 0$ та $\mathrm{P}(B) \neq 0$, то*

$$\mathrm{P}(A/B) = \mathrm{P}(A) \quad \text{і} \quad \mathrm{P}(B/A) = \mathrm{P}(B).$$

.

◀ З означення умовної ймовірності

$$\mathrm{P}(A/B) = \frac{\mathrm{P}(A \cdot B)}{\mathrm{P}(B)} \quad \text{і} \quad \mathrm{P}(B/A) = \frac{\mathrm{P}(A \cdot B)}{\mathrm{P}(A)}$$

Оскільки A і B незалежні, то з (1.7) маємо

$$\mathrm{P}(A/B) = \frac{\mathrm{P}(A) \cdot \mathrm{P}(B)}{\mathrm{P}(B)} = \mathrm{P}(A) \quad \text{і}$$

$$\mathrm{P}(B/A) = \frac{\mathrm{P}(A) \cdot \mathrm{P}(B)}{\mathrm{P}(A)} = \mathrm{P}(B). \quad ▶$$

Доведемо ще одну властивість незалежних подій.

Теорема. *Якщо події A і B незалежні, то незалежними є і події $\overline{A}$ і B та $\overline{A}$ і $\overline{B}$.*

◀ Досить довести, що $\overline{A}$ і B незалежні.

Розглянемо $\mathrm{P}(\overline{A} \cdot B) = \mathrm{P}(B) - \mathrm{P}(A \cdot B)$, бо $B = A \cdot B + \overline{A} \cdot B$ і $(A \cdot B) \cdot (\overline{A} \cdot B) = \emptyset$. Тому

$$\mathrm{P}(\overline{A} \cdot B) = \mathrm{P}(B) - \mathrm{P}(A \cdot B) = \\ = (1 - \mathrm{P}(A)) \cdot \mathrm{P}(B) = \mathrm{P}(\overline{A}) \cdot \mathrm{P}(B),$$

тобто, $\overline{A}$ і B незалежні. ▶

Перейдемо тепер від двох подій до кількох.

Означення. Випадкові події A_1, $A_2, \ldots,$ A_n називаються незалежними в сукупності, якщо для будь-якого набору індексів $1 \leq i_1 < i_2 < \cdots < i_k \leq n$ $(k = \overline{2, n})$

$$\mathrm{P}(A_{i_1} \cdot A_{i_2} \ldots A_{i_k}) = \mathrm{P}(A_{i_1}) \cdot \mathrm{P}(A_{i_2}) \ldots \mathrm{P}(A_{i_k}) \qquad (1.8)$$

Якщо рівність (1.8) має місце при $k = 2$, то події A_1, $A_2, \ldots,$ A_n називають попарно незалежними.

Легко встановити, що із попарної незалежності, взагалі кажучи, незалежність в сукупності не випливає. В цьому переконує нас такий приклад.

Приклад (Бернштейна). Нехай три грані однорідного тетраедра пофарбовані кожна в один з кольорів: синій, червоний, зелений. Четверта грань розфарбована так, що має всі три згадані кольори. Тетраедр підкидають і дивляться якою гранню він впаде на горизонтальну площину.

Нехай A_1={на грані, що випала, є синій колір}, A_2={на грані, що випала, є червоний колір}, A_3={на грані, що випала, є зелений колір}. Тетраедр однорідний, тому всі грані випадають з однаковими ймовірностями. Очевидно, що $\mathrm{P}(A_i) = \frac{1}{2}$, $\mathrm{P}(A_i \cdot A_j) = \frac{1}{4}$ при $i \neq j$. Тому події A_1, A_2, A_3 попарно незалежні. Але

$$\mathrm{P}(A_1 \cdot A_2 \cdot A_3) = \frac{1}{4} \neq \frac{1}{8} = \mathrm{P}(A_1) \cdot \mathrm{P}(A_2) \cdot \mathrm{P}(A_3),$$

і тому ці події не є незалежними в сукупності.

1.4 Формули повної ймовірності та Байєса

Розглянемо деякий ймовірнісний простір $(\Omega, \mathcal{F}, \mathrm{P})$.

Означення. Набір подій H_1, $H_2, \ldots, H_n$ будемо називати повною групою подій, якщо для них виконується таке:

1. $H_i \cdot H_j = \emptyset$, $i \neq j$;

2. $\bigcup\limits_{i=1}^{n} H_i = \Omega$.

Теорема (формула повної ймовірності). Нехай A — деяка подія, H_1, $H_2, \ldots, H_n$ — повна група подій і $\mathrm{P}(H_i) \neq 0$ при всіх $i = \overline{1, n}$. Тоді

$$\mathrm{P}(A) = \sum_{i=1}^{n} \mathrm{P}(A/H_i) \cdot \mathrm{P}(H_i).$$

◀ Очевидно, що подію A можна подати у вигляді $A = \bigcup\limits_{i=1}^{n} A \cdot H_i$ і $(A \cdot H_i) \cdot (A \cdot H_j) = \emptyset$ при $i \neq j$. Тому

$$\mathrm{P}(A) = \bigcup_{i=1}^{n} \mathrm{P}(A \cdot H_i).$$

За теоремою множення ймовірностей

$$\mathrm{P}(A \cdot H_i) = \mathrm{P}(A/H_i) \cdot \mathrm{P}(H_i)$$

при всіх $i = \overline{1, n}$. Отже

$$\mathrm{P}(A) = \sum_{i=1}^{n} \mathrm{P}(A/H_i) \cdot \mathrm{P}(H_i). \quad \blacktriangleright$$

В багатьох випадках система гіпотез не визначається однозначно. В таких випадках перевагу потрібно віддавати тій системі, для якої умовні ймовірності визначаються найбільш просто.

Приклад. З колоди 52 карт випадково послідовно і без повернення вибирають дві карти. Яка ймовірність того, що другою картою можна побити першу? (Друга карта повинна бути більш старшою тієї ж масті.)

Нехай A — подія, ймовірність якої нас цікавить. Розглянемо такі гіпотези: H_k ={в складі двох витягнутих карт є рівно k картинок}, $k = 0, 1, 2$ ("картинки— валет, дама, король і туз).

Очевидно, що набір H_k утворює повну групу подій. Але обчислення умовних ймовірностей $\mathrm{P}(A/H_k)$ є не менш складним ніж обчислення ймовірності події A. Це пояснюється тим, що зв'язок події A з даними гіпотезами не може бути достатньо просто описаний на мові операцій над подіями.

Розглянемо ще один набір, що утворює повну групу подій: H_k ={перша витягнута карта має вартість k очок}, $k = 2, 3, \ldots, 14$ ($k = 2$ — двійка, $k = 3$ — трійка,..., $k = 11$ — валет, $k = 12$ — дама, $k = 13$ — король, $k = 14$ — туз). Обчислимо умовні ймовірності.

$$\mathrm{P}(A/H_k) = \mathrm{P}\{\text{друга карта тієї ж масті та її вартість не менша, ніж } k+1 \text{ очко}\} = \tfrac{14-k}{51}$$

за класичним означенням ймовірності.

Безумовні ймовірності гіпотез $\mathrm{P}(H_k) = \frac{1}{13}$ через рівноймовірність подій H_k.

За формулою повної ймовірності

$$\mathrm{P}(A)=\sum_{k=2}^{14}\mathrm{P}(A/H_k)\mathrm{P}(H_k)=\frac{1}{13}\sum_{k=2}^{14}\frac{14-k}{51}=\frac{2}{17}.$$

Можна запропонувати ще більш вдалий підбір повної групи подій: $H_1 =${дві витягнуті карти однієї масті}, $H_2 =$ {витягнуті карти різних мастей}.

Очевидно, що $\mathrm{P}(A/H_2) = 0$, а $\mathrm{P}(A/H_1) = 0.5$ через рівноймовірність подій {друга карта старша від першої}, {перша карта старша від другої}. Крім того,

$$\mathrm{P}(H_1)=\frac{52\cdot 12}{52\cdot 51}=\frac{4}{17}.$$

Отже, за формулою повної ймовірності

$$\mathrm{P}(A)=\mathrm{P}(A/H_1)\cdot\mathrm{P}(H_1)=\frac{1}{2}\cdot\frac{4}{17}=\frac{2}{17}.$$

Теорема (формула Байєса). Нехай подія A така, що $\mathrm{P}(A)\neq 0$ і $H_1, H_2, \ldots, H_n$ — повна група подій, для яких $\mathrm{P}(H_i)\neq 0$ при всіх $i=\overline{1,n}$. Тоді

$$\mathrm{P}(H_i/A)=\frac{\mathrm{P}(A/H_i)\cdot\mathrm{P}(H_i)}{\mathrm{P}(A)}=\frac{\mathrm{P}(A/H_i)\cdot\mathrm{P}(H_i)}{\sum\limits_{i=1}^{n}\mathrm{P}(A/H_i)\cdot\mathrm{P}(H_i)}$$

при $i=\overline{1,n}$.

◀ Зауважимо, що друга рівність в твердженні теореми є наслідком формули повної ймовірності.

Доведемо першу рівність. Для цього розглянемо при $i=\overline{1,n}$

$$\mathrm{P}(A\cdot H_i)=\mathrm{P}(A/H_i)\cdot\mathrm{P}(H_i)=\mathrm{P}(H_i/A)\cdot\mathrm{P}(A).$$

Звідси
$$\mathrm{P}(H_i/A) = \frac{\mathrm{P}(A/H_i) \cdot \mathrm{P}(H_i)}{\mathrm{P}(A)}. \qquad \blacktriangleright$$

Зауваження.

1. Елементи повної групи подій ще називаються гіпотезами.

2. В формулах повної ймовірності та Байєса замість повної групи подій можна взяти набір подій (гіпотез) $H_1, H_2, \ldots, H_k$ таких, що
$$A \subset \bigcup_{i=1}^{n} H_i \text{ i } H_i \cdot H_j = \emptyset \text{ при } i \neq j.$$

3. Ймовірності $\mathrm{P}(H_i)$ в формулах Байєса часто називають апріорними[3] (до дослідними, заздалегідь відомими), а ймовірності $\mathrm{P}(H_i/A)$ — апостеріорними[4] (знайденими після того, як дослід виконано і відбулась подія A).

Особливе значення формула Байєса має для тих експериментів, в яких гіпотези H_k безпосередньо не спостерігаються, хоча апріорні ймовірності $\mathrm{P}(H_k)$ та відповідні умовні ймовірності $\mathrm{P}(A/H_k)$ відомі з додаткових досліджень. Така ситуація може виникнути, наприклад, якщо відсутній прилад, який дозволяє реєструвати події, що є гіпотезами, або реєстрація гіпотез приводить до знищення предмету спостереження. Для таких експериментів переоцінка ймовірностей гіпотез після досліду може бути здійснена на базі події A, що спостерігається і тісно пов'язана

[3] a priori — до випробування
[4] a posteriori — після випробування

з гіпотезами. Такий підхід часто використовується в задачах медичної та технічної діагностики.

Приклад. Вивчається три види дефектів запам'ятовуючих пристроїв, що виконані на інтегральних мікросхемах: дефекти схем (гіпотеза H_1), дефекти, пов'язані з паразитними зв'язками між чарунками (гіпотеза H_2), та дефекти адресних шин (гіпотеза H_3). Діагностика проводиться з допомогою деяких тестів. Нехай тест проведено і виявлено певний результат (подія A). З попередніх досліджень відомо, що одержаний результат спостерігається в 40%, 20% та 30% випадків при наявності дефектів, що складають зміст гіпотез H_1, H_2 та H_3, відповідно. Крім того, в тих же дослідженнях встановлено, що, в середньому, 10% несправних запам'ятовуючих пристроїв мають дефекти схем, 60% мають дефекти зв'язків та 30% мають дефекти адресних шин.

Встановити, яка з гіпотез має найбільшу апостеріорну ймовірність (який з дефектів найбільш ймовірний).

З умови зрозуміло, що

$$\mathrm{P}(A/H_1) = 0.4, \quad \mathrm{P}(A/H_2) = 0.2, \quad \mathrm{P}(A/H_3) = 0.3,$$
$$\mathrm{P}(H_1) = 0.1, \quad \mathrm{P}(H_1) = 0.6, \quad \mathrm{P}(H_1) = 0.3.$$

За формулою повної ймовірності

$$\mathrm{P}(A) = \sum_{k=1}^{3} \mathrm{P}(A/H_k)\mathrm{P}(H_k) = 0.25,$$

а за формулами Байєса

$$\mathrm{P}(H_1/A) = \tfrac{0.4\cdot 0.1}{0.25} = \tfrac{4}{25}, \quad \mathrm{P}(H_2/A) = \tfrac{0.2\cdot 0.6}{0.25} = \tfrac{12}{25},$$
$$\mathrm{P}(H_3/A) = \tfrac{0.3\cdot 0.3}{0.25} = \tfrac{9}{25}.$$

Отже, найбільш ймовірно, що запам'ятовуючий пристрій має дефекти зв'язку.

1.5 Повторні незалежні випробування

1.5.1 Схема Бернуллі

Нехай в деякому стохастичному експерименті може наблюдатися подія A, ймовірність якої дорівнює p. Розглянемо таку схему проведення випробувань (схема Бернуллі). Згаданий стохастичний експеримент проводиться в однакових умовах незалежно один від одного n раз. В кожному експерименті встановлюється, відбулась (успіх) чи не відбулась (невдача) подія A.

Розглянемо деякі питання, пов'язані зі схемою Бернуллі.

Спочатку знайдемо ймовірність $P_n(k)$ того, що в описаній схемі випробувань відбулось рівно k успіхів ($k = \overline{0, n}$). Для цього розглянемо події A_i ={в i-тому випробуванні відбувся успіх}. Через незалежність випробувань можна вважати, що події A_i незалежні в сукупності. Нехай подія B полягає в тому, що в схемі з n випробувань відбудеться рівно k успіхів. Очевидно, що

$$B = \sum A_{i_1} \cdot A_{i_2} \dots A_{i_k} \cdot \overline{A_{j_1}} \cdot \overline{A_{j_2}} \dots \overline{A_{j_{n-k}}},$$

де сума береться по наймождливіших наборах (невпорядкованих) індексів i_1, $i_2, \dots,$ i_k причому набір j_1, $j_2, \dots,$ j_{n-k} доповнює набір i_1, $i_2, \dots,$ i_k до невпорядкованого набору чисел 1, $2, \dots,$ n.

Всі події

$$A_{i_1} \cdot A_{i_2} \dots A_{i_k} \cdot \overline{A_{j_1}} \cdot \overline{A_{j_2}} \dots \overline{A_{j_{n-k}}}$$

попарно несумісні і мають ймовірності $p^k(1-p)^{n-k}$. Тому

$$P_n(k) = C_n^k p^k (1-p)^{n-k}. \tag{1.9}$$

1.5.2 Найімовірніша кількість успіхів в схемі Бернуллі

Дослідимо питання про поведінку чисел $P_n(k)$, заданих формулою (1.9), при зростанні k від 0 до n. Для зручності позначимо $q = 1 - p$ і розглянемо

$$\frac{P_n(k+1)}{P_n(k)} = \frac{C_n^{k+1} p^{k+1} q^{n-k-1}}{C_n^k p^k q^{n-k}} = \frac{n!k!(n-k)!}{(k+1)!(n-k-1)!n!} \cdot \frac{p}{q} = \\ = \frac{n-k}{k+1} \cdot \frac{p}{q}.$$

Звідси $P_n(k+1) > P_n(k)$, якщо $\frac{n-k}{k+1} \cdot \frac{p}{q} > 1$, тобто

$$(n-k)p > (k+1)q \Leftrightarrow np > k(p+q)+q = k+q \Leftrightarrow k < np-q.$$

Аналогічно $P_n(k+1) < P_n(k)$, при $k > np - q$.

Якщо $np - q$ ціле, то $P_n(k+1) = P_n(k)$ при $k = np - q$. Тут слід зауважити, що $-1 < np - q < n$ в усіх випадках.

Звідси можна зробити такий висновок щодо зміни $P_n(k)$ при рості k: якщо k змінюється від 0 до $[np - q]$, то $P_n(k)$ зростає; якщо k змінюється від $[np - q] + 1$ до n, то $P_n(k)$ спадає. Тому має місце таке твердження:

Теорема (про найімовірнішу кількість успіхів в схемі Бернуллі). Якщо $np - q$ $(q = 1 - p)$ ціле, то є два найбільш ймовірні значеня кількості успіхів в n випробуваннях Бернуллі: $np-q$ і $np-q+1$. Якщо ж $np-q$ не ціле, то найбільш ймовірна кількість успіхів дорівнює $[np - q] + 1$.

1.5.3 Теорема Пуассона

Формула (1.9) дає можливість обчислювати ймовірності $P_n(k)$ при будь-яких значеннях параметрів схеми Бернуллі. Але, особливо при великих n, використаня (1.9) веде за собою неоправдано складні обчислення. Розглянемо один спосіб наближеного обчислення чисел $P_n(k)$.

Розглянемо послідовність серій випробувань за схемою Бернуллі. В n-тій серії виконується n випробувань, ймовірність успіху в кожному з них дорівнює $p_n = \frac{\lambda}{n}$ ($\lambda = const$, $\lambda < n$).

Має місце теорема *Пуассона.*

Теорема Пуассона. Якщо $P_n(k)$ — ймовірність k успіхів у серії з n випробувань за схемою Бернуллі, в кожному з яких ймовірність успіху дорівнює $p_n = \frac{\lambda}{n}$, то

$$\lim_{n\to\infty} P_n(k) = \frac{\lambda^k}{k!}e^{-\lambda}.$$

◀ За формулою (1.9)

$$P_n(k) = C_n^k \left(\tfrac{\lambda}{n}\right)^k \left(1 - \tfrac{\lambda}{n}\right)^{n-k} =$$
$$= \tfrac{n(n-1)\dots(n-k+1)}{k!} \left(\tfrac{\lambda}{n}\right)^k \left(1 - \tfrac{\lambda}{n}\right)^{n-k} =$$
$$= \tfrac{\lambda^k}{k!} \left(1 - \tfrac{1}{n}\right) \dots \left(1 - \tfrac{k-1}{n}\right) \left(1 - \tfrac{\lambda}{n}\right)^{-k} \left(1 - \tfrac{\lambda}{n}\right)^{n}.$$

Тому

$$\lim_{n\to\infty} P_n(k) = \tfrac{\lambda^k}{k!} \lim_{n\to\infty} \left(1 - \tfrac{1}{n}\right) \cdot \dots \cdot \left(1 - \tfrac{k-1}{n}\right) \times$$
$$\times \left(1 - \tfrac{\lambda}{n}\right)^{-k} \lim_{n\to\infty} \left(1 - \tfrac{\lambda}{n}\right)^{n} =$$
$$= \tfrac{\lambda^k}{k!} \lim_{n\to\infty} \left(\left(1 - \tfrac{\lambda}{n}\right)^{-\frac{n}{\lambda}}\right)^{-\lambda} = \tfrac{\lambda^k}{k!} e^{-\lambda}. \quad ▶$$

З теореми Пуассона випливає, що в схемі Бернуллі з досить великою кількістю випробувань n та ймовірністю успіху в кожному випробуванні p

$$P_n(k) \approx \frac{\lambda^k}{k!}e^{-\lambda}, \tag{1.10}$$

де $\lambda = np$.

Формула (1.10) дає досить точне наближення при малих p (менших 0.1) і досить великих n (не менших кількох десятків) та $np \leq 10$. Хоча і ці побажання не можна вважати обов'язковими.

Приклад. Апаратура містить 2000 однаково надійних елементів, ймовірність відмови кожного з них дорівнює $p = 0.0005$. Знайти ймовірність відмови апаратури, якщо вона настає при відмові хоча б одного з елементів.

Оскільки кожен з елементів виходить з ладу зовсім незалежно від інших, то ми маємо справу із схемою Бернуллі з $n = 2000$ випробувань та ймовірністю успіху (вихід з ладу) в кожному випробуванні $p = 0.0005$. Тому ймовірність відмови апаратури дорівнює

$$1 - P_n(0) = 1 - C_{2000}^0 (0.0005)^0 (0.9995)^{2000} = \\ = 1 - (0.9995)^{2000} \approx 0.63227$$

Скористаємось теоремою Пуассона $P_n(0) \approx \frac{\lambda^0}{0!} e^{-\lambda}$, де $\lambda = np = 1$

Тому шукана ймовірність

$$1 - P_n(0) \approx 1 - e^{-1} \approx 0.63212.$$

Як бачимо, значення, одержане з допомогою наближеної формули за теоремою Пуассона, досить добре наближає точне значення ймовірності.

Розділ 2

Випадкові величини

2.1 Випадкові величини, їх розподіли

2.1.1 Функція розподілу

Нехай задано деякий ймовірнісний простір $(\Omega, \mathcal{F}, \mathrm{P})$. Нагадаємо, що це означає, що певному стохастичному експерименту відповідає простір елементарних подій Ω, σ-алгебра подій $\mathcal{F}$ та ймовірнісна міра P. Нехай в цьому експерименті вимірюється значення деякої величини ξ. Зрозуміло, що для кожної елементарної події $\omega \in \Omega$ одержуємо деяке значення ξ, тобто, ξ є функцією від ω

Означення. Функцію $\xi(\omega)$ зі значеннями в $\mathbb{R}$, задану на просторі елементарних подій Ω, будемо називати *випадковою величиною*, якщо $\{\omega : \xi(\omega) < x\} \in \mathcal{F}$ при всіх $x \in \mathbb{R}$.

Приклад. Двічі підкидається монета; ξ — кількість гербів, що випадуть.

Зрозуміло, що $\Omega = \{\text{гг, гц, цг, цц}\}$, $\mathcal{F}$ — σ-алгебра всіх підмножин Ω, P така, що $\mathrm{P}(\omega_i) = \frac{1}{4}$, $i = \overline{1,4}$.

Звідси величина ξ може приймати лише значення 0, 1, 2. Тому при $x \le 0$

$$\{\omega : \xi(\omega) < x\} = \emptyset \in \mathcal{F},$$

при $0 < x \le 1$

$$\{\omega : \xi(\omega) < x\} = \{\omega : \xi(\omega) = 0\} = \{\text{цц}\} \in \mathcal{F},$$

при $1 < x \le 2$

$$\begin{gathered}\{\omega : \xi(\omega) < x\} = \{\omega : \xi(\omega) = 0 \text{ або} \\ \xi(\omega) = 1\} = \{\text{цц, цг, гц}\} \in \mathcal{F},\end{gathered}$$

при $x > 2$

$$\begin{gathered}\{\omega : \xi(\omega) < x\} = \{\omega : \xi(\omega) = 0 \text{ або} \\ \xi(\omega) = 1 \text{ або } \xi(\omega) = 2\} = \\ = \{\text{цц, гц, цг, гг}\} = \Omega \in \mathcal{F}.\end{gathered}$$

Отже, ξ є випадковою величиною.

Зауваження. В більшості випадків при позначенні випадкової величини аргумент її, як функції на Ω, будемо пропускати і вважатимемо, що ξ і $\xi(\omega)$ це одне і те ж.

Означення. Функція $F(x)$, задана на $\mathbb{R}$, така, що

$$F(x) = \mathrm{P}(\xi < x),$$

називається функцією розподілу випадкової величини ξ.

Приклад. Для випадкової величини ξ , розглянутої в попередньому прикладі,

$$F(x) = \begin{cases} 0, & x \le 0 \\ \frac{1}{4}, & 0 < x \le 1 \\ \frac{3}{4}, & 1 < x \le 2 \\ 1, & x > 2 \end{cases}.$$

Розглянемо основні властивості функції розподілу:

1. Функція $F(x)$ неспадна, тобто, якщо $x_1 < x_2$, то $F(x_1) \leq F(x_2)$.

2. Для всіх $x \in \mathbb{R}$ $0 \leq F(x) \leq 1$ і $\lim\limits_{x\to-\infty} F(x) = 0$, а $\lim\limits_{x\to+\infty} F(x) = 1$.

3. Функція $F(x)$ неперервна зліва.

◀ 1) Очевидно, що

$$\{\omega : \xi(\omega) < x_1\} \subset \{\omega : \xi(\omega) < x_2\},$$

тому $\mathrm{P}(\xi < x_1) \leq \mathrm{P}(\xi < x_2)$ за властивістю монотонності ймовірності. Отже, $F(x_1) \leq F(x_2)$.

2) Оскільки $0 \leq \mathrm{P}(\xi < x) \leq 1$ за властивістю ймовірності, то $0 \leq F(x) \leq 1$.

Розглянемо послідовність $\{x_n : n \geq 1\}$ таку, що $x_1 > x_2 > \dots > x_n > \dots$ і $\lim\limits_{n\to\infty} x_n = -\infty$. Та розглянемо події $A_n = \{\omega : \xi(\omega) < x_n\}$. Очевидно, що
$A_1 \supset A_2 \supset \dots \supset A_n \supset \dots$ і $\bigcap\limits_{n=1}^{\infty} A_n = \emptyset$. Тому за теоремою неперервності ймовірності

$$\lim_{n\to\infty} F(x_n) = \lim_{n\to\infty} \mathrm{P}(A_n) = \mathrm{P}\left(\bigcap_{n=1}^{\infty} A_n\right) = \mathrm{P}(\emptyset) = 0.$$

Отже, $\lim\limits_{x\to-\infty} F(x) = 0$.

Розглянемо тепер послідовність $\{x_n : n \geq 1\}$ таку, що $x_1 < x_2 < \dots < x_n < \dots$ і $\lim\limits_{n\to\infty} x_n = +\infty$. Для подій $B_n = \{\omega : \xi(\omega) \geq x_n\}$, очевидно, виконується
$B_1 \supset B_2 \supset \dots \supset B_n \supset \dots$ і $\bigcap\limits_{n=1}^{\infty} B_n = \emptyset$. Тому за тією ж

властивістю неперервності ймовірності

$$\lim_{n\to\infty} F(x_n) = \lim_{n\to\infty} \mathrm{P}(\overline{B_n}) = 1 - \lim_{n\to\infty} \mathrm{P}(B_n) =$$
$$= 1 - \mathrm{P}\left(\bigcap_{n=1}^{\infty} B_n\right) = 1 - \mathrm{P}(\emptyset) = 1.$$

Отже, $\lim\limits_{x\to+\infty} F(x) = 1$.

3) Розглянемо довільну послідовність $\{x_n : n \geq 1\}$ таку, що $x_1 < x_2 < \dots < x_n < \dots$ і $\lim\limits_{n\to\infty} x_n = x_0$. Події $C_n = \{\omega : x_n \leq \xi(\omega) < x_0\}$ такі, що

$$C_1 \supset C_2 \supset \dots \supset C_n \supset \dots \text{ і } \bigcap_{n=1}^{\infty} C_n = \emptyset.$$

Тому за властивістю неперервності ймовірності

$$\lim_{n\to\infty} \mathrm{P}(x_n \leq \xi < x_0) = \mathrm{P}(\emptyset) = 0.$$

Тобто $\lim\limits_{n\to\infty} (\mathrm{P}(\xi < x_0) - \mathrm{P}(\xi < x_n)) = 0$ і
$\lim\limits_{n\to\infty} (F(x_0) - F(x_n)) = 0$, що еквівалентне рівності
$\lim\limits_{x\to x_0 - 0} F(x) = F(x_0)$.

Оскільки точка x_0 вибрана абсолютно довільно, то $F(x)$ неперервна зліва на $\mathbb{R}$. ►

2.1.2 Ймовірність потрапляння значення випадкової величини в проміжок

Нехай випадкова величина ξ має функцію розподілу $F(x)$. В багатьох випадках виникає необхідність обчислити ймовірність попадання випадкової величини ξ в один з інтервалів $(a; b)$, $[a; b)$, $(a; b]$, $[a; b]$ і т.д.

Доведемо таке твердження:

Теорема (про ймовірність потрапляння значення випадкової величини в проміжок). Для випадкової величини ξ з функцією розподілу $F(x)$ та чисел $a, b \in \mathbb{R}$ виконується

1. $\mathrm{P}(\xi \in (-\infty;\ b)) = F(b)$;
2. $\mathrm{P}(\xi \in (-\infty;\ b]) = F(b+0)$;
3. $\mathrm{P}(\xi \in (a;\ +\infty)) = 1 - F(a+0)$;
4. $\mathrm{P}(\xi \in [a;\ +\infty)) = 1 - F(a)$;
5. $\mathrm{P}(\xi \in (a;\ b)) = F(b) - F(a+0)$;
6. $\mathrm{P}(\xi \in (a;\ b]) = F(b+0) - F(a+0)$;
7. $\mathrm{P}(\xi \in [a;\ b)) = F(b) - F(a)$;
8. $\mathrm{P}(\xi \in [a;\ b]) = F(b+0) - F(a)$.

◀ Перше твердження не потребує доведення, бо співпадає, посуті, з означенням $F(b)$.

Нехай $\{x_n : n \geq 1\}$ довільна послідовність така, що $x_1 > x_2 > \cdots > x_n > \ldots$ і $\lim\limits_{n\to\infty} x_n = b$. Для подій $A_n = \{\xi < x_n\}$ виконується $A_1 \supset A_2 \supset \cdots \supset A_n \supset \ldots$ і $\bigcap\limits_{n=1}^{\infty} A_n = \{\xi \leq b\}$. Тому за теоремою неперервності ймовірності $\lim\limits_{n\to\infty} \mathrm{P}(\xi < x_n) = \mathrm{P}(\xi \leq b)$, тобто,

$$\mathrm{P}(\xi \in (-\infty;\ b]) = \mathrm{P}(\xi \leq b) = \lim_{n\to\infty} F(x_n) = F(b+0).$$

Отже, друге твердження доведено.

Оскільки $\{\xi \in (a;\ +\infty)\} = \overline{\{\xi \in (-\infty;\ a]\}}$, то з 2) маємо $\mathrm{P}(\xi \in (a;\ +\infty)) = 1 - F(a+0)$, що є змістом твердження 3).

Так само, оскільки $\{\xi \in [a;\ +\infty)\} = \overline{\{\xi \in (-\infty;\ a)\}}$, то з 1) маємо $\mathrm{P}(\xi \in [a;\ +\infty)) = 1 - F(a)$, як і стверджувалось в 4).

Подамо кожну з подій пунктів 5) — 8) у вигляді різниці

$$\{\xi \in (a;\ b)\} = \{\xi \in (-\infty;\ b)\} \setminus \{\xi \in (-\infty;\ a]\},$$

$$\{\xi \in (a;\ b]\} = \{\xi \in (-\infty;\ b]\} \setminus \{\xi \in (-\infty;\ a]\},$$

$$\{\xi \in [a;\ b)\} = \{\xi \in (-\infty;\ b)\} \setminus \{\xi \in (-\infty;\ a)\},$$

$$\{\xi \in [a;\ b]\} = \{\xi \in (-\infty;\ b]\} \setminus \{\xi \in (-\infty;\ a)\}.$$

Оскільки кожна перша подія в різниці є наслідком другої, то з 1) і 2) випливають твердження 5) — 8). ▶

Зауваження. Якщо функція розподілу $F(x)$ випадкової величини ξ неперервна, то, очевидно, всі ймовірності в 5) — 8) теореми дорівнюють $F(b) - F(a)$. Так само ймовірності в 1), 2) дорівнюють $F(b)$ а в 3) і 4) — $1 - F(a)$.

2.1.3 Неперервні та дискретні випадкові величини

Як ми бачили з попереднього параграфу, властивості випадкової величини багато в чому залежать від характеру її функції розподілу.

Означення. Випадкова величина ξ називається неперервною, якщо її функція розподілу $F(x)$ може бути подана у вигляді

$$F(x) = \int\limits_{-\infty}^{x} f(y)\mathrm{d}y \tag{2.1}$$

При цьому функція $f(x)$ називається щільністю розподілу величини ξ.

Зауваження.

1. З математичного аналізу відомо, що функція задана рівністю (2.1) неперервна, а у випадку неперервності функції $f(x)$ ще й диференційовна. Тому в точках, де існує похідна функції $F(x)$, $f(x) = \frac{\mathrm{d}F(x)}{\mathrm{d}x}$.

2. Іноді функцію $F(x)$ називають інтегральною, а функцію $f(x)$ — диференціальною функцією неперервної випадкової величини ξ.

3. Для неперервної випадкової величини ξ ймовірності з пунктів 5) — 8) теореми про ймовірність потрапляння значення випадкової величини в проміжок дорівнюють $\int\limits_a^b f(x)\mathrm{d}x$.

Наведемо найпростіші властивості щільності розподілу $f(x)$, що випливають з відповідних властивостей функції розподілу.

1. Функція $f(x)$ інтегровна в невласному розумінні на $(-\infty;\ +\infty)$ і
$$\int\limits_{-\infty}^{+\infty} f(x)\mathrm{d}x = 1 \tag{2.2}$$

2. Функція $f(x)$ невід'ємна.

◀ Властивість 1) випливає з того, що

$$\lim_{x\to+\infty} F(x) = 1$$

і має місце (2.1).

Для доведення 2) припустимо, що на деякому інтервалі $(\alpha;\beta)$ $f(x) < 0$. Тоді $\mathrm{P}(\xi \in (\alpha;\beta)) = \int\limits_{\alpha}^{\beta} f(x)\mathrm{d}x < 0$, що неможливо. Тому $f(x) \geq 0$. ►

Зауважимо, що в окремих точках (якщо їх зліченна кількість) $f(x)$ може бути від'ємною. Хоча ми будемо такі функції змінювати так, щоб $f(x)$ було невід'ємним при всіх значеннях аргумента. При цьому, значення інтеграла в (2.1) не змінюється.

Таким чином, завжди будемо вважати, що щільність розподілу $f(x)$ невід'ємна і для неї виконується (2.2). І навпаки, кожна невід'ємна функція, для якої виконується (2.2), може бути щільністю розподілу.

Означення. Випадкова величина називається дискретною, якщо вона може приймати не більше, ніж зліченну кількість значень.

Нехай дискретна випадкова величина ξ приймає значення x_1, x_2, $\ldots$ і $\mathrm{P}(\xi = x_k) = p_k$. Набір пар чисел $(x_1;\ p_1)$, $(x_2;\ p_2), \ldots$ будемо називати розподілом дискретної випадкової величини ξ. Ми будемо відкидати ті значення x_k випадкової величини ξ, для яких $p_k = 0$. Тому в розподілі дискретної випадкової величини всі $p_k > 0$ і, очевидно, $\sum\limits_k p_k = 1$.

Функція розподілу дискретної випадкової величини ξ задається рівністю

$$F(x) = \sum_{k:x_k<x} p_k.$$

Зауваження.

1. Вище було встановлено, що функція розподілу неперервної випадкової величини неперервна. Виявляє-

ться, що зовсім не обов'язково випадкова величина з неперервною функцією розподілу є неперервною. Такі величини називаються сингулярними.

2. Надалі ми будемо розглядати тільки неперервні та дискретні випадкові величини.

2.1.4 Розподіл функцій від випадкових величин

Нехай випадкова величина ξ має функцію розподілу $F_\xi(x)$ чи щільність розподілу $f_\xi(x)$. Розглянемо функцію $g(x)$, $x \in \mathbb{R}$ таку, що $\eta = g(\xi)$ є випадковою величиною. Нагадаємо, що для цього необхідно і досить щоб

$$\{\omega : \eta(\omega) < x\} \in \mathcal{F} \text{ при всіх } x \in \mathbb{R}.$$

Якщо випадкова величина ξ дискретна і має розподіл $(x_i; p_i)$, то випадкова величина $\eta = g(\xi)$ має розподіл $(y_k; q_k)$, де $y_k = g(x_{i_k})$, а $q_k = \sum\limits_{i:g(x_i)=y_k} p_i$.

Нехай тепер ξ — неперервна випадкова величина. Припустимо, для простоти, що функція $g(x)$ монотонно зростаюча. Тоді, оскільки

$$\mathrm{P}(\eta < x) = \mathrm{P}(g(\xi) < x) = \mathrm{P}(\xi < g^{-1}(x)), \text{ то}$$

$$F_\eta(x) = F_\xi(g^{-1}(x)). \tag{2.3}$$

Якщо, крім того, функція $g(x)$ неперервно диференційовна ($g'(x) \neq 0$), то випадкова величина η неперервна та її щільність розподілу

$$f_\eta(x) = \frac{\mathrm{d}F_\xi(g^{-1}(x))}{\mathrm{d}x} = f_\xi(g^{-1}(x)) \cdot \frac{1}{g'(g^{-1}(x))}. \tag{2.4}$$

Зауваження. Якщо ж функція $g(x)$ не задовольняє перераховані умови (доречі досить обмежливі), то, зрозуміло, користуватися формулами (2.3) чи (2.4) не можна. Тому в конкретних задачах більш прийнятним є проведення всіх викладок, які були використані при виведенні згаданих формул.

Приклад. Нехай випадкова величина ξ задана щільністю розподілу $f_\xi(x) = \frac{1}{\pi} \cdot \frac{1}{1+x^2}$, $x \in \mathbb{R}$. Знайти розподіл випадкової величини $\eta = \xi^2$.

Знайдемо

$$F_\eta(x) = \mathrm{P}(\eta < x) = \mathrm{P}(\xi^2 < x) =$$
$$= \begin{cases} 0, & x \le 0 \\ \mathrm{P}(-\sqrt{x} < \xi < \sqrt{x}), & x > 0 \end{cases} .$$

Нехай $x > 0$ і розглянемо

$$\mathrm{P}(-\sqrt{x} < \xi < \sqrt{x}) = \frac{1}{\pi} \int\limits_{-\sqrt{x}}^{\sqrt{x}} \frac{1}{1+y^2} \mathrm{d}y =$$
$$= \frac{1}{\pi} \operatorname{arctg} y \Big|_{-\sqrt{x}}^{\sqrt{x}} = \frac{1}{\pi} \operatorname{arctg} \sqrt{x}.$$

Звідси

$$f_\eta(x) = \begin{cases} 0, & x \le 0 \\ \frac{1}{\pi} \cdot \frac{1}{(1+x)\sqrt{x}} & x > 0 \end{cases} .$$

2.2 Випадкові вектори

2.2.1 Розподіли випадкового вектора та його елементів

Почнемо з означення:

Означення. Впорядкований набір випадкових величин $(\xi_1; \ldots; \xi_n)$, заданих на одному й тому ж ймовірнісному просторі $(\Omega,\ \mathcal{F},\ \mathrm{P})$, будемо називати n-вимірним випадковим вектором.

Надалі для спрощення записів розглядатимемо тільки двовимірні випадкові вектори.

Означення. Функція двох змінних $F(x,y)$, називається функцією розподілу випадкового вектора $(\xi;\eta)$, якщо для всіх $x,\ y \in \mathbb{R}$

$$F(x,y) = \mathrm{P}(\xi < x,\ \eta < y).$$

Означення. Фунція двох змінних $f(x,y)$, називається щільністю розподілу випадкового вектора $(\xi;\eta)$, якщо його функція розподілу може бути подана у вигляді

$$F(x,y) = \int\limits_{-\infty}^{x} \mathrm{d}s \int\limits_{-\infty}^{y} f(s,t)\mathrm{d}t.$$

В цьому випадку випадковий вектор $(\xi;\eta)$ називається неперервним.

Означення. Випадковий вектор $(\xi;\eta)$ називається дискретним, якщо множина його значень не більше ніж злічення.

Нехай дискретний випадковий вектор $(\xi;\eta)$ приймає значення $(x_i; y_j)$. Розглянемо набір чисел

$$p_{ij} = \mathrm{P}(\xi = x_i, \eta = y_j).$$

Набір трійок чисел $(x_i, y_j; p_{ij})$ будемо називати розподілом дискретного випадкового вектора $(\xi;\eta)$. Аналогічно до того, що ми мали для випадкових величин сформулюємо характерні властивості функцій та щільностей розподілів, а також розподілів (в дискретному випадку) випадкових векторів

1. (a) $\lim\limits_{x\to+\infty} F(x,y) = F_\eta(y)$, $\lim\limits_{y\to+\infty} F(x,y) = F_\xi(x)$, $\lim\limits_{x\to+\infty}\lim\limits_{y\to+\infty} F(x,y) = 1$;

   (b) $\int\limits_{-\infty}^{+\infty} f(x,y)\mathrm{d}x = f_\eta(y)$, $\int\limits_{-\infty}^{+\infty} f(x,y)\mathrm{d}y = f_\xi(x)$, $\int\limits_{-\infty}^{+\infty}\mathrm{d}x\int\limits_{-\infty}^{+\infty} f(x,y)\mathrm{d}y = 1$;

   (c) $\sum\limits_i p_{ij} = p_{\cdot j}$, $\sum\limits_j p_{ij} = p_{i\cdot}$, $\sum\limits_i\sum\limits_j p_{ij} = 1$.

   Тут $F_\xi(x)$, $F_\eta(y)$ — функції розподілу; $f_\xi(x)$, $f_\eta(y)$ — щільності розподілу; $(x_i; p_{i\cdot})$, $(y_j; p_{\cdot j})$ — розподіли випадкових величин ξ і η, відповідно.

2. Для всіх $x, y \in \mathbb{R}$ $0 \le F(x,y) \le 1$, $f(x,y) \ge 0$, а також для всіх i, j $0 \le p_{ij} \le 1$.

3. Функція $F(x,y)$ неспадна і неперервна зліва по кожній із змінних.

4. $\mathrm{P}((\xi;\eta) \in G) = \iint\limits_G f(x,y)\mathrm{d}x\mathrm{d}y$.

5. Для неперервного випадкового вектора в точках двічідиференційовності функції розподілу $F(x,y)$

$$f(x,y) = \frac{\partial^2 F(x,y)}{\partial x \partial y}.$$

Доведемо деякі з цих властивостей.

◀ Для послідовності $x_1 < x_2 < \dots < x_n < \dots$ такої, що $\lim\limits_{n\to\infty} x_n = +\infty$ розглянемо події

$$A_n = \{\xi \ge x_n, \eta < y\}.$$

$$A_1 \supset A_2 \supset \dots \supset A_n \supset \dots \text{ і } \bigcap_{n=1}^{\infty} A_n = \emptyset.$$

Тому $\lim\limits_{n\to\infty} \mathrm{P}(A_n) = \mathrm{P}(\emptyset) = 0$.

Отже,

$$\lim_{n\to\infty} F(x_n, y) = \lim_{n\to\infty} \mathrm{P}(\xi < x_n, \eta < y) =$$
$$= \lim_{n\to\infty} \mathrm{P}(\{\eta < y\} \setminus \{\xi \geq x_n, \eta < y\}) =$$
$$= \lim_{n\to\infty} (\mathrm{P}(\eta < y) - \mathrm{P}(\xi \geq x_n, \eta < y)) =$$
$$= \mathrm{P}(\eta < y) = F_\eta(y),$$

тому $\lim\limits_{x\to+\infty} F(x, y) = F_\eta(y)$.

Аналогічно доводиться, що $\lim\limits_{y\to+\infty} F(x, y) = F_\xi(x)$, і тому $\lim\limits_{x\to+\infty} \lim\limits_{y\to+\infty} F(x, y) = 1$.

Якщо випадковий вектор $(\xi; \eta)$ нерерервний, то, продиференціювавши по y рівність (тільки-що доведену)

$$\int_{-\infty}^{+\infty} \mathrm{d}s \int_{-\infty}^{y} f(s, t)\mathrm{d}t = F_\eta(y),$$

одержимо $f_\eta(y) = \int\limits_{-\infty}^{+\infty} f(s, y)\mathrm{d}s$.

Аналогічно $f_\xi(x) = \int\limits_{-\infty}^{+\infty} f(x, t)\mathrm{d}t$.

Розглянемо події $A_{ij} = \{\xi = x_i, \eta = y_j\}$, що утворюють повну групу подій. Оскільки $\bigcup\limits_i A_{ij} = \{\eta = y_j\}$, а $\bigcup\limits_j A_{ij} = \{\xi = x_i\}$, то

$$\sum_i p_{ij} = p_{\cdot j} = \mathrm{P}(\eta = y_j), \ \sum_j p_{ij} = p_{i\cdot} = \mathrm{P}(\xi = x_i).$$

Крім того, $\bigcup\limits_{ij} A_{ij} = \Omega$ і тому $\sum\limits_{ij} p_{ij} = 1$. Отже, 1) доведено.

Властивості 2) і 3) легко випливають з означень.

Нехай G — прямокутник ($G = [a;b] \times [c;d]$). Тоді (див. рис. 2.1)

$$\begin{gathered}\mathrm{P}((\xi;\eta) \in G) = \mathrm{P}((\xi;\eta) \in A \cup B \cup G) - \\ -\mathrm{P}((\xi;\eta) \in A \cup B) = \\ = \mathrm{P}((\xi;\eta) \in A \cup B \cup G) - (\mathrm{P}((\xi;\eta) \in A) + \\ +\mathrm{P}((\xi;\eta) \in B) - \mathrm{P}((\xi;\eta) \in A \cap B)).\end{gathered}$$

або

$$\begin{gathered}\mathrm{P}((\xi;\eta) \in G) = \mathrm{P}(a \le \xi \le b, c \le \eta \le d) = \\ = \mathrm{P}(\xi \le b, \eta \le d) - \mathrm{P}(\xi \le b, \eta < c) - \\ -\mathrm{P}(\xi < a, \eta \le d) + \mathrm{P}(\xi < a, \eta < c).\end{gathered}$$

З врахуванням неперервності $F(x,y)$ за обома змінними можна записати

$$\begin{gathered}\mathrm{P}((\xi;\eta) \in G) = F(b,d) - F(b,c) - F(a,d) + F(a,c) = \\ = (F(b,d) - F(b,c)) - (F(a,d) - F(a,c)) = \\ = \int\limits_a^b \mathrm{d}x \int\limits_c^d f(x,y)\mathrm{d}y.\end{gathered}$$

Якщо ж G не є прямокутником, то, використовуючи метод побудови подвійного інтегралу та неперервність ймовірнісної міри, можна встановити справедливість властивості 4).

Для доведення останньої властивості достатньо зауважити, що

$$\frac{\partial F(x,y)}{\partial x} = \int\limits_{-\infty}^{y} f(x,t)\mathrm{d}t \quad \text{і} \quad \frac{\partial^2 F(x,y)}{\partial x^2} = f(x,y). \quad \blacktriangleright$$

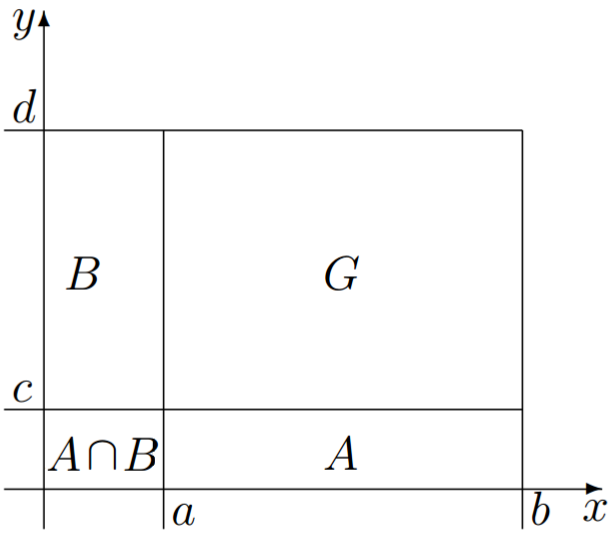

Рис. 2.1

Зауваження. Розподіли елементів випадкового вектора іноді називають маргінальним. Як і для випадкових величин розподіли випадкових векторів не обмежується неперервними та дискретними розподілами, але тільки їх ми будемо розглядати.

Крім маргінальних розподілів випадкового вектора, часто приходиться розглядати і умовні розподіли. Розглянемо розподіл елемента, наприклад, ξ випадкового вектора $(\xi;\eta)$ при умові, що інший елемент цього вектора $\eta < y_0$ (у випадку неперервного випадкового вектора), чи $\eta = y_0$ (у випадку дискретного випадкового вектора). Тут y_0 — деяке фіксоване значеня, вважаємо, що $\mathrm{P}(\eta < y_0) > 0$ або $\mathrm{P}(\eta = y_0) > 0$. Зрозуміло, що

$$\mathrm{P}(\xi < x/\eta < y_0) = \frac{\mathrm{P}(\xi < x, \eta < y_0)}{\mathrm{P}(\eta < y_0)} = \frac{F(x, y_0)}{F_\eta(y_0)}$$

чи

$$\mathrm{P}(\xi = x_i/\eta = y_0) = \frac{\mathrm{P}(\xi = x_i, \eta = y_0)}{\mathrm{P}(\eta = y_0)} = \frac{p_{ij}}{p_{\cdot j}},$$

де j таке, що $y_j = y_0$. Ці рівності і задають умовний розподіл елемента ξ випадкового вектора $(\xi; \eta)$ при умові, що $\eta < y_0$ чи $\eta = y_0$.

2.2.2 Незалежні випадкові величини

Означення. Випадкові величини ξ і η називаються *незалежними*, якщо для будь-яких $x, y \in \mathbb{R}$ події $\{\xi < x\}$, $\{\eta < y\}$ незалежні, тобто

$$\mathrm{P}(\xi < x, \eta < y) = \mathrm{P}(\xi < x) \cdot \mathrm{P}(\eta < y).$$

Зауваження. Для дискретних випадкових величин ξ і η незалежність означає

$$\mathrm{P}(\xi = x_i, \eta = y_j) = \mathrm{P}(\xi = x_i) \cdot \mathrm{P}(\eta = y_j)$$

при всіх i та j.

Знайдемо розподіл випадкового вектора $(\xi; \eta)$, елементи якого є незалежними випадковими величинами. Нехай $F_\xi(x)$ і $F_\eta(y)$ — функції розподілів випадкових величин ξ і η, відповідно. Для функції розподілу $F(x, y)$ випадкового вектора $(\xi; \eta)$

$$F(x, y) = \mathrm{P}(\xi < x, \eta < y) = \mathrm{P}(\xi < x) \cdot \mathrm{P}(\eta < y) = \\ = F_\xi(x) \cdot F_\eta(y).$$

Звідси — у випадку неперервних випадкових величин ξ і η — щільність розподілу випадкового вектора $(\xi; \eta)$ має вигляд

$$f(x, y) = \frac{\partial^2 F(x, y)}{\partial x \partial y} = f_\xi(x) \cdot f_\eta(y),$$

де $f_\xi(x)$, $f_\eta(y)$ — щільності розподілів випадкових величин ξ і η, відповідно.

Якщо випадкові величини ξ і η дискретні і мають розподіли (x_i, p_i) та (y_j, q_j), відповідно, то випадковий вектор $(\xi; \eta)$ має розподіл $(x_i, y_j; p_{ij})$, де $p_{ij} = p_i \cdot q_j$ при всіх можливих значеннях i та j.

2.2.3 Композиція розподілів

Нехай ξ і η незалежні випадкові величини. Знайдемо розподіл їх суми, який будемо називати композицією розподілів. Припустимо, що випадкові величини ξ і η неперервні і мають щільності розподілу $f_\xi(x)$, $f_\eta(x)$, відповідно. Знайдемо щільність розподілу випадкової величини $\psi = \xi + \eta$.

Відомо, що щільність розподілу випадкового вектора $(\xi; \eta)$ має вигляд $f(x, y) = f_\xi(x) \cdot f_\eta(y)$. Тому, функція розподілу випадкової величини ψ

$$F_\psi(x) = \iint\limits_{s+t<x} f(s,t)dsdt = \int\limits_{-\infty}^{+\infty} ds \int\limits_{-\infty}^{x-s} f(s,t)dt = $$
$$= \int\limits_{-\infty}^{+\infty} ds \int\limits_{-\infty}^{x-s} f_\xi(s) f_\eta(t)dt$$

Звідси, обчисливши похідну від $F_\psi(x)$, одержимо

$$f_\psi(x) = \int_{-\infty}^{+\infty} f_\xi(s) f_\eta(x-s)ds. \tag{2.5}$$

Зауваження. Інтеграл в правій частині формули (2.5) називається згорткою інтегровних на $(-\infty; +\infty)$ функцій $f_\xi(x)$, $f_\eta(x)$.

Отже, щільність розподілу композиції розподілів є згорткою щільностей розподілів випадкових величин, що утворюють композицію.

Якщо ж випадкові величини ξ і η залежні, то, аналогічно попередньому,

$$F_{\psi}(x) = \iint\limits_{s+t<x} f(s,t)dsdt = \int\limits_{-\infty}^{+\infty} ds \int\limits_{-\infty}^{x-s} f(s,t)dt$$

і $f_{\psi}(x) = \int\limits_{-\infty}^{+\infty} f(s, x-s)ds.$

2.3 Числові характеристики випадкових величин

Нехай задано ймовірнісний простір $(\Omega,\ \mathcal{F},\ \mathrm{P})$. Будемо вважати, що всі випадкові величини задані на цьому просторі.

2.3.1 Математичне сподівання

Означення. Математичним сподіванням випадкової величини ξ називається число $\mathrm{M}\xi = \sum\limits_{i} x_i p_i$, якщо ξ — дискретна випадкова величина з розподілом $(x_i; p_i)$ чи $\mathrm{M}\xi = \int\limits_{-\infty}^{\infty} xf(x)dx$, якщо ξ — неперервна випадкова величина з щільністю розподілу $f(x)$.

Зауваження.

1. Ряд чи інтеграл з означення математичного сподівання може і розбігатися. Для таких випадкових величин

математичного сподівання не існує. Надалі, якщо будемо говорити про математичне сподівання випадкової величини, то вважатимемо, що воно існує.

2. У наведеному тут означенні, поняття математичного сподівання вводиться лише для дискретних та неперервних випадкових величин. В загальному випадку математичне сподівання випадкової величини ξ (не обов'язково дискретної чи неперервної) можна визначити у вигляді інтеграла Стільтьєса

$$\mathrm{M}\xi = \int_{-\infty}^{\infty} xF(dx),$$

де $F(x)$ — функція розподілу випадкової величини ξ.

Розглянемо основні властивості математичного сподівання.

1. Для будь-якого числа $C \in \mathbb{R}$

$$\mathrm{M}(C \cdot \xi) = C \cdot \mathrm{M}\xi \text{ і } \mathrm{M}C = C.$$

2. Для двох випадкових величин ξ і η

$$\mathrm{M}(\xi + \eta) = \mathrm{M}\xi + \mathrm{M}\eta.$$

3. Для двох незалежних випадкових величин ξ і η

$$\mathrm{M}(\xi \cdot \eta) = \mathrm{M}\xi \cdot \mathrm{M}\eta.$$

4. Для функцій $g(x)$ і $h(x, y)$ таких, що $g(\xi)$, $h(\xi, \eta)$ — випадкові величини

$$\mathrm{M}g(\xi) = \int_{-\infty}^{+\infty} g(x)f(x)dx \quad \left(= \sum_k g(x_k)p_k\right)$$

$$\mathrm{M}h(\xi,\eta) = \int\limits_{-\infty}^{+\infty} dx \int\limits_{-\infty}^{+\infty} h(x,y)f(x,y)dy$$
$$\left(= \sum_k \sum_j g(x_k, y_j)p_{kj} \right),$$

де $f(x)$ — щільність розподілу ($(x_k; p_k)$ — розподіл) випадкової величини ξ, $f(x,y)$ ($(x_k, y_j; p_{kj})$ — розподіл) випадкового вектора $(\xi;\eta)$.

5. Якщо $\mathrm{P}(\xi \geq 0) = 1$, то $\mathrm{M}\xi \geq 0$.

6. Якщо $\mathrm{P}(\xi \geq 0) = 1$ і $\mathrm{M}\xi = 0$, то $\mathrm{P}(\xi = 0) = 1$.

◀ Доведемо спочатку властивість 4). Для простоти обмежимось дискретним випадком. Якщо не об'єднувати однакові значення $g(x_k)$ і $h(x_k, y_j)$, то випадкові величини $g(\xi)$ та $h(\xi,\eta)$ мають такі розділи $(g(x_k), p_k)$ і $(h(x_k, y_j), p_{kj})$, відповідно . Тому, очевидно,
$\mathrm{M}g(\xi) = \sum\limits_k g(x_k) \cdot p_k$ і $\mathrm{M}h(\xi,\eta) = \sum\limits_k \sum\limits_j h(x_k, y_j) \cdot p_{kj}$,
що і потрібно було довести.

Після цього доведення 1) є, наприклад, таким

$$\mathrm{M}(C\cdot\xi) = \sum_k Cx_k p_k = C\sum_k x_k p_k = C\cdot \mathrm{M}\xi \text{ і}$$
$$\mathrm{M}C = C\cdot 1 = C.$$

Так само в 2)

$$\mathrm{M}(\xi+\eta) =$$
$$= \sum_k \sum_j (x_k + y_j)p_{kj} = \sum_k \sum_j x_k p_{kj} + \sum_k \sum_j y_j p_{kj} =$$
$$= \sum_k x_k p_{k\cdot} + \sum_j y_j p_{\cdot j} = \mathrm{M}\xi + \mathrm{M}\eta.$$

Тут, як і раніше, $(x_k; p_{k\cdot})$, $(y_j; p_{\cdot j})$ — розподіли випадкових величин ξ та η, відповідно .

Для доведення 3) згадаємо, що коли ξ та η незалежні, то $p_{kj} = p_{k\cdot} \cdot p_{\cdot j}$. Тому

$$\mathrm{M}(\xi \cdot \eta) = \sum_k \sum_j x_k y_j p_{kj} =$$
$$= \sum_k \sum_j x_k y_j p_{k\cdot} p_{\cdot j} = \sum_k x_k p_{k\cdot} \sum_j y_j p_{\cdot j} = \mathrm{M}\xi \cdot \mathrm{M}\eta.$$

Властивості 5) і 6) очевидні, бо всі $x_k \geq 0$ і в сумі $\sum\limits_k x_k p_k$ всі доданки невід'ємні.

Аналогічно властивості 1) — 6) можна довести і для неперервного випадку. ▶

Зауваження. Враховуючи, що ймовірності p_k є деякою вагою подій $\{\xi = x_k\}$ і $\sum\limits_k p_k = 1$, можна назвати математичне сподівання зваженим середнім значенням випадкової величини або середнім значенням випадкової величини.

2.3.2 Дисперсія та середньоквадратичне відхилення

Означення. Дисперсією випадкової величини ξ називається число $\mathrm{D}\xi = \mathrm{M}(\xi - \mathrm{M}\xi)^2$.

Зауважимо, що дисперсію можуть мати тільки ті випадкові величини, що мають математичне сподівання. Далі в цьому пункті будемо розглядати випадкові величини, які мають дисперсію.

З означення випливають такі найпростіші властивості дисперсії випадкової величини.

1. Для будь-якої випадкової величини ξ

$$\mathrm{D}\xi \geq 0 \text{ і } \mathrm{D}\xi = \mathrm{M}\xi^2 - (\mathrm{M}\xi)^2.$$

2. Дисперсія невипадкової (сталої) величини

$$\mathrm{D}C = 0.$$

3. Для будь-якої випадкової величини ξ

$$\mathrm{D}(C \cdot \xi) = C^2\mathrm{D}\xi.$$

4. Для незалежних випадкових величин ξ та η

$$\mathrm{D}(\xi + \eta) = \mathrm{D}\xi + \mathrm{D}\eta.$$

5. Для випадкової величини ξ з щільністю розподілу $f(x)$ чи розподілом $(x_k; p_k)$,

$$\mathrm{D}\xi = \int\limits_{-\infty}^{+\infty} (x - \mathrm{M}\xi)^2 f(x)dx = \int\limits_{-\infty}^{+\infty} x^2 f(x)dx - (\mathrm{M}\xi)^2$$

чи

$$\mathrm{D}\xi = \sum_k (x_k - \mathrm{M}\xi)^2 p_k = \sum_k x_k^2 p_k - (\mathrm{M}\xi)^2.$$

◀ Властивість 1) легко випливає з властивостей 1), 2) і 5) математичного сподівання. Дійсно, оскільки $(\xi - \mathrm{M}\xi)^2 \geq 0$ з ймовірністю 1, то

$$\mathrm{D}\xi = \mathrm{M}(\xi - \mathrm{M}\xi)^2 \geq 0.$$

Крім того,

$$\mathrm{D}\xi = \mathrm{M}(\xi - \mathrm{M}\xi)^2 = \mathrm{M}(\xi^2 - 2\cdot\xi\mathrm{M}\xi + (\mathrm{M}\xi)^2) =$$
$$= \mathrm{M}\xi^2 - 2\cdot\mathrm{M}\xi\cdot\mathrm{M}\xi + (\mathrm{M}\xi)^2 = \mathrm{M}\xi^2 - (\mathrm{M}\xi)^2.$$

Далі, враховуючи, що $\mathrm{M}C = C$, одержимо

$$\mathrm{D}C = \mathrm{M}(C - C)^2 = \mathrm{M}0 = 0$$

(властивість 2)). Записавши

$$\mathrm{D}(C\cdot\xi) = \mathrm{M}(C\cdot\xi - \mathrm{M}(C\cdot\xi))^2 =$$
$$= \mathrm{M}(C\cdot\xi - C\cdot\mathrm{M}\xi)^2 = \mathrm{M}C^2(\xi - \mathrm{M}\xi)^2 =$$
$$= C^2\mathrm{M}(\xi - \mathrm{M}\xi)^2 = C^2\mathrm{D}\xi,$$

одержимо властивість 3).

Якщо випадкові величини ξ і η незалежні, то $\mathrm{M}(\xi\cdot\eta) = \mathrm{M}\xi\cdot\mathrm{M}\eta$. Тому

$$\mathrm{D}(\xi+\eta) = \mathrm{M}(\xi+\eta)^2 - (\mathrm{M}(\xi+\eta))^2 =$$
$$= \mathrm{M}(\xi^2 + 2\xi\eta + \eta^2) - (\mathrm{M}\xi + \mathrm{M}\eta)^2 =$$
$$= \mathrm{M}\xi^2 + 2\mathrm{M}(\xi\cdot\eta) + \mathrm{M}\eta^2 - 2\mathrm{M}\xi\mathrm{M}\eta - (\mathrm{M}\eta)^2 =$$
$$= \mathrm{M}\xi^2 - (\mathrm{M}\xi)^2 + \mathrm{M}\eta^2 - (\mathrm{M}\eta)^2 = \mathrm{D}\xi + \mathrm{D}\eta,$$

що і стверджувалось в 4).

Властивість 5) повністю випливає з властивості 4) математичного сподівання. ▶

Означення. Середньоквадратичним відхиленням випадкової величини ξ називається число $\sigma(\xi) = \sqrt{\mathrm{D}\xi}$.

Зауваження.

1. Властивість 1) дисперсії забезпечує існувавння середньоквадратичного відхилення у всіх випадкових величин, що мають дисперсію.

2. Оскільки математичне сподівання є середнім значенням випадкової величини, то дисперсія є середнім квадрату відхилення випадкової величини від свого середнього значення. Отже, дисперсія, а тому і середньоквадратичне відхилення, є характеристиками розкиду (розсіювання) значень випадкової величини відносно її середнього значення.

Приклад. Випадкова величина ξ задана розподілом

x_k	1	2	3
p_k	$\frac{1}{4}$	$\frac{1}{4}$	$\frac{1}{2}$

.

Знайти $\mathrm{M}\xi$, $\mathrm{D}\xi$, $\sigma(\xi)$.

Проведемо обчислення $\mathrm{M}\xi = 1\cdot\frac{1}{4}+2\cdot\frac{1}{4}+3\cdot\frac{1}{2} = \frac{9}{4}$; $\mathrm{D}\xi = \mathrm{M}\xi^2-(\mathrm{M}\xi)^2 = 1\cdot\frac{1}{4}+4\cdot\frac{1}{4}+9\cdot\frac{1}{2}-\frac{81}{16} = \frac{11}{16}$; $\sigma(\xi) = \sqrt{\mathrm{D}\xi} = \frac{\sqrt{11}}{4}$.

2.3.3 Характеристики випадкових векторів

Як і для випадкових величин для випадкових векторів можна ввести поняття математичного сподівання.

Означення. Вектор $(\mathrm{M}\xi; \mathrm{M}\eta)$ називається *математичним сподіванням* випадкового вектора $(\xi;\eta)$.

Аналогом дисперсії випадкової величини для випадкового вектора є коваріаційна матриця.

Означення. Коваріацією випадкових величин ξ і η називається число $\mathrm{cov}(\xi,\eta) = \mathrm{M}(\xi-\mathrm{M}\xi)(\eta-\mathrm{M}\eta)$. Коефіцієнтом кореляції двох випадкових величин ξ і η називається число $r(\xi,\eta) = \frac{\mathrm{cov}(\xi,\eta)}{\sigma(\xi)\cdot\sigma(\eta)}$.

Означення. Коваріаційною матрицею випадкового вектора $(\xi;\eta)$ називається матриця

$$K = \begin{pmatrix} \mathrm{D}\xi & \mathrm{cov}(\xi,\eta) \\ \mathrm{cov}(\xi,\eta) & \mathrm{D}\eta \end{pmatrix}.$$

Розглянемо деякі властивості коефіцієнта кореляції (коваріації) випадкових величин

1. $\operatorname{cov}(\xi, \eta) = \operatorname{cov}(\eta, \xi)$, $r(\xi, \eta) = r(\eta, \xi)$.

2. $|\operatorname{cov}(\xi, \eta)| \leq \sigma(\xi) \cdot \sigma(\eta)$, $|r(\xi, \eta)| \leq 1$.

3. Випадкові величини ξ і η лінійно залежні (тобто $\eta = a \cdot \xi + b$) тоді і тільки тоді, коли $|r(\xi, \eta)| = 1$.

4. Якщо випадкові величини ξ та η незалежні, то $r(\xi, \eta) = 0$.

◀ Властивість 1) елементарно випливає з означення. Для доведення властивості 2) розглянемо

$$0 \leq \mathrm{M}(\eta - \mathrm{M}\eta - t(\xi - \mathrm{M}\xi))^2 = t^2 \mathrm{D}\xi - 2t\operatorname{cov}(\xi, \eta) + \mathrm{D}\eta.$$

Тому дискримінант одержаного квадратного тричлена $D \leq 0$, тобто,

$$4\operatorname{cov}^2(\xi, \eta) - 4\mathrm{D}\xi\mathrm{D}\eta \leq 0 \text{ або } \operatorname{cov}^2(\xi, \eta) - \mathrm{D}\xi\mathrm{D}\eta \leq 0$$

і тому $|\operatorname{cov}(\xi, \eta)| \leq \sigma(\xi)\sigma(\eta)$. Звідси

$$|r(\xi, \eta)| = \frac{|\operatorname{cov}(\xi, \eta)|}{\sigma(\xi) \cdot \sigma(\eta)} \leq 1.$$

Розглянутий квадратний тричлен має корінь тоді і тільки тоді, коли $D = 0$, тобто, $|r(\xi, \eta)| = 1$. Тому

$$\mathrm{M}(\eta - \mathrm{M}\eta - a(\xi - \mathrm{M}\xi))^2 = 0$$

і, отже, з ймовірністю 1

$$\eta - \mathrm{M}\eta - a(\xi - \mathrm{M}\xi) = 0.$$

Позначивши $\mathrm{M}\eta - a\mathrm{M}\xi = b$, одержимо $\eta = a\xi + b$, що і становить зміст властивості 3).

Властивість 4) випливає з того, що

$$\operatorname{cov}(\xi, \eta) = \mathrm{M}(\xi\eta) - \mathrm{M}\xi\mathrm{M}\eta. \quad \blacktriangleright$$

Означення. Випадкові величини ξ і η називаються некорельованими, якщо $r(\xi, \eta) = 0$ $(\operatorname{cov}(\xi, \eta) = 0)$.

Зауваження.

1. Враховуючи, що $\mathrm{D}\xi = \operatorname{cov}(\xi, \xi)$ і так само $\mathrm{D}\eta = \operatorname{cov}(\eta, \eta)$, коваріаційну матрицю випадкового вектора $(\xi; \eta)$ можна записати у вигляді

$$K = \begin{pmatrix} \operatorname{cov}(\xi, \xi) & \operatorname{cov}(\xi, \eta) \\ \operatorname{cov}(\eta, \xi) & \operatorname{cov}(\eta, \eta) \end{pmatrix},$$

що оправдує її назву.

2. Якщо ж $|r(\xi, \eta)|$ близький до 1, то кажуть про майже лінійну залежність випадкових величин ξ і η, а коли близький до нуля, то говорять про слабку залежність (близькість до некорельованості) цих випадкових величин.

3. Взагалі кажучи, із некорельованості двох випадкових величин не випливає їх незалежність. Це підтверджує такий приклад.

Приклад. Нехай випадкова величина ξ задана щільністю розподілу

$$f(x) = \begin{cases} \frac{1}{2}, & \text{якщо } x \in [-1; 1] \\ 0, & \text{якщо } x \notin [-1; 1] \end{cases},$$

а випадкова величина $\eta = \xi^2$. Тоді

$$\operatorname{cov}(\xi, \eta) = \mathrm{M}(\xi\eta) - \mathrm{M}\xi\mathrm{M}\eta = \mathrm{M}\xi^3 - \mathrm{M}\xi\mathrm{M}\xi^2 =$$
$$= \tfrac{1}{2}\int\limits_{-1}^{1} x^3 dx - \tfrac{1}{2}\int\limits_{-1}^{1} x dx = 0$$

2.3.4 Інші характеристики випадкових величин

Дамо означення деяких інших характеристик випадкової величини, що дещо відрізняються від розглянутих вище.

Означення. Модою випадкової величини ξ називається число d, що є точкою максимуму щільності розподілу випадкової величини ξ, якщо вона неперервна, чи є найбільшймовірним можливим значенням випадкової величини ξ, якщо вона дискретна.

Мода випадкової величини може бути єдиною (унімодальний розподіл) чи мати кілька значень (мультимодальний розподіл).

Означення. Медіаною випадкової величини ξ з функцією розподілу $F(x)$ називається значення h, для якого $F(h) \leq \frac{1}{2}$, $F(h+0) \geq \frac{1}{2}$.

Для неперервної випадкової величини медіана може бути визначена як таке значення h, для якого $F(h) = \frac{1}{2}$. Зауважимо, що медіана є значенням, яке "ділить"розподіл на дві рівні (по «вазі») частини і тому є аналогом математичного сподівання для тих розподілів, для яких математичне сподівання не існує.

Означення. Квантилем порядку α, $\alpha \in (0;1)$ випадкової величини (розподілу) з функцією розподілу $F(x)$ називається число p_α, для якого $F(p_\alpha) \leq \alpha$, $F(p_\alpha + 0) \geq \alpha$.

Очевидно, що медіана є квантилем порядку $\frac{1}{2}$. Якщо ж випадкова величина неперервна, то квантиль порядку α є розв'язком рівняння $F(p_\alpha) = \alpha$.

Розглянемо на завершення ще кілька характеристик випадкової величини (моментні характеристики).

Означення. Моментом (початковим моментом) порядку k випадкової величини ξ називається число $\mathrm{M}\xi^k$.

Означення. Центральним моментом порядку k випадкової величини ξ називається число $\mathrm{M}(\xi - \mathrm{M}\xi)^k$.

Зауважимо, що математичне сподівання є початковим моментом першого порядку, а дисперсія є центральним моментом другого порядку.

З моментними характеристиками пов'язані такі характеристики, як коефіцієнт асиметрії $\frac{\mathrm{M}(\xi-\mathrm{M}\xi)^3}{\sqrt{(\mathrm{D}\xi)^3}}$ (характеризує асиметрію в розподілі випадкової величини відносно її математичного сподівання), коефіцієнт ексцесу $\frac{\mathrm{M}(\xi-\mathrm{M}\xi)^4}{(\mathrm{D}\xi)^2}-3$ (характеризує ступінь «крутості» щільності розподілу випадкової величини).

2.4 Приклади розподілів випадкових величин

2.4.1 Біноміальний розподіл

Нехай ξ — кількість успіхів в схемі Бернуллі з n випробувань. Відомо, що

$$\mathrm{P}(\xi = k) = C_n^k p^k (1-p)^{n-k},$$

де $k = 0, 1, \ldots, n$, $0 < p < 1$.

Означення. Розподіл $(x_k; p_k)$, $k = 0, 1, \ldots, n$, де $x_k = k$, $p_k = C_n^k p^k (1-p)^{n-k}$ називається біноміальним з параметрами n і p, а випадкова величина з цим розподілом — біноміально розподіленою.

Переконаємось, що набір чисел p_k дійсно задає розподіл ймовірностей. Для цього досить встановити, що $p_k \geq 0$ (це очевидно) і $\sum\limits_{k=0}^{n} p_k = 1$.

Дійсно

$$\sum_{k=0}^{n} p_k = \sum_{k=0}^{n} C_n^k p^k (1-p)^{n-k} = (p + 1 - p)^n = 1.$$

Знайдемо тепер математичне сподівання та дисперсію біноміального розподілу. Для цього розглянемо набір випадкових величин X_k — індикатор успіху в k-ому випробуванні схеми Бернуллі ($k = 1, 2, \ldots, n$).

$$X_k = \begin{cases} 1, & \text{якщо в } k\text{-ому випробуванні відбувся успіх,} \\ 0, & \text{в іншому випадку.} \end{cases}$$

Всі випадкові величини X_k однаково розподілені з розподілом $(1; p)$, $(0; 1-p)$ та попарно незалежні.

Оскільки $\xi = \sum\limits_{k=0}^{n} X_k$ і

$$\begin{gathered} \mathrm{M}X_k = 1 \cdot p + 0 \cdot (1-p) = p, \\ \mathrm{D}X_k = \mathrm{M}X_k^2 - (\mathrm{M}X_k)^2 = \\ = 1 \cdot p + 0 \cdot (1-p) - p^2 = p(1-p), \end{gathered}$$

то

$$\mathrm{M}\xi = \sum_{k=0}^{n} \mathrm{M}X_k = np, \quad \mathrm{D}\xi = \sum_{k=0}^{n} \mathrm{D}X_k = np(1-p).$$

2.4.2 Розподіл Пуассона

Означення. Розподіл (x_k, p_k), $k = 0, 1, 2, \ldots$, де $x_k = k$, $p_k = \frac{\lambda^k}{k!} e^{-\lambda}$, $\lambda > 0$, називається розподілом Пуассона з параметром λ.

З розподілом Пуассона ми вже зустрічались при вивченні схеми Бернуллі. Цей розподіл згідно теореми Пуассона може бути одержаний з біноміального розподілу шляхом граничного переходу при $n \to +\infty$.

Очевидно, що $p_k > 0$ при всіх $k \geq 0$. Крім того

$$\sum_{k=0}^{\infty} p_k = \sum_{k=0}^{\infty} \frac{\lambda^k}{k!} e^{-\lambda} = e^{\lambda} \cdot e^{-\lambda} = 1.$$

Тому набір чисел p_k дійсно може бути розподілом ймовірностей.

Математичне сподівання розподіленої за законом Пуассона випадкової величини ξ

$$\mathrm{M}\xi = \sum_{k=0}^{\infty} k \tfrac{\lambda^k}{k!} e^{-\lambda} = \lambda \sum_{k=1}^{\infty} \tfrac{\lambda^{k-1}}{(k-1)!} e^{-\lambda} = \\ = \lambda e^{\lambda} \cdot e^{-\lambda} = \lambda.$$

Дисперсія

$$\mathrm{D}\xi = \mathrm{M}\xi^2 - (\mathrm{M}\xi)^2 = \sum_{k=0}^{\infty} k^2 \tfrac{\lambda^k}{k!} e^{-\lambda} - \lambda^2 = \\ = \sum_{k=1}^{\infty} (k - 1 + 1) \tfrac{\lambda^k}{(k-1)!} e^{-\lambda} - \lambda^2 = \\ = \sum_{k=2}^{\infty} \tfrac{\lambda^k}{(k-2)!} e^{-\lambda} + \sum_{k=1}^{\infty} \tfrac{\lambda^k}{(k-1)!} e^{-\lambda} - \lambda^2 = \\ = \lambda^2 e^{\lambda} \cdot e^{-\lambda} + \lambda e^{\lambda} \cdot e^{-\lambda} - \lambda^2 = \lambda.$$

Таким чином параметр λ розподілу Пуассона є одночасно математичним сподіванням та дисперсією випадкової величини, що має цей розподіл.

Розглянемо одну цікаву властивість розподілу Пуассона: композиція двох розподілів Пуассона з параметрами λ_1 і λ_2 є розподілом Пуассона з параметром $\lambda_1 + \lambda_2$. Дійсно, нехай незалежні випадкові величини ξ і η мають розподіл Пуассона з параметрами λ_1 і λ_2, відповідно. За формулою повної ймовірності

$$\mathrm{P}(\xi+\eta=k)=\sum_{i=0}^{k}\mathrm{P}(\xi=i)\cdot\mathrm{P}(\eta=k-i)=$$
$$=\sum_{i=0}^{k}\tfrac{\lambda_1^i}{i!}e^{-\lambda_1}\tfrac{\lambda_2^{k-i}}{(k-i)!}e^{-\lambda_2}=\tfrac{1}{k!}\sum_{i=0}^{k}C_k^i\lambda_1^i\lambda_2^{k-i}e^{-(\lambda_1+\lambda_2)}=$$
$$=\tfrac{(\lambda_1+\lambda_2)^k}{k!}e^{-(\lambda_1+\lambda_2)},$$

що і треба було довести.

2.4.3 Геометричний розподіл

Розглянемо нескінченну кількість незалежних випробувань з ймовірністю успіху в кожному випробуавнні p. Нехай ξ — кількість випробувань до першого успіху.

Тоді $\mathrm{P}(\xi=k)=p(1-p)^k$.

Означення. Розподіл $(x_k;p_k)$, $k=0,1,2,\dots$, де $x_k=k$, $p_k=p(1-p)^k$, називається геометричним розподілом з параметром p.

Перевіримо характерні властивості розподілу

1. $p_k>0$,

2. $\sum\limits_{k=0}^{\infty}p_k=\sum\limits_{k=0}^{\infty}p(1-p)^k=\frac{p}{1-(1-p)}=1$, бо послідовність p_k

утворює нескінченну спадну геометричну прогресію із знаменником $q = 1 - p$.

Для обчислення числових характеристик випадкової величини ξ розглянемо функцію

$$f(x) = \sum_{k=0}^{\infty} x^k = \frac{1}{1-x}, \quad |x| < 1.$$

Тоді

$$\mathrm{M}\xi = \sum_{k=0}^{\infty} kp(1-p)^k = p(1-p)\sum_{k=1}^{\infty} k(1-p)^{k-1} =$$
$$= p(1-p)f'(1-p) = p(1-p)\tfrac{1}{(1-(1-p))^2} = \tfrac{1-p}{p}.$$
$$\mathrm{D}\xi = \mathrm{M}\xi^2 - (\mathrm{M}\xi)^2 = \sum_{k=0}^{\infty} k^2 p(1-p)^k - \tfrac{(1-p)^2}{p^2} =$$
$$= p(1-p)^2 \sum_{k=2}^{\infty} k(k-1)(1-p)^{k-2} +$$
$$+ \sum_{k=0}^{\infty} kp(1-p)^k - \tfrac{(1-p)^2}{p^2} =$$

$$= p(1-p)^2 f''(1-p) + \tfrac{1-p}{p} - \tfrac{(1-p)^2}{p^2} =$$
$$= p(1-p)^2 \tfrac{2}{(1-(1-p))^3} + \tfrac{1-p}{p} - \tfrac{(1-p)^2}{p^2} = \tfrac{1-p}{p^2}.$$

2.4.4 Рівномірний розподіл

Розглянемо неперервну випадкову величину ξ, яка може приймати будь-яке значення з відрізка $[a; b]$. Нехай ймовірність того, що ξ попаде в довільний інтервал, який повністю міститься в $[a; b]$, залежить тільки від довжини цього інтервала. Зрозуміло, що щільність розподілу цієї випадкової величини повинна мати вигляд

$$f(x) = \begin{cases} c, & \text{якщо } x \in [a; b], \\ 0, & \text{якщо } x \notin [a; b]. \end{cases}$$

З умови $\int\limits_{-\infty}^{+\infty} f(x)dx = 1$ знайдемо параметр c:

$$\int\limits_{-\infty}^{+\infty} f(x)dx = \int\limits_{a}^{b} cdx = c(b-a),$$

тому $c = \frac{1}{b-a}$.

Означення. Випадкова величина ξ має рівномірний розподіл на відрізку $[a; b]$, якщо її щільність розподілу

$$f(x) = \begin{cases} \frac{1}{b-a}, & \text{якщо } x \in [a; b], \\ 0, & \text{якщо } x \notin [a; b]. \end{cases}$$

Знайдемо числові характеристики рівномірно розподіленої випадкової величини.

Математичне сподівання

$$\mathrm{M}\xi = \int\limits_{-\infty}^{+\infty} xf(x)dx = \int\limits_{a}^{b} x\tfrac{1}{b-a}dx =$$
$$= \tfrac{b^2-a^2}{2} \cdot \tfrac{1}{b-a} = \tfrac{a+b}{2}.$$

Дисперсія

$$\mathrm{D}\xi = \mathrm{M}\xi^2 - (\mathrm{M}\xi)^2 = \int\limits_{-\infty}^{+\infty} x^2 f(x)dx - (\mathrm{M}\xi)^2 =$$
$$= \int\limits_{a}^{b} x^2 \tfrac{1}{b-a}dx - \tfrac{(a+b)^2}{4} =$$
$$= \tfrac{b^3-a^3}{3} \cdot \tfrac{1}{b-a} - \tfrac{(a+b)^2}{4} = \tfrac{(a-b)^2}{12}.$$

На завершення розглянемо спосіб моделювання розподілів маючи рівномірний розподіл.

Нехай випадкова величина η має неперервну зростаючу функцію розподілу $F(x)$. Розглянемо випадкову величину $\xi = F(\eta)$

$$\mathrm{P}(\xi < x) = \mathrm{P}(F(\eta) < x) =$$
$$= \begin{cases} 0, & x \leq 0 \\ \mathrm{P}(\eta < F^{-1}(x)), & 0 < x \leq 1 \\ 1, & x > 1 \end{cases} =$$
$$= \begin{cases} 0, & x \leq 0 \\ F(F^{-1}(x)), & 0 < x \leq 1 \\ 1, & x > 1 \end{cases} = \begin{cases} 0, & x \leq 0 \\ x, & 0 < x \leq 1 \\ 1, & x > 1 \end{cases}.$$

Тому щільність розподілу випадкової величини ξ

$$f(x) = \begin{cases} 1, & x \in [0;1] \\ 0, & x \notin [0;1] \end{cases},$$

тобто, ξ рівномірно розподілена на $[0;1]$.

Звідси випливає спосіб моделювання будь-якого неперервного розподілу, що задається функцією розподілу на $F(x)$. Досить взяти рівномірно розподілену на відрізку $[0;1]$ випадкову величину ξ і розглянути випадкову величину $\eta = F^{-1}(\xi)$. Вона матиме функцію розподілу $F(x)$.

2.4.5 Показниковий розподіл

Означення. Випадкова величина ξ має показниковий (експоненціальний) розподіл з параметром λ, якщо її щільність розподілу

$$f(x) = \begin{cases} \lambda e^{-\lambda x}, & x \geq 0, \\ 0, & x < 0. \end{cases}$$

Перевіримо, що функція $f(x)$ може бути щільністю розподілу:

1. $f(x) > 0$;
2. $\int\limits_{-\infty}^{+\infty} f(x)dx = \int\limits_{0}^{+\infty} \lambda e^{-\lambda x}dx = -e^{-\lambda x}\Big|_0^{+\infty} = 1.$

Математичне сподівання

$$\mathrm{M}\xi = \int\limits_{-\infty}^{+\infty} xf(x)dx = \int\limits_{0}^{+\infty} x\lambda e^{-\lambda x}dx =$$
$$= -xe^{-\lambda x}\Big|_0^{+\infty} + \int\limits_{0}^{+\infty} e^{-\lambda x}dx = -\tfrac{1}{\lambda}e^{-\lambda x}\Big|_0^{+\infty} = \tfrac{1}{\lambda}.$$

Дисперсія

$$\mathrm{D}\xi = \mathrm{M}\xi^2 - (\mathrm{M}\xi)^2 = \int\limits_{-\infty}^{+\infty} x^2 f(x)dx - (\mathrm{M}\xi)^2 =$$
$$= \int\limits_{0}^{+\infty} x^2 \lambda e^{-\lambda x}dx - \tfrac{1}{\lambda^2} =$$
$$= -x^2 e^{-\lambda x}\Big|_0^{+\infty} + 2\int\limits_{0}^{+\infty} x\lambda e^{-\lambda x}dx - \tfrac{1}{\lambda^2} =$$
$$= \tfrac{2}{\lambda^2} - \tfrac{1}{\lambda^2} = \tfrac{1}{\lambda^2}.$$

Розглянемо на завершення

$$\mathrm{P}(\xi \in [k; k+1]) = \int\limits_{k}^{k+1} f(x)dx = \int\limits_{k}^{k+1} \lambda e^{-\lambda x}dx =$$
$$= -e^{-\lambda x}\Big|_k^{k+1} = (1 - e^{-\lambda}) \cdot e^{-k\lambda}.$$

Звідси видно, що показниковий розподіл є неперервним аналогом геометричного розподілу з параметром $p = 1 - e^{-\lambda}$.

2.4.6 Нормальний розподіл

Нормальний роподіл відіграє виключно важливу роль в теорії ймовірностей та її застосуваннях. Важливість цього розподілу пояснюється тим, що при досить широких припущеннях розподіл суми великої кількості випадкових величин виявляється близьким до нормального розполілу. Це має місце, коли на значення деякої величини впливає велика кількість рівноправних випадкових факторів.

Означення. Випадкова величина ξ має нормальний розподіл з параметрами a та σ^2 (позначається $N(a;\sigma^2)$), якщо її щільність розподілу

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left\{-\frac{(x-a)^2}{2\sigma^2}\right\}, \quad x \in \mathbb{R}.$$

Перш ніж доводити, що функція $f(x)$ може бути щільністю розподілу випадкової величини знайдемо (інтеграл Пуассона)

$$\mathcal{I} = \int\limits_{-\infty}^{+\infty} e^{-t^2/2} dt.$$

Зауважимо, між іншим, що невизначений інтеграл від функції $e^{-t^2/2}$ не виражається через елементарні функції.

Розглянемо

$$\mathcal{I}^2 = \int\limits_{-\infty}^{+\infty} e^{-t^2/2} dt \int\limits_{-\infty}^{+\infty} e^{-s^2/2} ds =$$

$$= \left| \begin{array}{c} s = \rho\cos\phi \\ t = \rho\sin\phi \end{array} \right| = \int\limits_{0}^{2\pi} d\phi \int\limits_{0}^{+\infty} e^{-\rho^2/2} \rho d\rho =$$

$$= 2\pi e^{-\rho^2/2} \Big|_{0}^{+\infty} = 2\pi.$$

Тому $\mathcal{I} = \sqrt{2\pi}$.

Звідси

$$\int\limits_{-\infty}^{+\infty} f(x)dx = \frac{1}{\sqrt{2\pi}\sigma} \int\limits_{-\infty}^{+\infty} \exp\left\{-\frac{(x-a)^2}{2\sigma^2}\right\} dx =$$
$$= \left| \begin{array}{c} \frac{x-a}{\sigma} = t \\ dx = \sigma dt \end{array} \right| = \frac{1}{\sqrt{2\pi}} \int\limits_{-\infty}^{+\infty} e^{-t^2/2} dt = 1$$

Крім того, $f(x) > 0$ при $x \in \mathbb{R}$, тому $f(x)$ може бути щільністю розподілу випадкової величини.

З'ясуємо суть параметрів нормального розподілу. Для цього знайдемо його математичне сподівання і дисперсію.

$$\mathrm{M}\xi = \int\limits_{-\infty}^{+\infty} x f(x)dx =$$
$$= \frac{1}{\sqrt{2\pi}\sigma} \int\limits_{-\infty}^{+\infty} x \exp\left\{-\frac{(x-a)^2}{2\sigma^2}\right\} dx =$$
$$= \frac{1}{\sqrt{2\pi}\sigma} \int\limits_{-\infty}^{+\infty} \frac{x-a}{\sigma} \exp\left\{-\frac{(x-a)^2}{2\sigma^2}\right\} dx + a =$$
$$= -\frac{\sigma}{\sqrt{2\pi}} \exp\left\{-\frac{(x-a)^2}{2\sigma^2}\right\}\Bigg|_{-\infty}^{+\infty} + a = a;$$
$$\mathrm{D}\xi = \int\limits_{-\infty}^{+\infty} (x - \mathrm{M}\xi)^2 f(x)dx =$$
$$= \frac{1}{\sqrt{2\pi}\sigma} \int\limits_{-\infty}^{+\infty} (x-a)^2 \exp\left\{-\frac{(x-a)^2}{2\sigma^2}\right\} dx =$$
$$= -\frac{(x-a)\sigma}{\sqrt{2\pi}} \exp\left\{-\frac{(x-a)^2}{2\sigma^2}\right\}\Bigg|_{-\infty}^{+\infty} +$$
$$+\sigma^2 \frac{1}{\sqrt{2\pi}\sigma} \int\limits_{-\infty}^{+\infty} \exp\left\{-\frac{(x-a)^2}{2\sigma^2}\right\} dx = \sigma^2$$

Отже, параметрами нормального розподілу $N(a;\sigma^2)$ є його математичне сподівання a та дисперсія σ^2.

Для багатьох практичних задач є важливим обчислення ймовірностей попадання нормально розподіленої випадкової величини в деякий інтервал

$$\mathrm{P}(\xi \in (\alpha;\beta)) = \int\limits_{\alpha}^{\beta} f(x)dx =$$

$$= \tfrac{1}{\sqrt{2\pi}\sigma} \int\limits_{\alpha}^{\beta} \exp\left\{-\tfrac{(x-a)^2}{2\sigma^2}\right\} dx = \left| \begin{array}{c} \frac{x-a}{\sigma} = t \\ dx = \sigma dt \end{array} \right| =$$

$$= \tfrac{1}{\sqrt{2\pi}} \int\limits_{\frac{\alpha-a}{\sigma}}^{\frac{\beta-a}{\sigma}} e^{-t^2/2}dt = \Phi(\tfrac{\beta-a}{\sigma}) - \Phi(\tfrac{\alpha-a}{\sigma}),$$

де $\Phi(x) = \frac{1}{\sqrt{2\pi}} \int\limits_0^x e^{-t^2/2}dt$ (функція Лапласа).

Функція Лапласа табульована, що дає можливість обчислювати розглянуті ймовірності.

Наведемо деякі властивості функції Лапласа:

1. $\lim\limits_{x\to+\infty} \Phi(x) = \frac{1}{2}$;
2. $\Phi(-x) = -\Phi(x)$;
3. $F_0(x) = \frac{1}{2} + \Phi(x)$ є функцією розподілу $N(0;1)$.

Властивість 1) випливає з властивості 3), яка є очевидною, бо щільність розподілу $N(0;1)$

$$f_0(x) = \frac{1}{\sqrt{2\pi}} e^{-t^2/2}$$

є парною функцією і

$$F_0(x) = \frac{1}{2} + \frac{1}{\sqrt{2\pi}} \int\limits_0^x e^{-t^2/2}dt.$$

Для доведення властивості 2) запишемо

$$\Phi(-x) = \tfrac{1}{\sqrt{2\pi}} \int\limits_0^{-x} e^{-t^2/2} dt = \left| \begin{array}{c} t = -u \\ dt = -du \end{array} \right| =$$
$$= -\tfrac{1}{\sqrt{2\pi}} \int\limits_0^{x} e^{-u^2/2} du = -\Phi(x).$$

Зауваження. Нормальний розподіл $N(0;1)$ ще називають стандартним нормальним розподілом. Його функцію розподілу іноді також можна зустріти табульованою. Властивість 2) для неї запишеться у вигляді

$$F_0(-x) = 1 - F_0(x).$$

На завершення розглянемо ще одну властивість нормального розподілу. Нехай ξ має нормальний розподіл $N(a;\sigma^2)$. Тоді випадкова величина $\eta = k\xi + b$ $(k \neq 0)$ також має нормальний розподіл $N(ka + b; k^2\sigma^2)$.

Дійсно, нехай для визначеності $k > 0$, тоді

$$\mathrm{P}(\eta < x) = \mathrm{P}(k\xi + b < x) = \mathrm{P}(\xi < \tfrac{x-b}{k}) =$$
$$= \tfrac{1}{\sqrt{2\pi}\sigma} \int\limits_{-\infty}^{\frac{x-b}{k}} \exp\left\{-\tfrac{(t-a)^2}{2\sigma^2}\right\} dt =$$
$$= \left| \begin{array}{c} \frac{t-a}{\sigma} = s \\ dt = \sigma ds \end{array} \right| = \tfrac{1}{\sqrt{2\pi}k\sigma} \int\limits_{-\infty}^{x} \exp\left\{-\tfrac{\left(\frac{s-b}{k} - a\right)^2}{2\sigma^2}\right\} ds =$$
$$= \tfrac{1}{\sqrt{2\pi}k\sigma} \int\limits_{-\infty}^{x} \exp\left\{-\tfrac{(s-(ka+b))^2}{2(\sigma k)^2}\right\} ds,$$

що є функцією розподілу $N(ka + b; k^2\sigma^2)$.

Отже, шляхом перетворення $\eta = \sigma\xi + a$ з випадкової величини ξ, що має стандартний нормальний розподіл, можна одержати випадкову величину η з розподілом $N(a;\sigma^2)$.

2.4.7 Двовимірний нормальний розподіл

Означення. Випадковий вектор $(\xi;\eta)$ називається нормально розподіленим $(N(a_1;a_2;\sigma_1^2;\sigma_2^2;\rho))$, якщо його щільність розподілу $(x,y\in\mathbb{R})$

$$f(x,y)=\tfrac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}}\times$$
$$\times\exp\left\{-\tfrac{1}{2(1-\rho^2)}\left(\tfrac{(x-a_1)^2}{\sigma_1^2}-\tfrac{2\rho(x-a_1)(y-a_2)}{\sigma_1\sigma_2}+\tfrac{(y-a_2)^2}{\sigma_2^2}\right)\right\},$$

якщо $\sigma_1^2\sigma_2^2(1-\rho^2)\neq 0$.

Спеціально не будемо перевіряти, що вказана функція є щільністю розподілу (хоч її невід'ємність очевидна). Знайдемо маргінальні розподіли випадкового вектора $(\xi;\eta)$.

$$f_\xi(x)=\int\limits_{-\infty}^{+\infty}f(x;y)dy=\tfrac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}}\times$$
$$\times\int\limits_{-\infty}^{+\infty}\exp\left\{-\tfrac{1}{2(1-\rho^2)}\left(\tfrac{(x-a_1)^2}{\sigma_1^2}-\tfrac{2\rho(x-a_1)(y-a_2)}{\sigma_1\sigma_2}+\right.\right.$$
$$\left.\left.+\tfrac{(y-a_2)^2}{\sigma_2^2}\right)\right\}dy=\tfrac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}}\times$$
$$\times\int\limits_{-\infty}^{+\infty}\exp\left\{-\tfrac{1}{2(1-\rho^2)}\left(\rho\tfrac{x-a_1}{\sigma_1}-\tfrac{y-a_2}{\sigma_2}\right)^2-\tfrac{(x-a_1)^2}{2\sigma_1^2}\right\}dy=$$
$$=\left|\begin{array}{c}\tfrac{1}{\sqrt{1-\rho^2}}\left(\rho\tfrac{x-a_1}{\sigma_1}-\tfrac{y-a_2}{\sigma_2}\right)=t\\ dy=-\sigma_2\sqrt{1-\rho^2}dt\end{array}\right|=$$
$$=\tfrac{-\sigma_2\sqrt{1-\rho^2}}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}}\int\limits_{-\infty}^{+\infty}e^{-t^2/2}dt\cdot\exp\left\{-\tfrac{(x-a_1)^2}{2\sigma_1^2}\right\}=$$
$$=\tfrac{1}{\sqrt{2\pi}\sigma_1}\exp\left\{-\tfrac{(x-a_1)^2}{2\sigma_1^2}\right\}.$$

Тобто ξ має нормальний розподіл $N(a_1;\sigma_1^2)$ і, аналогічно, η має нормальний розподіл $N(a_2;\sigma_2^2)$. Звідси, між іншим, випливає, що $f(x;y)$ є дійсно щільністю розподілу випадкового вектора. Крім того, параметри двовимірного нормального розподілу мають такий зміст $a_1 = \mathrm{M}\xi$, $a_2 = \mathrm{M}\eta$, $\sigma_1^2 = \mathrm{D}\xi$, $\sigma_2^2 = \mathrm{D}\eta$.

Знайдемо коефіцієнт кореляції між випадковими величинами ξ і η.

$$\mathrm{cov}(\xi,\eta) = \mathrm{M}(\xi - \mathrm{M}\xi)(\eta - \mathrm{M}\eta) =$$

$$= \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}} \int\limits_{-\infty}^{+\infty} dx \int\limits_{-\infty}^{+\infty} (x-a_1)(y-a_2) \times$$

$$\times \exp\left\{-\frac{1}{2(1-\rho^2)}\left(\frac{(x-a_1)^2}{\sigma_1^2} - \frac{2\rho(x-a_1)(y-a_2)}{\sigma_1\sigma_2} + \frac{(y-a_2)^2}{\sigma_2^2}\right)\right\} dy =$$

$$= \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}} \int\limits_{-\infty}^{+\infty} dx \int\limits_{-\infty}^{+\infty} (x-a_1)(y-a_2) \times$$

$$\times \exp\left\{-\frac{1}{2(1-\rho^2)}\left(\frac{x-a_1}{\sigma_1} - \rho\frac{y-a_2}{\sigma_2}\right)^2 - \frac{(y-a_2)^2}{2\sigma_2^2}\right\} dy =$$

$$= \left| \begin{array}{c} \frac{1}{\sqrt{1-\rho^2}}\left(\frac{x-a_1}{\sigma_1} - \rho\frac{y-a_2}{\sigma_2}\right) = t \\ dx = \sigma_1\sqrt{1-\rho^2}dt \end{array} \right| =$$

$$= \frac{1}{2\pi\sigma_2} \int\limits_{-\infty}^{+\infty} dt \int\limits_{-\infty}^{+\infty} \left(\sqrt{1-\rho^2}\sigma_1 t + \rho\frac{\sigma_1}{\sigma_2}(y-a_2)\right)(y-a_2) \times$$

$$\times \exp\left\{-\frac{t^2}{2} - \frac{(y-a_2)^2}{2\sigma_2^2}\right\} dy =$$

$$= \frac{\sqrt{1-\rho^2}\sigma_1}{2\pi\sigma_2} \int\limits_{-\infty}^{+\infty} te^{-t^2/2}dt \int\limits_{-\infty}^{+\infty} (y-a_2)\exp\left\{-\frac{(y-a_2)^2}{2\sigma_2^2}\right\} dy +$$

$$+\rho\frac{\sigma_1}{2\pi\sigma_2^2} \int\limits_{-\infty}^{+\infty} e^{-t^2/2}dt \int\limits_{-\infty}^{+\infty} (y-a_2)\exp\left\{-\frac{(y-a_2)^2}{2\sigma_2^2}\right\} dy =$$

$$= 0 + \rho\frac{\sigma_1}{2\pi\sigma_2^2}\sqrt{2\pi}\sigma_2^2\sqrt{2\pi}\sigma_2 = \rho\sigma_1\sigma_2.$$

Тому кофіцієнт кореляції

$$r(\xi, \eta) = \frac{\operatorname{cov}(\xi, \eta)}{\sigma(\xi)\sigma(\eta)} = \frac{\rho\sigma_1\sigma_2}{\sigma_1\sigma_2} = \rho$$

і, отже, параметр ρ двовимірного нормального розподілу є коефіцієнтом кореляції між його компонентами ξ та η.

Як відомо, коли $\rho = 0$, то випадкові величини ξ і η некорельовані. Тоді

$$f(x, y) = \tfrac{1}{2\pi\sigma_1\sigma_2} \exp\left\{-\tfrac{(x-a_1)^2}{2\sigma_1^2} - \tfrac{(y-a_2)^2}{2\sigma_2^2}\right\} = \\ = f_\xi(x) \cdot f_\eta(y),$$

що свідчить про незалежність випадкових величин ξ і η. Отже, для нормально розподілених випадкових величин поняття некорельованості та незалежності тотожні.

Зауваження. Нормальний розподіл на площині з параметром $\rho = 0$ називається канонічним. При цьому головний еліпс розсіювання цього розподілу

$$\tfrac{1}{2(1-\rho^2)} \left(\tfrac{(x-a_1)^2}{\sigma_1^2} - \tfrac{2\rho(x-a_1)(y-a_2)}{\sigma_1\sigma_2} + \tfrac{(y-a_2)^2}{\sigma_2^2} \right) = \lambda^2$$

має канонічне рівняння відносно змінних $x - a_1$ та $y - a_2$. Ймовірність попадання нормального випадкового вектора в основний еліпс розсіювання дорівнює $1 - e^{-\lambda^2}$.

Встановимо це для канонічного нормального розподілу. Ймовірність попадання випадкового вектора з цим розподілом в головний еліпс розсіювання

$$\frac{1}{2}\left(\frac{(x-a_1)^2}{\sigma_1^2} + \frac{(y-a_2)^2}{\sigma_2^2}\right) = \lambda^2$$

дорівнює

$$p_\lambda = \iint\limits_D \tfrac{1}{2\pi\sigma_1\sigma_2} \exp\left\{-\tfrac{(x-a_1)^2}{\sigma_1^2} - \tfrac{(y-a_2)^2}{\sigma_2^2}\right\} dxdy =$$

$$= \left| \begin{array}{c} x = a_1 + \sigma_1 r\cos\phi \\ y = a_2 + \sigma_2 r\sin\phi \end{array} \right| =$$

$$= \int\limits_0^{2\pi} d\phi \int\limits_0^{\sqrt{2}\lambda} \tfrac{1}{2\pi\sigma_1\sigma_2} e^{-r^2/2} r\sigma_1\sigma_2 dr =$$

$$= 2\pi \tfrac{1}{2\pi}\left(-e^{-r^2/2}\right)\Big|_0^{\sqrt{2}\lambda} = 1 - e^{-\lambda^2},$$

де $D = \left\{(x;y) : \frac{1}{2}\left(\frac{(x-a_1)^2}{\sigma_1^2} + \frac{(y-a_2)^2}{\sigma_2^2}\right) \le \lambda^2\right\}$.

Зауважимо, що півосі головного еліпса розсіювання дорівнюють $\sqrt{2}\lambda\sigma_1$ і $\sqrt{2}\lambda\sigma_2$.

Обчислення ймовірностей p_λ при $\lambda = 1, 2, 3$ дають такі результати $p_1 = 0.632$, $p_2 = 0.982$, $p_3 = 0.999$. Звідси видно, що в канонічному випадку нормально розподілений випадковий вектор майже достовірно попадає в еліпс з центром $(a_1; a_2)$ (центр розсіювання) і півосями $3\sqrt{2}\sigma_1$ і $3\sqrt{2}\sigma_2$.

2.5 Характеристичні функції

2.5.1 Означення на властивості

Розглянемо на ймовірніносному просторі $(\Omega, \mathcal{F}, \mathrm{P})$ двовимірний випадковий вектор $(\xi; \eta)$. Задамо комплексно-значну випадкову величину ψ рівністю $\psi = \xi + i\eta$.

Якщо існують математичні сподівання випадкових величин ξ і η, то математичне сподівання випадкової величини ψ визначимо так: $\mathrm{M}\psi = \mathrm{M}\xi + i\mathrm{M}\eta$.

Зауважимо, що введені поняття задовольняють всі звичайні для них умови. Єдине, що розподілом комплекснозначної випадкової величини слід вважати розподіл відповідного випадкового вектора.

Означення. Характеристичною функцією $\phi(t)$ випадкової величини ξ називається комплекснозначна функція, задана при $t \in R$ співвідношенням

$$\phi(t) = \mathrm{M}e^{it\xi} = \mathrm{M}(\cos t\xi + i \sin t\xi).$$

Якщо випадкова величина ξ неперервна і $f(x)$ — її щільність розподілу, то

$$\phi(t) = \int_{-\infty}^{+\infty} e^{itx} f(x) dx.$$

Тобто, характеристична функція є перетворенням Фур'є щільності розподілу неперервної випадкової величини. Для дискретної випадкової величини з розподілом $(x_k; p_k)$ характеристична функція

$$\phi(t) = \sum_k e^{itx_k} p_k.$$

Розглянемо деякі властивості характеристичних функцій:

1. $\phi(0) = 1$, $\phi(t) \leq 1$ для всіх $t \in R$;
2. $\phi(-t) = \overline{\phi(t)}$
3. Для всіх $n \in N$, будь-яких дійсних чисел $t_1, t_2, \ldots, t_n$ і будь-яких комплексних чисел $c_1, c_2, \ldots, c_n$

$$\sum_{k=1}^{n} \sum_{m=1}^{n} \phi(t_m - t_k) c_m \overline{c_k} \geq 0;$$

4. Характеристична функція суми двох незалежних випадкових величин дорівнює добутку характеристичних функцій цих величин.

5. Характеристична функція однозначно визначає розподіл випадкової величини.

6. Якщо для випадкової величини ξ існує $\mathrm{M}|\xi|^n$, то її характеристична функція n раз неперервно диференційовна і $\phi^{(n)}(0) = i^n \mathrm{M}\xi^n$;

7. Нехай $F_n(x)$ — послідовність функцій розподілу, $F(x)$ — неперервна функція розподілу, $\phi_n(t)$ і $\phi(t)$ відповідні їм характеристичні функції. Якщо

$$\lim_{n\to\infty} \phi_n(t) = \phi(t)$$

для всіх $t \in R$, то

$$\lim_{n\to\infty} \sup_x |F_n(x) - F(x)| = 0,$$

тобто $F_n(x)$ рівномірно по x збігається до $F(x)$.

Доведемо деякі з цих властивостей, обмежившись випадком неперервних випадкових величин.

◀ Очевидно, що $\phi(0) = \mathrm{M}e^{i0\xi} = \mathrm{M}1 = 1$. Далі

$$|\phi(t)| = \left|\mathrm{M}e^{it\xi}\right| = \left|\int\limits_{-\infty}^{+\infty} e^{itx} f(x)dx\right| \leq$$

$$\leq \int\limits_{-\infty}^{+\infty} |e^{itx}|\, f(x)dx = \int\limits_{-\infty}^{+\infty} f(x)dx = 1,$$

бо $|e^{itx}| = 1$ і $f(x) \geq 0$, що і стверджувалось в 1).

Враховуючи, що

$$\phi(-t) = \mathrm{M}e^{-it\xi} = \mathrm{M}(\cos t\xi - i\sin t\xi) =$$
$$= \mathrm{M}\cos t\xi - i\mathrm{M}\sin t\xi = \overline{\mathrm{M}\cos t\xi + i\mathrm{M}\sin t\xi} =$$
$$= \overline{\mathrm{M}e^{it\xi}} = \overline{\phi(t)},$$

одержимо властивість 2).

Розглянемо

$$\sum_{k=1}^{n}\sum_{m=1}^{n}\phi(t_m - t_k)c_m\overline{c_k} =$$
$$= \mathrm{M}\left(\sum_{k=1}^{n}\sum_{m=1}^{n} e^{i(t_m-t_k)\xi}c_m\overline{c_k}\right) =$$
$$= \mathrm{M}\left(\sum_{m=1}^{n} e^{it_m\xi}c_m\overline{\sum_{m=1}^{n} e^{it_k\xi}c_k}\right) =$$
$$= \mathrm{M}\left|\sum_{m=1}^{n} e^{it_m\xi}c_m\right|^2 \geq 0,$$

що становить зміст властивості 3).

Нехай $\phi_1(t)$, $\phi_2(t)$ — характеристичні функції незалежних випадкових величин ξ_1 і ξ_2 відповідно. Тоді характеристична функція $\phi(t)$ їх суми (властивість 4))

$$\phi(t) = \mathrm{M}e^{it(\xi_1+\xi_2)} = \mathrm{M}\left(e^{it\xi_1}e^{it\xi_2}\right) =$$
$$= \mathrm{M}e^{it\xi_1}\cdot\mathrm{M}e^{it\xi_2} = \phi_1(t)\cdot\phi_2(t)$$

через незалежність випадкових величин $e^{it\xi_1}$ і $e^{it\xi_2}$

Властивості 5) — 7) залишимо без доведення. Тільки формально обчислимо (для властивості 6))

$$\phi^{(n)}(t) = \int\limits_{-\infty}^{+\infty}(ix)^n e^{itx} f(x)dx.$$

Тому

$$\phi^{(n)}(0) = i^n \int\limits_{-\infty}^{+\infty} x^n f(x) dx = i^n \mathrm{M}\xi^n. \qquad \blacktriangleright$$

Зауваження. Для того, щоб неперервна функція $f(t)$ задана на R була характеристичною функцією деякого розподілу, необхідно і досить, щоб вона задовольняла розглянуті вище властивості 1) і 3) (теорема Бохнера-Хінчина).

2.5.2 Характеристична функція нормального розподілу

Нехай випадкова величина ξ має нормальний розподіл $N(a, \sigma^2)$. Тоді її характеристична функція визначається як

$$\mathrm{M}e^{it\xi} = \tfrac{1}{\sqrt{2\pi}\sigma} \int\limits_{-\infty}^{+\infty} e^{itx} \exp\left\{-\tfrac{(x-a)^2}{2\sigma^2}\right\} dx =$$

$$= \tfrac{1}{\sqrt{2\pi}\sigma} \int\limits_{-\infty}^{+\infty} \exp\left\{-\tfrac{1}{2}\left(\tfrac{(x-a)^2}{\sigma^2} - 2it(x-a)\right) + ita\right\} dx =$$

$$= e^{ita} \tfrac{1}{\sqrt{2\pi}\sigma} \int\limits_{-\infty}^{+\infty} \exp\left\{-\tfrac{(x-a-it\sigma^2)^2}{2\sigma^2} - \tfrac{t^2\sigma^2}{2}\right\} dx =$$

$$= e^{ita-\frac{t^2\sigma^2}{2}} \tfrac{1}{\sqrt{2\pi}\sigma} \int\limits_{-\infty}^{+\infty} \exp\left\{-\tfrac{(x-a-it\sigma^2)^2}{2\sigma^2}\right\} dx =$$

$$= e^{ita-\frac{t^2\sigma^2}{2}}.$$

Тобто $\phi(t) = e^{ita-\frac{t^2\sigma^2}{2}}$.

Зауваження. Для наступних застосувань випишемо характеристичну функцію стандартного нормального розподілу: $\phi(t) = e^{-\frac{t^2}{2}}$.

На завершення розглянемо корисне твердження: сума двох незалежних нормально розподілених випадкових величин нормально розподілена. Дійсно, використовуючи одну з властивостей характеристичних функцій матимемо, що характеристична функція суми незалежних випадкових величин з розподілами $N(a_1, \sigma_1^2)$ і $N(a_2, \sigma_2^2)$ така

$$\phi(t) = e^{ita_1 - \frac{t^2\sigma_1^2}{2}} \cdot e^{ita_2 - \frac{t^2\sigma_2^2}{2}} = \\ = e^{it(a_1+a_2) - \frac{t^2(\sigma_1^2+\sigma_2^2)}{2}},$$

що є характеристичною функцією номального розподілу $N(a_1 + a_2, \sigma_1^2 + \sigma_2^2)$. Оскільки характеристична функція однозначно визначає розподіл, то згадане твердження доведено.

Легко також встановити, що добуток нормально розподіленої випадкової величини на число має нормальний розподіл.

Дійсно, його характеристична функція

$$\mathrm{M}e^{it(c\xi)} = \mathrm{M}e^{i(tc)\xi} = e^{i(tc)a - \frac{(tc)^2\sigma^2}{2}} = \\ = e^{it(ac) - \frac{t^2(\sigma c)^2}{2}}$$

є характеристичною функцією розподілу $N(ac, (\sigma c)^2)$.

Звідси можна зробити висновок, що лінійна комбінація незалежних в сукупності нормально розподілених випадкових величин має нормальний розподіл.

Розділ 3

Граничні теореми теорії ймовірностей

3.1 Збіжність послідовностей випадкових величин. Закони великих чисел

3.1.1 Нерівності Чебишева

Граничні теореми пов'язують викладену раніше теорію та її застосування. Одна з таких теорем — теорема Пуассона — вже була розглянута. В цьому розділі ми розглянемо більш загальні теореми. А почнемо з нерівностей, які будуть відігравати важливу роль, як в доведенні граничних теорем так і в розв'язанні багатьох практичних задач.

Зафіксуємо деякий ймовірнісний простір $(\Omega, \mathcal{F}, \mathrm{P})$; будемо вважати, що всі випадкові величини, які будуть розглядатися, задані на цьому просторі.

Теорема (узагальнена нерівність Чебишева). Нехай ξ — довільна випадкова величина і $g(x)$ — невід'ємна парна і неспадна на $[0; +\infty)$ функція. Тоді, якщо існує $\mathrm{M}g(\xi)$, то при всіх $\varepsilon > 0$ таких, що $g(\varepsilon) \neq 0$,

$$\mathrm{P}(|\xi| \geq \varepsilon) \leq \frac{\mathrm{M}g(\xi)}{g(\varepsilon)}. \tag{3.1}$$

◀ Доведемо цю теорему в припущенні, що випадкова величина ξ неперервна. Нехай $f(x)$ — її щільність розподілу. Тоді

$$\mathrm{M}g(\xi) = \int\limits_{-\infty}^{+\infty} g(x)f(x)dx =$$

$$= \int\limits_{-\infty}^{-\varepsilon} g(x)f(x)dx + \int\limits_{-\varepsilon}^{\varepsilon} g(x)f(x)dx + \int\limits_{\varepsilon}^{+\infty} g(x)f(x)dx \geq$$

$$\geq \int\limits_{-\infty}^{-\varepsilon} g(x)f(x)dx + \int\limits_{\varepsilon}^{+\infty} g(x)f(x)dx$$

бо $g(x)f(x) \geq 0$ при всіх $x \in \mathbb{R}$. Звідси

$$\mathrm{M}g(\xi) \geq g(\varepsilon) \int_{-\infty}^{-\varepsilon} f(x)dx + g(\varepsilon) \int_{\varepsilon}^{+\infty} f(x)dx,$$

бо функція $g(x)$ парна і $g(x) \geq g(\varepsilon)$ при $x \geq \varepsilon$ (неспадна).

Отже, $\mathrm{M}g(\xi) \geq g(\varepsilon) \cdot \mathrm{P}(|\xi| \geq \varepsilon)$. Звідси і випливає твердження теореми. ▶

Розглянемо наслідки цієї теореми.

Наслідок (перша нерівність Чебишева). Якщо випадкова величина ξ має скінченний момент $\mathrm{M}|\xi|$, то для всіх $\varepsilon > 0$

$$\mathrm{P}(|\xi| \geq \varepsilon) \leq \frac{\mathrm{M}|\xi|}{\varepsilon}.$$

При умові, що $\mathrm{P}(\xi \geq 0) = 1$,

$$\mathrm{P}(\xi \geq \varepsilon) \leq \frac{\mathrm{M}\xi}{\varepsilon}.$$

◀ Застосуємо нерівність 3.1 з $g(x) = |x|$. ▶

Наслідок (друга нерівність Чебишева в нецентрованій формі). Якщо існує $\mathrm{M}\xi^2$ (чи $\mathrm{D}\xi$), то при всіх $\varepsilon > 0$

$$\mathrm{P}(|\xi| \geq \varepsilon) \leq \frac{\mathrm{M}\xi^2}{\varepsilon^2}.$$

◀ Візьмемо в нерівності (3.1) функцію $g(x) = x^2$. ▶

Наслідок (друга нерівність Чебишева в центрованій формі). Якщо існує $\mathrm{M}\xi^2$ (чи $\mathrm{D}\xi$), то при всіх $\varepsilon > 0$

$$\mathrm{P}(|\xi - \mathrm{M}\xi| \geq \varepsilon) \leq \frac{\mathrm{D}\xi}{\varepsilon^2}.$$

◀ Досить застосувати до випадкової величини $\xi - \mathrm{M}\xi$ і функції $g(x) = x^2$ нерівність (3.1). ▶

Розглянемо приклад застосування нерівності Чебишева.

Приклад. Скільки вимірів потрібно зробити, щоб середнє арифметичне результатів дало вимірювану величину з точністю 0.05 та надійністю 90%, якщо дисперсії (задаються точністю приладу) результатів вимірів (як випадкових величин) не більші ніж 0.2?

Перш ніж розв'язувати цю задачу, дамо деякі пояснення. Припустимо, що вимірюється деяка величина, значення якої l. При кожному вимірі неминучі помилки. Але ми будемо вважати, що всі вони випадкові, тобто, систематичні помилки відсутні. Ці помилки призводять до того, що в результаті вимірів ми одержимо значення n випадкових величин $\xi_1, \xi_2, \ldots, \xi_n$. Природно вважати їх однаково розподіленими з $\mathrm{M}\xi_i = l$ $(i = \overline{1,n})$ та незалежними. Тому $\frac{1}{n}\sum\limits_{i=1}^{n}\mathrm{M}\xi_i = l$.

Якщо $\mathrm{P}\left(\left|\frac{1}{n}\sum\limits_{i=1}^{n}\mathrm{M}\xi_i - l\right| < \varepsilon\right) \geq \gamma$, то кажуть, що величина (випадкова) $\frac{1}{n}\sum\limits_{i=1}^{n}\mathrm{M}\xi_i$ наближає значення l з точністю ε та надійністю γ (іноді надійність виражають у процентах).

В нашому випадку $\varepsilon = 0.05$, $\gamma = 0.9$. Скористаємося другою нерівністю Чебишева, в центрованій формі, для випадкової величини $\eta = \frac{1}{n}\sum\limits_{i=1}^{n}\xi_i$.

$$\mathrm{M}\eta = \frac{1}{n}\sum_{i=1}^{n}\mathrm{M}\xi_i = l, \quad \mathrm{D}\eta = \frac{1}{n^2}\sum_{i=1}^{n}\mathrm{D}\xi_i \leq \frac{0.2\cdot n}{n^2} = \frac{0.2}{n}.$$

Тому

$$\mathrm{P}\left(\left|\tfrac{1}{n}\sum_{i=1}^{n}\mathrm{M}\xi_i - l\right| < \varepsilon\right) =$$
$$= 1 - \mathrm{P}\left(\left|\tfrac{1}{n}\sum_{i=1}^{n}\mathrm{M}\xi_i - l\right| \geq \varepsilon\right) \geq 1 - \tfrac{0.2}{n^2\cdot\varepsilon^2}.$$

Знайдемо n з умови $1 - \frac{0.2}{n^2\cdot\varepsilon^2} \geq \gamma$. Отже,

$$n \geq \frac{0.2}{(1-\gamma)\varepsilon^2} = \frac{0.2}{0.1\cdot 0.05^2} = 800.$$

Тому досить зробити 800 вимірів.

3.1.2 Закон великих чисел

Розглянемо послідовність випадкових величин $\{\xi_n\}$, $(n \geq 1)$, що мають скінченні математичні сподівання. Нас цікавитиме відповідь на питання: чи будуть середні арифметичні $\frac{1}{n}\sum\limits_{i=1}^{n}\xi_i$ і $\frac{1}{n}\sum\limits_{i=1}^{n}\mathrm{M}\xi_i$ в якомусь розумінні наближатися одне до одного при рості n. З таким питанням ми вже зустрічалися в прикладі попереднього пункта. Відповіді на це питання і становлять суть, так званих, законів великих чисел.

Теорема Чебишева. Нехай для послідовності незалежних випадкових величин $\{\xi_n\}$, $(n \geq 1)$ існують $\mathrm{D}\xi_n$ і $\mathrm{D}\xi_n \leq C$ при всіх n. Тоді

$$\mathrm{P}\left(\left|\frac{1}{n}\sum_{i=1}^{n}\xi_i - \frac{1}{n}\sum_{i=1}^{n}\mathrm{M}\xi_i\right| \geq \varepsilon\right) \to 0,\ n \to \infty$$

◀ Розглянемо випадкові величини $\eta_n = \frac{1}{n}\sum\limits_{i=1}^{n}\xi_i$.

$$\mathrm{M}\eta_n = \frac{1}{n}\sum_{i=1}^{n}\mathrm{M}\xi_i, \quad \mathrm{D}\eta_n = \frac{1}{n^2}\sum_{i=1}^{n}\mathrm{D}\xi_i \leq \frac{nC}{n^2} = \frac{C}{n}$$

Тому за другою нерівністю Чебишева в центрованій формі

$$\mathrm{P}\left(\left|\frac{1}{n}\sum_{i=1}^{n}\xi_i - \frac{1}{n}\sum_{i=1}^{n}\mathrm{M}\xi_i\right| \geq \varepsilon\right) \leq \frac{\mathrm{D}\eta_n}{\varepsilon^2} = \frac{C}{n\varepsilon^2}$$

Оскільки $\lim\limits_{n\to\infty}\frac{C}{n\varepsilon^2} = 0$, а ймовірність невід'ємна, то має місце твердження теореми. ▶

Наслідок.

1. Твердження теореми Чебишева називається законом великих чисел.

2. З доведення теореми Чебишева видно, що якщо відмовитись від незалежності випадкових величин ξ_n, то

для виконаня закону великих чисел досить вимагати щоб $\frac{1}{n^2}\mathrm{D}\left(\sum\limits_{i=1}^{n}\xi_i\right) \to 0$, $n \to \infty$ (теорема Маркова).

В теоремі Чебишева ми зустрілися з однією з форм збіжності послідовності випадкових величин, а саме із збіжністю за ймовірністю.

Означення. Послідовність випадкових величин $\{\xi_n\}$, $(n \geq 1)$ збігається за ймовірністю до випадкової величини ξ, якщо для всіх $\varepsilon > 0$

$$\mathrm{P}(|\xi_n - \xi| \geq \varepsilon) \to 0,\ n \to \infty.$$

Позначається це так: $\xi_n \xrightarrow{\mathrm{P}} \xi$, $n \to \infty$.

Таким чином закон великих чисел стверджує, що різниця $\frac{1}{n}\sum\limits_{i=1}^{n}\xi_i - \frac{1}{n}\sum\limits_{i=1}^{n}\mathrm{M}\xi_i$ збігається за ймовірністю до 0.

Наведемо один наслідок теореми Чебишева.

Наслідок (теорема Бернуллі). Відносна частота успіхів в n випробуваннях за схемою Бернуллі збігається за ймовірністю при $n \to \infty$ до ймовірності успіху в одному випробуванні.

◄ Нехай X_k — індикатор успіху в k-ому випробуванні. Тоді $\nu_n = \frac{1}{n}\sum\limits_{k=1}^{n} X_k$ — відносна частота успіху в n випробуваннях. Нехай p — ймовірність успіху в кожному випробуванні. Тоді

$$\mathrm{P}(X_k = 1) = p, \quad \mathrm{P}(X_k = 0) = 1 - p,$$
$$\mathrm{M}X_k = p, \quad \mathrm{D}X_k = p - p^2 = p(1-p).$$

Звідси $\frac{1}{n}\sum\limits_{k=1}^{n}\mathrm{M}X_k = p$ і за теоремою Чебишева

$$\mathrm{P}(|\nu_n - p| \geq \varepsilon) \to 0,\ n \to \infty,$$

або $\nu_n \xrightarrow{\mathrm{P}} p$, $n \to \infty$. ▶

Зауваження. Теорема Бернуллі оправдує, так зване, емпіричне чи частотне означення ймовірності. Коли ймовірністю випадкової події вважається число до якого наближається відносна частота появи цієї події при великій кількості незалежних повторень експерименту в незмінних умовах.

Збіжність за ймовірністю $\xi_n \xrightarrow{\mathrm{P}} \xi$, $n \to \infty$, яка виникає, наприклад, в законі великих чисел, ще не забезпечує рівності $\lim\limits_{n\to\infty} \xi_n = \xi$ на всьому Ω, або хоч на його частині з ймовірністю 1. Більш сильним поняттям збіжності є таке.

Означення. Послідовність випадкових величин $\{\xi_n\}$, $(n \geq 1)$ збігається з ймовірністю 1 (майже напевно) до випадкової величини ξ, якщо $\mathrm{P}(\lim\limits_{n\to\infty} \xi_n = \xi) = 1$.

Позначається це так $\xi_n \xrightarrow{\mathrm{P1}} \xi$, $n \to \infty$, або $\xi_n \xrightarrow{\text{м.н.}} \xi$, $n \to \infty$.

Твердження, аналогічне закону великих чисел, в якому збіжність за ймовірністю замінене на збіжність з ймовірністю 1, називається підсиленим законом великих чисел. Сформулюємо відповідну теорему.

Теорема Колмогорова. Нехай $\{\xi_n\}$, $(n \geq 1)$ послідовність незалежних випадкових величин, для яких існує $\mathrm{D}\xi_n$. Якщо $\sum\limits_{k=1}^{\infty} \frac{1}{k^2}\mathrm{D}\xi_k < +\infty$, то

$$\mathrm{P}\left(\lim_{n\to\infty}\left(\frac{1}{n}\sum_{i=1}^{n}\xi_i - \frac{1}{n}\sum_{i=1}^{n}\mathrm{M}\xi_i\right) = 0\right) = 1.$$

Залишимо цю теорему без доведення, яке можна знайти, наприклад, в [8].

Наслідок (теорема Бореля). Відносна частота успіхів в n випробуваннях за схемою Бернуллі збігається з ймовірністю 1 при $n \to \infty$ до ймовірності успіху в одному випробовуванні.

◀ Розглянувши ті ж випадккові величини X_k, що і в доведенні теореми Бернуллі, запишемо

$$\sum_{k=1}^{\infty} \frac{1}{k^2} \mathrm{D} X_k = \sum_{k=1}^{\infty} \frac{p(1-p)}{k^2} < +\infty.$$

Тому твердження теореми Бореля випливає з теореми Колмогорова. ▶

Теорема Бореля ще раз свідчить на користь емпіричного поняття ймовірності.

3.2 Центральна гранична теорема

3.2.1 Центральна гранична теорема

Цей параграф присвячений одному з найбільш важливих тверджень теорії ймовірностей: при досить широких припущеннях суми великої кількості незалежних малих випадкових доданків мають розподіл близький до нормального. Цей результат є, наприклад, теоретичною основою застосування нормального розподілу при розв'язуванні багатьох практичних задач. Так помилки вимірів, що дають вимірювальні прилади, швидкості руху молекул газу мають розподіл, що мало відрізняється від нормального.

Розглянемо одну з найпростіших теорем, що узаконює згадане твердження.

Центральна гранична теорема. Нехай $\{\xi_n\}$, $(n \geq 1)$ — послідовність незалежних однаково розподілених випадкових величин із скінченною дисперсією $(\mathrm{D}\xi_k = \sigma^2 > 0)$

і $s_n = \sum\limits_{k=1}^{n} \xi_k$. Тоді при $n \to \infty$ для всіх $x \in R$

$$\mathrm{P}\left(\frac{s_n - \mathrm{M}s_n}{\sqrt{\mathrm{D}s_n}} < x\right) \to \Phi(x) = \frac{1}{\sqrt{2\pi}} \int\limits_{-\infty}^{x} e^{-t^2/2} dt. \qquad (3.2)$$

◀ Нехай $\mathrm{M}\xi_k = a$. Тоді

$$\mathrm{M}s_n = \sum_{k=1}^{n} \mathrm{M}\xi_k = n \cdot a, \quad \mathrm{D}s_n = \sum_{k=1}^{n} \mathrm{D}\xi_k = n \cdot \sigma^2.$$

Розглянемо випадкові величини

$$\psi_n = \frac{s_n - \mathrm{M}s_n}{\sqrt{\mathrm{D}s_n}} = \frac{\sum\limits_{k=1}^{n} \xi_k - n \cdot a}{\sqrt{n}\sigma} = \frac{1}{\sqrt{n}} \sum_{k=1}^{n} \frac{\xi_k - a}{\sigma}$$

і $\eta_k = \frac{\xi_k - a}{\sigma}$. Зауважимо, що $\psi_n = \frac{1}{\sqrt{n}} \sum\limits_{k=1}^{n} \eta_k$.

Оскільки

$$\mathrm{M}\eta_k = \tfrac{1}{\sigma}(\mathrm{M}\xi_k - a) = 0,$$
$$\mathrm{D}\eta_k = \tfrac{1}{\sigma^2}\mathrm{D}(\xi_k - a) = \tfrac{1}{\sigma^2}\mathrm{M}(\xi_k - a)^2 = \tfrac{1}{\sigma^2}\mathrm{D}\xi_k = 1,$$

то характеристичні функції випадкових величин η_k

$$\phi(t) = 1 - \frac{t^2}{2} + o(t^2).$$

Тому характеристичні функції випадкових величин ψ_n

$$\phi_{\psi_n}(t) = \left(\phi\left(\frac{t}{\sqrt{n}}\right)\right)^n = \left(1 - \frac{t^2}{2n} + o\left(\frac{t^2}{n}\right)\right)^n.$$

Оскільки

$$\lim_{n\to\infty} \phi_{\psi_n}(t) = \left(1 - \frac{t^2}{2n} + o\left(\frac{t^2}{n}\right)\right)^n = e^{-t^2/2},$$

то при $n \to \infty$ $\mathrm{P}(\psi_n < x) \to \Phi(x)$, бо функція $e^{-t^2/2}$ є характеристичною функцією стандартного нормального розподілу. ►

Сформулюємо (без доведення) центральну граничну теорему для послідовності неоднаково розподілених випадкових величин.

Теорема Ляпунова. Якщо для послідовності незалежних випадкових величин $\{\xi_n\}$, $(n \geq 1)$, що мають скінченні моменти третього порядку, виконується

$$\lim_{n\to\infty} \frac{\sum\limits_{k=1}^{n} \mathrm{M}|\xi_k - \mathrm{M}\xi_k|^3}{\left(\sum\limits_{k=1}^{n} \mathrm{D}\xi_k\right)^{3/2}} = 0,$$

то має місце співвідношення (3.2).

Зауваження. Зміст умов центральної граничної теореми (в будь-якому варіанті) полягає в тому, що в сумі $\sum\limits_{k=1}^{n} \xi_k$ жоден з доданків не домінує над іншими.

Приклад. Розв'яжемо задачу, сформульовану в попередній лекції. Скільки вимірів потрібно зробити, щоб з точністю 0.05 та надійністю 90% середнє арифметичне вимірів дало вимірювану величину, якщо дисперсії результатів вимірів не перевищують 0.2?

Будемо вважати, що результати вимірів ξ_k є незалежними однаково розподіленими випадковими величинами.

Нехай $\mathrm{M}\xi_k = a$, $\mathrm{D}\xi_k = \sigma^2$. Тоді при великих n за центральною граничною теоремою

$$\mathrm{P}\left(\frac{\sum\limits_{k=1}^{n}\xi_k - n\cdot a}{\sqrt{n}\sigma} < x\right) \approx \Phi(x) = \frac{1}{\sqrt{2\pi}}\int\limits_{-\infty}^{x} e^{-t^2/2}dt.$$

Тому

$$\mathrm{P}\left(\left|\tfrac{1}{n}\sum_{k=1}^{n}\xi_k - a\right| < \varepsilon\right) =$$
$$= \mathrm{P}\left(-\tfrac{\varepsilon\sqrt{n}}{\sigma} < \tfrac{\frac{1}{n}\sum\limits_{k=1}^{n}\xi_k - a}{\sigma/\sqrt{n}} < \tfrac{\varepsilon\sqrt{n}}{\sigma}\right) \approx$$
$$\approx \Phi\left(\tfrac{\varepsilon\sqrt{n}}{\sigma}\right) - \Phi\left(-\tfrac{\varepsilon\sqrt{n}}{\sigma}\right) = 2\Phi\left(\tfrac{\varepsilon\sqrt{n}}{\sigma}\right) - 1$$

Вибравши n таким, що $2\Phi\left(\frac{\varepsilon\sqrt{n}}{\sigma}\right) - 1 \geq \gamma$, одержимо, що середнє арифметичне результатів вимірів з точністю ε і надійністю γ дає вимірювану величину.

Звідси $\Phi\left(\frac{\varepsilon\sqrt{n}}{\sigma}\right) \geq \frac{\gamma+1}{2}$ і $\frac{\varepsilon\sqrt{n}}{\sigma} \geq u_{\frac{\gamma+1}{2}}$, де $u_{\frac{\gamma+1}{2}}$ квантиль порядку $\frac{\gamma+1}{2}$ стандартного нормального розподілу. Тому $n \geq \frac{\sigma^2}{\varepsilon^2}u^2_{\frac{\gamma+1}{2}}$. В нашому випадку $\gamma = 0.9$, $\varepsilon = 0.05$, $\sigma^2 = 0.2$ і з таблиці функції стандартного нормального розподілу $u_{0.95} = 1.645$.

Отже, $n \geq \frac{0.2}{0.0025}\cdot(1.645)^2 \approx 220$.

Зауваження. В прикладі ми одержали значення значно менше, ніж у попередній лекції. Це свідчить про те, що центральна гранична теорема є більш точним твердженням, ніж закон великих чисел чи нерівність Чебишева що була застосована раніше.

3.2.2 Теореми Муавра-Лапласа

Розглянемо деякі наслідки центральної граничної теореми стосовно схеми Бернуллі.

Інтегральна теорема Муавра-Лапласа. Нехай ν_n — кількість успіхів в n випробовуваннях за схемою Бернуллі з ймовірністю успіху в кожному випробовуванні p $(0 < p < 1)$. Тоді при $a < b$

$$\lim_{n\to\infty} \mathrm{P}\left(a \le \frac{\nu_n - n\cdot p}{\sqrt{np(1-p)}} < b\right) = \Phi(b) - \Phi(a)$$

(тут, як і раніше, $\Phi(x) = \frac{1}{\sqrt{2\pi}} \int\limits_{-\infty}^{x} e^{-t^2/2}dt$).

◀ Випадкові величини $\nu_n = \sum\limits_{k=1}^{n} X_k$, де X_k — індикатор успіху в k-ому випробуванні (незалежні однаково розподілені випадкові величини).

$$\mathrm{M}\nu_n = np, \quad \mathrm{D}\nu_n = np(1-p).$$

Тому з центральної граничної теореми випливає твердження даної теореми, бо випадкові величини $\frac{\nu_n - n\cdot p}{\sqrt{np(1-p)}}$ мають граничний стандартний нормальний розподіл. ▶

Зауваження. Можна довести, що

$$\mathrm{P}(m_1 \le \nu_n < m_2) = \Phi\left(\tfrac{m_2 - np}{\sqrt{np(1-p)}}\right) - \Phi\left(\tfrac{m_1 - np}{\sqrt{np(1-p)}}\right) + \\ + O\left(\tfrac{1}{\sqrt{np(1-p)}}\right),\ n \to \infty. \tag{3.3}$$

Це твердження також називають інтегральною теоремою Муавра-Лапласа.

Звідси одержимо таку теорему.

Локальна теорема Муавра-Лапласа. Для ν_n — кількості успіхів в n випробовуваннях за схемою Бернуллі

$$P(\nu_n = m) = \frac{1}{\sqrt{2\pi np(1-p)}} \exp\left\{-\frac{(m-np)^2}{2np(1-p)}\right\} +$$
$$+O\left(\frac{1}{\sqrt{np(1-p)}}\right), \; n \to \infty.$$

◀ Оскільки

$$P(\nu_n = m) = P(m \le \nu_n < m+1),$$

то з (3.3) та диференційовності функції $\Phi(x)$

$$P(\nu_n = m) = \Phi\left(\frac{m-np}{\sqrt{np(1-p)}} + \frac{1}{\sqrt{np(1-p)}}\right) -$$
$$-\Phi\left(\frac{m-np}{\sqrt{np(1-p)}}\right) + O\left(\frac{1}{\sqrt{np(1-p)}}\right) =$$
$$= \phi\left(\frac{m-np}{\sqrt{np(1-p)}}\right) \cdot \frac{1}{\sqrt{np(1-p)}} + o\left(\frac{1}{\sqrt{np(1-p)}}\right) +$$
$$+O\left(\frac{1}{\sqrt{np(1-p)}}\right) =$$
$$= \frac{1}{\sqrt{2\pi np(1-p)}} \exp\left\{-\frac{(m-np)^2}{2np(1-p)}\right\} + O\left(\frac{1}{\sqrt{np(1-p)}}\right),$$

бо $\Phi'(x) = \phi(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}$. ▶

Зауваження. При великих n можна використовувати наближені рівності

$$P(m_1 \le \nu_n < m_2) \approx \Phi\left(\frac{m_2-np}{\sqrt{np(1-p)}}\right) - \Phi\left(\frac{m_1-np}{\sqrt{np(1-p)}}\right) \quad (3.4)$$

$$P(\nu_n = m) \approx \frac{1}{\sqrt{2\pi np(1-p)}} \exp\left\{-\frac{(m-np)^2}{2np(1-p)}\right\} \quad (3.5)$$

Слід зауважити, що рівність (3.5) може бути застосована і у випадку досить великих p на відміну від того, що в аналогічному випадку відома нам формула Пуассона вимагає малості параметра p. Проте, якщо ймовірність p близька до 0 чи 1, то для підвищення точності слід використовувати теорему Пуассона.

Розділ 4

Оцінювання параметрів розподілів

4.1 Точкові оцінки

4.1.1 Статистики та оцінки

Задача математичної статистики — отримати конкретні висновки з експерементальних даних. Вихідним матеріалом для статистичного дослідження реального явища служить набір результатів спостережень над ним або ж результатів спеціально поставлених експериментів. У випадку, коли результати спотережень є числовими, їх можна розглядати як значення випадкової величини. Ця випадкова величина може бути як дискретною, так і неперервною.

Нехай в стохастичному експерименті спостерігається деяка випадкова величина. Повторимо цей експеримент в незмінних умовах n разів. З організованим таким чином дослідженням пов'язаний n-вимірний випадковий вектор

$$\mathbf{X} = (X_1, X_2, \ldots, X_n),$$

де випадкові величини X_i, $i = \overline{1, n}$, незалежні в сукупності і однаково розподілені.

Означення. Випадковий вектор $\mathbf{X} = (X_1, X_2, \ldots, X_n)$ зі значеннями в просторі $\mathbb{R}^n$, утворений послідовністю незалежних, однаково розподілених випадкових величин X_1, $X_2, \ldots, X_n$, кожна з яких має розподіл G, називають вибіркою обсягом n із розподілу G.

Простір $\mathbb{R}^n$, у якому вибірка набуває значень, будемо називати вибірковим простором.

Далі матимемо справу з вибірками, розподіли (функції розподілу) яких залежать від невідомого параметра θ, який може набувати довільних значень з деякої множини можливих значень $\Theta \subset \mathbb{R}$.

Задача оцінювання параметрів розподілів. Нехай $\mathbf{x} = (x_1, x_2, \ldots, x_n)$ — реалізація вибірки $\mathbf{X} = (X_1, X_2, \ldots, X_n)$ з розподілу $F(\cdot; \theta)$. Значення параметра θ в розподілі $F(\cdot; \theta)$ невідоме, і його необхідно оцінити (наближено визначити) за реалізацією $\mathbf{x}$ вибірки $\mathbf{X}$. У цьому й полягає задача оцінювання параметрів розподілів.

Сформульовану задачу можна розглядати як задачу відшукування такої функції $h : \mathbb{R}^n \to \Theta$, що $h(\mathbf{x}) \approx \theta$, де $\mathbf{x}$ — реалізація вибірки $\mathbf{X}$. Значення $h(\mathbf{x})$ ми й будемо використовувати як θ. Оскільки для кожної реалізації $\mathbf{x}$ значення $h(\mathbf{x})$, яке використовується як θ, буде своє, оцінку параметра θ можна вважати випадковою величиною.

Означення. Борелєву функцію $h(\cdot)$, задану на вибірковому просторі $\mathbb{R}^n$, зі значеннями в Θ — множині можливих значень параметра θ — будемо називати статистикою, а $\hat{\theta} = h(\mathbf{X}) = h(X_1, X_2, \ldots, X_n)$ — борелєву функцію від вибірки зі значеннями в Θ — оцінкою (точковою оцінкою) параметра θ.

4.1.2 Властивості оцінок. Нерівність Крамера-Рао

Основне питання задачі оцінювання параметрів розподілів — наскільки великою є похибка при використанні оцінки $\hat{\theta}$ як значення θ. Кількісно міру похибки при заміні θ на $\hat{\theta}$ (міру розсіювання $\hat{\theta}$ відносно θ) будемо описувати величиною

$$\delta^2 = \mathrm{M}|\hat{\theta} - \theta|^2,$$

яку називають середньоквадратичною похибкою оцінки $\hat{\theta}$.

Серед усіх оцінок з однією і тією ж дисперсією $\mathrm{D}\hat{\theta}$ найменшу середньоквадратичну похибку мають оцінки, для яких $\mathrm{M}\hat{\theta} = \theta$. Останнє випливає з рівностей

$$\delta^2 = \mathrm{M}|\hat{\theta} - \theta|^2 = \mathrm{M}(\hat{\theta}^2 - 2\hat{\theta}\theta + \theta^2) = \mathrm{M}\hat{\theta}^2 - 2\theta\mathrm{M}\hat{\theta} + \theta^2 =$$
$$= \mathrm{M}\hat{\theta}^2 - (\mathrm{M}\hat{\theta})^2 + (\mathrm{M}\hat{\theta})^2 - 2\theta\mathrm{M}\hat{\theta} + \theta^2 = \mathrm{D}\hat{\theta} + (\mathrm{M}\hat{\theta} - \theta)^2.$$

Відзначимо також, що якщо $\mathrm{M}\hat{\theta} = \theta$, то середньоквадратична похибка δ^2 перетворюється в дисперсію оцінки $\hat{\theta}$.

Означення. Оцінка $\hat{\theta}$ параметра θ називається незміщеною, якщо $\mathrm{M}\hat{\theta} = \theta$.

Наочно незміщеність оцінки $\hat{\theta}$ параметра θ означає, що при багаторазовому використанні оцінки $\hat{\theta}$ як значення θ, тобто, при багаторазовій заміні θ на $\hat{\theta}$, середнє значення похибки $\hat{\theta} - \theta$ дорівнює нулеві. Незміщеність вказує на відсутність систематичної похибки.

Для оцінювання параметра θ можна запропонувати багато оцінок $\hat{\theta}$ з $\mathrm{M}\hat{\theta} = \theta$. Із сукупності таких оцінок природно вибрати ті, що мають найменшу дисперсію.

Означення. Якщо в деякому класі незміщених оцінок параметра θ, які мають скінченні дисперсії, існує така оцінка $\hat{\theta}$, що $\mathrm{D}\hat{\theta} \le \mathrm{D}\tilde{\theta}$ для всіх оцінок $\tilde{\theta}$ із цього класу, то кажуть, що оцінка $\hat{\theta}$ є ефективною в даному класі оцінок.

Іншими словами, дисперсія ефективної оцінки параметра в деякому класі незміщених оцінок є мінімальною серед дисперсій всіх оцінок з цього класу.

Ефективну оцінку в класі всіх незміщених оцінок будемо називати просто ефективною оцінкою (без слів «в класі незміщених оцінок»). У літературі з математичної статистики замість терміну «ефективна оцінка» можна також зустріти такі назви, як «незміщена оцінка з мінімальною дисперсією» або «оптимальна оцінка».

Теорема (про єдиність ефективної оцінки). Якщо $\hat{\theta}$ і $\tilde{\theta}$ — ефективні оцінки параметра θ, то з ймовірністю 1 вони співпадають: $\mathrm{P}\{\hat{\theta} = \tilde{\theta}\} = 1$.

◀ За умовою теореми $\mathrm{D}\hat{\theta} = \mathrm{D}\tilde{\theta}$. Нехай $\theta^* = \frac{1}{2}(\hat{\theta} + \tilde{\theta})$. Тоді

$$\mathrm{D}\theta^* = \frac{1}{4}(\mathrm{D}\hat{\theta} + \mathrm{D}\tilde{\theta}) + \frac{1}{2}\mathrm{cov}(\hat{\theta}, \tilde{\theta}) = \frac{1}{2}(\mathrm{D}\hat{\theta} + \mathrm{cov}(\hat{\theta}, \tilde{\theta})).$$

Оскільки

$$|\mathrm{cov}(\hat{\theta}, \tilde{\theta})| \le \sqrt{\mathrm{D}\hat{\theta}\mathrm{D}\tilde{\theta}} = \mathrm{D}\hat{\theta},$$

виконується нерівність

$$\mathrm{D}\theta^* = \frac{1}{2}|\mathrm{D}\hat{\theta} + \mathrm{cov}(\hat{\theta}, \tilde{\theta})| \le \frac{1}{2}(\mathrm{D}\hat{\theta} + \mathrm{D}\hat{\theta}) \le \mathrm{D}\hat{\theta}.$$

З цієї нерівності і того, що $\hat{\theta}$ — ефективна оцінка, випливає, що $\mathrm{D}\theta^* = \mathrm{D}\hat{\theta} = \mathrm{D}\tilde{\theta}$. Тоді

$$\mathrm{cov}(\hat{\theta}, \tilde{\theta}) = \mathrm{D}\hat{\theta} = \mathrm{D}\tilde{\theta} = \sqrt{\mathrm{D}\hat{\theta}\mathrm{D}\tilde{\theta}}.$$

Остання рівність дозволяє нам стверджувати, що випадкові величини $\hat{\theta}$ і $\tilde{\theta}$ з ймовірністю 1 лінійно залежні, тобто $\mathrm{P}\{\hat{\theta} = k\tilde{\theta} + b\} = 1$ для деяких сталих k і b. Враховуючи те, що

$$\mathrm{D}\hat{\theta} = \mathrm{cov}(k\tilde{\theta} + b, \tilde{\theta}) = k\mathrm{D}\tilde{\theta} = k\mathrm{D}\hat{\theta},$$

отримуємо $k = 1$. З умови незміщеності оцінок випливає, що $b = 0$:

$$\mathrm{M}\tilde{\theta} = \mathrm{M}\hat{\theta} = \mathrm{M}(\tilde{\theta} + b) = \mathrm{M}\tilde{\theta} + b.$$

Таким чином, $\mathrm{P}\{\hat{\theta} = \tilde{\theta}\} = 1$. ▶

Нехай $\mathbf{X} = (X_1, X_2, \ldots, X_n)$ — вибірка з розподілу зі щільністю $f(x,\theta)$, $x \in \mathbb{R}$, $\theta \in \Theta$. Позначимо через

$$f_{\mathbf{X}}(\mathbf{x};\theta) = f_{\mathbf{X}}(x_1, x_2, \ldots, x_n; \theta)$$

щільність розподілу випадкового вектора $\mathbf{X}$. Далі, у цьому параграфі, при розгляді функції $f_{\mathbf{X}}(\mathbf{x},\theta)$, будемо застосовувати диференціювання за параметром під знаком інтеграла, залежного від параметру. Тому будемо вважати, що Θ — інтервал на прямій $\mathbb{R}$, а функція $f_{\mathbf{X}}(\mathbf{x},\theta)$, $\mathbf{x} \in \mathbb{R}^n$, $\theta \in \Theta$, задовольняє умови регулярності, які забезпечують законність вказаних операцій.

Означення. Функцію

$$I(\theta) = \mathrm{M}\left(\frac{\partial \ln f_{\mathbf{X}}(\mathbf{X};\theta)}{\partial\theta}\right)^2$$

(якщо вона визначена) називають кількістю інформації за Фішером.

Теорема (нерівність Крамера-Рао). Якщо $\hat{\theta}$ — незміщена оцінка параметра θ і виконуються умови регулярності, то має місце нерівність Крамера-Рао

$$\mathrm{D}\hat{\theta} \geq \frac{1}{I(\theta)}, \tag{4.1}$$

де $I(\theta)$ — кількість інформації за Фішером.

◀ Нехай $f(x;\theta) > 0$ при $x \in A \subset \mathbb{R}$ і $f(x;\theta) = 0$ при $x \notin A$. Тоді щільність

$$f_{\mathbf{X}}(\mathbf{x},\theta) = f_{\mathbf{X}}(x_1, x_2, \ldots, x_n; \theta) = \prod_{i=1}^{n} f(x_i;\theta)$$

розподілу вектора $\mathbf{X} = (X_1, X_2, \ldots, X_n)$ відмінна від нуля на множині $B = A \times A \times \ldots \times A \subset \mathbb{R}^n$. Оскільки

$$\int\limits_{\mathbb{R}^n} f_{\mathbf{X}}(\mathbf{x},\theta) d\mathbf{x} = \int\limits_{B} f_{\mathbf{X}}(\mathbf{x},\theta) d\mathbf{x} = 1,$$

то

$$\frac{\partial}{\partial\theta}\int\limits_{\mathbb{R}^n} f_{\mathbf{X}}(\mathbf{x},\theta) d\mathbf{x} = \frac{\partial}{\partial\theta}\int\limits_{B} f_{\mathbf{X}}(\mathbf{x},\theta) d\mathbf{x} = \int\limits_{B} \frac{\partial f_{\mathbf{X}}(\mathbf{x},\theta)}{\partial\theta} d\mathbf{x} =$$

$$= \int\limits_{B} \frac{\partial \ln f_{\mathbf{X}}(\mathbf{x},\theta)}{\partial\theta} f_{\mathbf{X}}(\mathbf{x},\theta) d\mathbf{x} = 0. \tag{4.2}$$

З умови незміщеності оцінки $\hat{\theta} = h(\mathbf{X}) = h(X_1,\ X_2, \ldots,$ $X_n)$ випливає, що

$$\mathrm{M}\hat{\theta} = \int\limits_{\mathbb{R}^n} h(\mathbf{x}) f_{\mathbf{X}}(\mathbf{x},\theta) d\mathbf{x} = \int\limits_{B} h(\mathbf{x}) f_{\mathbf{X}}(\mathbf{x},\theta) d\mathbf{x} = \theta.$$

Тоді

$$\frac{\partial}{\partial\theta}\int\limits_{\mathbb{R}^n} h(\mathbf{x}) f_{\mathbf{X}}(\mathbf{x},\theta) d\mathbf{x} = \frac{\partial}{\partial\theta}\int\limits_{B} h(\mathbf{x}) f_{\mathbf{X}}(\mathbf{x},\theta) d\mathbf{x} =$$

$$= \int\limits_{B} h(\mathbf{x}) \frac{\partial \ln f_{\mathbf{X}}(x,\theta)}{\partial\theta} f_{\mathbf{X}}(\mathbf{x},\theta) d\mathbf{x} = 1. \tag{4.3}$$

Віднявши від (4.3) рівність (4.2), помножену на параметр θ, отримуємо

$$\int\limits_{B} (h(\mathbf{x}) - \theta) \frac{\partial \ln f_{\mathbf{X}}(\mathbf{x},\theta)}{\partial\theta} f_{\mathbf{X}}(\mathbf{x},\theta) d\mathbf{x} = 1.$$

Скориставшись нерівністю Коші-Буняковського, встановлюємо, що

$$1 \leq \int\limits_{B} (h(\mathbf{x}) - \theta)^2 f_{\mathbf{X}}(\mathbf{x},\theta) d\mathbf{x} \times$$

$$\times \int\limits_{B} \left(\frac{\partial \ln f_{\mathbf{X}}(\mathbf{x},\theta)}{\partial\theta} \right)^2 f_{\mathbf{X}}(\mathbf{x},\theta) d\mathbf{x} = \mathrm{D}\hat{\theta} I(\theta). \tag{4.4}$$

Це доводить твердження теореми. ▶

За умов регулярності нерівність (4.1) визначає нижню межу дисперсій всіх незміщених оцінок параметра θ. Величину

$$e(\theta) = \frac{1}{I(\theta)\mathrm{D}\hat{\theta}}$$

називають показником ефективності в розумінні Крамера-Рао. Із (4.1) випливає, що $0 < e(\theta) \leq 1$ для будь-якої незміщеної оцінки параметра θ.

Означення. Незміщену оцінку $\hat{\theta}$ параметра θ називають ефективною в розумінні Крамера-Рао, якщо показник ефективності $e(\theta) = 1$.

Зауваження. Рівність

$$\mathrm{D}\hat{\theta} = \frac{1}{I(\theta)}$$

має місце тоді й тільки тоді, коли

$$\frac{\partial \ln f_{\mathbf{X}}(\mathbf{x}, \theta)}{\partial \theta} = c(\theta)(h(\mathbf{x}) - \theta) \tag{4.5}$$

при всіх $x \in \mathbb{R}^n$. Це випливає з необхідних і достатніх умов перетворення в рівність нерівності Коші-Буняковського. Останнє співвідношення, при виконанні умов регулярності, є критерієм ефективності незміщеної оцінки.

Зауваження. З теореми про єдиність ефективної оцінки випливає, що ефективна оцінка в розумінні Крамера-Рао є ефективною.

Зауваження. Нерівність Крамера-Рао має місце і в дискретному випадку. Якщо розподіл випадкового вектора $\mathbf{X} = (X_1, X_2, \ldots, X_n)$ дискретний, то $f_{\mathbf{X}}(\mathbf{x}, \theta)$ — це ймовірність події $\{\mathbf{X} = \mathbf{x}\}$. При цьому $f_{\mathbf{X}}(\mathbf{x}, \theta) > 0$ лише для скінченної або зліченної множини точок $\mathbf{x} \in \mathbb{R}^n$. Доведення нерівності (4.1) у цьому випадку є повторенням доведення останньої теореми з очевидними змінами.

Часто можна розглядати не одну оцінку $\hat{\theta} = h(\mathbf{X}) = h(X_1, X_2, \ldots, X_n)$, побудовану за вибіркою $\mathbf{X} = (X_1, X_2, \ldots, X_n)$, а послідовність оцінок $\hat{\theta}_n = h_n(X_1, X_2, \ldots, X_n)$, $n = 1, 2, \ldots$. У цій ситуації природно говорити про асимптотичну поведінку послідовності оцінок $\hat{\theta}_n$.

Означення. Послідовність оцінок $\hat{\theta}_n$, $n = 1, 2, \ldots$, параметра θ називається консистентною (слушною, змістовною), якщо вона збігається за ймовірністю до θ, тобто для кожного $\varepsilon > 0$ $\mathrm{P}\{|\hat{\theta}_n - \theta| > \varepsilon\} \to 0$ при $n \to \infty$.

Означення. Послідовність оцінок $\hat{\theta}_n$, $n = 1, 2, \ldots$, параметра θ називається асимптотично незміщеною, якщо $\mathrm{M}\hat{\theta}_n \to \theta$ при $n \to \infty$.

Означення. Послідовність незміщених оцінок $\hat{\theta}_n$, $n = 1, 2, \ldots$, параметра θ називається асимптотично ефективною, якщо $e(\theta) \to 1$ при $n \to \infty$.

Зауваження. Надалі поняття «оцінка» будемо використовувати також у широкому розумінні як послідовність оцінок.

Приклад (оцінка математичного сподівання нормального розподілу). Нехай $\mathbf{X} = (X_1, X_2, \ldots, X_n)$ — вибірка із нормального розподілу з невідомим математичним сподіванням a і відомою дисперсією σ^2. Перевірити незміщеність, консистентність та ефективність оцінки

$$\hat{a} = \bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$$

параметра a.

За умовою $\mathrm{M}X_i = a$, $i = \overline{1, n}$. Тоді, враховуючи лінійність математичного сподівання, отримуємо

$$\mathrm{M}\bar{X} = \mathrm{M}\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right) = \frac{1}{n}\sum_{i=1}^{n} \mathrm{M}X_i = \\ = \frac{1}{n}\sum_{i=1}^{n} a = \frac{1}{n} na = a, \tag{4.6}$$

а це й означає, що $\bar{X}$ є неміщеною оцінкою параметра a.

Оскільки $X_1, X_2, \ldots, X_n$ — незалежні та однаково розподілені випадкові величини зі скінченною дисперсією σ^2, то згідно з законом великих чисел у формі Чебишева для будь-якого $\varepsilon > 0$

$$\mathrm{P}\{|\bar{X} - a| > \varepsilon\} \to 0, \quad n \to \infty. \tag{4.7}$$

З останнього співвідношення випливає консистентність оцінки $\bar{X}$.

Покажемо тепер, що $\bar{X}$ є ефективною оцінкою параметра a. Для цього достатньо перевірити виконання рівності (4.5). Щільність розподілу випадкового вектора $\mathbf{X} = (X_1, X_2, \ldots, X_n)$ має вигляд

$$\begin{aligned} &f_{\mathbf{X}}(\mathbf{x}; a) = f_{\mathbf{X}}(x_1, x_2, \ldots, x_n; a) = \\ &= \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi}\sigma} \exp\left\{-\frac{(x_i - a)^2}{2\sigma^2}\right\} = \\ &= (2\pi\sigma^2)^{-\frac{n}{2}} \exp\left\{-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i - a)^2\right\}. \end{aligned}$$

Тоді

$$\begin{aligned} &\frac{\partial \ln f_{\mathbf{X}}(\mathbf{x}; a)}{\partial a} = \frac{\partial}{\partial a}\left(-\frac{n}{2}\ln(2\pi\sigma^2) - \frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i - a)^2\right) = \\ &= -\frac{1}{2\sigma^2}\sum_{i=1}^{n}\frac{\partial}{\partial a}(x_i - a)^2 = \frac{1}{\sigma^2}\sum_{i=1}^{n}(x_i - a) = \\ &= \frac{1}{\sigma^2}\left(\sum_{i=1}^{n} x_i - na\right) = \frac{n}{\sigma^2}(\bar{x} - a). \end{aligned}$$

Звідси випливає, що $\hat{a} = \bar{X}$ — ефективна оцінка параметра a.

Приклад (оцінка дисперсії нормального розподілу при відомому математичному сподіванні). Нехай

$\mathbf{X} = (X_1, X_2, \ldots, X_n)$ — вибірка із нормального розподілу $N(a, \sigma^2)$, і нехай математичне сподівання a — відоме, а дисперсія σ^2 — невідома. Перевірити незміщеність, консистентність та ефективність оцінки

$$\hat{\sigma^2} = \frac{1}{n}\sum_{i=1}^{n}(X_i - a)^2.$$

параметра σ^2.

Незміщеність оцінки $\hat{\sigma^2}$ випливає з наступних співвідношень:

$$\mathrm{M}\hat{\sigma^2} = \frac{1}{n}\sum_{i=1}^{n}\mathrm{M}(X_i - a)^2 = \frac{1}{n}\sum_{i=1}^{n}\mathrm{M}(X_i - \mathrm{M}X_i)^2 =$$
$$= \frac{1}{n}\sum_{i=1}^{n}\sigma^2 = \frac{1}{n}n\sigma^2 = \sigma^2.$$

Покажемо, що $\hat{\sigma^2}$ є консистентною оцінкою. Згідно з другою нерівністю Чебишева

$$\mathrm{P}\left\{|\hat{\sigma^2} - \sigma^2| > \varepsilon\right\} \le \frac{\mathrm{D}\hat{\sigma^2}}{\varepsilon^2}.$$

Знайдемо дисперсію оцінки $\hat{\sigma^2}$:

$$\mathrm{D}\hat{\sigma^2} = \mathrm{D}\left(\frac{1}{n}\sum_{i=1}^{n}(X_i - a)^2\right) =$$
$$= \frac{1}{n^2}\sum_{i=1}^{n}\mathrm{D}(X_i - a)^2 = \frac{1}{n^2}\sum_{i=1}^{n}\mathrm{D}(X_i - \mathrm{M}X_i)^2 =$$
$$= \frac{1}{n^2}\sum_{i=1}^{n}\Big(\mathrm{M}(X_i - \mathrm{M}X_i)^4 - \big(\mathrm{M}(X_i - \mathrm{M}X_i)^2\big)^2\Big) =$$

$$= \frac{1}{n^2} \sum_{i=1}^{n} (3\sigma^4 - \sigma^4) = \frac{2\sigma^4}{n}.$$

Тоді

$$\mathrm{P}\left\{|\hat{\sigma^2} - \sigma^2| > \varepsilon\right\} \le \frac{2\sigma^4}{n\varepsilon^2} \to 0, \quad n \to \infty.$$

Це доводить консистентність оцінки $\hat{\sigma^2}$.

Доведемо ефективність оцінки $\hat{\sigma^2}$. Для цього позначимо $\theta = \sigma^2$ і перевіримо виконання рівності (4.5) для щільності $f_{\mathbf{X}}(\mathbf{x};\theta)$, $\mathbf{x} \in \mathbb{R}^n$, випадкового вектора $\mathbf{X} = (X_1, X_2, \ldots, X_n)$. Враховуючи результат попереднього прикладу, маємо

$$\frac{\partial \ln f_{\mathbf{X}}(\mathbf{x};\theta)}{\partial\theta} = \frac{\partial}{\partial\theta}\left(-\frac{n}{2}\ln(2\pi\theta) - \frac{1}{2\theta}\sum_{i=1}^{n}(x_i - a)^2\right) =$$

$$= -\frac{n}{2\theta} + \frac{1}{2\theta^2}\sum_{i=1}^{n}(x_i - a)^2 = \frac{n}{2\theta^2}\left(\frac{1}{n}\sum_{i=1}^{n}(x_i - a)^2 - \theta\right).$$

З останньої рівності випливає, що $\hat{\sigma^2}$ — ефективна оцінка параметра σ^2.

Зауваження. Наведені у цьому параграфі властивості оцінок можна узагальнити на випадок, коли θ — векторний параметр, тобто $\theta = (\theta_1, \theta_2, \ldots, \theta_s) \in \Theta \subset \mathbb{R}^s$.

Вправа. Довести, що коли дисперсія незміщеної оцінки прямує до нуля при $n \to \infty$, то ця оцінка є консистентною.

4.1.3 Емпірична функція розподілу

Нехай $\mathbf{X} = (X_1, X_2, \ldots, X_n)$ — вибірка із розподілу F. Припустимо, що функція розподілу $F(x) = \mathrm{P}\{X_i < x\}$, $x \in \mathbb{R}$, $i = \overline{1, n}$, невідома і її необхідно оцінити.

Означення. Емпіричною функцією розподілу називається функція $\hat{F}_n$ визначена на $\mathbb{R}$ рівністю

$$\hat{F}_n(x) = \frac{1}{n}\sum_{i=1}^{n} \mathbb{I}_{\{X_i<x\}},$$

де $\mathbb{I}_{\{X_i<x\}}$ — індикатор події $\{X_i < x\}$.

Зауваження. При кожному фіксованому $x \in \mathbb{R}$ емпірична функція розподілу $\hat{F}_n(x)$ як функція випадкового вектора $(X_1, X_2, \ldots, X_n)$ є випадковою величиною.

У процесі побудови графіка емпіричної функції розподілу (а точніше її реалізації) ми стикаємося з задачею впорядкування точок її стрибків — вибіркових значень.

Нехай $(x_1, x_2, \ldots, x_n)$ — реалізація вибірки $(X_1, X_2, \ldots, X_n)$. Розташуємо числа $x_1, x_2, \ldots, x_n$ у порядку зростання:

$$x_{(1)} \le x_{(2)} \le \ldots \le x_{(n)}, \tag{4.8}$$

де $x_{(1)}$ — найменше, $x_{(n)}$ — найбільше серед значень x_1, $x_2, \ldots, x_n$.

Означення. Послідовність чисел $x_{(1)}, x_{(2)}, \ldots, x_{(n)}$, які задовольняють умову (4.8), називають варіаційним рядом послідовності $x_1, x_2, \ldots, x_n$ або просто варіаційним рядом. Число $x_{(i)}$, $i = \overline{1,n}$, називають i-м членом варіаційного ряду.

Позначимо через $X_{(i)}$, $i = \overline{1,n}$, випадкову величину, яка при кожній реалізації вибірки $(X_1, X_2, \ldots, X_n)$ набуває значення, що дорівнює i-му члену варіаційного ряду.

Означення. Послідовність випадкових величин $X_{(1)}$, $X_{(2)}, \ldots, X_{(n)}$ називають варіаційним рядом послідовності $X_1, X_2, \ldots, X_n$. При цьому $X_{(i)}$, $i = \overline{1,n}$, називають i-м членом варіаційного ряду вибірки.

Нехай серед членів варіаційного ряду $x_{(1)}, x_{(2)}, \ldots, x_{(n)}$ є r різних чисел, які також впорядковано за зростанням.

Позначимо їх $z_{(1)}, z_{(2)}, \ldots, z_{(r)}$. Припустимо, що кожне з них повторюється відповідно $n_1, n_2, \ldots, n_r$ разів, при цьому очевидно, що $\sum\limits_{i=1}^{r} n_i = n$. Тоді реалізацію емпіричної функції розподілу $\hat{F}_n(x)$ можна подати у вигляді:

$$\hat{F}_n(x) = \frac{1}{n} \sum_{i:\ z_{(i)} < x} n_i = \begin{cases} 0, & x \le z_{(1)}; \\ \frac{n_1}{n}, & z_{(1)} < x \le z_{(2)}; \\ \frac{n_1+n_2}{n}, & z_{(2)} < x \le z_{(3)}; \\ \ldots \\ 1, & x > z_{(r)}. \end{cases} \tag{4.9}$$

Із вищенаведеного випливає, що при фіксованих значеннях вибірки, $\hat{F}_n(x)$ як функція аргументу x є дискретною функцією розподілу. Розподіл $\left(z_{(i)}, \frac{n_i}{n}\right)$, $i = \overline{1, r}$, що їй відповідає, називається емпіричним.

Теорема (про властивості емпіричної функції розподілу). Для кожного $x \in \mathbb{R}$ значення емпіричної функції розподілу $\hat{F}_n(x)$ має такі властивості:

1) випадкова величина $n\hat{F}_n(x)$ має біноміальний розподіл з парараметрами $n,\ F(x)$:

$$\mathrm{P}\left\{\hat{F}_n(x) = \frac{k}{n}\right\} = C_n^k F^k(x)(1 - F(x))^{n-k}, \quad k = \overline{0, n};$$

2) $\hat{F}_n(x)$ є незміщеною оцінкою значення функції розподілу $F(x)$:
$$\mathrm{M}\hat{F}_n(x) = F(x);$$

3) $\hat{F}_n(x)$ є консистентною оцінкою значення функції розподілу $F(x)$: для кожного $\varepsilon > 0$

$$\mathrm{P}(|\hat{F}_n(x) - F(x)| > \varepsilon) \to 0, \quad n \to \infty.$$

◀ Зафіксуємо довільне $x \in \mathbb{R}$. Розглянемо послідовність із n випробувань, в яких i-м успіхом називається подія $\{X_i < x\}$, $i = \overline{1, n}$. Оскільки $X_1, X_2, \ldots, X_n$ незалежні та однаково розподілені випадкові величини, то дана послідовність є схемою випробувань Бернуллі, причому ймовірність успіху не залежить від номера випробування і дорівнює $p = \mathrm{P}\{X_i < x\} = F(x)$. Випадкова величина $n\hat{F}_n(x) = \sum\limits_{i=1}^{n} \mathbb{I}_{\{X_i<x\}}$ — це кількість успіхів в n випробуваннях Бернуллі. Отже,

$$\mathrm{P}\left\{n\hat{F}_n(x) = k\right\} = C_n^k p^k (1-p)^{n-k} =$$
$$= C_n^k F^k(x)(1-F(x))^{n-k}, \quad k = \overline{0, n},$$

що і доводить 1).

Властивість 2) виводиться з лінійності математичного сподівання та з формули для математичного сподівання індикаторної величини:

$$\mathrm{M}(\hat{F}_n(x)) = \frac{1}{n}\sum_{i=1}^{n} \mathrm{M}(\mathbb{I}_{\{X_i<x\}}) = \frac{1}{n}\sum_{i=1}^{n} p = p = F(x).$$

Доведемо властивість 3). За означенням величина $\hat{F}_n(x)$ є відносною частотою успіхів в n випробуваннях за схемою Бернуллі. Тоді згідно з теоремою Бернуллі закону великих чисел $\hat{F}_n(x)$ збігається за ймовірністю при $n \to \infty$ до ймовірності успіху в одному випробуванні $p = F(x)$. Це означає, що $\hat{F}_n(x)$ є консистентною оцінкою для $F(x)$. ▶

4.1.4 Вибіркові моменти

Нехай X — випадкова величина з функцією розподілу $F(x)$, $x \in \mathbb{R}$. Нагадаємо, що знаючи F, можна записати математичне сподівання функції $g(X)$ у вигляді

$$\mathrm{M}g(X) = \int\limits_{-\infty}^{\infty} g(x)F(dx),$$

де останній інтеграл є інтегралом Стільтьєса. У випадку, коли $g(X) = X^k$, або $g(X) = (X - \mathrm{M}X)^k$, $k \geq 0$, отримуємо відповідно початкові моменти m_k і центральні моменти m_k^0 порядку k випадкової величини X:

$$m_k = \mathrm{M}X^k = \int\limits_{-\infty}^{\infty} x^k F(dx),$$

$$m_k^0 = \mathrm{M}(X - \mathrm{M}X)^k = \int\limits_{-\infty}^{\infty} (x - \mathrm{M}X)^k F(dx).$$

Зокрема, при $g(X) = X$ і $g(X) = (X - \mathrm{M}X)^2$ отримуємо формули відповідно для математичного сподівання і дисперсії випадкової величини X.

В математичній статистиці всі ці числові характеристики, а також функцію розподілу випадкової величини X називають теоретичними.

Відзначимо, що між теоретичними центральними і початковими моментами існує простий зв'язок. Справді,

$$m_k^0 = \mathrm{M}(X - \mathrm{M}X)^k = \mathrm{M}\sum_{i=0}^{k} C_k^i X^i (-\mathrm{M}X)^{k-i} =$$

$$= \sum_{i=0}^{k} C_k^i \mathrm{M}X^i (-\mathrm{M}X)^{k-i} = \sum_{i=0}^{k} (-1)^{k-i} C_k^i m_i (m_1)^{k-i} =$$

$$= \sum_{i=2}^{k} (-1)^{k-i} C_k^i m_i (m_1)^{k-i} + (-1)^{k-1}(k-1)(m_1)^k. \quad (4.10)$$

Випишемо цей звязок між моментами для перших чотирьох значень k:

$$m_0^0 = 1, \quad m_1^0 = 0, \quad m_2^0 = m_2 - (m_1)^2,$$
$$m_3^0 = m_3 - 3m_2m_1 + 2(m_1)^2,$$
$$m_4^0 = m_4 - 4m_3m_1 + 6m_2(m_1)^2 - 3(m_1)^4.$$

Нехай $\mathbf{X} = (X_1, X_2 \ldots, X_n)$ — вибірка, $\hat{F}_n(x),\ x \in \mathbb{R},$ — емпірична функція розподілу, побудована за реалізацією $\mathbf{x} = (x_1, x_2, \ldots, x_n)$ вибірки $\mathbf{X}$. Позначимо через $\widetilde{X}$ випадкову величину з функцією розподілу $\hat{F}_n$. За означенням початкові і центральні моменти k-го порядку випадкової величини $\widetilde{X}$ відповідно дорівнюють

$$\hat{m}_k(\mathbf{x}) = \mathrm{M}\widetilde{X}^k = \frac{1}{n}\sum_{i=1}^{r} z_{(i)}^k n_i = \frac{1}{n}\sum_{i=1}^{n} x_i^k,$$
$$\hat{m}_k^0(\mathbf{x}) = \mathrm{M}(\widetilde{X} - \mathrm{M}\widetilde{X})^k =$$
$$= \frac{1}{n}\sum_{i=1}^{r}(z_{(i)} - \mathrm{M}\widetilde{X})^k n_i = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^k,$$

де $\bar{x} = \hat{m}_1(\mathbf{x}),\ z_{(i)},\ i = \overline{1, r}$ — можливі значення випадкової величини $\widetilde{X}$, а $\frac{n_i}{n}$ — ймовірності цих значень (див. (4.9)).

Означення. Випадкова величина

$$\hat{m}_k = \hat{m}_k(\mathbf{X}) = \frac{1}{n}\sum_{i=1}^{n} X_i^k$$

називається вибірковим початковим моментом порядку k. Зокрема, вибірковий початковий момент першого порядку $\bar{X} = \hat{m}_1(\mathbf{X})$ називається вибірковим середнім.

Означення. Випадкова величина

$$\hat{m}_k^0 = \hat{m}_k^0(\mathbf{X}) = \frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})^k$$

називається вибірковим центральним моментом порядку k. Зокрема, вибірковий центральний момент другого порядку $\hat{\sigma^2} = \hat{m}_2^0(\mathbf{X})$ називається вибірковою дисперсією.

Зауваження. При кожній реалізації $\mathbf{x}$ вибірки $\mathbf{X}$ моменти $\hat{m}_k$, $\hat{m}_k^0$ є теоретичними моментами порядку k випадкової величини $\widetilde{X}$, і тому вони задовольняють всі властивості теоретичних моментів.

З урахуванням останнього зауваження, наприклад, можна стверджувати, що для вибіркових початкових і центральних моментів виконуються співвідношення (4.10). Зокрема, має місце тотожність

$$\hat{\sigma^2} = \hat{m}_2 - (\hat{m}_1)^2 = \frac{1}{n}\sum_{i=1}^{n} X_i^2 - \bar{X}^2.$$

Також відзначимо, що центральні моменти є інваріантними відносно зсувів, тобто для довільної сталої c

$$\hat{m}_k^0 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})^k =$$
$$= \frac{1}{n}\sum_{i=1}^{n}\left((X_i - c) - \frac{1}{n}\sum_{i=1}^{n}(X_i - c)\right)^k. \tag{4.11}$$

Теорема (про властивості вибіркового середнього). Нехай $\mathbf{X} = (X_1, X_2, \ldots, X_n)$ — вибірка з розподілу зі скінченною дисперсією, яку позначимо через σ^2. Тоді оцінка $\bar{X}$ математичного сподівання $a = \mathrm{M}X_1$ є незміщеною, консистентною та ефективною в класі всіх лінійних оцінок, тобто оцінок вигляду

$$\widetilde{\theta} = \sum_{i=1}^{n} \alpha_i X_i,$$

де $\alpha_1, \alpha_2, \ldots, \alpha_n$ — довільні сталі такі, що $\sum\limits_{i=1}^{n} \alpha_i = 1$.

◀ Незміщеність і консистентність оцінки $\bar{X}$ випливає з рівності (4.6) і співвідношення (4.7) відповідно.

Доведемо, що вибіркове середнє $\bar{X}$ є ефективною оцінкою параметра a в класі всіх лінійних оцінок. Для цього достатньо показати, що

$$\mathrm{D}\widetilde{\theta} = \mathrm{D}\sum_{i=1}^{n} \alpha_i X_i = \sum_{i=1}^{n} \alpha_i^2 \mathrm{D} X_i = \sigma^2 \sum_{i=1}^{n} \alpha_i^2,$$

де $\sigma^2 \equiv \mathrm{D}X_1$, набуває свого мінімального значення при $\alpha_i = \frac{1}{n}$, тобто коли $\widetilde{\theta} = \bar{X}$.

Для знаходження умовного мінімуму функції

$$g(\alpha_1, \alpha_2, \ldots, \alpha_n) = \sum_{i=1}^{n} \alpha_i^2$$

при умові

$$\sum_{i=1}^{n} \alpha_i = 1$$

складемо функцію Лагранжа

$$L(\alpha_1, \alpha_2, \ldots, \alpha_n; \lambda) = \sum_{i=1}^{n} \alpha_i^2 + \lambda \left(\sum_{i=1}^{n} \alpha_i - 1 \right),$$

де λ — множник Лагранжа. Запишемо необхідні умови існування умовного екстремуму:

$$\begin{cases} \frac{\partial L}{\partial \alpha_i} = 2\alpha_i + \lambda = 0, \quad i = \overline{1, n}, \\ \frac{\partial L}{\partial \lambda} = \sum\limits_{i=1}^{n} \alpha_i - 1 = 0. \end{cases}$$

Розв'язавши цю систему, знаходимо $\lambda = -\frac{2}{n}$ і $\alpha_i = \frac{1}{n}$, $i = \overline{1, n}$, і переконуємося в тому, що при цих значеннях аргументів функція $g(\alpha_1, \alpha_2, \ldots, \alpha_n)$ має умовний мінімум. ►

Теорема (про властивості вибіркової дисперсії). Якщо $\mathbf{X} = (X_1, X_2, \ldots, X_n)$ — вибірка з розподілу зі скінченною дисперсією σ^2, то вибіркова дисперсія $\hat{\sigma^2}$ — зміщена консистентна оцінка σ^2.

◄ Враховуючи те, що оцінка $\hat{\sigma^2}$ інваріантна відносно зсувів (див. формулу (4.11)), зокрема, не змінюється при відніманні від X_i їх спільного середнього $a = \mathrm{M}X_i$, можна вважати, що $a = \mathrm{M}X_i = 0$. Тоді

$$\mathrm{M}\hat{\sigma^2} = \mathrm{M}\left(\frac{1}{n}\sum_{i=1}^{n} X_i^2 - \bar{X}^2\right) = \frac{1}{n}\sum_{i=1}^{n}\mathrm{M}X_i^2 - \mathrm{M}\bar{X}^2 =$$

$$= \frac{1}{n}\sum_{i=1}^{n}\mathrm{D}X_i - \mathrm{D}\bar{X} = \frac{1}{n}n\sigma^2 - \frac{\sigma^2}{n} = \frac{n-1}{n}\sigma^2 \neq \sigma^2,$$

тобто $\hat{\sigma^2}$ — зміщена оцінка для дисперсії.

Покажемо, що $\hat{\sigma^2}$ є консистентною оцінкою. Доведення проведемо для випадку, коли існують моменти даного розподілу до четвертого порядку включно. Згідно з другою нерівністю Чебишева

$$\mathrm{P}\left\{\left|\hat{\sigma^2} - \frac{n-1}{n}\sigma^2\right| > \varepsilon\right\} \leq \frac{\mathrm{D}\hat{\sigma^2}}{\varepsilon^2}. \qquad (4.12)$$

Для знаходження дисперсії оцінки $\hat{\sigma^2}$ скористаємося формулою

$$\mathrm{D}\hat{\sigma^2} = \mathrm{M}(\hat{\sigma^2})^2 - (\mathrm{M}\hat{\sigma^2})^2.$$

Оскільки

$$(\hat{\sigma^2})^2 = \left(\frac{1}{n}\sum_{i=1}^{n} X_i^2 - \bar{X}^2\right)^2 =$$

$$= \frac{1}{n^2}\left(\sum_{i=1}^{n} X_i^2\right)^2 - 2\bar{X}^2\frac{1}{n}\sum_{i=1}^{n} X_i^2 + \bar{X}^4,$$

то виконується рівність

$$\mathrm{M}(\hat{\sigma^2})^2 = \frac{1}{n^2}\mathrm{M}\left(\sum_{i=1}^{n} X_i^2\right)^2 -$$

$$- \frac{2}{n}\mathrm{M}\left(\bar{X}^2\sum_{i=1}^{n} X_i^2\right) + \mathrm{M}\bar{X}^4. \qquad (4.13)$$

Для першого доданка у формулі (4.13) маємо

$$\mathrm{M}\left(\sum_{i=1}^{n} X_i^2\right)^2 = \mathrm{M}\left(\sum_{i=1}^{n} X_i^4 + \sum_{i\neq j} X_i^2 X_j^2\right) =$$

$$= \sum_{i=1}^{n} \mathrm{M}X_i^4 + \sum_{i\neq j} \mathrm{M}X_i^2 \mathrm{M}X_j^2 =$$

$$= \sum_{i=1}^{n} m_4^0 + \sum_{i\neq j} \sigma^2\sigma^2 = nm_4^0 + n(n-1)\sigma^4. \qquad (4.14)$$

Спростимо другий доданок у формулі (4.13):

$$\mathrm{M}\left(\bar{X}^2\sum_{i=1}^{n} X_i^2\right) = \frac{1}{n^2}\mathrm{M}\left(\left(\sum_{j=1}^{n} X_j\right)^2 \sum_{i=1}^{n} X_i^2\right) =$$

$$= \frac{1}{n^2}\mathrm{M}\left(\left(\sum_{j=1}^{n} X_j^2 + \sum_{j\neq k} X_j X_k\right)\sum_{i=1}^{n} X_i^2\right) =$$

$$= \frac{1}{n^2}\mathrm{M}\left(\sum_{j=1}^{n} X_j^2 \sum_{i=1}^{n} X_i^2\right) + \frac{1}{n^2}\mathrm{M}\left(\sum_{i=1}^{n} X_i^2 \sum_{j\neq k} X_j X_k\right).$$

Враховуючи те, що

$$\mathrm{M}\sum_{i:\ i\neq k} X_i X_k = \sum_{i:\ i\neq k} \mathrm{M}X_i \mathrm{M}X_k = 0,$$

отримуємо

$$\sum_{i=1}^{n} X_i^2 \sum_{j\neq k} X_j X_k =$$

$$= \sum_{i\neq j, i\neq k, j\neq k} \mathrm{M}(X_i^2 X_j X_k) + 2\sum_{i\neq j} X_i^3 X_j = 0.$$

Звідси

$$\mathrm{M}\left(\bar{X}^2 \sum_{i=1}^{n} X_i^2\right) = \frac{1}{n^2}\mathrm{M}\left(\sum_{j=1}^{n} X_j^2 \sum_{i=1}^{n} X_i^2\right) =$$

$$= \frac{1}{n^2}\mathrm{M}\left(\sum_{i=1}^{n} X_i^4 + \sum_{i\neq j} X_i^2 X_j^2\right) =$$

$$= \frac{1}{n^2}\left(nm_4^0 + n(n-1)\sigma^4\right). \tag{4.15}$$

Аналогічно можна показати, що

$$\mathrm{M}\bar{X}^4 = \frac{1}{n^4}\left(\sum_{i=1}^{n} X_i\right)^4 = \frac{m_4^0 + 3(n-1)\sigma^4}{n^3}. \tag{4.16}$$

Підставивши (4.14), (4.15), (4.16) у (4.13), отримуємо

$$\mathrm{M}(\hat{\sigma^2})^2 = \frac{nm_4^0 + n(n-1)\sigma^4}{n^2} -$$

$$- \frac{2(nm_4^0 + n(n-1)\sigma^4)}{n^3} + \frac{m_4^0 - 3\sigma^4}{n^3} =$$

$$= \sigma^4 + \frac{m_4^0 - 3\sigma^4}{n} - \frac{2m_4^0 - 5\sigma^4}{n^2} + \frac{m_4 + 3(n-1)\sigma^4}{n^3}.$$

Оскільки $\mathrm{M}\hat{\sigma^2} = \sigma^2 - \frac{\sigma^2}{n}$, остаточно виводимо

$$\mathrm{D}\hat{\sigma^2} = \frac{m_4^0 - \sigma^4}{n} - \frac{2(m_4^0 - 2\sigma^4)}{n^2} + \frac{m_4^0 - 3\sigma^4}{n^3}.$$

Підставивши вираз для $\mathrm{D}\hat{\sigma^2}$ у другу нерівність Чебишева (4.12) і перейшовши до границі при $n \to \infty$, встановлюємо, що вибіркова дисперсія $\hat{\sigma^2}$ є консистентною оцінкою для σ^2. ▶

Зауваження. З теореми випливає, що оцінка

$$S^2 = \frac{n}{n-1}\hat{\sigma^2} = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2$$

є незміщеною і консистентною оцінкою дисперсії σ^2. Її називають виправленою вибірковою дисперсією.

Справді,

$$\mathrm{M}S^2 = \frac{n}{n-1}\mathrm{M}\hat{\sigma^2} = \frac{n}{n-1}\frac{n-1}{n}\sigma^2 = \sigma^2,$$
$$\mathrm{D}S^2 = \frac{n^2}{(n-1)^2}\mathrm{D}\hat{\sigma^2} \to 0, \quad n \to \infty,$$

а це й означає незміщеність і консистентність S^2.

Зауваження. Можна довести, що вибіркові початкові і центральні моменти є консистентними оцінками відповідних теоретичних моментів, якщо тільки вони існують. Однак ці оцінки, крім $\bar{X}$, є зміщеними.

4.1.5 Метод моментів

Метод моментів був запропонований англійським статистиком К. Пірсоном і є одним із перших загальних методів оцінювання невідомих параметрів розподілів. Цей метод полягає у наступному.

Нехай $\mathbf{X} = (X_1, X_2, \ldots, X_n)$ — вибірка з розподілу $F(\cdot, \theta)$. Розподіл $F(\cdot, \theta)$ залежить від невідомого векторного параметра $\theta = (\theta_1, \theta_2, \ldots, \theta_s)$, який набуває значень із множини $\Theta \subset \mathbb{R}^s$. Необхідно знайти оцінку параметра θ за реалізацією $\mathbf{x} = (x_1, x_2, \ldots, x_n)$ вибірки $\mathbf{X}$.

Будемо вважати, що для даного розподілу існують перші s моментів, які, очевидно, є функціями векторного аргументу θ:

$$m_k = m_k(\theta) = \int\limits_{-\infty}^{\infty} x^k F(dx; \theta), \quad k = \overline{1, s}.$$

Розглянемо вибіркові моменти

$$\hat{m}_k = \hat{m}_k(\mathbf{X}) = \frac{1}{n} \sum_{i=1}^{n} X_i^k.$$

Нагадаємо, що вибіркові моменти $\hat{m}_k$ є консистентними оцінками відповідних теоретичних моментів m_k. Це означає, що при великому обсязі вибірки n значення оцінок $\hat{m}_k = \hat{m}_k(\mathbf{x})$, де $\mathbf{x}$ — реалізація вибірки $\mathbf{X}$, є близькими до відповідних моментів m_k.

Згідно з методом моментів в якості точкової оцінки $\hat{\theta} = (\hat{\theta}_1, \hat{\theta}_2, \ldots, \hat{\theta}_n)$ векторного параметра θ беруть оцінку, значення якої для довільної реалізації $\mathbf{x}$ вибірки $\mathbf{X}$ отримують як розв'язок системи рівнянь

$$\hat{m}_k = m_k(\theta), \quad k = \overline{1, s}.$$

Можна показати, що за умови неперервної залежності розв'язку цієї системи від $\hat{m}_k,\ k = \overline{1,s}$, оцінка, отримана методом моментів, є консистентною і має асимптотично нормальний розподіл, тобто її розподіл при $n \to \infty$ прямує до нормального.

Приклад. Знайти оцінки параметрів a та σ^2 нормального розподілу

$$f(x; a, \sigma^2) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left\{-\frac{(x-a)^2}{2\sigma^2}\right\}.$$

Для нормального розподілу з параметрами a та σ^2 теоретичні моменти $m_1(a,\sigma^2)$ та $m_2(a,\sigma^2)$ відповідно дорівнюють

$$m_1(a,\sigma^2) = a, \quad m_2(a,\sigma^2) = m_2^0(a,\sigma^2) + (m_1(a,\sigma^2))^2 = \sigma^2 + a^2.$$

Нехай $\mathbf{x} = (x_1, x_2, \ldots x_n)$ — реалізація вибірки $\mathbf{X}$ із розподілу $N(a,\sigma^2)$. Тоді

$$\hat{m}_1 = \frac{1}{n}\sum_{i=1}^{n} x_i, \quad \hat{m}_2 = \frac{1}{n}\sum_{i=1}^{n} x_i^2.$$

Прирівнюючи значення вибіркових моментів $\hat{m}_1$ та $\hat{m}_2$ до відповідних теоретичних моментів $m_1(a,\sigma^2)$ та $m_2(a,\sigma^2)$, отримуємо систему рівнянь

$$\begin{cases} \frac{1}{n}\sum\limits_{i=1}^{n} x_i = a, \\ \frac{1}{n}\sum\limits_{i=1}^{n} x_i^2 = \sigma^2 + a^2, \end{cases}$$

звідки знаходимо значення оцінок

$$\hat{a} = \frac{1}{n}\sum_{i=1}^{n} x_i, \quad \hat{\sigma^2} = \frac{1}{n}\sum_{i=1}^{n} x_i^2 - \left(\frac{1}{n}\sum_{i=1}^{n} x_i\right)^2.$$

Отже, оцінками методу моментів для математичного a і дисперсії σ^2 нормального розподілу є відповідно вибіркове середнє

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

і вибіркова дисперсія

$$\hat{\sigma^2} = \frac{1}{n}\sum_{i=1}^{n} X_i^2 - \bar{X}^2.$$

4.1.6 Метод максимальної правдоподібності

Одним із найбільш універсальних методів оцінювання параметрів розподілів є метод максимальної правдоподібності, запропонований англійським статистиком Р. Фішером.

Нехай $\mathbf{X} = (X_1, X_2, \ldots, X_n)$ — вибірка з розподілу $F(\cdot, \theta)$, де $\theta = (\theta_1, \theta_2, \ldots, \theta_s) \in \Theta \subset \mathbb{R}^s$. Параметр θ — невідомий, і його необхідно оцінити за реалізацією $\mathbf{x}$ вибірки $\mathbf{X}$. Як і раніше, через $f_{\mathbf{X}}(\mathbf{x}; \theta) = f_{\mathbf{X}}(x_1, x_2, \ldots, x_n; \theta)$ будемо позначати щільність розподілу випадкового вектора $\mathbf{X}$ у випадку, коли він абсолютно неперервний, і ймовірність події $\{\mathbf{X} = \mathbf{x}\}$, якщо вектор $\mathbf{X}$ — дискретний.

Означення. Функцією правдоподібності вибірки $\mathbf{X} = (X_1, X_2, \ldots, X_n)$ будемо називати випадкову функцію $L(\theta)$ аргументу $\theta \in \Theta \subset \mathbb{R}^s$, що визначається рівністю

$$L(\theta) = f_{\mathbf{X}}(\mathbf{X}; \theta).$$

Метод максимальної правдоподібносі полягає в тому, що за оцінку параметра θ вибирається оцінка $\hat{\theta} = (\hat{\theta}_1,$

$\hat{\theta}_2, \ldots, \hat{\theta}_n)$, значення якої при кожній реалізації $\mathbf{x}$ вибірки $\mathbf{X}$ максимізує функцію правдоподібності $L(\theta)$:

$$L(\hat{\theta}) = \max_{\theta \in \Theta} L(\mathbf{x}, \theta).$$

Точкова оцінка, знайдена у такий спосіб, називається оцінкою максимальної правдоподібності.

Відзначимо, що функції $L(\theta)$ і $\ln L(\theta)$ досягають найбільшого значення в одній і тій самій точці. А знаходити точку, в якій функція $\ln L(\theta)$ досягає найбільшого значення, часто зручніше.

Логарифм від функції правдоподібності називають логарифмічною функцією правдоподібності.

Якщо функція $L(\theta)$ диференційовна при будь-якій реалізації $\mathbf{x}$ вибірки $\mathbf{X}$ і максимум $L(\theta)$ досягається у внутрішній точці множини Θ, то значення оцінки максимальної правдоподібності задовольняє рівняння (необхідні умови екстремуму)

$$\frac{\partial \ln L(\theta)}{\partial \theta_k} = 0, \quad k = \overline{1, s}. \tag{4.17}$$

Рівняння (4.17) називають рівняннями правдоподібності.

Можна показати, що за умов регулярності функції $f_{\mathbf{X}}(\mathbf{x}; \theta)$ оцінка максимальної правдоподібності $\hat{\theta}_n$ є консистентною, асимптотично ефективною і має асимптотично нормальний розподіл.

Приклад. Знайти оцінку максимальної правдоподібності ймовірності p успіху у схемі n випробувань Бернуллі.

Розглянемо вибірку $\mathbf{X} = (X_1, X_2, \ldots X_n)$ із дискретного розподілу $(0; 1-p)$, $(1; p)$ (X_i — індикатор успіху в i-му випробуванні Бернуллі). Для будь-якої реалізації $\mathbf{x}$ вибірки

X функція правдоподібності $L(p)$ має вигляд:

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

Знайшовши

$$\ln L(p) = \sum_{i=1}^{n} \left(x_i \ln p + (1-x_i) \ln(1-p) \right),$$

отримуємо рівняння правдоподібності

$$\frac{\partial \ln L(p)}{\partial p} = \sum_{i=1}^{n} \left(\frac{x_i}{p} - \frac{1-x_i}{1-p} \right) = 0.$$

Розв'язком цього рівняння є

$$\hat{p} = \frac{1}{n} \sum_{i=1}^{n} x_i = \frac{k}{n},$$

де k — кількість успіхів в схемі Бернуллі. Нескладно переконатися в тому, що $\hat{p}$ є точкою максимуму функції $L(p)$. Таким чином, оцінкою максимальної правдоподібності ймовірності p є відносна частота успіхів в n випробуваннях.

4.2 Інтервальні оцінки

4.2.1 Поняття про інтервальне оцінювання

Для оцінювання невідомих параметрів розподілів поряд із розглянутими вище точковими оцінками використовуються також інтервальні оцінки. На відміну від точкової оцінки, інтервальна оцінка дозволяє отримати ймовірнісну характеристику точності оцінювання невідомого параметра.

Нехай $\mathbf{X} = (X_1, X_2, \ldots, X_n)$ — вибірка з розподілу $F(\cdot, \theta)$. Значення параметра $\theta \in \mathbb{R}$ в розподілі $F(\cdot, \theta)$ невідоме і його необхідно оцінити.

Означення. Інтервальною оцінкою (надійним інтервалом) параметра θ з рівнем надійності $1-\alpha$ називається випадковий інтервал $(\hat{\theta}_1, \hat{\theta}_2)$, де $\hat{\theta}_1$, $\hat{\theta}_2$ є функціями вибірки $\mathbf{X}$, який з ймовірністю $1-\alpha$ містить значення оцінюваного параметра θ, тобто

$$\mathrm{P}\{\hat{\theta}_1 < \theta < \hat{\theta}_2\} = 1 - \alpha. \tag{4.18}$$

Ймовірнісною характеристикою точності інтервального оцінювання параметра θ є випадкова величина

$$l = \hat{\theta}_2 - \hat{\theta}_1,$$

яка для будь-якої реалізації $\mathbf{x}$ вибірки $\mathbf{X}$ є довжиною інтервалу $(\hat{\theta}_1, \hat{\theta}_2)$.

В деяких ситуаціях (наприклад, при розгляді дискретних випадкових величин) замість рівності (4.18) вдається лише отримати нерівність

$$\mathrm{P}\{\hat{\theta}_1 < \theta < \hat{\theta}_2\} \geq 1 - \alpha,$$

тобто побудувати інтервальну оцінку параметра θ з рівнем надійності не меншим, ніж $1 - \alpha$. Інколи виникає потреба оцінити параметр θ тільки зліва або тільки справа. При цьому, якщо

$$\mathrm{P}\{\hat{\theta}_1 < \theta\} = 1 - \alpha,$$

то інтервал $(\hat{\theta}_1, \infty)$ називають правостороннім надійним інтервалом, а у випадку, коли

$$P\{\theta < \hat{\theta}_2\} = 1 - \alpha,$$

інтервал $(-\infty, \theta_2)$ називають лівостороннім надійним інтервалом з рівнем надійності $1-\alpha$.

Зауваження. Поняття інтервальної оцінки тривіально поширюються на випадок векторного параметра $\theta = (\theta_1, \theta_2, \ldots, \theta_s) \in \Theta \subset \mathbb{R}^s$.

4.2.2 Загальний метод опорних величин

Загальним методом побудови інтервальних оцінок для θ є метод опорних величин. Він полягає у знаходженні опорної випадкової величини вигляду $h(\mathbf{X};\theta) = h(X_1, X_2, \ldots, X_n;\, \theta)$, що має такі властивості:

1) вона є функцією вибірки $\mathbf{X}$ та невідомого параметра θ,

2) має повністю відомий розподіл,

3) монотонно залежить від θ.

Виходячи з відомого розподілу цієї випадкової величини, знаходять такі значення h_1, h_2, що

$$\mathrm{P}\{h_1 < h(\mathbf{X};\theta) < h_2\} = 1-\alpha, \tag{4.19}$$

при всіх можливих значеннях θ. Далі, з урахуванням властивості 3), розв'язують відносно θ нерівність у фігурних дужках виразу (4.19) і, таким чином, знаходять випадкові величини $\hat{\theta}_1, \hat{\theta}_2$ такі, що

$$\mathrm{P}\{\hat{\theta}_1 < \theta < \hat{\theta}_2\} = \mathrm{P}\{h_1 < h(\mathbf{X};\theta) < h_2\} = 1-\alpha.$$

Отже, $(\hat{\theta}_1, \hat{\theta}_2)$ — шуканий надійний інтервал для θ з рівнем надійності $1-\alpha$.

Якщо розподіл опорної величини лише асимптотично наближається до відомого, то, розглядаючи вибірку великого обсягу n, відповідні асимптотичні надійні інтервали будують, виходячи з наближених рівностей

$$\mathrm{P}\{h_1 < h_n(\mathbf{X};\theta) < h_2\} \approx 1-\alpha.$$

4.2.3 Деякі спеціальні розподіли

Нижче будуть визначені деякі спеціальні розподіли, які мають важливе значення в математичній статистиці і використовуватимуться нами у параграфі 4.2.4 цього розділу для побудови інтервальних оцінок параметрів нормального розподілу, а також у наступних розділах даної праці.

Означення. Випадкова величина χ_n^2 має χ^2-розподіл або розподіл Пірсона з n ступенями свободи, якщо її можна подати у вигляді

$$\chi_n^2 = \sum_{i=1}^{n} X_i^2,$$

де $X_1, X_2, \ldots, X_n$ — незалежні в сукупності стандартні нормальні величини.

Щільність розподілу випадкової величини χ_n^2 задається формулою:

$$f_{\chi_n^2}(x) = \begin{cases} \frac{1}{2^{\frac{n}{2}}\Gamma\left(\frac{n}{2}\right)} x^{\frac{n}{2}-1} e^{-\frac{x}{2}}, & x > 0, \\ 0, & x \le 0, \end{cases}$$

де $\Gamma(p) = \int\limits_0^\infty t^{p-1}e^{-t}dt,\ p>0,$ — гамма-функція Ейлера.

Знайдемо математичне сподівання та дисперсію:

$$\mathrm{M}\chi_n^2 = \mathrm{M}\sum_{i=1}^{n} X_i^2 = \sum_{i=1}^{n} \mathrm{M}X_i^2 = n,$$

$$\mathrm{D}\chi_n^2 = \mathrm{D}\sum_{i=1}^{n} X_i^2 = \sum_{i=1}^{n} \mathrm{D}X_i^2 = \sum_{i=1}^{n} \left(\mathrm{M}X_i^4 - (\mathrm{M}X_i^2)^2\right) = 2n.$$

Означення. Випадкова величина τ_n має t-розподіл або розподіл Стьюдента з n ступенями свободи, якщо її можна подати у вигляді

$$\tau_n = \frac{X}{\sqrt{\chi_n^2/n}},$$

де X — стандартна нормальна величина, а χ_n^2 — незалежна від неї величина з χ^2-розподілом та n ступенями свободи.

Щільність розподілу випадкової величини τ_n має вигляд:

$$f_{\tau_n}(x) = \frac{\Gamma\left(\frac{n+1}{2}\right)}{\Gamma\left(\frac{n}{2}\right)\sqrt{\pi n}} \left(1 + \frac{x^2}{n}\right)^{-\frac{n+1}{2}}, \quad x \in \mathbb{R}.$$

Можна показати, що

$$\mathrm{M}\tau_n = 0, \quad n \geq 2,$$

$$\mathrm{D}\tau_n = \frac{n}{n-2}, \quad n > 2.$$

Означення. Випадкова величина $\phi_{n,m}$ має F-розподіл або розподіл Фішера з n, m ступенями свободи, якщо її можна подати у вигляді

$$\phi_{n,m} = \frac{\chi_n^2/n}{\chi_m^2/m},$$

де χ_n^2, χ_m^2 — незалежні величини з χ^2-розподілом та n, m ступенями свободи відповідно.

Для розподілу Фішера

$$f_{\phi_{n,m}}(x) = \begin{cases} \frac{\Gamma\left(\frac{n+m}{2}\right)}{\Gamma\left(\frac{n}{2}\right)\Gamma\left(\frac{m}{2}\right)} n^{\frac{n}{2}} m^{\frac{m}{2}} \frac{x^{\frac{n}{2}-1}}{(m+nx)^{\frac{n+m}{2}}}, & x > 0, \\ 0, & x \leq 0, \end{cases}$$

$$\mathrm{M}\phi_{n,m} = \frac{m}{m-2}, \quad m > 2,$$

$$\mathrm{D}\phi_{n,m} = \frac{2m^2(m+n-2)}{n(m-2)^2(m-4)}, \quad m > 4.$$

Теорема (про вибіркові моменти нормальної вибірки). Нехай $\mathbf{X} = (X_1, X_2, \ldots, X_n)$ — вибірка із розподілу $N(a, \sigma^2)$. Тоді вибіркове середнє $\bar{X}$ та виправлена вибіркова дисперсія S^2 — незалежні. До того ж, величина $\bar{X}$ має розподіл $N\left(a, \frac{\sigma^2}{n}\right)$, а величина $(n-1)\frac{S^2}{\sigma^2}$ — χ^2-розподіл з $n-1$ ступенями свободи.

Перш ніж доводити цю теорему, нагадаємо деякі поняття з лінійної алгебри.

Матриця A розміру $n \times n$ називається ортогональною, якщо вона невироджена та $A^T = A^{-1}$, де A^T — транспонована матриця. Це рівносильно тому, що $A^T A = A A^T = E$, а $|\det A| = 1$ з огляду на рівність

$$(\det A)^2 = \det A \det A = \det A^T \det A = \det(A^T A) = 1.$$

Якщо $\|\mathbf{x}\| = \left(\sum\limits_{i=1}^{n} x_i^2\right)^{\frac{1}{2}}$ — евклідова норма в $\mathbb{R}^n$, то $\|A\mathbf{x}\| = \|\mathbf{x}\|$. Справді,

$$\|A\mathbf{x}\| = (A\mathbf{x})^T A\mathbf{x} = \mathbf{x}^T A^T A\mathbf{x} = \mathbf{x}^T \mathbf{x} = \|\mathbf{x}\|.$$

Нехай тепер $\mathbf{X} = (X_1, X_2, \cdot, X_n)$ — вибірка з розподілу $N(0, \sigma^2)$. Тоді

$$f_{\mathbf{X}}(\mathbf{x}) = f_{\mathbf{X}}(x_1, x_2, \ldots, x_n) = (2\pi\sigma^2)^{-\frac{n}{2}} \exp\left\{-\frac{1}{2\sigma^2}\sum_{i=1}^{n} x_i^2\right\} =$$
$$= (2\pi\sigma^2)^{-\frac{n}{2}} \exp\left\{-\frac{1}{2\sigma^2}\|\mathbf{x}\|^2\right\},$$

а для випадкового вектора $Y = g(\mathbf{X}) = A\mathbf{X} + \mathbf{c}$, використавши відому формулу

$$f_{\mathbf{Y}}(\mathbf{y}) = f_{\mathbf{X}}(g^{-1}(\mathbf{y})) \cdot J_{g^{-1}}(y)$$

(J_h — якобіан відображення h, тобто визначник матриці $\left(\frac{\partial h_i}{\partial y_j}\right)_{i,j=1}^{n}$), отримуємо

$$\begin{aligned} f_{\mathbf{Y}}(\mathbf{y}) &= f_{\mathbf{X}}(A^{-1}(\mathbf{y}-\mathbf{c})) = f_{\mathbf{X}}(A^{T}(\mathbf{y}-\mathbf{c})) = \\ &= (2\pi\sigma^2)^{-\frac{n}{2}} \exp\left\{-\frac{1}{2\sigma^2}\|A^T(\mathbf{y}-\mathbf{c})\|^2\right\} = \\ &= (2\pi\sigma^2)^{-\frac{n}{2}} \exp\left\{-\frac{1}{2\sigma^2}\|\mathbf{y}-\mathbf{c}\|^2\right\} = \\ &= \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi}\sigma} \exp\left\{-\frac{(y_i-c_i)^2}{2\sigma^2}\right\}. \end{aligned} \tag{4.20}$$

Таким чином, компоненти $Y_1, Y_2, \ldots, Y_n$ випадкового вектора $\mathbf{Y}$ незалежні і мають розподіл $N(c_i, \sigma^2)$. Більше того, якщо $\mathbf{Z} = \mathbf{X} + \mathbf{d}$, а $\mathbf{Y} = g(\mathbf{Z}) = A(\mathbf{X} + \mathbf{d}) + \mathbf{c} = A\mathbf{X} + (A\mathbf{d} + \mathbf{c})$, то $f_{\mathbf{Y}}(\mathbf{y}) = f_{\mathbf{X}}(\mathbf{y} - (A\mathbf{d} + \mathbf{c}))$. Отже, ми довели наступне твердження.

Теорема (про ортогональне перетворення нормального вектора) Нехай $\mathbf{X} = (X_1, X_2, \ldots, X_n)$ — випадковий вектор, компоненти якого незалежні та нормально розподілені з однаковою дисперсією σ^2, а $\mathbf{Y} = (Y_1, Y_2, \ldots, Y_n) = A\mathbf{X} + \mathbf{c}$ з ортогональною матрицею $A = (a_{ij})$. Тоді $Y_1, Y_2, \ldots, Y_n$ незалежні та мають нормальний розподіл з математичним сподіванням $\mathrm{M}Y_i = \sum\limits_{j=1}^{n} a_{ij}\mathrm{M}X_j + c_i$ та дисперсією σ^2.

Доведемо тепер теорему про вибіркові моменти нормальної вибірки.

◀ Побудуємо яку-небудь ортогональну матрицю A з першим рядком

$$\left(\frac{1}{\sqrt{n}}, \frac{1}{\sqrt{n}}, \ldots, \frac{1}{\sqrt{n}}\right)$$

і покладемо $\mathbf{Y} = (Y_1, Y_2, \ldots, Y_n) = A\mathbf{X}$. Тоді за теоремою про ортогональне перетворення нормального вектора випадкові величини $Y_1, Y_2, \ldots, Y_n$ незалежні і нормально розподілені з математичним сподіванням $\mathrm{M}Y_i = a\sum\limits_{j=1}^{n} a_{ij}$ та дисперсією σ^2. Оскільки матриця A ортогональна, при $i > 1$ маємо

$$\frac{1}{\sqrt{n}}\sum_{j=1}^{n} a_{ij} = \sum_{j=1}^{n} a_{1j}a_{ij} = 0,$$

а тому

$$\mathrm{M}Y_1 = a\sqrt{n}, \quad \mathrm{M}Y_2 = \mathrm{M}Y_3 = \ldots = \mathrm{M}Y_n = 0.$$

Розглянемо випадкову величину $\frac{1}{\sigma^2}\sum\limits_{i=2}^{n} Y_i^2 = \sum\limits_{i=2}^{n}\left(\frac{Y_i}{\sigma}\right)^2$. Ця величина має χ^2-розподіл з $n-1$ ступенями свободи, як сума квадратів $n-1$ стандартних нормальних величин. З рівності

$$\bar{X} = \frac{1}{\sqrt{n}}\left(\frac{1}{\sqrt{n}}, \frac{1}{\sqrt{n}}, \ldots, \frac{1}{\sqrt{n}}\right)(X_1, X_2, \ldots X_n)^T = \frac{Y_1}{\sqrt{n}}.$$

випливає, що випадкові величини $\bar{X}$ та $\frac{1}{\sigma^2}\sum\limits_{i=2}^{n} Y_i^2$ незалежні, причому вибіркове середнє $\bar{X}$ має розподіл $N\left(a, \frac{\sigma^2}{n}\right)$. Для того, щоб завершити доведення теореми, достатньо показати, що $(n-1)S^2 = \sum\limits_{i=2}^{n} Y_i^2$, а це неважко перевірити з огляду на ортогональність матриці A:

$$(n-1)S^2 = \sum_{i=1}^{n}(X_i - \bar{X})^2 = \sum_{i=1}^{n} X_i^2 - n\bar{X}^2 = \|X\|^2 - Y_1^2 =$$

$$= \|AX\|^2 - Y_1^2 = \sum_{i=1}^{n} Y_i^2 - Y_1^2 = \sum_{i=2}^{n} Y_i^2. \quad \blacktriangleright$$

Теорема (про статистику Стьюдента від нормальної вибірки) Нехай $\mathbf{X} = (X_1, X_2, \ldots, X_n)$ — вибірка із розподілу $N(a, \sigma^2)$, а S^2 — виправлена вибіркова дисперсія. Тоді випадкова величина

$$\tau_{n-1} = \sqrt{n}\frac{\bar{X} - a}{S}$$

має розподіл Стьюдента з $n-1$ ступенями свободи.

◀ Випадкову величину τ_{n-1} запишемо у вигляді

$$\tau_{n-1} = \sqrt{n}\frac{\bar{X} - a}{S} = \frac{\frac{\bar{X}-a}{\sigma/\sqrt{n}}}{\sqrt{S^2/\sigma^2}} = \frac{\frac{\bar{X}-a}{\sigma/\sqrt{n}}}{\sqrt{\frac{(n-1)S^2}{\sigma^2}/(n-1)}}.$$

З теореми про вибіркові моменти нормальної вибірки випливає, що випадкові величини $\frac{\bar{X}-a}{\sigma/\sqrt{n}}$ та $\frac{(n-1)S^2}{\sigma^2}$ — незалежні, до того ж, вони мають стандартний нормальний розподіл та χ^2-розподіл з $n-1$ ступенями свободи відповідно. Це й означає, що випадкова величина τ_{n-1} має розподіл Стьюдента з $n-1$ ступенями свободи. ▶

4.2.4 Інтервальні оцінки параметрів нормального розподілу

Нехай $\mathbf{X} = (X_1, X_2, \ldots, X_n)$ — вибірка з нормального розподілу $N(a, \sigma^2)$. Розглянемо різні варіанти побудови інтервальних оцінок для математичного сподівання a і дисперсії σ^2.

Надійний інтервал для математичного сподівання при відомій дисперсії. Точковою оцінкою математичного сподівання є вибіркове середнє $\bar{X}$. За теоремою про вибіркові моменти нормальної вибірки ця випадкова величина має розподіл $N\left(a, \frac{\sigma^2}{n}\right)$. Тому випадкова величина

$$Z = h(\mathbf{X}; a) = \frac{\bar{X} - a}{\sigma/\sqrt{n}} = \sqrt{n}\frac{\bar{X} - a}{\sigma}$$

має стандартний нормальний розподіл $N(0, 1)$. Це означає, що

$$\mathrm{P}\left\{-t < \sqrt{n}\frac{\bar{X} - a}{\sigma} < t\right\} = F_0(t) - F_0(-t) = 2F_0(t) - 1,$$

де $F_0(t) = \frac{1}{\sqrt{2\pi}} \int\limits_{-\infty}^{t} \exp\{-\frac{z^2}{2}\}dz$ — функція розподілу $N(0, 1)$. Інакше кажучи,

$$\mathrm{P}\left\{\bar{X} - \frac{\sigma}{\sqrt{n}}t < a < \bar{X} + \frac{\sigma}{\sqrt{n}}t\right\} = 2F_0(t) - 1.$$

Зафіксуємо рівень надійності $1 - \alpha$ і знайдемо таке t, щоб виконувалася рівність $2F_0(t) - 1 = 1 - \alpha$. Отримуємо $t = u_{1-\frac{\alpha}{2}}$ — квантиль рівня $1 - \frac{\alpha}{2}$ стандартного нормального розподілу. Тоді

$$\mathrm{P}\left\{\bar{X} - \frac{\sigma}{\sqrt{n}}u_{1-\frac{\alpha}{2}} < a < \bar{X} + \frac{\sigma}{\sqrt{n}}u_{1-\frac{\alpha}{2}}\right\} = 1 - \alpha.$$

Отже, надійний інтервал для математичного сподівання нормального розподілу при відомій дисперсії σ^2 має вигляд:

$$\left(\bar{X} - \frac{\sigma}{\sqrt{n}} u_{1-\frac{\alpha}{2}},\ \bar{X} + \frac{\sigma}{\sqrt{n}} u_{1-\frac{\alpha}{2}}\right).$$

Надійний інтервал для математичного сподівання при невідомій дисперсії. Поряд з точковою оцінкою $\bar{X}$ математичного сподівання a розглянемо оцінку S^2 для дисперсії σ^2. За теоремою про статистику Стьюдента від нормальної вибірки випадкова величина

$$\tau_{n-1} = \sqrt{n}\frac{\bar{X} - a}{S}$$

має розподіл Стьюдента з $n-1$ ступенями свободи. Позначимо через $F_{\tau_{n-1}}(t)$ функцію розподілу величини τ_{n-1}. Тоді

$$\mathrm{P}\left\{-t < \sqrt{n}\frac{\bar{X} - a}{S} < t\right\} = F_{\tau_{n-1}}(t) - F_{\tau_{n-1}}(-t) =$$
$$= 2F_{\tau_{n-1}}(t) - 1;$$
$$\mathrm{P}\left\{\bar{X} - \frac{S}{\sqrt{n}} t < a < \bar{X} + \frac{S}{\sqrt{n}} t\right\} = 2F_{\tau_{n-1}}(t) - 1.$$

Розв'язавши рівняння $2F_{\tau_{n-1}}(t) - 1 = 1 - \alpha$, де $1-\alpha$ — рівень надійності, отримуємо $t = t_{1-\frac{\alpha}{2}}(n-1)$ — квантиль рівня $1-\frac{\alpha}{2}$ розподілу Стьюдента з $n-1$ ступенями свободи. Тоді

$$\mathrm{P}\left\{\bar{X} - \frac{S}{\sqrt{n}} t_{1-\frac{\alpha}{2}}(n-1) < a < \bar{X} + \frac{S}{\sqrt{n}} t_{1-\frac{\alpha}{2}}(n-1)\right\} = 1-\alpha.$$

Таким чином, надійний інтервал для математичного сподівання нормального розподілу при невідомій дисперсії має вигляд:

$$\left(\bar{X} - \frac{S}{\sqrt{n}} t_{1-\frac{\alpha}{2}}(n-1),\ \bar{X} + \frac{S}{\sqrt{n}} t_{1-\frac{\alpha}{2}}(n-1)\right).$$

Надійний інтервал для дисперсії при відомому математичному сподіванні. Найкращою точковою оцінкою дисперсії σ^2 при відомому математичному сподіванні a є $\hat{\sigma^2} = \frac{1}{n}\sum\limits_{i=1}^{n}(X_i - a)^2$ (див. приклади параграфу 4.1.2). Розглянемо випадкову величину

$$n\frac{\hat{\sigma^2}}{\sigma^2} = \sum_{i=1}^{n}\left(\frac{X_i - a}{\sigma}\right)^2.$$

Ця випадкова величина є сумою квадратів n стандартних нормальних величин і тому має χ^2-розподіл з n ступенями свободи. Тоді при заданому рівні надійності $1-\alpha$ маємо

$$1-\alpha = \mathrm{P}\left\{\sigma_1^2 < \sigma^2 < \sigma_2^2\right\} = \mathrm{P}\left\{n\frac{\hat{\sigma^2}}{\sigma_2^2} < n\frac{\hat{\sigma^2}}{\sigma^2} < n\frac{\hat{\sigma^2}}{\sigma_1^2}\right\} =$$
$$= F_{\chi^2_n}\left(n\frac{\hat{\sigma^2}}{\sigma_1^2}\right) - F_{\chi^2_n}\left(n\frac{\hat{\sigma^2}}{\sigma_2^2}\right), \tag{4.21}$$

де $F_{\chi^2_n}(t)$ — функція χ^2-розподілу з n ступенями свободи. Виберемо σ_1^2 і σ_2^2 так, щоб $F_{\chi^2_n}\left(n\frac{\hat{\sigma^2}}{\sigma_1^2}\right) = 1-\frac{\alpha}{2}$, $F_{\chi^2_n}\left(n\frac{\hat{\sigma^2}}{\sigma_2^2}\right) = \frac{\alpha}{2}$. Отримаємо: $\sigma_1^2 = \frac{n\hat{\sigma^2}}{\chi^2_{1-\frac{\alpha}{2}}(n)}$, $\sigma_2^2 = \frac{n\hat{\sigma^2}}{\chi^2_{\frac{\alpha}{2}}(n)}$, де $\chi^2_{1-\frac{\alpha}{2}}(n)$ та $\chi^2_{\frac{\alpha}{2}}(n)$ — квантилі χ^2 розподілу з n ступенями свободи рівня $1-\frac{\alpha}{2}$ та $\frac{\alpha}{2}$ відповідно.

Отже, надійний інтервал для дисперсії нормального розподілу при відомому математичному сподіванні a має вигляд:

$$\left(\frac{n\hat{\sigma^2}}{\chi^2_{1-\frac{\alpha}{2}}(n)},\ \frac{n\hat{\sigma^2}}{\chi^2_{\frac{\alpha}{2}}(n)}\right).$$

Надійний інтервал для дисперсії при невідомому математичному сподіванні. Точковою оцінкою дисперсії σ^2 при невідомому математичному сподіванні є виправлена вибіркова дисперсія S^2. За теоремою про вибіркові моменти нормальної вибірки випадкова величина

$$(n-1)\frac{S^2}{\sigma^2} = \sum_{i=1}^{n}\left(\frac{X_i-\bar{X}}{\sigma}\right)^2$$

має χ^2-розподіл з $n-1$ ступенями свободи. Тоді

$$\begin{aligned}1-\alpha &= \mathrm{P}\left\{\sigma_1^2 < \sigma^2 < \sigma_2^2\right\} = \\ &= \mathrm{P}\left\{(n-1)\frac{S^2}{\sigma_2^2} < (n-1)\frac{S^2}{\sigma^2} < (n-1)\frac{S^2}{\sigma_1^2}\right\} = \\ &= F_{\chi^2_{n-1}}\left((n-1)\frac{S^2}{\sigma_1^2}\right) - F_{\chi^2_{n-1}}\left((n-1)\frac{S^2}{\sigma_2^2}\right).\end{aligned}$$

Межі надійного інтервалу σ_1^2 і σ_2^2 виберемо так, щоб $F_{\chi^2_{n-1}}\left((n-1)\frac{S^2}{\sigma_1^2}\right) = 1-\frac{\alpha}{2}$, $\quad F_{\chi^2_{n-1}}\left((n-1)\frac{S^2}{\sigma_2^2}\right) = \frac{\alpha}{2}$.
Отримаємо: $\quad \sigma_1^2 = \frac{(n-1)S^2}{\chi^2_{1-\frac{\alpha}{2}}(n-1)}$, $\quad \sigma_2^2 = \frac{(n-1)S^2}{\chi^2_{\frac{\alpha}{2}}(n-1)}$.

Таким чином, надійний інтервал для дисперсії нормального розподілу при невідомому математичному сподіванні має вигляд:

$$\left(\frac{(n-1)S^2}{\chi^2_{1-\frac{\alpha}{2}}(n-1)},\ \frac{(n-1)S^2}{\chi^2_{\frac{\alpha}{2}}(n-1)}\right).$$

Вправа. Побудувати лівосторонні та правосторонні надійні інтервали для параметрів нормального розподілу.

Розділ 5

Перевірка статистичних гіпотез

5.1 Задачі перевірки статистичних гіпотез

В цьому розділі розглядається ще один клас задач математичної статистики, пов'язаних з перевіркою статистичних гіпотез.

У теорії перевірки статистичних гіпотез прийнято наступні означення та домовленості. Статистичною гіпотезою називається довільне припущення про розподіл випадкової величини, яке перевіряється за результатами спостережень цієї величини. Гіпотези щодо розподілів, які однозначно їх визначають, називаються простими, у протилежному разі — складними. Гіпотеза, яка підлягає перевірці, називається основною чи нульовою і позначається H_0. Гіпотеза, що суперечить H_0, називається альтернативною чи конкуруючою гіпотезою. Вона позначається H_1. Правило, за яким приймається рішення на користь однієї із вказаних гіпотез, називається статистичним критерієм.

Нехай $\mathbf{X} = (X_1, X_2, \ldots, X_n)$ — вибірка із розподілу G, стосовно якого висувається гіпотеза H_0 при альтернативній гіпотезі H_1. Перевірка гіпотези H_0 зводиться до наступного. За реалізацією $\mathbf{x}$ вибірки $\mathbf{X}$ підраховується значення деякої борелєвої функції $h(\mathbf{x})$, заданої на вибірковому просторі $\mathbb{R}^n$. Ця функція називається статистикою критерію. Вона вибирається так, щоб за припущення H_0 розподіл величини $h(\mathbf{X})$ був відомим. Нехай D — множина значень статистики h, а $D_k \subset D$ — така множина, що, у випадку істинності H_0, ймовірність потрапляння значення випадкової величини $h(\mathbf{X})$ в D_k дорівнює наперед заданому числу α, яке називається рівнем значущості критерію, тобто

$$\mathrm{P}_{H_0}\{h(\mathbf{X}) \in D_k\} = \alpha. \tag{5.1}$$

При цьому рівень значущості α вибирається достатньо малим, щоб за припущення H_0 подію, яка полягає в тому, що $h(\mathbf{X}) \in D_k$ можна було б вважати такою, що в одноразовому експерименті практично не відбувається. Якщо підраховане за реалізацією вибірки значення $h(\mathbf{x})$ все ж потрапило в область D_k, то це можна пояснити тим, що нульова гіпотеза хибна і тому її треба відхилити. У цьому випадку приймається альтернативна гіпотеза H_1. У випадку, коли значення $h(\mathbf{x})$ не потрапило в D_k, гіпотеза H_0 не відхиляється.

Множина D_k, на якій відхиляється нульова гіпотеза, називається критичною областю статистики критерію. У загальному випадку з критичною областю D_k статистики критерію можна пов'язати критичну область вибірки

$$W_k = \{\mathbf{x} \in \mathbb{R}^n : \ h(\mathbf{x}) \in D_k\}.$$

При потраплянні реалізації $\mathbf{x}$ вибірки $\mathbf{X}$ у критичну область W_k нульова гіпотеза відхиляється, у протилежному випадку — не відхиляється.

Зауваження. У зв'язку зі статистичним характером перевірки гіпотез зауважимо таке. Перевірка гіпотези ні в якому разі не є доведенням того, що вона істинна чи хибна. Невдача у відхиленні H_0 означає лише те, що немає достатньо вагомих підстав для відхилення H_0.

При використанні будь-якого статистичного критерію можливі помилки. Розрізняють помилки двох видів.

Означення. Помилка, яка полягає в тому, що гіпотеза H_0 відхиляєтья, коли вона справджується, називається помилкою першого роду.

Означення. Помилка, яка полягає в тому, що гіпотеза H_0 не відхиляєтья, коли правильною є альтернативна гіпотеза H_1, називається помилкою другого роду.

Ймовірність помилки першого роду співпадає з рівнем значущості критерію α і тому може бути зроблена як завгодно малою. Ймовірність помилки другого роду дорівнює

$$\beta = \mathrm{P}_{H_1}\{h(\mathbf{X}) \in D \setminus D_k)\} = \mathrm{P}_{H_1}\{\mathbf{X} \in \mathbb{R}^n \setminus W_k\}.$$

Означення. Потужністю критерію називається ймовірність відсутності помилки другого роду, тобто ймовірність відхилення хибної гіпотези H_0:

$$1 - \beta = \mathrm{P}_{H_1}\{h(\mathbf{X}) \in D_k)\} = \mathrm{P}_{H_1}\{\mathbf{X} \in W_k\}.$$

Відзначимо, що і рівень значущості, і потужність критерію одночасно монотонно (у напрямку зростання) залежать від критичної області W_k вибірки $\mathbf{X}$. Зокрема, при $W_k = \varnothing$ ймовірність помилки першого роду нульова, другого роду дорівнює одиниці, а потужність — нульова. При

$W_k = \mathbb{R}^n$ має місце протилежне: похибка першого роду і потужність одночасно дорівнюють одиниці. Тому задачу відшукування оптимального критерію можна звести до задачі умовної оптимізації, яка полягає у знаходженні найбільшої потужності при обмеженні на ймовірність помилки першого роду.

Зауваження. Кожен критерій перевірки статистичної гіпотези можна задавати парою $(h(\mathbf{X}), D_k)$, що утворена статистикою критерію h та її критичною областю D_k (або ж критичною областю W_k вибірки $\mathbf{X}$).

5.2 Найбільш потужні критерії

5.2.1 Критерій відношення правдоподібностей

Нехай $\mathbf{X} = (X_1, X_2, \ldots, X_n)$ — вибірка із невідомого розподілу F, стосовно якого висувається гіпотеза H_0 при альтернативній гіпотезі H_1.

Означення Статистичний критерій із критичною областю W_k^* вибірки є найбільш потужним критерієм рівня α, якщо його рівень значущості дорівнює α, причому довільний критерій із критичною областю W_k і рівнем значущості α має не більшу потужність, ніж W_k^*:

$$\mathrm{P}_{H_1}\{\mathbf{X} \in W_k\} \le \mathrm{P}_{H_1}\{\mathbf{X} \in W_k^*\}.$$

Задача відшукування найбільш потужного критерію не завжди має розв'язок. Однак у випадку простих основної та альтернативної гіпотез найбільш потужний критерій існує. Припустимо, що основна гіпотеза та альтернативна гіпотеза є простими:

$$H_0 : F = G_0, \quad H_1 : F = G_1,$$

де обидва розподіли G_0 та G_1 повністю визначені. Нехай $G_{\mathbf{X},i}(B) = \mathrm{P}_{H_i}\{\mathbf{X} \in B\},\ B \subset \mathbb{R}^n,\ i = 0, 1,$ — відповідні розподіли вибірок. За умови абсолютної неперервності міри $G_{\mathbf{X},1}(B)$ відносно міри $G_{\mathbf{X},0}(B)$, позначення $G_{\mathbf{X},1} \ll G_{\mathbf{X},0}$, існує щільність $l_{01} \geq 0$ така, що

$$G_{\mathbf{X},1}(B) = \int\limits_B l_{01}(\mathbf{x}) G_{\mathbf{X},0}(d\mathbf{x}). \tag{5.2}$$

До того ж, випадкову величину $l_{01}(\mathbf{X})$ можна подати у вигляді відношення правдоподібностей

$$l_{01}(\mathbf{X}) = \frac{L_1(\mathbf{X})}{L_0(\mathbf{X})},$$

де $L_i(\mathbf{X})$ — функція правдоподібності вибірки $\mathbf{X}$ із розподілу $G_i,\ i = 0, 1$.

Теорема (критерій Неймана-Пірсона). Нехай основна гіпотеза та альтернативна гіпотеза є простими і нехай $G_{\mathbf{X},1} \ll G_{\mathbf{X},0}$. Якщо при заданому рівні значущості $\alpha \in (0, 1)$ існує стала l_α така, що

$$\mathrm{P}_{H_0}\{l_{01}(\mathbf{X}) \geq l_\alpha\} = \alpha,$$

то критерій відношення правдоподібностей $(l_{01}(\mathbf{X}), [l_\alpha, \infty))$ із критичною областю вибірки

$$W_k^* = \{\mathbf{x} \in \mathbb{R}^n :\ l_{01}(\mathbf{x}) \geq l_\alpha\}$$

є найбільш потужним критерієм рівня α.

◀ Розглянемо довільний критерій із критичною областю W_k і рівнем значущості α:

$$\mathrm{P}_{H_0}\{\mathbf{X} \in W_k\} = G_{\mathbf{X},0}(W_k) = \alpha.$$

Обчислимо його потужність з урахуванням формули (5.2):

$$\begin{aligned}
&\mathrm{P}_{H_1}\{\mathbf{X} \in W_k\} = G_{\mathbf{X},1}(W_k) = \\
&= G_{\mathbf{X},1}(W_k \setminus W_k^*) + G_{\mathbf{X},1}(W_k \cap W_k^*) = \\
&= G_{\mathbf{X},1}(W_k \setminus W_k^*) + G_{\mathbf{X},1}(W_k^*) - G_{\mathbf{X},1}(W_k^* \setminus W_k) = \\
&= \int\limits_{W_k \setminus W_k^*} l_{01}(\mathbf{x}) G_{\mathbf{X},0}(d\mathbf{x}) + G_{\mathbf{X},1}(W_k^*) - \int\limits_{W_k^* \setminus W_k} l_{01}(\mathbf{x}) G_{\mathbf{X},0}(d\mathbf{x}).
\end{aligned}$$

За означенням критичної області W_k^* при всіх $x \in W_k^*$ виконується нерівність $l_{01}(x) \geq l_\alpha$. Це означає, що якщо $x \notin W_k^*$, то $l_{01}(x) < l_\alpha$. Тоді

$$\begin{aligned}
&\mathrm{P}_{H_1}\{\mathbf{X} \in W_k\} \leq l_\alpha G_{\mathbf{X},0}(W_k \setminus W_k^*) + \\
&+ G_{\mathbf{X},1}(W_k^*) - l_\alpha G_{\mathbf{X},0}(W_k^* \setminus W_k) = \\
&= l_\alpha G_{\mathbf{X},0}(W_k) + G_{\mathbf{X},1}(W_k^*) - l_\alpha G_{\mathbf{X},0}(W_k^*) = \\
&= \alpha l_\alpha + G_{\mathbf{X},1}(W_k^*) - \alpha l_\alpha = G_{\mathbf{X},1}(W_k^*) = \mathrm{P}_{H_1}\{\mathbf{X} \in W_k^*\}.
\end{aligned}$$

Теорему доведено. ▶

5.2.2 Приклад відшукування найбільш потужного критерію

Нехай $\mathbf{X} = (X_1, X_2, \ldots, X_n)$ — вибірка із нормального розподілу з невідомим математичним сподівання a і відомою дисперсією σ^2. Побудувати найбільш потужний критерій для випадку двох простих основної і альтернативної гіпотез

$$H_0 : a = a_0, \quad H_1 : a = a_1,$$

де a_0 і a_1 деякі задані значення.

Для визначеності будемо вважати, що $a_0 < a_1$.

Відношення правдоподібностей має вигляд

$$l_{01}(\mathbf{X}) = \frac{\prod\limits_{i=1}^{n} \frac{1}{\sqrt{2\pi}\sigma} \exp\left\{-\frac{(X_i-a_1)^2}{2\sigma^2}\right\}}{\prod\limits_{i=1}^{n} \frac{1}{\sqrt{2\pi}\sigma} \exp\left\{-\frac{(X_i-a_0)^2}{2\sigma^2}\right\}} =$$

$$= \frac{\exp\left\{-\frac{1}{2\sigma^2}\sum\limits_{i=1}^{n}(X_i-a_1)^2\right\}}{\exp\left\{-\frac{1}{2\sigma^2}\sum\limits_{i=1}^{n}(X_i-a_0)^2\right\}} =$$

$$= \exp\left\{-\frac{1}{2\sigma^2}\sum_{i=1}^{n}\left((X_i-a_1)^2-(X_i-a_0)^2\right)\right\} =$$

$$= \exp\left\{\frac{1}{2\sigma^2}\sum_{i=1}^{n}(a_1-a_0)(2X_i-(a_1+a_0))\right\} =$$

$$= \exp\left\{\frac{n}{\sigma^2}(a_1-a_0)\bar{X} - \frac{n}{2\sigma^2}(a_1^2-a_0^2)\right\},$$

де $\bar{X}$ — вибіркове середнє. Нерівність

$$l_{01}(\mathbf{X}) = \exp\left\{\frac{n}{\sigma^2}(a_1-a_0)\bar{X} - \frac{n}{2\sigma^2}(a_1^2-a_0^2)\right\} \geq l_\alpha$$

еквівалентна нерівності

$$\sqrt{n}\frac{\bar{X}-a_0}{\sigma} \geq \frac{\sigma}{\sqrt{n}(a_1-a_0)}\ln l_\alpha + \frac{\sqrt{n}}{2\sigma}(a_1-a_0). \qquad (5.3)$$

З теореми про вибіркові моменти нормальної вибірки випливає, що випадкова величина

$$Z = \sqrt{n}\frac{\bar{X}-a_0}{\sigma}$$

у лівій частині нерівності (5.3), за припущення H_0, має стандартний нормальний розподіл. Праву частину нерівності позначимо через C. Розв'язавши рівняння $\mathrm{P}_{H_0}\{Z \geq C\} = \alpha$, отримуємо $C = u_{1-\alpha}$ — квантиль рівня $1-\alpha$ розподілу $N(0,1)$. Тоді стала l_α, для якої

$$\mathrm{P}_{H_0}\{l_{01} \geq l_\alpha\} = \alpha,$$

знаходиться з умови

$$\frac{\sigma}{\sqrt{n}(a_1 - a_0)} \ln l_\alpha + \frac{\sqrt{n}}{2\sigma}(a_1 - a_0) = u_{1-\alpha}$$

і вона дорівнює

$$l_\alpha = \exp\left\{\sqrt{n}\frac{a_1 - a_0}{\sigma}\left(u_{1-\alpha} - \frac{\sqrt{n}}{2\sigma}(a_1 - a_0)\right)\right\}$$

Таким чином, ми знайшли сталу l_α для якої

$$\mathrm{P}_{H_0}\{l_{0,1}(\mathbf{X}) \geq l_\alpha\} = \mathrm{P}_{H_0}\{Z \geq u_{1-\alpha}\} = \alpha$$

і, отже, критерій $(Z,\ [u_{1-\alpha}, +\infty))$ із статистикою Z та критичною областю $D_k = [u_{1-\alpha}, +\infty)$ є найбільш потужним критерієм рівня α для перевірки гіпотези H_0 при альтернативній гіпотезі H_1.

5.3 Перевірка гіпотез про значення параметрів і надійні інтервали

Нехай $\mathbf{X} = (X_1, X_2, \ldots, X_n)$ — вибірка з розподілу $F(\cdot;\theta)$, який залежить від невідомого параметра θ. Розглянемо гіпотезу $H_0 : \theta = \theta_0$, де θ_0 — гіпотетичне значення параметра

θ. Альтернативною будемо вважати гіпотезу $H_1 : \theta \neq \theta_0$. Необхідно за реалізацією $\mathbf{x}$ вибірки $\mathbf{X}$ зробити висновок про гіпотезу H_0, а саме: відхиляти її чи ні.

Перевірку гіпотези H_0 можна проводити, базуючись на інтервальних оцінках невідомого параметра θ.

Між знаходженням інтервальних оцінок параметрів і перевіркою гіпотез про їх можливі значення існує тісний взаємозв'язок. Це, по суті, два різні способи формулювання однієї й тієї ж задачі. Надійний інтервал для невідомого параметра можна розглядати як множину прийнятних значень для його гіпотетичного значення, а доповнення цього інтервалу — як відповідну критичну область для нульової гіпотези. Тобто, для перевірки гіпотези про значення невідомого параметра достатньо побудувати надійний інтервал для параметра і перевірити, чи потрапляє в нього гіпотетичне значення параметра.

Проілюструємо сказане, розглянувши надійний інтервал $(\hat{\theta}_1, \hat{\theta}_2)$ для параметра θ з рівнем надійності $1-\alpha$. Будемо вважати, що цей інтервал побудовано загальним методом опорних величин. Тоді для будь-якої реалізації $\mathbf{x}$ вибірки $\mathbf{X}$ виконується співвідношення

$$\hat{\theta}_1 < \theta < \hat{\theta}_2 \Longleftrightarrow h_1 < h(\mathbf{x};\theta) < h_2,$$

де $h(\mathbf{x};\theta)$ — значення деякої опорної величини $h(\mathbf{X};\theta)$, а h_1, h_2 — значення, знайдені з умови

$$\mathrm{P}\{h_1 < h(\mathbf{X};\theta) < h_2\} = 1-\alpha.$$

Тоді $h(\mathbf{X};\theta_0)$ та $D_k = D \setminus (h_1, h_2)$ є відповідно статистикою критерію та критичною областю для перевірки гіпотези $H_0 : \theta = \theta_0$ при заданому рівні значущості α і при альтернативній гіпотезі $H_1 : \ \theta \neq \theta_0$. Це означає, що у випадку, коли $h(\mathbf{x};\theta_0) \in D_k$ (або $\theta_0 \notin (\hat{\theta}_1, \hat{\theta}_2)$), гіпотеза H_0 відхиляється і приймається альтернативна гіпотеза H_1.

На завершення, для гіпотези H_0 розглянемо ще два варіанти альтернативних гіпотез: $H_1' : \theta < \theta_0$, $H_1'' : \theta > \theta_0$. Критерії з альтернативними гіпотезами H_1' та H_1'' називаються односторонніми — лівостороннім та правостороннім відповідно. Якщо альтернативною є гіпотеза H_1', то основна гіпотеза H_0 відхиляється у випадку, коли гіпотетичне значення θ_0 невідомого параметра θ не потрапляє у лівосторонній надійний інтервал $(-\infty, \hat{\theta}_2)$ для θ з рівнем надійності $1-\alpha$; якщо альтернативною є гіпотеза H_1'', то H_0 відхиляється у випадку, коли θ_0 не потрапляє у відповідний правосторонній надійний інтервал $(\hat{\theta}_1, \infty)$.

Зауваження. Наведені у цьому параграфі міркування поширюються на випадок векторного параметра $\theta = (\theta_1, \theta_2, \ldots, \theta_s) \in \Theta \subset \mathbb{R}^s$.

5.4 Перевірка гіпотез про значення параметрів нормального розподілу

Нехай $\mathbf{x} = (x_1, x_2, \ldots, x_n)$ — реалізація вибірки $\mathbf{X} = (X_1, X_2, \ldots, X_n)$ із розподілу $N(a, \sigma^2)$. Побудуємо критерії для перевірки різних варіантів гіпотез про значення математичного сподівання a і дисперсії σ^2.

Перевірка гіпотези про значення математичного сподівання при відомій дисперсії. Перевіримо гіпотезу $H_0 : a = a_0$ при альтернативній гіпотезі $H_1 : a \neq a_0$. Точковою оцінкою невідомого математичного сподівання a є вибіркове середнє $\bar{X}$. З теореми про вибіркові моменти нормальної вибірки випливає, що величина

$$Z = \sqrt{n}\frac{\bar{X} - a_0}{\sigma},$$

за припущення H_0, має стандартний нормальний розподіл $N(0,1)$. Це означає, що

$$\mathrm{P}_{H_0}\left\{|Z| \geq u_{1-\frac{\alpha}{2}}\right\} = \alpha,$$

де $u_{1-\frac{\alpha}{2}}$ — квантиль рівня $1-\frac{\alpha}{2}$ стандартного нормального розподілу.

Тоді критерій для перевірки гіпотези H_0 при альтернативній гіпотезі H_1 полягає в тому, що гіпотеза H_0 відхиляється, якщо значення $|Z|$, підраховане за реалізацією $\mathbf{x}$ вибірки $\mathbf{X}$, потрапляє у критичну область $D_k = [u_{1-\frac{\alpha}{2}}, +\infty)$, і не відхиляється у протилежному разі. При цьому, ймовірність помилки при відхиленні H_0 дорівнює α.

Якщо альтернативною є гіпотеза $H_1' : a < a_0$ чи $H_1'' : a > a_0$, то критична область статистики Z має вигляд $D_k' = (-\infty, u_\alpha]$ чи $D_k'' = [u_{1-\alpha}, +\infty)$ відповідно.

Перевірка гіпотези про значення математичного сподівання при невідомій дисперсії. Перевіримо тепер сформульовану вище гіпотезу H_0 при альтернативній гіпотезі H_1 у випадку, коли дисперсія σ^2 невідома. Поряд з точковою оцінкою $\bar{X}$ математичного сподівання a розглянемо оцінку S^2 для дисперсії σ^2. Згідно з теоремою про статистику Стьюдента від нормальної вибірки, випадкова величина

$$\tau_{n-1} = \sqrt{n}\frac{\bar{X} - a_0}{S},$$

за гіпотези H_0, має розподіл Стьюдента з $n-1$ ступенями свободи. Тому

$$\mathrm{P}_{H_0}\left\{|\tau_{n-1}| \geq t_{1-\frac{\alpha}{2}}(n-1)\right\} = \alpha, \tag{5.4}$$

де $t_{1-\frac{\alpha}{2}}(n-1)$ — квантиль рівня $1-\frac{\alpha}{2}$ розподілу Стьюдента з $n-1$ ступенями свободи.

Таким чином, критерій з рівнем значущості α для перевірки гіпотези H_0 при альтернативній гіпотезі H_1 можна сформулювати так: гіпотеза H_0 відхиляється, якщо значення $|\tau_{n-1}|$ потрапляє у критичну область $D_k = [t_{1-\frac{\alpha}{2}}(n-1), +\infty)$, і не відхиляється у протилежному разі.

У випадку, коли альтернативною є одна з гіпотез $H_1' : a < a_0$, $H_1'' : a > a_0$, критична область статистики Z є відповідно однією з областей: $D_k' = (-\infty, t_\alpha(n-1)]$, $D_k'' = [t_{1-\alpha}(n-1), +\infty)$.

Перевірка гіпотези про значення дисперсії при відомому математичному сподіванні. Перевіримо гіпотезу $H_0 : \sigma^2 = \sigma_0^2$ при альтернативній гіпотезі $H_1 : \sigma^2 \neq \sigma_0^2$. Нехай математичне сподівання a — відоме. Найкращою точковою оцінкою невідомої дисперсії σ^2 при відомому математичному сподіванні a є $\hat{\sigma^2} = \frac{1}{n}\sum\limits_{i=1}^{n}(X_i - a)^2$. За гіпотези H_0 випадкова величина

$$\chi_n^2 = n\frac{\hat{\sigma^2}}{\sigma_0^2}$$

має χ^2-розподіл з n ступенями свободи, як сума квадратів n стандартних нормальних величин. Тому

$$\mathrm{P}_{H_0}\left\{\chi^2_{\frac{\alpha}{2}}(n) < \chi_n^2 < \chi^2_{1-\frac{\alpha}{2}}(n)\right\} = 1-\alpha,$$

або

$$\mathrm{P}_{H_0}\{\chi_n^2 \in D_k\} = \alpha, \quad D_k = [0, \chi^2_{\frac{\alpha}{2}}(n)] \cup [\chi^2_{1-\frac{\alpha}{2}}(n), +\infty),$$

де $\chi^2_\gamma(n)$ — квантиль рівня γ розподілу χ^2 з n ступенями свободи.

Отже, критерій з рівнем значущості α для перевірки перевірки гіпотези H_0 при альтернативній гіпотезі H_1 полягає в тому, що гіпотеза H_0 відхиляється, якщо значення статистики χ_n^2 потрапляє у критичну область D_k, і не відхиляється у протилежному разі.

Якщо альтернативною є гіпотеза $H_1' : \sigma^2 < \sigma_0^2$ або гіпотеза $H_1'' : \sigma^2 > \sigma_0^2$, то відповідні критичні області статистики χ_n^2 виглядають так: $D_k' = [0, \chi_\alpha^2(n)],\ D_k'' = [\chi_{1-\alpha}^2(n), +\infty)$.

Перевірка гіпотези про значення дисперсії при невідомому математичному сподіванні. Перевіримо тепер гіпотезу H_0 при альтернативній гіпотезі H_1 у випадку, коли a — невідоме. Точковою оцінкою дисперсії σ^2 при невідомому математичному сподіванні є виправлена вибіркова дисперсія S^2. Якщо справджується гіпотеза H_0, то за теоремою про вибіркові моменти нормальної вибірки випадкова величина

$$\chi_{n-1}^2 = (n-1)\frac{S^2}{\sigma_0^2}$$

має χ^2-розподіл з $n-1$ ступенями свободи. Тоді

$$\mathrm{P}_{H_0}\left\{\chi_{n-1}^2 \in D_k\right\} = \alpha,$$
$$D_k = [0, \chi_{\frac{\alpha}{2}}^2(n-1)] \cup [\chi_{1-\frac{\alpha}{2}}^2(n-1), +\infty).$$

Таким чином, правило перевірки гіпотези H_0 при альтернативній гіпотезі H_1 можна сформулювати так: гіпотеза H_0 відхиляється, якщо значення статистики χ_n^2 потрапляє у критичну область D_k, і не відхиляється у протилежному разі. При цьому, ймовірність помилки першого роду дорівнює α.

При альтернативній гіпотезі $H_1' : \sigma^2 < \sigma_0^2$ чи $H_1'' : \sigma^2 > \sigma_0^2$ критична область статистики критерію є однією з областей $D_k' = [0, \chi_\alpha^2(n-1)],\ D_k'' = [\chi_{1-\alpha}^2(n-1), +\infty)$ відповідно.

5.5 Критерій Стьюдента

Нехай $X = (X_1, X_2, \ldots, X_n)$ та $Y = (Y_1, Y_2, \ldots, Y_m)$ — незалежні вибірки з розподілів $N(a_1, \sigma^2)$ та $N(a_2, \sigma^2)$ відповідно. Математичні сподівання a_1, a_2 — невідомі, дисперсія σ^2 — однакова для обидвох розподілів і також невідома. Щодо параметрів a_1 та a_2 висувається гіпотеза $H_0 : a_1 - a_2 = a_0$, де a_0 — деяка відома стала. Зокрема, при $a_0 = 0$, гіпотеза H_0 полягає в тому, що математичні сподівання a_1, a_2 рівні між собою, або, що те саме, $\mathbf{X}$ та $\mathbf{Y}$ є вибірками з одного й того ж нормального розподілу. Альтернативною будемо вважати гіпотезу $H_1 : a_1 - a_2 \neq a_0$. Необхідно за реалізаціями $\mathbf{x}$ та $\mathbf{y}$ цих вибірок зробити висновок: відхиляти чи не відхиляти нульову гіпотезу.

Побудуємо критерій Стьюдента для перевірки гіпотези H_0. Нехай $\bar{X}, \bar{Y}$ — відповідні вибіркові середні для $\mathbf{X}$ та $\mathbf{Y}$, а $S_1^2,\ S_2^2$ — виправлені вибіркові дисперсії. Тоді з незалежності вибірок і теореми про вибіркові моменти нормальної вибірки випливає, що величини $\bar{X}, \bar{Y}, S_1^2, S_2^2$ незалежні. До того ж, величини $\bar{X}, \bar{Y}$ мають нормальний розподіл $N\left(a_1, \frac{\sigma^2}{n}\right)$, $N\left(a_2, \frac{\sigma^2}{m}\right)$ відповідно, а величини $(n-1)\frac{S_1^2}{\sigma^2}$, $(m-1)\frac{S_2^2}{\sigma^2}$ — розподіли χ^2 з $n-1$ та $m-1$ ступенями свободи відповідно. Тому випадкова величина

$$\sqrt{\frac{nm}{n+m}}\frac{\bar{X} - \bar{Y} - a_0}{\sigma}$$

за припущення H_0 має стандартний нормальний розподіл і не залежить від суми

$$(n-1)\frac{S_1^2}{\sigma^2} + (m-1)\frac{S_2^2}{\sigma^2},$$

яка за означенням має χ^2-розподіл з $(n-1) + (m-1) = n + m - 2$ ступенями свободи.

За означенням розподілу Стьюдента величина, що утворена відношенням

$$\tau_{n+m-2} = \sqrt{\frac{nm}{n+m}} \frac{\bar{X} - \bar{Y} - a_0}{(\sqrt{((n-1)S_1^2 + (m-1)S_2^2)/(n+m-2)}},$$

у випадку істинності гіпотези H_0, має розподіл Стьюдента з $n+m-2$ ступенями свободи і тому

$$\mathrm{P}_{H_0}\left\{|\tau_{n+m-2}| \geq t_{1-\frac{\alpha}{2}}(n+m-2)\right\} = \alpha,$$

де $t_{1-\frac{\alpha}{2}}(n+m-2)$ — квантиль рівня $1-\frac{\alpha}{2}$ розподілу Стьюдента з $n+m-2$ ступенями свободи.

Отже, критерій Стьюдента переверки гіпотези H_0 можна сформулювати так: гіпотеза H_0 відхиляється, якщо значення $|\tau_{n+m-2}|$, підраховане за реалізаціями $\mathbf{x}, \mathbf{y}$ вибірок $\mathbf{X}, \mathbf{Y}$ відповідно, потрапляє у критичну область $D_k = \left[t_{1-\frac{\alpha}{2}}(n+m-2), +\infty\right)$, і не відхиляється у протилежному разі. При цьому, ймовірність помилки при відхиленні H_0 (помилки першого роду) дорівнює α.

Вправа. Побудувати критерій для перевірки гіпотези про рівність математичних сподівань двох незалежних нормальних вибірок з відомими дисперсіями, що є різними.

5.6 Критерій Фішера

Нехай $X = (X_1, X_2, \ldots, X_n)$ та $Y = (Y_1, Y_2, \ldots, Y_m)$ — незалежні вибірки з розподілів $N(a_1, \sigma_1^2)$ та $N(a_2, \sigma_2^2)$ відповідно. Параметри $a_1, a_2, \sigma_1^2, \sigma_2^2$ — невідомі. Щодо дисперсій σ_1^2 та σ_2^2 висувається гіпотеза $H_0 : \sigma_1^2 = \sigma_2^2$, тобто гіпотеза, яка полягає в тому, що $\mathbf{X}$ та $\mathbf{Y}$ є вибірками з нормального розподілу з однією й тією самою дисперсією. Альтернативною будемо вважати гіпотезу $H_1 : \sigma_1^2 \neq \sigma_2^2$. Необхідно за реалізаціями $\mathbf{x}$ та $\mathbf{y}$ цих вибірок зробити висновок: відхиляти чи не відхиляти нульову гіпотезу.

Побудуємо критерій Фішера для перевірки гіпотези H_0. Нехай S_1^2 та S_2^2 — виправлені вибіркові дисперсії для вибірок $\mathbf{X}$ та $\mathbf{Y}$ відповідно. Тоді з незалежності вибірок випливає, що величини S_1^2, S_2^2 незалежні. За теоремою про вибіркові моменти нормальної вибірки величини $(n-1)\frac{S_1^2}{\sigma_1^2}$ та $(m-1)\frac{S_2^2}{\sigma_2^2}$ мають розподіли χ^2 з $n-1$ та $m-1$ ступенями свободи відповідно і тому, у випадку істинності H_0, відношення

$$\phi_{n-1,m-1} = \frac{S_1^2}{S_2^2}$$

має розподіл Фішера з $n-1,\ m-1$ ступенями свободи. Це означає, що

$$\mathrm{P}_{H_0}\left\{f_{\frac{\alpha}{2}}(n-1,m-1) < \phi_{n-1,m-1} < f_{1-\frac{\alpha}{2}}(n-1,m-1)\right\} = \\ = 1-\alpha,$$

або

$$\mathrm{P}_{H_0}\left\{\phi_{n-1,m-1} \in D_k\right\} = \alpha,$$
$$D_k = [0, f_{\frac{\alpha}{2}}(n-1,m-1)] \cup [f_{1-\frac{\alpha}{2}}(n-1,m-1), +\infty),$$

де $f_\gamma(n-1,m-1)$ — квантиль рівня γ розподілу Фішера з $n-1,\ m-1$ ступенями свободи.

Отже, критерій Фішера з рівнем значущості α полягає в тому, що якщо значення $\phi_{n-1,m-1}$, підраховане за реалізаціями $\mathbf{x}, \mathbf{y}$ вибірок $\mathbf{X}, \mathbf{Y}$ відповідно, потрапляє у критичну область D_k, то гіпотеза H_0 відхиляється і не відхиляється у протилежному разі.

Вправа. Сформулювати критерій Фішера для випадку, коли a_1, a_2 — відомі.

5.7 Критерій χ^2

Нехай $\mathbf{X} = (X_1, X_2, \ldots, X_n)$ — вибірка із невідомого розподілу F, стосовно якого висувається гіпотеза $H_0 : F = G$, де G належить заданому класу розподілів (зокрема, G може бути повністю визначеним розподілом). Інакше кажучи, гіпотезу H_0 можна сформулювати так: $\mathbf{X}$ є вибіркою з розподілу G. Необхідно за реалізацією $\mathbf{x}$ вибірки $\mathbf{X}$ зробити висновок про гіпотезу H_0, а саме: відхиляти її чи ні.

5.7.1 Критерій χ^2 (гіпотетичний розподіл визначений)

Ідея побудови критерію для перевірки гіпотези H_0 грунтується на тому, що емпіричний розподіл $\hat{F}_n(x)$, побудований за реалізацією $\mathbf{x}$ вибірки $\mathbf{X}$ з F, мало відрізняється від розподілу F. Справді, за теоремою про властивості емпіричної функції розподілу, для кожного $x \in \mathbb{R}$ значення емпіричної функції розподілу $\hat{F}_n(x)$ є незміщеною і консистентною оцінкою $F(x)$. Тому, якщо ввести відхилення $D(\hat{F}_n, G)$ емпіричного розподілу від гіпотетичного, причому так, щоб воно набувало малих значень, коли гіпотеза H_0 справджується, і великих, коли вона не справджується, то гіпотезу H_0 природно відхиляти або не відхиляти залежно від того, якого значення набуло відхилення $D(\hat{F}_n, G)$ для конкретної реалізації $\mathbf{x}$ вибірки $\mathbf{X}$ — великого чи малого.

Відхилення $D(\hat{F}_n, G)$ можна будувати різними способами. Критерій χ^2 побудований на підставі відхилення $\hat{F}_n$ від G, запропонованого К. Пірсоном. Це відхилення будується так: область значень гіпотетичого розподілу G розбивається на скінченну кількість неперетинних множин Δ_i, $i = \overline{1,k}$, і як відхилення розглядається

$$D(\hat{F}_n, G) = \sum_{i=1}^{k} \frac{n}{p_i}\left(\frac{\nu_i}{n} - p_i\right) = \sum_{i=1}^{k} \frac{(\nu_i - np_i)^2}{np_i},$$

де p_i — ймовірність того, що випадкова величина X_j потрапить до множини Δ_i обчислена за гіпотетичним розподілом G, тобто $p_i = \mathrm{P}\{X_j \in \Delta_i\} = G(\Delta_i)$; $\nu_i = \nu_i(\mathbf{X}) = \sum_{j=1}^{n} \mathbb{I}_{\{X_j \in \Delta_i\}}$ — випадкова величина, яка при кожній реалізації $\mathbf{x}$ вибірки $\mathbf{X}$ дорівнює кількості значень серед x_1, $x_2, \ldots, x_n$, що потрапили до Δ_i.

Розглянемо випадкову величину $\hat{p}_i = \frac{\nu_i}{n} = \frac{1}{n}\sum_{j=1}^{n} I_{\{X_j \in \Delta_i\}}$, $i = \overline{1,k}$. При фіксованих значеннях вибірки $\hat{p}_i$ — це ймовірність того, що випадкова величина X_j потрапить до множини Δ_i, обчислена за емпіричним розподілом $\hat{F}_n$. Цілком аналогічно доведенню теореми про властивості емпіричної функції розподілу встановлюється, що випадкова величина $\hat{p}_i$ є незміщеною та консистентною оцінкою ймовірності $\mathrm{P}\{X_j \in \Delta_i\} = F(\Delta_i)$. Це означає, що у випадку істинності H_0 ця оцінка є незміщеною та консистентною для p_i, і тому відхилення $D(\hat{F}_n, G)$ у цьому випадку мінімальне (порівняно з відхиленнями $D(\hat{F}_n, G^*)$, коли розподіли G^* відмінні від F).

Теорема Пірсона. *Якщо справджується гіпотеза H_0, то для всіх $x > 0$*

$$\lim_{n\to\infty} \mathrm{P}\{D(\hat{F}_n, G) < x\} = F_{\chi^2_{k-1}}(x),$$

де $F_{\chi^2_{k-1}}(x)$ — функція χ^2-розподілу з $k-1$ ступенями свободи.

◀ Доведення проведемо для випадку $k = 2$. У цьому випадку $\nu_2 = n - \nu_1$, $p_2 = 1 - p_1$. Випадкову величину $D(\hat{F}_n, G)$ подамо у вигляді

$$D(\hat{F}_n, G) = \frac{(\nu_1 - np_1)^2}{np_1} + \frac{(\nu_2 - np_2)^2}{np_2} = \frac{(\nu_1 - np_1)^2}{np_1} +$$
$$+ \frac{(n - \nu_1 - n(1-p_1))^2}{n(1-p_1)} = \frac{(\nu_1 - np_1)^2}{np_1} + \frac{(-\nu_1 + np_1)^2}{n(1-p_1)} =$$
$$= \frac{(\nu_1 - np_1)^2}{n}\left(\frac{1}{p_1} + \frac{1}{1-p_1}\right) =$$
$$= \frac{(\nu_1 - np_1)^2}{np_1(1-p_1)} = \left(\frac{\nu_1 - np_1}{\sqrt{np_1(1-p_1)}}\right)^2.$$

Оскільки величина ν_1 є сумою n незалежних однаково розподілених випадкових величин: $\nu_1 = \sum\limits_{j=1}^{n} I_{\{X_j \in \Delta_1\}}$ і її моменти

$$\mathrm{M}\nu_1 = \sum_{j=1}^{n} \mathrm{M}I_{\{X_j \in \Delta_1\}} = np_1,$$
$$\mathrm{D}\nu_1 = \sum_{j=1}^{n} \mathrm{D}I_{\{X_j \in \Delta_1\}} = np_1(1-p_1),$$

за центральною граничною теоремою для всіх $x \in \mathbb{R}$

$$\mathrm{P}\left\{\frac{\nu_1 - np_1}{\sqrt{np_1(1-p_1)}} < x\right\} \longrightarrow F_0(x) = \frac{1}{\sqrt{2\pi}}\int\limits_{-\infty}^{x} e^{-\frac{t^2}{2}}dt,$$

при $n \to \infty$. Останнє співвідношення означає, що граничний розподіл випадкової величини $Z_n = \frac{\nu_1 - np_1}{\sqrt{np_1(1-p_1)}}$ співпадає зі стандартним нормальним розподілом. Тоді граничний розподіл випадкової величини Z_n^2 співпадає із χ^2 розподілом з одним ступенем свободи. Отже, для всіх $x > 0$

$$\mathrm{P}\{D(\hat{F}_n, G) < x\} = \mathrm{P}\{Z_n^2 < x\} \longrightarrow F_{\chi^2_{k-1}}(x),$$

при $n \to \infty$, а це й треба було довести. ▶

З теореми Пірсона випливає, що при великих обсягах вибірки

$$\mathrm{P}\{D(\hat{F}_n, G) \geq \chi^2_{1-\alpha}(k-1)|H_0\} \approx \alpha,$$

де α — заданий рівень значущості, $\chi^2_{1-\alpha}(k-1)$ — квантиль рівня $1-\alpha$ розподілу χ^2 з $k-1$ ступенями свободи. Отже критерій χ^2 з рівнем значущості α полягає у тому, що гіпотеза H_0 відхиляється, якщо значення $D(\hat{F}_n, G)$, підраховане за реалізацією $\mathbf{x}$ вибірки $\mathbf{X}$, потрапляє у критичну область $[\chi^2_{1-\alpha}(k-1), +\infty)$, і не відхиляється у протилежному разі.

Зауваження. Критерієм χ^2 можна користуватися у випадку, коли $np_i \geq 10$, $i = \overline{1,k}$. Якщо для деяких i ця умова не виконується, то відповідні множини Δ_i потрібно об'єднати з сусідніми.

5.7.2 Критерій χ^2 (гіпотетичний розподіл невизначений)

Нехай розподіл G визначений з точністю до невідомого векторного параметра $\theta = (\theta_1, \theta_2, \ldots, \theta_s)$, який набуває значень із множини $\Theta \subset \mathbb{R}^s$. У цьому випадку гіпотеза $H_0 : F(\cdot) = G(\cdot; \theta)$ означає, що $\mathbf{X}$ є вибіркою із розподілу, що належить до класу розподілів $G(\cdot; \theta)$, $\theta \in \Theta$.

Оскільки параметр θ невідомий, замість нього будемо використовувати значення його оцінки $\hat{\theta} = (\hat{\theta}_1, \hat{\theta}_2, \ldots, \hat{\theta}_s)$, підраховане за реалізацією $\mathbf{x}$ вибірки $\mathbf{X}$, і отже, як гіпотетичний розглядати розподіл $G(\cdot;\hat{\theta})$. Р. Фішер встановив, що коли гіпотеза H_0 справджується і оцінка $\hat{\theta}$ є такою, що її значення при кожній реалізації x вибірки X мінімізує відхилення

$$D(\hat{F}_n, G) = \sum_{i=1}^{k} \frac{(\nu_i - np_i(\hat{\theta}))^2}{np_i(\hat{\theta})},$$

між $\hat{F}_n$ і G, то розподіл $D(\hat{F}_n, G)$ при $n \to \infty$ прямує до χ^2-розподілу з $(k-1-s)$ ступенями свободи, де $s = \dim(\theta)$ — розмірність векторного параметра θ.

Отже, у випадку, коли шукану оцінку $\hat{\theta}$ знайдено, критерієм χ^2 з рівнем значущості α можна користуватися у такому формулюванні: гіпотеза H_0 відхиляється, якщо значення $D(\hat{F}_n, G)$, підраховане за реалізацією $\mathbf{x}$ вибірки $\mathbf{X}$, потрапляє у критичну область $[\chi^2_{1-\alpha}(k-1-s), +\infty)$, і не відхиляється у протилежному разі.

Зауваження. Оцінку $\hat{\theta} = (\hat{\theta}_1, \hat{\theta}_2, \ldots, \hat{\theta}_s)$, що мінімізує відхилення $D(\hat{F}_n, G)$ можна отримати як оцінку максимальної правдоподібності для θ, побудовану за частотами $\nu = (\nu_1, \nu_2, \ldots, \nu_k)$. Функція правдоподібності має вигляд:

$$L(\theta) = \prod_{i=1}^{k} (p_i(\theta))^{\nu_i},$$

де

$$\sum_{i=1}^{k} p_i(\theta) = 1, \quad \sum_{i=1}^{k} \nu_i(\theta) = n.$$

5.7.3 Критерій χ^2 як критерій незалежності

Нехай X та Y — дві дискретні випадкові величини, які можуть набувати значення $a_1, a_2, \dots a_s$ та $b_1, b_2, \dots b_r$ відповідно. За результатами n спостережень $(x_1, y_1), (x_2, y_2), \dots (x_n, y_n)$ випадкового вектора (X, Y) потрібно перевірити гіпотезу H_0 : випадкові величини X та Y — незалежні. Позначимо

$$p_{ij} = \mathrm{P}\{X = a_i, Y = b_j\}, \quad i = \overline{1, s},\ j = \overline{1, r},$$

$$p_{i\bullet} = \mathrm{P}\{X = a_i\} = \sum_{j=1}^{r} p_{ij}, \quad p_{\bullet j} = \mathrm{P}\{Y = b_j\} = \sum_{i=1}^{s} p_{ij}.$$

Тоді гіпотезу H_0 можна сформулювати так:

$$H_0 : p_{ij} = p_{i\bullet} p_{\bullet j}, \quad i = \overline{1, s},\ j = \overline{1, r},$$

або так: вибірка[1]

$$(\mathbf{X}, \mathbf{Y}) = ((X_1, Y_1), (X_2, Y_2), \dots (X_n, Y_n))$$

є вибіркою із розподілу G : $((a_i, b_j); p_{i\bullet} p_{\bullet j})$, $i = \overline{1, s}$, $j = \overline{1, r}$, визначеного з точністю до $(s+r-2)$-вимірного векторного параметра

$$\theta = (p_{1\bullet}, p_{2\bullet}, \dots p_{s-1\bullet}, p_{\bullet 1}, p_{\bullet 2}, \dots, p_{\bullet r-1})$$

(значення $p_{s\bullet}$ та $p_{\bullet r}$ знаходяться з рівностей $p_{s\bullet} = 1 - \sum\limits_{i=1}^{s-1} p_{i\bullet}$, $p_{\bullet r} = 1 - \sum\limits_{j=1}^{r-1} p_{\bullet j}$).

---

[1] Під багатовимірною вибіркою розуміємо впорядкований набір незалежних і однаково розподілених випадкових векторів $X_1, X_2, \dots, X_n$. Усі вищенаведені результати поширюються на випадок багатовимірної вибірки $\mathbf{X}$.

Для перевірки цієї гіпотези скористаємося критерієм χ^2. Область значень гіпотетичного розподілу розіб'ємо на множини Δ_{ij}, $i=\overline{1,s}$, $j=\overline{1,r}$, кожна з яких складається лише з однієї точки (a_i, b_j). Нехай ν_{ij} — кількість значень серед (x_1, y_1), $(x_2, y_2), \ldots (x_n, y_n)$, що потрапили до Δ_{ij}. За частотами (ν_{ij}) знайдемо оцінку максимальної правдоподібності невідомого векторного параметра θ. У випадку істинності H_0, функція правдоподібності $L(\theta)$ має вигляд:

$$L(\theta) = \prod_{i=1}^{s} \prod_{j=1}^{r} p_{ij}^{\nu_{ij}} = \prod_{i=1}^{s} p_{i\bullet}^{\nu_{i\bullet}} \prod_{j=1}^{r} p_{\bullet j}^{\nu_{\bullet j}},$$

де

$$\nu_{i\bullet} = \sum_{j=1}^{r} \nu_{ij}, \quad \nu_{\bullet j} = \sum_{i=1}^{s} \nu_{ij}.$$

Тоді

$$\begin{aligned}
\ln L(\theta) &= \sum_{i=1}^{s} \nu_{i\bullet} \ln p_{i\bullet} + \sum_{j=1}^{r} \nu_{\bullet j} \ln p_{\bullet j} = \\
&= \sum_{i=1}^{s-1} \nu_{i\bullet} \ln p_{i\bullet} + \sum_{j=1}^{r-1} \nu_{\bullet j} \ln p_{\bullet j} + \nu_{s\bullet} \ln p_{s\bullet} + \nu_{\bullet r} \ln p_{\bullet r} = \\
&= \sum_{i=1}^{s-1} \nu_{i\bullet} \ln p_{i\bullet} + \sum_{j=1}^{r-1} \nu_{\bullet j} \ln p_{\bullet j} + \\
&+ \left(n - \sum_{i=1}^{s-1} \nu_{i\bullet}\right) \ln \left(1 - \sum_{i=1}^{s-1} p_{i\bullet}\right) + \\
&+ \left(n - \sum_{j=1}^{r-1} \nu_{\bullet j}\right) \ln \left(1 - \sum_{j=1}^{r-1} p_{\bullet j}\right).
\end{aligned}$$

Продиференціювавши останню рівність за $p_{1\bullet}, p_{2\bullet}, \ldots, p_{s-1\bullet}$, $p_{\bullet 1}, p_{\bullet 2}, \ldots, p_{\bullet r-1}$, отримуємо систему рівнянь

$$\frac{\partial \ln L(\theta)}{\partial p_{i\bullet}} = \frac{\nu_{i\bullet}}{p_{i\bullet}} - \frac{n - \sum\limits_{i=1}^{s-1} \nu_{i\bullet}}{1 - \sum\limits_{i=1}^{s-1} p_{i\bullet}} = 0, \quad i = \overline{1, s-1},$$

$$\frac{\partial \ln L(\theta)}{\partial p_{\bullet j}} = \frac{\nu_{\bullet j}}{p_{\bullet j}} - \frac{n - \sum\limits_{j=1}^{r-1} \nu_{\bullet j}}{1 - \sum\limits_{j=1}^{r-1} p_{\bullet j}} = 0, \quad j = \overline{1, r-1}.$$

Розв'язавши цю систему, встановлюємо, що оцінка максимальної правдоподібності параметра θ має вигляд

$$\hat{p}_{i\bullet} = \frac{\nu_{i\bullet}}{n}, \quad \hat{p}_{\bullet j} = \frac{\nu_{\bullet j}}{n},$$

тобто

$$\hat{\theta} = \frac{1}{n}(\nu_{1\bullet}, \nu_{2\bullet}, \ldots, \nu_{s-1\bullet}, \nu_{\bullet 1}, \nu_{\bullet 2}, \ldots, \nu_{\bullet r-1}).$$

Отже, у випадку, коли гіпотеза H_0 справджується, розподіл відхилення

$$D(\hat{F}_n, G) = \sum_{i=1}^{s} \sum_{i=1}^{r} \frac{(\nu_{ij} - n\hat{p}_{i\bullet}\hat{p}_{\bullet j})^2}{n\hat{p}_{i\bullet}\hat{p}_{\bullet j}} =$$

$$= n \sum_{i=1}^{s} \sum_{i=1}^{r} \frac{(\nu_{ij} - \hat{\nu}_{i\bullet}\hat{\nu}_{\bullet j}/n)^2}{\hat{\nu}_{i\bullet}\hat{\nu}_{\bullet j}}$$

прямує до χ^2-розподілу з $rs-1-(s+r-2) = (s-1)(r-1)$ ступенями свободи. Відповідно до критерію χ^2, гіпотеза H_0 відхиляється, якщо значення $D(\hat{F}_n, G)$ потрапляє у критичну область $[\chi^2_{1-\alpha}((s-1)(r-1)), +\infty)$, і не відхиляється у протилежному разі. При цьому, ймовірність помилки першого роду дорівнює α.

Розділ 6

Лінійна регресія

6.1 Метод найменших квадратів

У процесі практичного застосування методів математичної статистики часто виникає задача, найпростіший варіант якої сформульовано нижче.

Припустимо, що між величинами y та x існує лінійна залежність виду

$$y = \alpha + \beta x, \tag{6.1}$$

де x — деяка невипадкова змінна, α, β — невідомі коефіцієнти, які необхідно визначити експериментально. Описану залежність величини y від x називають простою лінійною регресією, а коефіцієнти α, β — коефіцієнтами регресії.

Нехай відбувається n незалежних експериментів. У кожному з них величина x набуває деяких цілком визначених значень x_i, що є відомими і, взагалі кажучи, різними для різних експериментів. Припустимо, що в кожному експерименті спостерігається нормальна випадкова величина Y_i, $i = \overline{1, n}$, з математичним сподіванням

$$\mathrm{M}Y_i = \alpha + \beta x_i, \quad i = \overline{1, n}, \tag{6.2}$$

і невідомою дисперсією σ^2. Таким чином, розглядається вибірка $\mathbf{Y} = (Y_1, Y_2, \ldots, Y_n)$, утворена послідовністю незалежних нормально розподілених випадкових величин з однаковими дисперсіями σ^2 і математичними сподіваннями вигляду (6.2).

Запишемо рівняння регресії у вигляді

$$y = a + b(x - \bar{x}),$$

де

$$a = \alpha + b\bar{x}, \quad b = \beta, \quad \bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i.$$

Оцінимо коефіцієнти a, b.

Функція правдоподібності вибірки $\mathbf{Y}$ для довільної її реалізації $\mathbf{y} = (y_1, y_2, \ldots, y_n)$ має вигляд

$$L(a, b, \sigma^2) = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi}\sigma} \exp\left\{-\frac{(y_i - \mathrm{M}Y_i)^2}{2\sigma^2}\right\} =$$

$$= (2\pi\sigma^2)^{-\frac{n}{2}} \exp\left\{-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(y_i - a - b(x_i - \bar{x}))^2\right\}.$$

Оцінки трьох невідомих параметрів a, b, σ^2 побудуємо методом максимальної правдоподібності. Для того, щоб відшукати максимум функції правдоподібності, спочатку мінімізуємо квадратичну форму

$$Q(a, b) = \sum_{i=1}^{n}(y_i - a - b(x_i - \bar{x}))^2,$$

а потім, знайдемо максимум за θ функції

$$\varphi(\theta) = (2\pi\theta)^{-\frac{n}{2}} \exp\left\{-\frac{Q^*}{2\theta}\right\},$$

де $Q^* = \min\limits_{a,b} Q(a,b)$. Оскільки

$$\begin{aligned} Q(a,b) &= \sum_{i=1}^{n} \left((\bar{y}-a) + (y_i - \bar{y}) - b(x_i - \bar{x})\right)^2 = \\ &= n\left((\bar{y}-a)^2 + b^2 s_1^2 + s_2^2 - 2bk_{12}\right) = \\ &= n\left((\bar{y}-a)^2 + \left(s_1 b - \frac{k_{12}}{s_1}\right)^2 + s_2^2 - \frac{k_{12}^2}{s_1^2}\right), \end{aligned}$$

де

$$s_1^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2, \quad s_2^2 = \frac{1}{n}\sum_{i=1}^{n}(y_i - \bar{y})^2,$$

$$k_{12} = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y}),$$

то мінімум Q^* функції $Q(a,b)$ досягається при

$$a = \hat{a} = \bar{y}, \quad b = \hat{b} = \frac{k_{12}}{s_1^2}$$

і дорівнює

$$Q^* = \min_{a,b} Q(a,b) = \sum_{i=1}^{n}(y_i - \hat{a} - \hat{b}(x_i - \bar{x}))^2 = n s_2^2 (1 - \hat{r}^2),$$

де

$$\hat{r} = \frac{k_{12}}{s_1 s_2} = \frac{\sum\limits_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum\limits_{i=1}^{n}(x_i - \bar{x})^2}\sqrt{\sum\limits_{i=1}^{n}(y_i - \bar{y})^2}}.$$

Знайдемо тепер значення θ, при якому функція $\varphi(\theta)$ досягає максимуму. З огляду на те, що функції $\varphi(\theta)$ і $\ln\varphi(\theta)$ досягають максимуму в одній і тій самій точці, шукане значення θ є розв'язком рівняння (необхідна умова екстремуму):

$$(\ln\varphi(\theta))' = 0.$$

Спростивши похідну у правій частині цього рівняння, отримуємо:

$$(\ln\varphi(\theta))' = \left(-\frac{n}{2}\ln(2\pi\theta) - \frac{Q^*}{2\theta}\right)' = -\frac{n}{2\theta} + \frac{Q^*}{2\theta^2} = 0.$$

Звідси

$$\hat{\theta} = \hat{\sigma^2} = \frac{Q^*}{n} = s_2^2(1 - \hat{r}^2).$$

Отже, оцінками невідомих параметрів a, b, σ^2 є

$$\hat{a} = \bar{Y} = \frac{1}{n}\sum_{i=1}^{n} Y_i, \quad \hat{b} = \frac{\sum\limits_{i=1}^{n}(x_i - \bar{x})(Y_i - \bar{Y})}{\sum\limits_{i=1}^{n}(x_i - \bar{x})^2}, \tag{6.3}$$

$$\hat{\sigma^2} = \frac{1}{n}\sum_{i=1}^{n}(Y_i - \hat{a} - \hat{b}(x_i - \bar{x}))^2.$$

Зауваження. Метод оцінювання параметрів, що полягає у мінімізації квадратичної форми $Q(a, b)$, називається методом найменших квадратів, а відповідні оцінки — оцінками методу найменших квадратів.

6.2 Властивості оцінок методу найменших квадратів

Дослідимо властивості побудованих вище оцінок $\hat{a}$, $\hat{b}$, $\hat{\sigma^2}$. Відзначимо, що

$$\mathrm{M}\hat{a} = \frac{1}{n}\sum_{i=1}^{n}\mathrm{M}Y_i = \frac{1}{n}\sum_{i=1}^{n}(a + b(x_i - \bar{x})) = a,$$

$$\mathrm{M}\hat{b} = \frac{1}{ns_1^2}\sum_{i=1}^{n}(x_i - \bar{x})\mathrm{M}(Y_i - \bar{Y}) =$$

$$= \frac{1}{ns_1^2}\sum_{i=1}^{n}(x_i - \bar{x})\mathrm{M}(Y_i - \hat{a}) =$$

$$= \frac{b}{ns_1^2}\sum_{i=1}^{n}(x_i - \bar{x})^2 = \frac{b}{ns_1^2}ns_1^2 = b,$$

а тому оцінки $\hat{a}$ та $\hat{b}$ є незміщеними. Оскільки величини Y_i незалежні,

$$\mathrm{D}\hat{a} = \frac{1}{n^2}\sum_{i=1}^{n}\mathrm{D}Y_i = \frac{\sigma^2}{n}.$$

Враховуючи те, що

$$\hat{b} = \frac{1}{ns_1^2}\sum_{i=1}^{n}(x_i - \bar{x})(Y_i - \bar{Y}) = \frac{1}{ns_1^2}\sum_{i=1}^{n}(x_i - \bar{x})Y_i,$$

отримуємо

$$\mathrm{D}\hat{b} = \frac{1}{n^2s_1^4}\sum_{i=1}^{n}(x_i - \bar{x})^2\mathrm{D}Y_i = \frac{1}{n^2s_1^4}\sum_{i=1}^{n}(x_i - \bar{x})^2\sigma^2 = \frac{\sigma^2}{ns_1^2}.$$

З отриманих формул для дисперсій оцінкок $\hat{a}, \hat{b}$ і з другої нерівності Чебишева випливає, що ці оцінки є консистентними для параметрів a, b відповідно.

Далі, розглянемо квадратичну форму $Q(a,b)$. Оскільки випадкові величини $U_i = Y_i - a - b(x_i - \bar{x})$ незалежні і мають розподіл $N(0,\sigma^2)$, то величина $\frac{Q(a,b)}{\sigma^2}$ має χ^2-розподіл з n ступенями свободи. Записавши $Q(a,b)$ у вигляді

$$Q(a,b) = \sum_{i=1}^{n}(Y_i - \hat{a} - \hat{b}(x_i - \bar{x}) + (\hat{a} - a) + (\hat{b} - b)(x_i - \bar{x}))^2,$$

нескладно перевірити, що

$$Q(a,b) = \sum_{i=1}^{n}(Y_i - \hat{a} - \hat{b}(x_i - \bar{x}))^2 + n(\hat{a} - a)^2 + ns_1^2(\hat{b} - b)^2 =$$
$$= n\widehat{\sigma^2} + n(\hat{a} - a)^2 + ns_1^2(\hat{b} - b)^2.$$

Тому величини

$$\chi^2_{n-2} = \frac{n\widehat{\sigma^2}}{\sigma^2}, \quad Z_1 = \sqrt{n}\frac{\hat{a} - a}{\sigma}, \quad Z_2 = \sqrt{n}\frac{s_1(\hat{b} - b)}{\sigma}$$

незалежні, причому величина χ^2_{n-2} має χ^2-розподіл з $n-2$ ступенями свободи, а величини Z_1 та Z_2 мають нормальний розподіл, як лінійні комбінації незалежних нормальних величин.

Тепер, спираючись на властивості χ^2-розподілу (див. 2.3), знайдемо математичне сподівання і дисперсію оцінки $\widehat{\sigma^2}$:

$$\mathrm{M}\widehat{\sigma^2} = \frac{\sigma^2 \mathrm{M}\chi^2_{n-2}}{n} = \frac{(n-2)\sigma^2}{n},$$
$$\mathrm{D}\widehat{\sigma^2} = \frac{\sigma^4 \mathrm{D}\chi^2_{n-2}}{n^2} = \frac{2(n-2)\sigma^4}{n^2}.$$

З отриманих формул випливає, що оцінка $\widehat{\sigma^2}$ є консистентною, але зміщеною, і тому замість неї для оцінювання невідомого параметра σ^2 використовують її незміщений варіант

$$\tilde{\sigma^2} = \frac{n}{n-2}\hat{\sigma^2}. \tag{6.4}$$

Розглянемо випадкові величини Z_1, Z_2. З наведених вище формул для математичного сподівання і дисперсії оцінок $\hat{a}$ та $\hat{b}$ випливає, що Z_1 та Z_2 є стандартними нормальними величинами. Тоді величини

$$T_1 = \frac{Z_1}{\sqrt{\chi^2_{n-2}/(n-2)}} = \sqrt{n}\frac{\hat{a}-a}{\tilde{\sigma}},$$

$$T_2 = \frac{Z_2}{\sqrt{\chi^2_{n-2}/(n-2)}} = s_1\sqrt{n}\frac{\hat{b}-b}{\tilde{\sigma}}$$

за означенням мають розподіл Стьюдента з $n-2$ ступенями свободи і, отже,

$$\mathrm{P}\left\{|T_i| < t_{1-\frac{\alpha}{2}}\right\} = 1-\alpha, \quad i = 1,2.$$

Останнє співвідношення дозволяє як перевірити гіпотезу щодо значень коефіцієнтів регресії a, b, так і побудувати надійні інтервали для них.

Отримані результати сформулюємо у вигляді теореми:

Теорема (про властивості оцінок методу найменших квадратів). Оцінки $\hat{a}, \hat{b}, \tilde{\sigma^2}$ з (6.3),(6.4) параметрів a, b, σ^2 є незміщеними, консистентними та незалежними між собою. До того ж, надійнійні інтервали для коефіцієнтів a, b з рівнем надійності $1-\alpha$, мають вигляд

$$\hat{a} - \frac{\tilde{\sigma}}{\sqrt{n}}t_{1-\frac{\alpha}{2}}(n-2) < a < \hat{a} + \frac{\tilde{\sigma}}{\sqrt{n}}t_{1-\frac{\alpha}{2}}(n-2),$$

$$\hat{b} - \frac{\tilde{\sigma}}{s_1\sqrt{n}}t_{1-\frac{\alpha}{2}}(n-2) < b < \hat{b} + \frac{\tilde{\sigma}}{s_1\sqrt{n}}t_{1-\frac{\alpha}{2}}(n-2),$$

де $t_{1-\frac{\alpha}{2}}(n-2)$ — квантиль рівня $1-\alpha$ розподілу Стьюдента з $n-2$ ступенями свободи.

Вправа. Побудувати надійний інтервал для невідомої дисперсії σ^2.

Література

[1] *Бернштейн С.Н. Теория вероятностей. 4-е изд.* – М.: Гостехиздат, 1946.

[2] *Большев Л.Н., Смирнов Н.В. Таблицы математической статистики.* – М.: Наука, 1983.

[3] *Борель Э. Вероятность и достоверность.* – М.: Физматгиз, 1961.

[4] *Боровков А.А. Теория вероятностей.* – М.: Наука, 1986.

[5] *Боровков А.А. Математическая статистика.* – М.: Наука, 1984.

[6] *Ван дер Варден Б.Л. Математическая статистика.* – М.: Изд-во иностр. лит., 1960.

[7] *Венцель Е.С. Теория вероятностей.* – М.: Высшая школа, 1998.

[8] *Гихман И.И., Скороход А.В., Ядренко М.Й. Теория вероятностей и математическая статистика.* – К.: Вища школа, 1979.

[9] *Гмурман В.Е. Теория вероятностей и математическая статистика.* – М.: Высшая школа, 2003.

[10] *Гнеденко Б.В. Курс теории вероятностей.* – М.: Едиториал УРСС, 2005.

[11] *Гнеденко Б.В., Хинчин А.Я. Элементарное введение в теорию вероятностей.* – М.: Наука, 1976.

[12] *Горяинов В.Б., Павлов И.В., Цветкова Г.М. и др. Математическая статистика.* – М.: Изд-во МГТУ им. Н.Э. Баумана, 2001.

[13] *Ивашев-Мусатов О.С. Теория вероятностей и математическая статистика.* – М.: Наука, 1979.

[14] *Ивченко Г.И., Медведев Ю.И. Математическая статистика.* – М.: Высшая школа, 1992.

[15] *Карташов М.В. Імовірність, процеси, статистика.* – К.: Видавничо-поліграфічний центр «Київський університет», 2008.

[16] *Колмогоров А.Н. Основные понятия теории вероятностей.* – М.: ОНТИ, 1936.

[17] *Королюк В.С., Портенко Н.И., Скороход А.В., Турбин А.Ф. Справочник по теории вероятностей и математической статистике.* – М.: Наука, 1985.

[18] *Крамер Г. Математические методы статистики.* – М.: Мир, 1975.

[19] *Ламперти Дж. Вероятность.* – М.: Наука, 1973.

[20] *Леман Э. Проверка статистических гипотез.* – М.: Наука, 1979.

[21] *Леман Э. Теория точечного оценивания.* – М.: Наука, 1991.

[22] *Лоэв М. Теория вероятностей.* – М.: ИЛ, 1962.

[23] *Лукач Е. Характеристические функции.* – М.: Наука, 1979.

[24] *Неве Ж. Математические основы теории вероятностей.* – М.: Мир, 1969.

[25] *Нейман Ю. Вводный курс теории вероятностей и математической статистики.* – М.: Наука, 1968.

[26] *Петров В.В. Суммы независимых случайных величин.* – М.: Наука, 1972.

[27] *Порохов Ю.В., Розанов Ю.А. Теория вероятностей. 2-е изд.* – М.: Наука, 1985.

[28] *Севастьянов Б.А. Курс теории вероятностей и математической статистики.* – М.: Наука, 1982.

[29] *Скороход А.В. Елементи теорії ймовірностей та випадкових процесів.* – К.: Вища шко- ла, 1975.

[30] *Скороход А.В. Лекції з теорії випадкових процесів: Навч. посібник.* – К.: Либідь, 1990.

[31] *Турчин В.М. Математична статистика.* – К.: Академія, 1999.

[32] *Феллер В. Введение в теорию вероятностей и ее приложения.* – М.: Мир, 1984.

[33] *Шефтель З.Г. Теорія ймовірностей.* – К.: Вища школа, 1995.

[34] *Ширяев А.Н. Вероятность.* – М.: Наука, 1980.

Зміст