

МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ
КИЇВСЬКИЙ НАЦІОНАЛЬНИЙ УНІВЕРСИТЕТ ІМЕНІ ТАРАСА ШЕВЧЕНКА

В. С. Донченко, М. В.-С. Сидоров

ТЕОРІЯ ЙМОВІРНОСТЕЙ ТА МАТЕМАТИЧНА СТАТИСТИКА для соціальних наук

Навчальний посібник

*Рекомендовано Міністерством освіти і науки України
як навчальний посібник для студентів нематематичних напрямів підготовки*



УДК 519.2(075.8)
ББК 22.17я73
Д67

Рецензенти:

д-р філос. наук, проф. Є. І. Головаха,
д-р фіз.-мат. наук, проф. Н. О. Вірченко,
д-р фіз.-мат. наук, проф. А. В. Гладкий,
д-р фіз.-мат. наук, проф. Д. А. Ключин

*Рекомендовано до друку вченою радою
факультету кібернетики
(протокол № 2 від 28 жовтня 2013 року)*

*Ухвалено науково-методичною радою
Київського національного університету імені Тараса Шевченка
29 листопада 2013 року*

Д67 **Донченко В. С.**
Теорія ймовірностей та математична статистика для соціальних наук : навч. посіб. / В. С. Донченко, М. В.-С. Сидоров. – К. : ВПЦ "Київський університет", 2015. – 400 с.

ISBN 978-966-439-850-0

Викладено основні поняття і методи теорії ймовірностей та математичної статистики.

Для студентів вищих навчальних закладів. Стане у пригоді студентам соціологічних, психологічних, політичних, економічних та інших нематематичних спеціальностей та викладачам.

УДК 519.2(075.8)
ББК 22.17я73

*Рекомендовано Міністерством освіти і науки України
як навчальний посібник для вищих навчальних закладів
(лист № 1/11-12626 від 07.08.14)*

ISBN 978-966-439-850-0

© Донченко В. С., Сидоров М. В.-С., 2015
© Київський національний університет імені Тараса Шевченка,
ВПЦ "Київський університет", 2015

Вступ

Сучасна соціологічна наука вимагає від дослідника досконалого володіння математико-статистичним апаратом обробки та аналізу соціологічних даних. На сьогодні важко собі уявити фундаментальне соціологічне дослідження без побудови описових та пояснювальних моделей, що теж передбачає глибоку обізнаність фахівця-соціолога щодо засобів реалізації такого моделювання. Такі засоби вимагають від професіонала знань не тільки елементарних основ математики, але і поглибленого розуміння теорії ймовірностей та статистичного аналізу даних, без чого неможлива розробка нових та вдосконалення існуючих методів обробки й аналізу соціологічної інформації. Більшість посібників та підручників з теоретичних основ теорії ймовірностей та математичної статистики розраховані на те, що ця дисципліна є елементом комплексу дисциплін вищої математики і вивчається після того, як студент оволодіває основами математичного аналізу, теорії міри, комбінаторики, дискретної математики та ін.

Виклад матеріалу в цьому посібнику є нетрадиційним – він подається з погляду частотних характеристик вибірки, що значно спрощує та прискорює розуміння головних принципів теорії ймовірностей та математичної статистики, а також дозволяє використовувати отримані знання на практиці.

Посібник підготовлено відповідно до навчальної програми дисципліни, що викладається на факультеті соціології Київського національного університету імені Тараса Шевченка. Він містить матеріал, необхідний студентам для оволодіння знаннями з цієї навчальної дисципліни і складається з чотирьох розділів: "Первісні статистичні характеристики", де викладено принципи сприйняття та аналізу числових даних; "Теорія ймовірностей", тут наведено головні поняття, правила та закони теорії ймовірностей; "Математична статистика", де сформульовано принципи статистичного аналізу даних та перевірки гіпотез; "Статистичні таблиці". До розділів запропоновано перелік задач для самопідготовки.

Розрахований на студентів нематематичних спеціальностей і побудований так, що опанування викладеного матеріалу не вимагає від читача попередніх глибоких знань із вищої математики.

Зміст

Вступ	3
1. ПЕРВІСНІ СТАТИСТИЧНІ ХАРАКТЕРИСТИКИ.....	9
1.1. Вибірка. Подія. Частота події	9
1.1.1. Масові явища. Їх випадковість.....	9
1.1.2. Спостереження масових явищ.....	11
1.1.3. Подія. Абсолютна та відносна частота події за фіксованою вибіркою a_1, \dots, a_N	11
1.1.4. Частоти як функції подій	13
1.2. Операції над подіями. Властивості операцій. Властивості частот	14
1.2.1. Іменовані події	14
1.2.2. Події, операції над подіями; несумісні та попарно несумісні події.....	14
1.2.3. Ілюстрація подій, операцій над ними та властивостей діаграмами Ейлера – Венна	17
1.2.4. Властивості операцій над подіями.....	19
1.2.5. Характеристичні властивості абсолютних та відносних частот. Адитивність	20
1.3. Принцип групування даних.....	24
1.3.1. Принцип групування у вивченні масових явищ	24
1.3.2. Принцип групування для дискретних даних.....	25
1.3.3. Принцип групування для неперервних даних.....	27
1.3.4. Зображення результату групування для дискретного та неперервного випадків: діаграми та функції	36
1.3.5. Випадкова величина, розподіли частот випадкових величин	53
1.3.6. Запис умов групування за допомогою випадкової величини	54
1.4. Умовна частота. Формула Байєса для умовних частот.....	55
1.4.1. Умовна частота, зв'язок із безумовною	55
1.4.2. Теорема добутку для частот	57
1.4.3. Формула Байєса для частот	58
1.5. Інтегральні характеристики вибірки	59
1.5.1. Особливості застосування інтегральних характеристик вибірки	63
1.5.2. Обчислення інтегральних характеристик вибірки.	69

2. ТЕОРІЯ ЙМОВІРНСТЕЙ.....	82
2.1. Коротка історична довідка	82
2.2. Визначення ймовірності.	
Характеристичні властивості ймовірності.....	83
2.2.1. Закон стійкості частот, частотне визначення ймовірності	83
2.2.2. Передбачувальна властивість імовірності.....	84
2.2.3. Умовні ймовірності	84
2.2.4. Незалежність події A від події B , еквівалентне визначення незалежності. Незалежні в сукупності події, незалежні випадкові величини	85
2.2.5. Характеристичні властивості ймовірності	87
2.3. Задання ймовірностей.....	88
2.3.1. Задання ймовірностей для дискретних даних	89
2.3.2. Задання розподілів імовірностей для неперервних даних	90
2.4. Інтегральні характеристики розподілу ймовірностей.....	92
2.4.1. Інтегральні характеристики розподілу ймовірностей на базі середнього.....	92
2.4.2. Математичне сподівання випадкової величини	93
2.4.3. Дисперсія випадкової величини. Нерівність Чебишова	94
2.4.4. Обчислення математичного сподівання та дисперсії.....	96
2.5. Розподіли ймовірностей	97
2.5.1. Задання розподілів імовірностей	97
2.5.2. Основні дискретні розподіли.....	97
2.5.3. Основні неперервні розподіли.....	107
2.6. Основна задача теорії ймовірності	116
2.6.1. Формулювання основної задачі теорії ймовірності (ОЗТЙ).....	116
2.6.2. Постановка, схема, базові задачі ОЗТЙ.....	118
2.6.3. Формула повної ймовірності (ФПЙ).....	122
2.6.4. Формула Байеса	124
2.6.5. Використання графів для розв'язування задач на ФПЙ та формулу Байеса	126
2.6.6. Дерево прийняття рішень	133
2.6.7. Обчислення ряду розподілу функції від незалежних дискретних випадкових величин.....	136

2.6.8. Схема незалежних випробувань Бернуллі (СНВБ)	148
2.6.9. Схема незалежних випробувань до першої появи події А (СНВППА)	159
2.6.10. Розв'язування ОЗТЙ	160
2.6.11. Розв'язування ОЗТЙ через рівну можливість.....	161
2.6.12. Геометричні ймовірності	164
2.7. Багатовимірні розподіли.....	181
2.7.1. Дискретні та неперервні багатовимірні розподіли.....	181
2.7.2. Задання багатовимірних розподілів.....	182
2.7.3. Багатовимірні гауссівські розподіли.....	184
2.7.4. Системи випадкових величин, розподіли ймовірностей багатовимірних випадкових величин	186
2.8. Інтегральні характеристики багатовимірного розподілу або системи випадкових величин	189
2.8.1. Інтегральні характеристики системи випадкових величин на базі середнього.....	189
2.8.2. Незалежність випадкових величин. Властивості незалежності.....	192
2.8.3. Дисперсія суми та різниці випадкових величин.....	195
2.9. Однаково розподілені випадкові величини	196
2.9.1. Однакові розподіли, однаково розподілені випадкові величини	196
2.9.2. Незалежні однаково розподілені випадкові величини	196
2.10. Граничні теореми	197
2.10.1. Закон великих чисел.....	198
2.10.2. Центральна гранична теорема.....	199
2.10.3. Обчислення функції розподілу нормального розподілу через функцію Лапласа	201
2.10.4. Інтегральна теорема Муавра – Лапласа.....	202
2.10.5. Теорема Пуассона.....	205
2.10.6. Наближене обчислення ймовірностей біноміального розподілу	206
3. МАТЕМАТИЧНА СТАТИСТИКА	211
3.1. Основна задача математичної статистики	211
3.2. Задача точкового оцінювання (точкова оцінка параметра).....	212
3.2.1. Оцінювання параметрів дискретних розподілів	213
3.2.2. Оцінювання параметрів неперервних розподілів	214

3.3. Стандартні розподіли математичної статистики.	
Критичні величини	218
3.3.1. χ_n^2 -розподіл	218
3.3.2. t_n -розподіл Стьюдента	220
3.3.3. $F_{m,n}$ -розподіл Фішера.....	221
3.3.4. Симетричні та несиметричні розподіли ймовірностей.....	223
3.3.5. Нижні та верхні критичні величини	225
3.3.6. Основні співвідношення між критичними величинами	226
3.3.7. Критичні величини для основних розподілів	227
3.3.8. Довірчі області.....	232
3.4. Задача точності точкового оцінювання (інтервальне оцінювання параметра)	236
3.4.1. Другий тип ОЗМС (задача ТТО)	236
3.4.2. Точність точкового оцінювання параметрів нормального $N(m, \sigma^2)$ -розподілу	239
3.4.3. Нормальне наближення в обчисленні ТТО інтегральних характеристик на базі середнього	248
3.5. Перевірка статистичних гіпотез.....	258
3.5.1. Третій тип ОЗМС: перевірка гіпотез типу H_0 та типу альтернативного вибору	258
3.5.2. Помилки першого та другого роду. Формалізація третього типу ОЗМС	259
3.5.3. Загальний алгоритм розв'язання задачі перевірки статистичних гіпотез типу H_0 : критерій згоди чи критерій значущості.....	262
3.5.4. Критерій згоди для параметра m нормального розподілу	265
3.5.5. Інтерпретація похибок першого та другого роду. Потужність критерію.....	275
3.5.6. Перевірка гіпотези про значення параметра σ^2 нормального розподілу.....	279
3.5.7. Довірчі інтервали для параметрів нормального розподілу.....	283
3.5.8. Перевірка гіпотези про значення параметрів (таблиця критеріїв).....	286
3.5.9. Критерій Пірсона χ^2	290
3.5.10. Критерій згоди Колмогорова.....	305

3.5.11. Перевірка гіпотез однорідності двох вибірок	313
3.6. Лінійна регресія.....	333
3.6.1. Проста (парна) лінійна регресія.	335
3.6.2. Основна задача регресії	336
3.6.3. Емпірична регресія для найпростішої регресії	342
3.6.4. Прогноз за допомогою регресії	343
3.6.5. Точність точкового оцінювання параметрів регресії та прогнозу за емпіричною регресією	345
3.6.6. Довірчі смуги.	346
3.6.7. Множинна та поліноміальна лінійні регресії.....	351
4. СТАТИСТИЧНІ ТАБЛИЦІ	355
4.1. Таблиця значень гамма-функції.....	355
4.2. Таблиця значень функції Лапласа	362
4.3. Таблиця значень верхніх критичних величин розподілу χ_n^2	363
4.4. Таблиця значень верхніх критичних величин t_n -розподілу Стьюдента	369
4.5. Таблиця значень верхніх критичних величин $F_{m,n}$ -розподілу Фішера	372
4.6. Таблиця значень z -перетворень Фішера	393
4.7. Таблиця значень функції Колмогорова $K(\lambda)$	397
Список літератури	399

1. ПЕРВІСНІ СТАТИСТИЧНІ ХАРАКТЕРИСТИКИ

1.1. Вибірка. Подія. Частота події

1.1.1. Масові явища. Їх випадковість

У науці вивчають ті явища, які задовольняють умови відтворюваності, тобто можуть багаторазово повторюватись у часі та просторі.

Масовими називають явища, умови спостереження яких можуть багаторазово відтворюватись у часі чи просторі.

Випадковими масовими явищами називають такі, умови спостереження яких неоднозначно пов'язані з результатом цих спостережень.

Якщо неможливо однозначно пов'язати результати спостереження з умовами, в яких явище спостерігається, для вивчення випадкових масових явищ застосовують принцип групування, який полягає

1) у визначенні множини всіх можливих результатів спостережень, які спостерігались чи можуть спостерігатися;

2) у визначенні частоти появи тих чи інших результатів.

Кожне відтворення умов спостережень та фіксацію його результатів називають ще дослідом, експериментом та ін.

Послідовність N (натуральне число) результатів спостережень досліджуваного випадкового масового явища називають **вибіркою** і записують у вигляді:

$$a_1, a_2, \dots, a_N,$$

де a_1 – результат першого спостереження, перший елемент вибірки;

a_2 – результат другого спостереження, другий елемент вибірки, ...;

a_N – результат N -го спостереження, N -й елемент вибірки.

Приклад 1.1

1) Стать новонародженої дитини:

X, Д, X, X, Д, $N = 5$.

2) Темперамент (холерик, сангвінік, меланхолік, флегматик):

X, M, M, C, C, M, Ф, $N = 7$.

3) Зріст (см):

$$145, 150, 137, 141, N = 4.$$

Кількість елементів вибірки N називають **об'ємом вибірки**.

Приклад 1.2

1) $\underbrace{Д\ X\ X\ Д\ X\ \dots\ X}_{42}$ – вибірка об'єму 42;

$$X=22, Д=20, n=2, N=42.$$

2) $\underbrace{Ф\ C\ X\ M\ M\ \dots\ Ф}_{20}$ – вибірка об'єму 20;

$$Ф=4, C=3, X=7, M=6, n=4, N=20.$$

З умов випадковості, тобто неоднозначності результатів, спостереження, перш за все, визначають за всіма можливими результатами, що спостерігалися чи могли спостерігатися. Цю множину позначають через U і називають **простором елементарних подій** чи **достовірною подією**. Результати, з яких складається множина U , називають **елементарними подіями**.

Для прикладу 1.1 маємо таке:

1) стать новонародженої дитини: $U = \{X, Д\}$;

2) темперамент: $U = \{X, M, C, Ф\}$, якщо вважати, що існують 4 типи темпераменту (холерик, меланхолік, сангвінік, флегматик);

3) зріст (см): $U = [50, 250]$ є множиною дійсних чисел, хоча в цьому випадку діапазон можливих значень зросту можна обрати й іншим шляхом.

Приклад 1.3

Простори елементарних подій та їх потужності для наведених прикладів випадкових масових явищ мають такий вигляд:

1) стать новонародженої дитини

$$\begin{array}{cc} U : u_1 & u_2 \Rightarrow n = 2; \\ \Downarrow & \Downarrow \\ X & Д \end{array}$$

2) темперамент

$$\begin{array}{cccc} U : u_1 & u_2 & u_3 & u_4 \Rightarrow n = 4, \\ \Downarrow & \Downarrow & \Downarrow & \Downarrow \\ M & Ф & C & X \end{array}$$

$U = \{u_1, u_2, \dots, u_n\}$ – множина можливих результатів спостереження (достовірна подія);

3) зріст (см) $U = [70; 255]$.

1.1.2. Спостереження масових явищ

Простори елементарних подій розрізняють за "обсягом" можливих результатів, що є суттєвим, коли задають основні елементи, які потрібні для опису і дослідження випадкових масових явищ. За цим принципом розрізняють два основні типи даних в описі простору елементарних подій: дискретний та неперервний.

Дискретним в описі U називають тип даних, який відповідає випадку, коли множина U є скінченною або зліченною. **Неперервним** в описі U називають тип даних для випадку, коли множина U є більш ніж зліченною (континуальною) і є відрізком, півпрямую чи всією числовою прямою.

Дискретною називають множину, елементи якої можна пронумерувати, використовуючи скінченну кількість номерів (натуральних чисел) чи зліченну – всі натуральні числа. Нумерацію позначають записом елементів множини зліва направо: перший номер відповідає першому ліворуч, другий другому і т. д. Для скінченних множин рядок запису обривається, для нескінченних – у цьому випадку злічених – продовжується нескінченно.

Якщо елементи множини неможливо занумерувати, то говорять, що множина є **більш ніж зліченною** або **континуальною** (наприклад, інтервал дійсних чисел, півпряма або вся числова вісь).

Отже, до дискретного типу належать простори елементарних подій, що розглянуті у пунктах 1) та 2) прикладів 1.1 та 1.2; до неперервного типу – наведені у пунктах 3) цих прикладів.

1.1.3. Подія. Абсолютна та відносна частота події за фіксованою вибіркою a_1, \dots, a_N

Початковою інформацією для дослідження випадкового масового явища є та, що міститься в тій чи іншій вибірці, в якій зафіксовані спостереження цього явища. Подальше використання таких даних залежить від їх типу (дискретного чи неперервного).

Важливою характеристикою масового явища, яку розглядають для наявної вибірки, є абсолютна і відносна частоти, які визначають за цією вибіркою.

Найпростіше цю характеристику в абсолютному чи відносному випадку визначають для дискретних даних. У цьому разі абсолютна та відносна частоти характеризують окремі елементарні події. Так, це характеристика ваги елементарної події серед спостережень: для абсолютної частоти це кількість появ досліджуваної елементарної події серед елементів вибірки, для відносної – частка елементів вибірки, що мають досліджуване значення результату. **Абсолютну та відносну частоти** елементарної події u позначають відповідно m_u та f_u .

Розглянемо приклади підрахунку абсолютних та відносних частот результатів дискретного типу за відповідними вибірками a_1, \dots, a_N :

Приклад 1.4

1) a_1, \dots, a_N : Д Х Д Х Х Х Д Д Д

Об'єм вибірки	$N=10$	
Абсолютні частоти	$m_X = 5$	$m_D = 5$
Відносні частоти	$f_X = \frac{m_X}{N} = \frac{5}{10} = 0,5$	$f_D = \frac{m_D}{N} = \frac{5}{10} = 0,5$

2) a_1, \dots, a_N : Д Д Д Д Д Д Д Д Х

Об'єм вибірки	$N=10$	
Абсолютні частоти	$m_X = 1$	$m_D = 9$
Відносні частоти	$f_X = \frac{m_X}{N} = \frac{1}{10} = 0,1$	$f_D = \frac{m_D}{N} = \frac{9}{10} = 0,9$

Абсолютні та відносні частоти результатів залежать від вибірки, за якою їх підраховують;

3) a_1, \dots, a_N :

Ф Х С М М С Х Х С С.

Об'єм вибірки: $N=10$;

P – помірковані темпераменти {Ф, М};

L – лабільні темпераменти {Х, С};

$m_P = 3$
 $m_L = 7$ } абсолютні частоти .

Загалом абсолютні та відносні частоти за вибіркою a_1, \dots, a_N для досліджуваних масових явищ можна визначати не тільки для елементарних подій (окремих результатів), а й для груп (множин таких результатів) або для результатів, що характеризуються тією чи іншою властивістю.

Будь-яку групу результатів $A \subseteq U$ чи умову (властивість, вимогу), що її визначає, називають **подією**. Події позначають великими літерами латинського алфавіту.

Абсолютною частотою події A (m_A) називають кількість елементів вибірки, значення яких належать A (мають властивість A).

Якщо у спостереженні результат належить A , то кажуть, що у відповідному спостереженні відбулась подія A .

Так у прикладі 1.4 пункту 3) у 1-му, 4-му, 5-му спостереженнях відбулась подія P (спостерігалися помірковані темпераменти), а у 2-му, 3-му, 6-му, 7-му, 8-му, 9-му, 10-му випадках – подія L (спостерігались лабільні темпераменти).

Відносною частотою події A (f_A) називають частку елементів вибірки, значення яких належать A (мають властивість A), тобто є відношенням абсолютної частоти до об'єму вибірки:

$$f_A = \mathbf{Ч}_N(A) = \frac{m_A}{N}. \quad (1.1)$$

Так у прикладі 1.4 пункту 3) відносні частоти появи поміркованих P та лабільних L темпераментів складають відповідно $f_P = 3/10, f_L = 7/10$.

1.1.4. Частоти як функції подій

Частоту приписують кожній події – певній підмножині множини можливих результатів чи умові, що її визначає, – певне число (кількість елементів вибірки з A для абсолютної частоти чи питому вагу таких елементів у вибірці для відносної). Це означає, що частота є числовою характеристикою, тобто функцією цих подій. Отже, важливою рисою випадкових масових явищ, представлених вибіркою, є поява функції множин, тобто частот, які є числовими характеристиками (функціями) цих подій.

1.2. Операції над подіями. Властивості операцій. Властивості частот

1.2.1. Іменовані події

Серед усіх подій спеціальними позначеннями виділяють дві: весь простір елементарних подій U , та порожню множину \emptyset .

\emptyset є групою результатів, яка не містить жодного елемента. Якщо цю подію характеризують як умову, то її називають **неможливою подією**, оскільки вона не містить жодного результату і жоден елемент вибірки цій події не належить. Отже, поява цієї події неможлива для будь-якої вибірки, що характеризується множиною елементарних подій U . Тому абсолютна частота цієї події m_{\emptyset} дорівнює нулю: $m_{\emptyset} = 0$. Оскільки $f_{\emptyset} = m_{\emptyset}/N$, то $f_{\emptyset} = 0$.

Простір елементарних подій U , якщо його характеризують як умову, називають **достовірною подією**, оскільки вона з'являється в кожному спостереженні і всі елементи вибірки належать простору елементарних подій.

Через те, що достовірна подія з'являється в кожному спостереженні, її абсолютна та відносна частоти відповідно такі:

$$m_U = N, \quad f_U = 1.$$

Виокремлення неможливої та достовірної подій-умов пов'язано з тим, що вони мають наперед відомі значення абсолютної та відносної частот.

1.2.2. Події, операції над подіями; несумісні та попарно несумісні події

Операцією над подіями називають перетворення однієї чи більше подій, результатом якого є нова подія.

Розглядають, як правило, чотири стандартні операції над подіями: додавання (сумування), добутку (множення), заперечення (доповнення) та різниці. Три перші називають основними, остання – різниці – похідною.

Сума, добуток та різниця в початковому визначенні перетворюють дві або більше подій, тому їх називають бінарними і позначають так само, як для арифметичних відповідників. Запере-

чення застосовується до однієї події, тому його називають унарним (табл. 1.1).

Таблиця 1.1

Операції над подіями

Позначення операції	Запис операції	Прочитання операції	Визначення операції
+	$A+B$	Сума подій A, B . A плюс B . A чи B . Хоча б одна з двох: A чи B	Сумою двох чи більше подій ($A+B, A_1+A_2+\dots+A_n$) називають таку складену подію, яка виконується тоді і тільки тоді, коли виконується хоча б одна умовна складова
\times	$A \times B$	Добуток подій A, B . A помножити на B . A та B . Одночасно A та B .	Добутком двох чи більше подій (записується $A \cdot B, A_1 \cdot A_2 \cdot \dots \cdot A_n$) називають таку складену подію, яка виконується тоді й тільки тоді, коли виконуються одночасно всі умовні складові
—	\bar{A}	Заперечення події A . Доповнення до події A . Не A . Все, крім A .	Запереченням події (застосовують до однієї події) називають таку похідну подію, яка виконується тоді, коли не виконується A
\setminus	$A \setminus B$	Різниця подій A, B . A мінус B . A , але не B . A крім B .	Різницею двох подій A та B називають таку складену подію, яка виконується тоді і тільки тоді, коли виконується подія A і не виконується подія B

У прочитанні записів операцій відбито обидва варіанти розгляду події: як множини та як умови. Так, наприклад, прочитання операції додавання як "хоча б одне з двох: A чи B " або " A чи B " відповідає розгляду подій як умов, водночас інші два варіанти прочитання скоріше властиві розгляду подій як множин. Те саме стосується й інших операцій.

Для визначення операцій посилаються на визначеність умови, яка характеризує результат за виконанням чи невиконанням умови, що описує кожну з подій операції. Таке посилання здійснюють за таблицями, які називають **таблицями істинності для операцій**, у них відображають значення результату за значеннями компонент (табл. 1.2а)

Таблиця 1.2а

Таблиця істинності для операції додавання

A	B	$A+B$
В	В	в
В	Н	в
Н	В	в
Н	Н	н

Користування таблицею: за настанням ("В" – відбулася) чи ненастанням ("Н" – не відбулася) подій A та B у відповідному рядку визначаємо, відбулася ("в") чи не відбулася ("н") при цьому подія " $A+B$ ". Із табл. 1.2а видно, що подія " $A+B$ " не відбувається тільки тоді, коли не відбуваються одночасно обидві складові (останній рядок таблиці: Н–Н–н).

Таблиця 1.2б

Таблиця істинності для операції множення

A	B	$A \times B$
В	В	в
В	Н	н
Н	В	н
Н	Н	н

Із табл. 1.2б видно, що подія " $A \times B$ " відбувається тільки тоді, коли відбуваються одночасно обидві складові (перший рядок таблиці: В–В–в).

Таблиця 1.2в

Таблиця істинності для операції заперечення

A	\bar{A}
в	н
н	в

Події A та B називають **несумісними**, якщо вони не можуть виконуватися одночасно. Прикладом двох несумісних подій є подія A та її заперечення (табл. 1.2в). Формальний запис несумісності: $A \times B = \emptyset$. Отже, неможливість одночасного виконання означає несумісні події.

Події A_1, \dots, A_n називають **попарно несумісними**, якщо жодна пара подій (умов) не виконується одночасно, тобто будь-яка пара такого набору є несумісною: $A_i A_j = \emptyset, i \neq j$.

1.2.3. Ілюстрація подій, операцій над ними та властивостей діаграмами Ейлера – Венна

Діаграмою Ейлера – Венна називають рисунок, на якому

- простір елементарних подій зображено внутрішньою частиною деякого прямокутника на площині. Елементарною подією є кожна точка площини цього прямокутника;
- подією як множиною є внутрішня чи зовнішня частини довільного замкненого неперервного контуру без перетину усередині цього прямокутника. Подією як умовою є умова належності внутрішньої чи зовнішньої частин цього контуру. Як правило, якщо спеціально не обумовлено інше, мають на увазі внутрішність.

На таких рисунках зручно ілюструвати характеристизацію подій. Так несумісність подій ілюструють "неперетином" відповідних контурів (рис. 1.1).

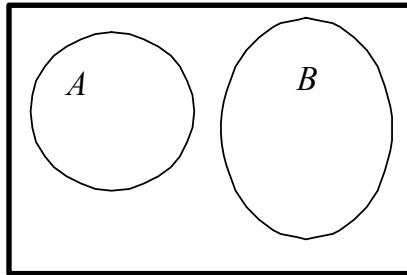


Рис. 1.1. Несумісні події

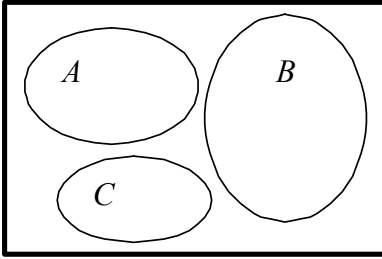


Рис. 1.2. Парно несумісні події

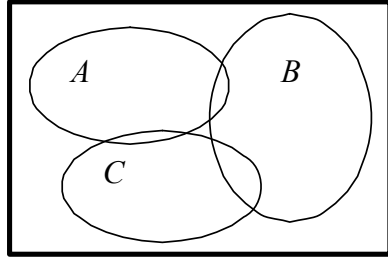


Рис. 1.3. Несумісні в сукупності події

Попарну несумісність ілюструють так, як це показано на рис. 1.2.

Рис. 1.3 ілюструє ситуацію, коли події є несумісними в сукупності, тобто вони не мають жодного спільного елемента, але тим не менше, вони не є попарно несумісними (кожна пара має спільні елементи).

Інші діаграми ілюструють операції добутку (рис. 1.4), додавання (рис. 1.5), віднімання (рис. 1.6) та заперечення (рис. 1.7).

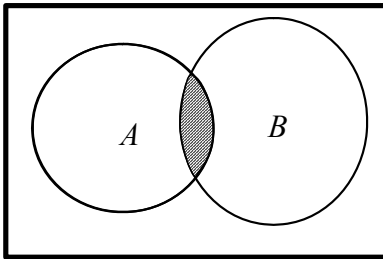


Рис. 1.4. Добуток подій

$$A \times B$$

За визначенням добутку, цю множину утворено тими точками прямокутника (простору елементарних подій), які одночасно належать внутрішній частині обох контурів (складових операції).

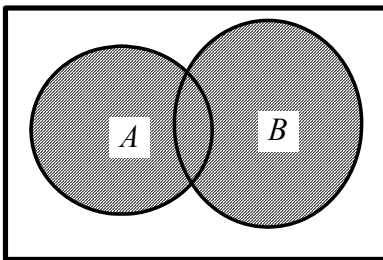


Рис. 1.5. Сума подій

$$A+B$$

За визначенням суми цю множину утворено тими точками прямокутника (простору елементарних подій), які належать внутрішній області контуру, утвореного зовнішніми частинами контурів (складових).

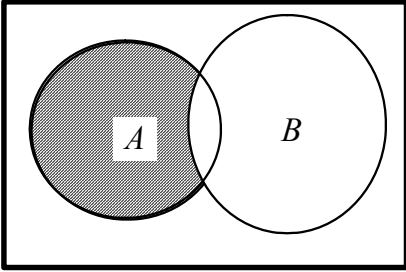


Рис. 1.6. Різниця подій

$$A \setminus B$$

За визначенням різниці, цю множину утворено тими точками прямокутника (простору елементарних подій), які належать області події A і не належать області події B .

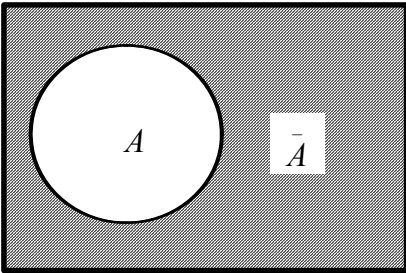


Рис. 1.7. Доповнення (заперечення) події

$$\bar{A}$$

За визначенням заперечення, цю множину утворено тими точками прямокутника – простору елементарних подій, які не належать області події A .

1.2.4. Властивості операцій над подіями

Під властивостями операцій розумітимемо співвідношення, які фіксують поведінку операцій щодо інших операцій і до себе, а також до іменованих подій.

Властивості операцій над множинами загалом збігаються з відповідниками для операцій над числами. Зокрема, операції додавання та множення є *асоціативними* (для довільних подій A , B та C справедливі співвідношення $A+(B+C) = (A+B)+C$ та $A \times (B \times C) = (A \times B) \times C$), *комутативними* (для довільних подій A і B мають місце співвідношення $A+B = B+A$ та $A \times B = B \times A$) та *дистрибутивними* стосовно одна одної (для довільних подій A , B та C справедливі співвідношення $A \times (B+C) = (A \times B) + (A \times C)$ та $A+(B \times C) = (A+B) \times (A+C)$). Останнє співвідношення розширює

властивість дистрибутивності для подій у тому розумінні, в якому вона властива числам.

Виконуються співвідношення (поведінка операцій щодо самих себе):

$$\begin{aligned} A \times A &= A, \\ A + A &= A, \\ \overline{\overline{A}} &= A. \end{aligned}$$

Важливими є також дві властивості: закон виключеного третього та закони де Моргана.

Закон виключеного третього має вигляд

$$A \times \overline{A} = \emptyset, \quad A + \overline{A} = U.$$

Закони де Моргана (перший та другий) у найпростішому варіанті:

$$\begin{aligned} \overline{A + B} &= \overline{A} \times \overline{B}, \\ \overline{A \times B} &= \overline{A} + \overline{B}. \end{aligned}$$

Формулу першого закону де Моргана прочитаємо так: "Заперечення суми є добутком заперечень", другий закон як "Заперечення добутку є сумою заперечень". У цьому варіанті законів де Моргана не важливо, про яку кількість подій ідеться, отже, формульними відповідниками можуть бути:

$$\begin{aligned} \overline{A_1 + A_2 + \dots + A_n} &= \overline{A_1} \times \overline{A_2} \times \dots \times \overline{A_n}, \\ \overline{A_1 \times A_2 \times \dots \times A_n} &= \overline{A_1} + \overline{A_2} + \dots + \overline{A_n}. \end{aligned}$$

Ці узагальнені співвідношення теж справедливі і також називаються, відповідно, першим та другим законом де Моргана.

Поведінка операцій щодо іменованих подій характеризується співвідношеннями:

$$\begin{aligned} A \times \emptyset &= \emptyset, & A \times U &= A, \\ A + \emptyset &= A, & A + U &= U. \end{aligned}$$

1.2.5. Характеристичні властивості абсолютних та відносних частот. Адитивність

До характеристичних властивостей довільного об'єкта належать властивості, які вичерпно характеризують його як об'єкт

досліджуваного типу. До таких властивостей абсолютної та відносної частот відносять такі, виконання яких дає змогу стверджувати, що можна побудувати послідовність, для якої об'єкт із цими властивостями буде частотою.

Характеристичними властивостями як функціями подій абсолютної і відносної частот є три: невід'ємність; адитивність та нормованість.

Невід'ємність означає, що для будь-якої події значення частоти (відносної чи абсолютної) невід'ємне:

$$m_A \geq 0, \quad f_A \geq 0.$$

Адитивність, скінченна адитивність та зліченна адитивність означає властивість подій, яка дозволяє записувати частоту (абсолютну чи відносну) події як суму несумісних чи попарно несумісних подій, на які розбита сама подія, незалежно від типу розбиття. Різниця між варіантами цієї властивості полягає в кількості подій у сумі: для адитивності – це сума двох несумісних подій, для скінченної адитивності – сума скінченної кількості попарно несумісних подій, для зліченної – сума зліченної кількості попарно несумісних подій.

Маємо формалізовані записи варіантів адитивності:

а) адитивність для двох несумісних ($A \times B = \emptyset$) подій A та B

$$m_{A+B} = m_A + m_B,$$

$$f_{A+B} = f_A + f_B;$$

б) скінченна адитивність для довільних попарно несумісних ($A_i \times A_j = \emptyset, i \neq j$) подій A_1, \dots, A_n

$$m_{A_1+\dots+A_n} = m_{A_1} + \dots + m_{A_n},$$

$$f_{A_1+\dots+A_n} = f_{A_1} + \dots + f_{A_n};$$

в) зліченна адитивність для довільних попарно несумісних ($A_i \times A_j = \emptyset, i \neq j$) подій A_1, \dots, A_n, \dots

$$m_{A_1+\dots+A_n+\dots} = m_{A_1} + \dots + m_{A_n} + \dots,$$

$$f_{A_1+\dots+A_n+\dots} = f_{A_1} + \dots + f_{A_n} + \dots$$

Властивості адитивності та скінченної адитивності еквівалентні, зліченна адитивність є сильнішою, оскільки адитивність та скінченна адитивність є її наслідками.

Нормованість означає сумарне одиничне значення відносних частот усіх елементарних подій із множини U . Вона впливає зі зліченної адитивності, визначення відносної частоти (згідно з формулою (1.1)) та визначення простору елементарних подій U і стосується тільки відносної частоти. Якщо $U = \{u_1, u_2, u_3, \dots, u_n, \dots\}$, то

$$f_U = f_{u_1 + \dots + u_n + \dots} = f_{u_1} + \dots + f_{u_n} + \dots = 1.$$

Приклад 1.5

Операції над подіями

Підкидають симетричний гральний кубик. Побудувати множини результатів цього стохастичного експерименту. Описати події: A – випало число очок, кратне 3; B – випало парне число очок; події $A+B$, $A \times B$.

Розв'язання

За множини результатів цього стохастичного експерименту слід брати $U = \{1, 2, 3, 4, 5, 6\}$. Події A і B як підмножини U описують таким чином: $A = \{3, 6\}$, $B = \{2, 4, 6\}$. Подія $A+B$ полягає в тому, що випаде або парне число очок, або число очок, кратне трьом. Звідси маємо $A+B = \{2, 3, 4, 6\}$. Подія $A \times B$ передбачає, що випадає число, яке одночасно ділиться на 2, і 3, а отже, містить тільки один елемент $A \times B = \{6\}$.

ЗАДАЧІ

№ 1

Вказати події, що протилежні таким подіям:

- A – поява герба при двох підкиданнях монети;
- B – три влучення при трьох пострілах;
- C – принаймні одне влучення при трьох пострілах.

№ 2

Нехай A , B , C – три випадкові події. Записати події, які полягають у тому, що з A , B , C відбулися:

- тільки подія A ;
- події A та B і не відбулася подія C ;
- усі три події;
- принаймні одна подія;
- одна й тільки одна подія;

- е) не більше двох подій;
- є) не відбулося жодної події.

№ 3

Два гральні кубики підкидають один раз. Побудувати множину результатів цього стохастичного експерименту (появи цифр на гранях кубиків і випадкові події): A – сума цифр на гранях кубиків буде кратна двом; B – сума цифр буде кратна п'яти. Описати події $A+B$, $A \times B$, $A-B$.

№ 4

Монету підкидають двічі. Описати множину результатів цього стохастичного експерименту та події: A – принаймні один раз з'явиться герб; B – при другому підкиданні з'явиться герб.

№ 5

З усіх сімей із трьома дітьми вибрали одну. Описати множину результатів цього стохастичного експерименту і такі випадкові події: A – у сім'ї є хлопчик і дівчинка; B – у сім'ї не більше однієї дівчинки. Описати $A+B$, $A \times B$.

№ 6

Задано множину чисел $U = \{1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12\}$. Подія A полягає в тому, що навмання взяте число з U буде меншим за 7, подія B – таке число більше за 4. Побудувати події $A+B$, $A \times B$, $A-B$.

№ 7

Монету підкидають 5 разів. Описати множину результатів цього стохастичного експерименту (U) та знайти кількість елементів цієї множини.

№ 8

10 кульок пронумеровані від 1 до 10 та містяться в урні. Кульки з урни виймають по одній без повернення. Описати множину результатів цього стохастичного експерименту (U) та знайти кількість елементів цієї множини.

№ 9

У трьох урнах містяться кольорові кулі. У першій – 2 червоні та 1 синя, у другій – 2 зелені й 1 жовта, у третій – 1 зелена, 1 червона та 1 синя. Із кожної урни беруть по одній кулі. Описати

множину результатів цього стохастичного експерименту (U) і такі випадкові події: A – з'являться дві червоні кулі, B – дві кулі будуть одного кольору, C – три кулі будуть одного кольору. Описати $A \times C$, $A \times B$, $B \times C$, $A \times B \times C$.

№ 10

Симетричний гральний кубик підкидають 6 разів. Описати множину результатів цього стохастичного експерименту (U). Описати подію A – випадуть усі шість граней.

№ 11

Під час маркетингового опитування щодо вживання солодоців було опитано 100 осіб. Виявилось, що 60 осіб вживають солодоці українського виробництва, 50 – імпортного, а 10 – взагалі не вживають солодоців. За діаграмами Ейлера – Венна знайти кількість людей, які вживають як імпортні, так і українські солодоці.

№ 12

Обчислити вирази:

$$1) \overline{(A+B)} \times \overline{(C+B)} \times \overline{(A+C)},$$

$$2) \overline{(A \setminus B)} \times \overline{(B \setminus A)} \times \overline{(C \setminus A)} \times \overline{(A \setminus C)} \times \overline{(C \setminus B)} \times \overline{(B \setminus C)}.$$

1.3. Принцип групування даних

1.3.1. Принцип групування у вивченні масових явищ

Оскільки випадкове масове явище проявляється через вибірку (записи результатів спостережень у тій чи іншій кількості), то інформація про нього вимушено є тою, яку можна здобути з вибіркою. У статистиці така основна інформація про випадкове масове явище міститься в частотах подій, пов'язаних із досліджуваним масовим явищем. Оскільки події пов'язують із множинами результатів (просторами елементарних подій), загальним є метод, за яким

1) визначають усі можливі результати, які спостерігалися (представлені у наявній вибірці) чи можуть спостерігатися, тобто явно описують множину U ;

2) обчислюють частоти всіх можливих подій та досліджують співвідношення між ними.

Частоти можливих подій можна знайти обчисленням частот "мінімального" запасу подій та описом того, як частоти решти подій обчислюють через частоти подій із "мінімального" запасу. "Мінімальні" запаси подій, так само як і правила обчислення частот решти подій, залежать від типу даних (дискретного чи неперервного). Отже, принцип групування може бути уточнений згідно з методом, за яким визначають усі можливі результати, а також частоти "мінімального" запасу подій. Це уточнення, можна конкретизувати через "мінімальні" запаси та правила обчислення частот усіх можливих подій за частотами подій із "мінімального" запасу.

Загальним змістом принципу групування є вивчення частоти $\mathcal{C}_M(A)$ для подій A (груп результатів) із множини можливих значень спостережень U .

Частоти для числових груп результатів називають розподілами частот. Отже, частота $\mathcal{C}_M(A)$ як функція події A є розподілом частот, коли аргументи A є числовими, тобто U – числова множина.

1.3.2. Принцип групування для дискретних даних

Конкретизація принципу групування пов'язана перш за все з визначенням "мінімального" запасу подій, якими для дискретних даних є сукупність усіх одноелементних подій. Знання частот цих подій забезпечує можливість обчислення частот решти подій (усіх можливих груп результатів) з U . Принцип групування для дискретних даних такий:

1) визначення множини $U = \{u_1, u_2, \dots, u_n(\dots)\}$ усіх можливих результатів спостережень (простору елементарних подій);

2) знаходження частот f_u всіх елементарних подій $u \in U$;

3) обчислення частоти будь-якої групи результатів з U (будь-яких подій) як суми частот тих елементарних подій, з яких складається A , тобто за формулою

$$f_A = \sum_{u \in A} f_u .$$

Можливість обчислення ймовірностей будь-яких подій за ймовірностями елементарних подій (точніше одноелементних подій) впливає із скінченної або зліченної адитивності, якщо взяти до уваги, що будь-яка подія з дискретного простору елементарних подій є сумою своїх одноелементних підмножин:

$$f_A = \mathcal{P}_N(A) = \mathcal{P}_N\left(\bigcup_{u \in A} \{u\}\right) = \sum_{u \in A} \mathcal{P}_N(\{u\}) = \sum_{u \in A} f_u .$$

Варіаційним рядом вибірки є та сама вибірка, упорядкована за зростанням.

Варіаційний ряд і вибірка мають однакову кількість елементів. Якщо вихідна вибірка позначається u_1, \dots, u_N , то елементи варіаційного ряду u_1^*, \dots, u_N^* .

Розмах (R) вибірки є різницею між максимальним u_N^* і мінімальним u_1^* значеннями елементів вибірки

$$R = u_N^* - u_1^* .$$

Множину всіх можливих результатів та відповідних їм частот називають **рядом розподілу**. Ряд розподілу представляють за допомогою таблиці з двох стовпчиків чи рядків, в одному з яких перераховано всі можливі результати, в іншому – відповідні їм частоти:

U	u_1	u_2	...	u_n	(...)
f_u	f_1	f_2	...	f_n	(...)

Характеристичною властивістю ряду розподілу частот є невід'ємність та нормованість. Це впливає з того, що достовірна подія має одиничну частоту, і, за формулою обчислення частот для дискретних даних, є сумою частот тих елементарних подій, з яких складається:

$$1 = \mathcal{P}_N(U) = \mathcal{P}_N\left(\bigcup_{u \in U} \{u\}\right) = \sum_{u \in U} \mathcal{P}_N(\{u\}) = \sum_{u \in U} f_u .$$

Ряд розподілу частот як результат застосування принципу групування, утворює той "мінімальний" запас інформації про події та частоти, який дозволяє підрахувати частоту будь-якої події за формулою

$$\mathcal{C}_N(A) = \mathcal{C}_N\left(\bigcup_{u \in A} \{u\}\right) = \sum_{u \in A} \mathcal{C}_N(u).$$

Ознака визначення мінімально необхідного запасу подій, за частотами яких можна обчислити решту, є загальною у застосуванні принципу групування.

Ці "мінімальні" запаси подій залежать від типу даних (дискретного чи неперервного). У прикладі 1.4 описано випадок дискретних даних, для яких "мінімальним" запасом є одноелементні множини.

Зазначимо, що коли U є числовою множиною, то ряд розподілу частот задає значення, які називають **розподілом частот**.

1.3.3. Принцип групування для неперервних даних

Для дискретних даних "мінімально" необхідним запасом подій є всі одноелементні події. У неперервному випадку – це групи, сформовані за принципом "менше ніж" або за "інтервальним принципом".

Для неперервних даних простором елементарних подій є вся числова вісь:

$$U = (-\infty; +\infty).$$

Оскільки U – числова множина, частоту у випадку неперервних даних називають розподілом частот.

Цей варіант визначення U , охоплює й інші варіанти визначення для неперервних даних: $U = [a, b]$, $U = [0, \infty)$ і т. ін.

Результати застосування принципу групування для неперервних даних можуть бути представлені:

- за результатами, згрупованими за принципом "менше ніж";
- за результатами, згрупованими за "інтервальним" принципом.

У першому випадку результат застосування принципу групування називають кумулятивною функцією розподілу частот (КФРЧ), у другому – групуваною вибіркою.

Кумулятивна функція розподілу частот, її графік

Принцип "менше ніж" у групуванні даних та визначення КФРЧ. Принцип "менше ніж" у формуванні "мінімального" запасу множин означає, що розглядаються всі можливі події, які складаються зі значень, що не перевищують x для того чи іншого дійсного значення x , тобто всі групи результатів вигляду $(-\infty; x)$, для довільного значення x .

Принцип групування "менше ніж" для неперервних даних такий:

1) множину U всіх можливих результатів спостережень вважають множиною дійсних чисел $U = (-\infty; +\infty)$;

2) розглядають частоту всіх можливих подій, сформованих за принципом "менше ніж": $\mathcal{C}_N(-\infty; x)$ для довільного дійсного x .

Частоту подій, сформованих за принципом "менше ніж" як функцію правого кінця інтервалу групування, називають **кумулятивною функцією розподілу частот і позначають $F_N(x)$** :

$$F_N(x) = \mathcal{C}_N(-\infty; x).$$

Приклад 1.6

Для вибірки 58, 71, 72, 71, 73, 72

$N = 6$;

$\mathcal{C}_N(-\infty; x)$ для значень $x = 5, 60, 70, 80$ має такий вигляд:

$$\mathcal{C}_N(-\infty; 5) = 0, \quad \mathcal{C}_N(-\infty; 60) = \frac{m_A}{N} = \frac{1}{6},$$

$$\mathcal{C}_N(-\infty; 70) = \frac{1}{6}, \quad \mathcal{C}_N(-\infty; 80) = 1.$$

Загалом визначення КФРЧ вимагає обчислення її значення для довільного дійсного значення аргументу. Для значень аргументу КФРЧ між послідовними спостережуваними значеннями результатів значення самої КФРЧ залишається одним і тим самим. Це означає, що графік КФРЧ є східчастою лінією (рис. 1.8):

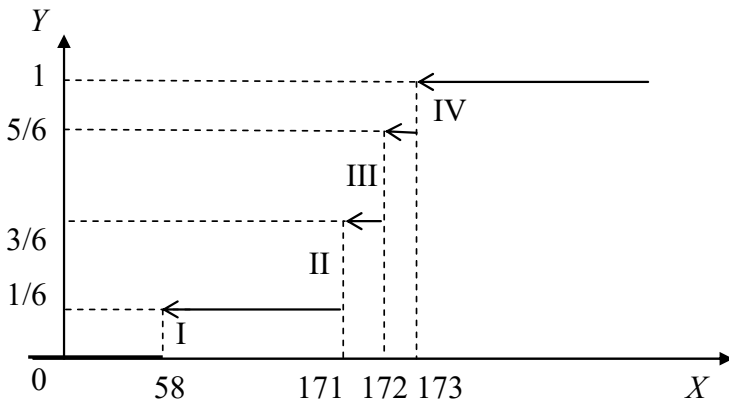


Рис. 1.8. КФРЧ

Побудова такого графіка є прикладом зображення результатів застосування принципу "менше ніж" для неперервних даних.

Побудова графіка КФРЧ. Графік КФРЧ – це східчаста лінія. При проходженні числової осі зліва праворуч змінює свої значення від **нульового до одиничного**.

Для побудови графіка КФРЧ виконують такі кроки (розглянемо це на прикладі вибірки 58, 171, 172, 171, 173, 172):

1) визначаємо всі можливі **різні** результати, представлені у вибірці, і нумеруємо в порядку зростання:

58	171	172	173
II	II	II	II
x_1^*	x_2^*	x_3^*	x_4^*

2) усі значення x_1^*, \dots, x_m^* наносимо на вісь OX ;

3) визначаємо абсолютні та відносні частоти всіх відповідних результатів

$$m_{x_1}^* \dots m_{x_m}^* :$$

$$m_{58} = 1 \qquad m_{171} = 2 \qquad m_{172} = 2 \qquad m_{173} = 1$$

$$f_{x_1}^* \dots f_{x_m}^* :$$

$$f_{58} = 1/6 \qquad f_{171} = 2/6 \qquad f_{172} = 2/6 \qquad f_{173} = 1/6$$

4) значення відносних частот наносимо під відповідними значеннями результату на вісь OY ;

5) графік КФРЧ маємо у вигляді східців у процесі руху зліва праворуч віссю OX від нульового (лівіше найменшого x_1^*), змінюючи висоту східця при проходженні кожного чергового результату на відносну частоту цього результату (рис 1.7).

Інтервальний принцип групування для неперервних даних

Для неперервних даних існує ще один принцип групування за інтервалами можливих даних. Таке групування називатимемо **інтервальним принципом**. Це групування означає, що множини досліджуваних результатів утворює інтервал $[a; b)$.

Групованою вибіркою називають представлення вихідної вибірки у вигляді розбиття на групи за інтервалами однакової довжини з описом ознак груп (границь інтервалів групування) та кількісного складу кожної з груп (абсолютних чи відносних частот). Цю вибірку, як і ряд розподілу, подають у вигляді таблиці з двох стовпчиків чи рядків, в одному з яких вказують ознаки груп, а в іншому – абсолютні чи відносні частоти кожної з груп. Довжину спільного інтервалу групування позначають через h .

Границі послідовних інтервалів позначають x_0, x_1, \dots, x_M , кількість інтервалів позначають M , а самі інтервали:

$$[x_0, x_1) [x_1, x_2) , \dots, [x_{M-1}, x_M),$$

де $x_0 < x_1 < \dots < x_M$.

Оскільки всі інтервали мають однакову довжину, то

$$x_i - x_{i-1} = h, i = 1, 2, \dots, M,$$

чи

$$x_i = x_{i-1} + h, i = 1, 2, \dots, M.$$

Отже, груповану вибірку представлено в таблиці, до якої іноді вводять стовпчик середин інтервалів групування $x_i^*, i = 1, \dots, M$:

$[x_{i-1}, x_i)$	m_i	f_i	x_i^*
$[x_0, x_1)$	m_1	f_1	x_1^*
$[x_1, x_2)$	m_2	f_2	x_2^*
...
$[x_{M-1}, x_M)$	m_M	f_M	x_M^*

Загалом, довжина інтервалів групування може бути будь-якою. Кожен інтервал має містити "достатню" кількість спостережень (як правило, на практиці, якщо інтервал містить менше трьох спостережень, то його об'єднують з одним із сусідніх).

Для визначення кількості інтервалів часто використовують емпіричну формулу, яку називають **формулою Стерджеса**, за якою кількість груп M визначають через об'єм вибірки:

$$M = [1 + 3,322 \lg N]. \quad (1.2)$$

Квадратні дужки означають, що беруть цілу частину від значення того виразу, який стоїть у квадратних дужках.

Наприклад, при об'ємі вибірки $N = 100$ кількість інтервалів складатиме

$$M = [1 + 3,322 \cdot \lg 100] = [1 + 3,322 \cdot 2] = [1 + 6,644] = [7,644] = 7.$$

За кількістю груп визначають довжину інтервалу групування h через мінімальний a_{\min} та максимальний a_{\max} елементи вибірки відповідно до виразу

$$h = \frac{a_{\max} - a_{\min}}{M - 1} = \frac{R}{M - 1}. \quad (1.3)$$

Границі інтервалів групування знаходять за співвідношеннями

$$\begin{aligned} x_0 &= a_{\min} - h/2, \\ x_1 &= x_0 + h, \\ x_2 &= x_1 + h, \\ &\dots \\ x_i &= x_{i-1} + h, \\ &\dots \\ x_M &= x_{M-1} + h = a_{\max} + h/2. \end{aligned} \quad (1.4)$$

Тобто будується розбиття на інтервали таким чином, щоб найменше значення вибірки (перше значення варіаційного ряду) потрапляло у середину першого інтервалу, а найбільше значення вибірки (останнє значення варіаційного ряду) – у середину останнього інтервалу.

Отже, алгоритм формування інтервалів групування полягає у визначенні:

- 1) кількості інтервалів групування;
- 2) довжини інтервалу групування;
- 3) границь інтервалів групування.

Побудова всієї групованої вибірки після формування інтервалів передбачає підрахунок абсолютних чи відносних частот появи спостережень із кожного інтервалу групування.

Враховуючи всю множину можливих значень $U = (-\infty; +\infty)$, крім інтервалів групування з довжиною h розглядаються ще два інтервали $(-\infty; x_0)$, $[x_M; +\infty)$, частота появи результатів з яких дорівнює нулю.

Прикладом групованої вибірки, що ілюструє інтервальний принцип у застосуванні до неперервних даних є групована вибірка, представлена у вигляді таблиці:

$[x_{i-1}, x_i)$	m_A
[55; 60)	1
[60; 65)	0
[65; 70)	0
[70; 75)	5

Ця таблиця представляє груповану вибірку зросту з прикладу 1.6.

Застосування групованої вибірки до вихідної саме в такому вигляді, як у цій таблиці, є невдалим, тому що у групованій вибірці містяться групи, не представлені жодним спостереженням: це друга та третя групи, а перша група містить лише один елемент.

Наступний приклад групованої вибірки ілюструє більшу наповненість груп, що її складають. У цьому прикладі в таблиці, що представляє груповану вибірку є ще один стовпчик, який вводять одразу або пізніше. Цей стовпчик показує середини інтервалів групування, які позначають x_i^* , $i = 1, \dots, M$.

Приклад 1.7

$[x_{i-1}, x_i)$	m_i	f_i	x_i^*
[100; 150)	2	2/100	125
[150; 200)	97	97/100	175
[200; 250)	1	1/100	225

$N = 100$, $h = x_{i+1} - x_i = 50$.

Зуваження. Оскільки групована вибірка не розрізняє результатів, що потрапляють в одну й ту саму групу, за потреби

вважають, що всі результати (m_i) з i -ї групи $[x_{i-1}; x_i)$, де $i=1,2, \dots, M$, ототожнюються із серединою x_i^* інтервалу групування:

$$x_i^* = \frac{x_{i-1} + x_i}{2}.$$

Приклад 1.8 (побудова групуваної вибірки)

Нехай маємо вибірку:

2, 3, 2, 1, 3, 4, 5, 4, 10, 6, 7, 3, 4, 0, 3, 8, 9, 4, 3, 5, 6, 3, 8, 2, 6, 6, 3, 4, 6, 3, 0, 10, 5, 1, 5, 4, 7, 8, 3, 1.

За нею потрібно побудувати групувану вибірку.

Розв'язання

Спочатку побудуємо для зручності варіаційний ряд. Отримаємо 0, 0, 1, 1, 1, 2, 2, 2, 3, 3, 3, 3, 3, 3, 3, 3, 3, 3, 3, 4, 4, 4, 4, 4, 4, 5, 5, 5, 5, 6, 6, 6, 6, 6, 6, 7, 7, 8, 8, 8, 9, 10, 10.

Об'єм вибірки $N = 40$, $a_{\min} = 0$, $a_{\max} = 10$.

За формулою Стерджеса (1.2) одержимо кількість інтервалів групування

$$M = [1 + 3,322 \lg 40] = [6,322] = 6.$$

Довжину інтервалів знаходимо за формулою (1.3)

$$h = (a_{\max} - a_{\min}) / (M - 1) = (10 - 0) / (6 - 1) = 10 / 5 = 2.$$

Тепер потрібно знайти початок першого інтервалу групування, тобто $x_0 = a_{\min} - h/2 = 0 - 2/2 = -1$. Для визначення інших границь інтервалів групування скористаємось формулою (1.4), тобто заповнимо таблицю інтервалів:

$[x_{i-1}, x_i)$	m_i	f_i	x_i^*
$[-1; 1)$			
$[1; 3)$			
$[3; 5)$			
$[5; 7)$			
$[7; 9)$			
$[9; 11)$			

Якщо ми вірно порахували всі інтервали, то найбільше значення вибірки має потрапити у середину останнього інтервалу. Дійсно, $a_{\max} = 10$, а з іншого боку, середина шостого (останнього) інтервалу дорівнює $(9+11)/2=10$. Отже, самі інтервали ми порахували вірно.

У стовпчику таблиці m_i потрібно вказати абсолютні частоти потрапляння елементів вибірки у кожен з інтервалів, або, інакше, – кількість елементів вибірки, що потрапляють до кожного з інтервалів.

До першого інтервалу можуть потрапити тільки ті елементи вибірки, значення яких не менші за -1 і менші за 1 . Такими значеннями є 0 та 0 . Отже, тільки 2 елементи вибірки потрапляють до першого інтервалу.

До другого інтервалу потрапляють елементи вибірки зі значеннями від 1 включно до 3 , не включаючи 3 , тобто з нашої вибірки тільки $1, 1, 1, 2, 2, 2$. Отже, до другого інтервалу потрапляє 6 елементів вибірки

Аналогічно, до третього інтервалу потрапляють тільки елементи вибірки $3, 3, 3, 3, 3, 3, 3, 3, 3, 4, 4, 4, 4, 4, 4$ – їх 15, до четвертого – $5, 5, 5, 5, 6, 6, 6, 6, 6$ (усього їх 9), до п'ятого – $7, 7, 8, 8, 8$ (їх 5), до шостого інтервалу потрапляють елементи вибірки $9, 10, 10$, тобто лише 3. Запишемо ці результати в таблицю:

$[x_{i-1}, x_i)$	m_i	f_i	x_i^*
$[-1; 1)$	2		
$[1; 3)$	6		
$[3; 5)$	15		
$[5; 7)$	9		
$[7; 9)$	5		
$[9; 11)$	3		

Відносні частоти f_i обчислюють за формулою $f_i = m_i/N$. Отримаємо

$[x_{i-1}, x_i)$	m_i	f_i	x_i^*
$[-1; 1)$	2	2/40	
$[1; 3)$	6	6/40	
$[3; 5)$	15	15/40	
$[5; 7)$	9	9/40	
$[7; 9)$	5	5/40	
$[9; 11)$	3	3/40	

x_i^* обчислюють як середини відповідних інтервалів:

$[x_{i-1}, x_i)$	m_i	f_i	x_i^*
$[-1; 1)$	2	2/40	0
$[1; 3)$	6	6/40	2
$[3; 5)$	15	15/40	4
$[5; 7)$	9	9/40	6
$[7; 9)$	5	5/40	8
$[9; 11)$	3	3/40	10

Так для вказаної вибірки, побудовано групувану вибірку.

Характеристичні властивості КФРЧ та групуваної вибірки

До характеристичних належать властивості, які визначають об'єкт як такий. Запишемо **властивості КФРЧ**:

1) ступінчастий характер графіка, кускова сталість КФРЧ;

2) невід'ємність: $F_M(x) = \mathcal{C}_M(-\infty; x) = \frac{m_A}{N} \geq 0$;

3) монотонне неспадання: $x_2 > x_1 \Rightarrow F_M(x_2) \geq F_M(x_1)$;

4) неперервність зліва (графік): на правих кінцях східців значення КФРЧ збігається зі значенням самого східця;

5) нормованість: $F_N(+\infty)=1$, $F_N(-\infty)=0$: граничні значення КФРЧ, коли аргумент прямує відповідно до $+\infty$ чи $-\infty$.

Групувана вибірка в неперервному випадку за змістом, представленням та властивостями аналогічна ряду розподілу частот для дискретних даних. Тому її **характеристичні властивості** є аналогами для ряду розподілу частот.

До таких характеристичних властивостей належать невід'ємність та нормованість (сумарне одиничне значення у стовпчику відносних частот).

Результати застосування принципу групування для наочного представлення зображують у вигляді графіків у певній системі координат чи там, де це доцільно, у вигляді діаграм. Так здійснюється наочне представлення ряду розподілу для випадку дискретних даних та групуваної вибірки і КФРЧ – для неперервних.

Через графіки можуть бути зображені всі результати застосування принципу групування, коли результати спостереження числові, тобто для розподілів частот, у вигляді діаграм (ряд розподілу та групувана вибірка).

1.3.4. Зображення результату групування для дискретного та неперервного випадків: діаграми та функції

Діаграма – рисунок, на якому результати для дискретних даних чи групи результатів групуваної вибірки для неперервних даних представлено однотипними геометричними об'єктами, а їхні абсолютні та відносні частоти – геометричними характеристиками цих об'єктів.

Розрізняють три основні види діаграм: стовпчикова, секторна та лінійчаста.

1. Стовпчикова діаграма – це рисунок, на якому результати з ряду розподілу чи інтервали групування групуваної вибірки представлено горизонтальними або вертикальними стовпчиками, відкладеними від одного рівня, а довжини цих стовпчиків пропорційні абсолютним чи відносним частотам цих результатів. Коефіцієнт пропорційності визначається вибором масштабної одиниці та, іноді, додатковими міркуваннями, як наприклад у випадку графічного представлення гістограми.

Приклад 1.9

Для рядів розподілу, що відповідають розподілу за статтю та темпераментом (дискретні дані) маємо такі дані.

Стать новонароджених

Д	Х
0,49	0,51

Темпераменти

Ф	М	Х	С
0,48	0,12	0,2	0,2

Стовпчикові діаграми представлено на рис. 1.9, а і 1.9, б відповідно.

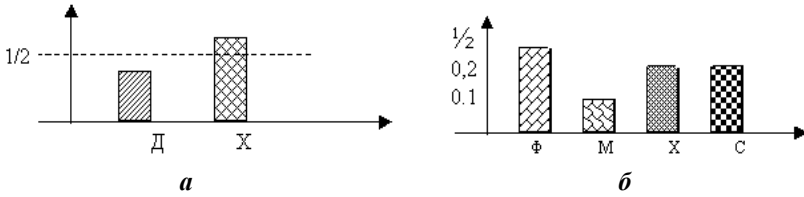


Рис. 1.9. Стовпчикові діаграми для дискретного типу даних

Подамо груповану вибірку таблицею

$[x_{i-1}, x_i)$	f_i
[50; 80)	0,1
[80; 110)	0,3
[110; 140)	0,5
[140; 170)	0,1

Далі зобразимо її стовпчиковою діаграмою (рис. 1.10).

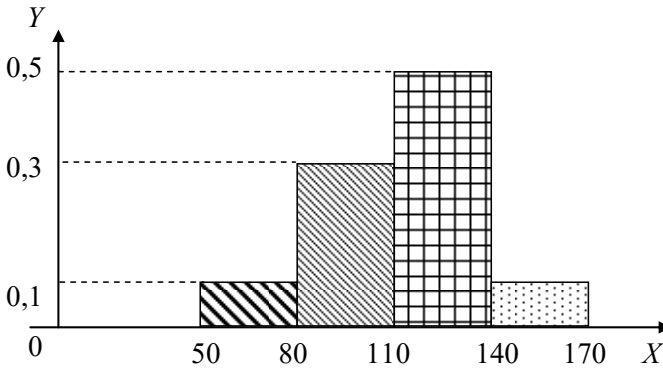
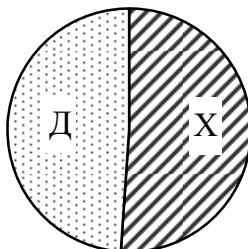


Рис. 1.10. Стовпчикова діаграма для групованої вибірки

2. Секторна діаграма – зображення ряду розподілу чи групованої вибірки, в якому результати (дискретні дані) чи групи результатів (неперервні дані – групована вибірка) представлено секторами одного кола (рис. 1.11) чи циліндра, а кути цих сек-

торів пропорційні абсолютним (відносним) частотам результатів або груп результатів:

$$\alpha_i = 360^\circ \times f_i.$$



$$\alpha_{\text{Д}} = 360^\circ \cdot 0,49; \quad \alpha_{\text{Х}} = 360^\circ \cdot 0,51$$

Рис. 1.11. Секторна діаграма

3. Лінійчаста діаграма. На лінійчастій діаграмі результати (дискретні дані) чи групи результатів (неперервні дані – групувана вибірка) зображено послідовними частинами одного й того самого відрізка завдовжки L .

Довжини відрізків l_i , (рис. 1.12) що зображують результати чи групи результатів, пропорційні абсолютним та відносним частотам відповідних результатів чи груп результатів: $l_i = L \cdot f_i$, де індекс i набуває значень порядкових номерів усіх результатів ряду розподілу чи групуваної вибірки.

Так, для рядів розподілу та групуваної вибірки, наведених у прикладі 1.9, відповідні діаграми мають вигляд:

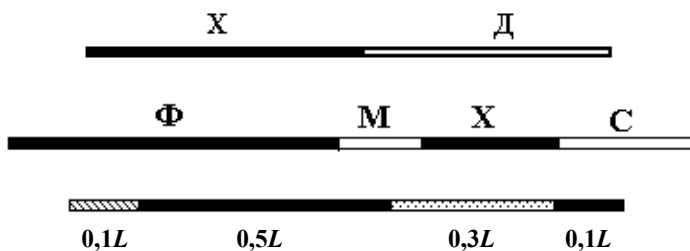


Рис. 1.12. Лінійчаста діаграма

Представлення результатів групування функціями та їхнє зображення

Графічним представленням ряду розподілу у випадку дискретних числових даних є полігон частот, у випадку неперервних – графік КФРЧ, а також графік, що зображає групувану вибірку – гістограма.

Ряд розподілу для числових даних можна в прямокутній системі координат представити дискретним набором точок на площині з координатами (x_i, f_i) , де індекс i набуває значень номерів усіх результатів ряду розподілу. Для наочності ці точки з'єднують відрізками і отриманий рисунок називають **полігоном частот**. На рис. 1.13 зображено полігон частот для числових результатів:

3	4	7	8	9
0,1	0,2	0,3	0,1	0,3

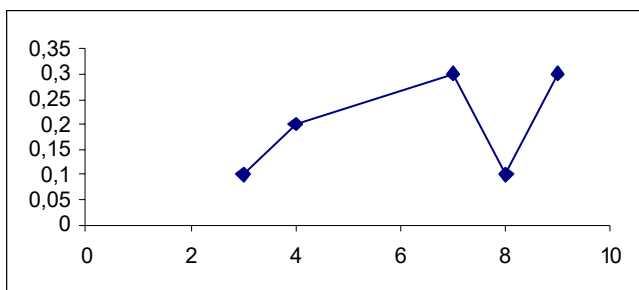


Рис. 1.13. Полігон частот

Гістограма як функція, характеристика гістограми. Гістограма (позначається $f_M(x)$) – це функція, визначена на всій числовій осі, є кусково-сталою зі значенням $\frac{f_i}{h}$ на інтервалах групування і нульовою – поза інтервалами групування.

Гістограма зображується графіком, що є ступінчастою лінією, яка складається з верхніх границь стовпчикової діаграми, коли стовпчики розташовані над відповідними інтервалами групування, а їх висотою є нормовані відносні частоти $\frac{f_i}{h}$, де

h – довжина інтервалу групування. За такого вибору площа стовпчика збігається з відносною частотою відповідного інтервалу.

Якщо поділити відносну частоту на h , то звичайна стовпчикова діаграма перетвориться на нормовану стовпчикову діаграму.

Приклад 1.10

Для групуваної вибірки з довжиною інтервалу $h = 30$

$[x_{i-1}, x_i)$	f_i	$\frac{f_i}{h}$
[50; 80)	0,1	$\frac{0,1}{30}$
[80; 110)	0,3	$\frac{0,3}{30}$
[110; 140)	0,5	$\frac{0,5}{30}$
[140; 170)	0,1	$\frac{0,1}{30}$

виконаємо нормовану стовпчикову діаграму і гістограму (рис. 1.14).

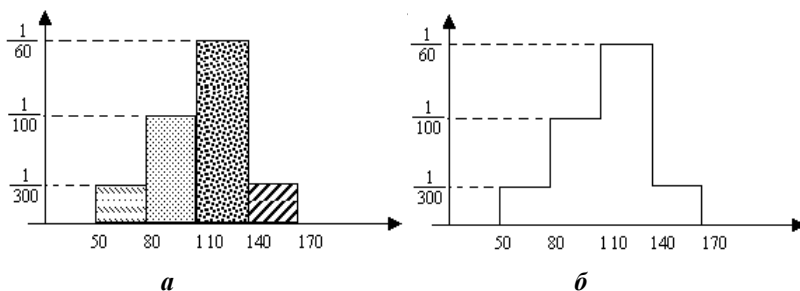


Рис. 1.14. Нормована стовпчикова діаграма (а) і гістограма (б)

Множення відповідних частот на $\frac{1}{h}$ не змінює загальної пропорційності висоти стовпчиків до абсолютних чи відносних частот, а змінює тільки коефіцієнт пропорційності.

Характеристичні властивості гістограми визначаються характеристичними властивостями групуваної вибірки:

- 1) кускова сталість;
- 2) невід'ємність: $f_N(x) \geq 0$;
- 3) сумарна площа під графіком гістограми є одиничною.

**Важливі властивості
КФРЧ і гістограми**

Різницева властивість КФРЧ. Частота появи результатів групувань за інтервалами $[a; b]$ дорівнює різниці значень КФРЧ на границях інтервалу:

$$\mathcal{C}_N [a; b] = F_N(b) - F_N(a).$$

Площинна властивість гістограми. Частота появи результатів із множини B , отриманої об'єднанням частини інтервалів групування, дорівнює площі під графіком гістограми над відповідною множиною:

$$\mathcal{C}_N (B) = S_B^{f_N}.$$

Доведення

Для довільного b $(-\infty; b) = (-\infty; a) + [a; b]$ є сумою двох несумісних подій (не мають спільних елементів). За адитивністю

$$\begin{aligned} \mathcal{C}_N (-\infty; b) &= \mathcal{C}_N ((-\infty; a) + [a; b]) = \mathcal{C}_N (-\infty; a) + \mathcal{C}_N [a; b], \\ F_N(a) &= \mathcal{C}_N(-\infty; a), \quad F_N(b) = \mathcal{C}_N(-\infty; b), \\ F_N(b) &= F_N(a) + \mathcal{C}_N[a; b]. \end{aligned}$$

Таким чином,

$$\mathcal{C}_N [a; b] = F_N(b) - F_N(a).$$

Доведення площинної властивості проілюструємо на прикладі (доведення у загальному випадку проводиться так само).

Приклад 1.11

Нехай $B = [50; 80) + [140; 170)$. Тоді за адитивністю $\mathcal{C}_N (B) = \mathcal{C}_N([50; 80) + [140; 170))$: події $[50; 80)$ та $[140; 170)$ несумісні (не мають спільних елементів).

Оскільки $\mathcal{C}_N[140; 170) = h \cdot f_4 / h = S_4$ та $\mathcal{C}_N[50; 80) = h \cdot f_1 / h = S_1$ і $S_1 + S_4 = S_B^{f_N}$, то

$$\mathcal{C}_N (B) = S_B^{f_N}.$$

Приклад 1.12 (групування дискретних даних та зображення результатів групування)

Маємо вибірку: 10, 14, 12, 14, 17, 16, 13, 15, 12, 14 (вік членів шахового гуртка). Записати цю вибірку у вигляді варіаційного ряду, знайти розмах вибірки. Скласти ряд розподілу частот вибірки. Зобразити результати групування у вигляді секторної та стовпчикової діаграм.

Розв'язання

Упорядковуємо елементи вибірки за зростанням і отримаємо варіаційний ряд:

10, 12, 12, 13, 14, 14, 14, 15, 16, 17.

Розмах вибірки $R = 17 - 10 = 7$.

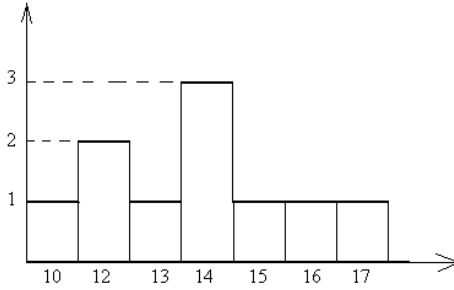
Різними в цій вибірці є елементи $u_1=10, u_2=12, u_3=13, u_4=14, u_5=15, u_6=16, u_7=17$, їх абсолютні частоти $m_1=1, m_2=2, m_3=1, m_4=3, m_5=1, m_6=1, m_7=1$. Об'єм вибірки $N=10$. Отже, ряд розподілу частот вихідної вибірки має вигляд

u_i	10	12	13	14	15	16	17
f_i	$\frac{1}{10}$	$\frac{1}{5}$	$\frac{1}{10}$	$\frac{3}{10}$	$\frac{1}{10}$	$\frac{1}{10}$	$\frac{1}{10}$

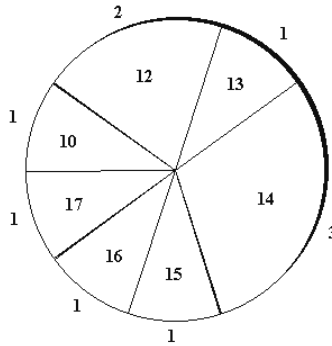
Тут відносні частоти $f_i = \frac{m_i}{N}$. Для контролю правильності знаходимо $\sum_{i=1}^7 m_i = N = 10, \sum_{i=1}^7 f_i = 1$. Будуємо стовпчикову діаграму (рис. 1.15, а).

Для побудови секторної діаграми (рис. 1.15, б) виконаємо додаткові розрахунки: $\alpha = \frac{360^\circ}{N} = \frac{360^\circ}{10} = 36^\circ$ і кожному елементу u_i ряду розподілу частот відповідає сектор діаграми з кутом $\alpha_i = \alpha \cdot m_i$, тобто для u_1, u_3, u_5, u_6 та u_7 $\alpha_1 = \alpha_3 = \alpha_5 = \alpha_6 = \alpha_7 = 36^\circ \cdot 1 = 36^\circ$; для u_2 маємо $\alpha_2 = 36^\circ \cdot 2 = 72^\circ$; для u_4 маємо $\alpha_4 = 36^\circ \cdot 3 = 108^\circ$.

Перевірка: $\sum_{i=1}^7 \alpha_i = 360^\circ$.



a



б

Рис. 1.15. Стовпчикова (а) і секторна (б) діаграми

Приклад 1.13 (групування неперервних даних та зображення результатів групування)

Представити вибірку з 55 спостережень у вигляді таблиці розподілу частот, використовуючи 7 інтервалів:

20,3	15,4	17,2	19,2	23,3	17,9	21,9
15,3	16,8	13,2	20,4	16,5	19,7	10,5
14,3	20,1	16,8	14,7	20,8	19,5	15,3
19,3	17,8	16,2	15,7	22,8	21,9	12,5
10,1	21,1	17,8	14,7	14,5	18,0	18,4
13,9	19,1	18,5	20,2	23,8	16,7	20,4
17,5	17,2	17,6	17,8	21,3	17,5	19,4
17,8	13,5	17,8	11,8	18,6	17,1.	

Результати групування зобразити у вигляді гістограми та об'їдної цїї гістограми.

Розв'язання

$a_{\min}=10,1$, $a_{\max}=23,8$, $N=55$; кїлькїсть їнтервалїв $M = 7$, розмах вибїрки $R = 23,8-10,1=13,7$. Довжина їнтервалу групування

$$h = \frac{13,7}{7-1} \approx 2,28.$$

$$x_0 = a_{\min} - h/2 \approx 8,96,$$

$$x_1 = x_0 + h \approx 11,24,$$

$$x_2 = x_1 + h \approx 13,52 \dots$$

Результати групування заносимо в табл. 1.3.

Таблиця 1.3

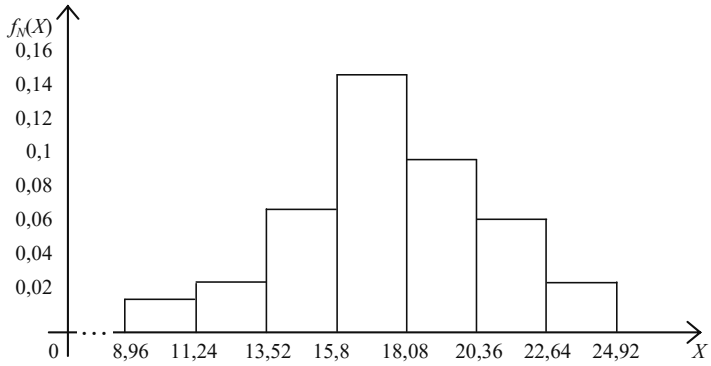
Результати їнтервального групування

№ п/п	їнтервал	Середина їнтервалу	Частота m_i	Вїдносна частота f_i	Висота стовпчика h_i
1	[8,96; 11,24)	10,1	2	0,03636	0,015949
2	[11,24; 13,52)	12,38	3	0,54545	0,023923
3	[13,52; 15,8)	14,66	9	0,163636	0,07177
4	[15,8; 18,08)	16,94	18	0,327273	0,143541
5	[18,08; 20,36)	19,22	12	0,218182	0,095694
6	[20,36; 22,64)	21,5	8	0,145455	0,063796
7	[22,64; 24,92]	23,78	3	0,054545	0,023923

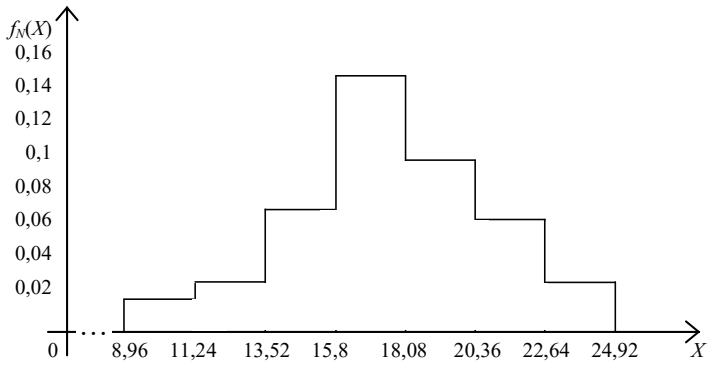
Оскїльки довжини їнтервалїв $h = 2,28$, то в гїстограмї висоти стовпчикїв $h_i = \frac{n_i}{hN}$ (тодї сума площ усїх стовпчикїв

$$S = \sum_{i=1}^7 h \cdot h_i = 1).$$

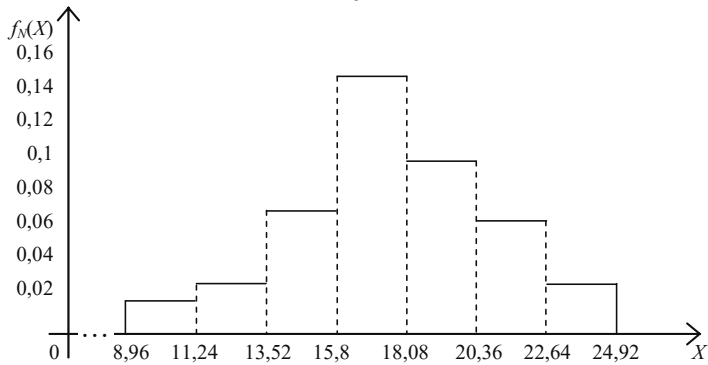
Будуємо нормовану стовпчикову дїаграму, гїстограму ї гїстограму як функцїю (рис. 1.16).



a



b



c

Рис. 1.16. Нормована стовпчикова діаграма (а), гістограма (б), гістограма як функція (в)

Приклад 1.14 (КФРЧ та полігон частот)

У результаті опитування 55 пересічних громадян про їхній сумарний місячний прибуток отримано вибірку:

203, 154, 172, 192, 230, 181, 219, 153, 168, 132, 204, 165, 197, 205, 143, 201, 168, 147, 208, 145, 153, 193, 178, 162, 157, 228, 219, 125, 110, 211, 183, 147, 145, 181, 184, 139, 191, 185, 202, 230, 167, 204, 195, 172, 196, 178, 213, 175, 194, 178, 135, 178, 118, 186, 191.

Побудувати для неї полігон частот та КФРЧ.

Розв'язання

Спочатку згрупуємо дані. Оскільки розмах вибірки $R = 230 - 110 = 120$, то за довжину інтервалу групування візьмемо

$h = \frac{120}{7-1} = 20$, за перший інтервал візьмемо $[100; 120)$. Результати запишемо у табл. 1.4.

Таблиця 1.4

Групована вибірка

№ п/п	Інтервал	Середина інтервалу z_i	Частота m_i	Накопичена частота $\sum_{j=1}^i m_j$	Відносна частота $\frac{m_i}{N}$	Накопичена відносна частота $\sum_{j=1}^i \frac{m_j}{N}$
1	[100; 120)	110	2	2	0,0364	0,0364
2	[120; 140)	130	4	6	0,0727	0,1091
3	[140; 160)	150	8	14	0,1455	0,2546
4	[160; 180)	170	12	26	0,2182	0,4728
5	[180; 200)	190	16	42	0,2909	0,7637
6	[200; 220)	210	10	52	0,1818	0,9455
7	[220; 240]	230	3	55	0,0545	1,0000

Полігон частот будуюмо як ламану з вершинами в точках (z_i, m_i) (рис. 1.17, а).

Оскільки середина першого інтервалу групування $z_1=110$, то КФРЧ $F_N(x) = 0$ при $x \leq 110$ і аналогічно $F_N(x) = 1$ при $x > 230$.

На (110, 230] графік КФРЧ будемо за третім та останнім стовпчиками табл. 1.3. у вигляді сідців (рис. 1.17, б).

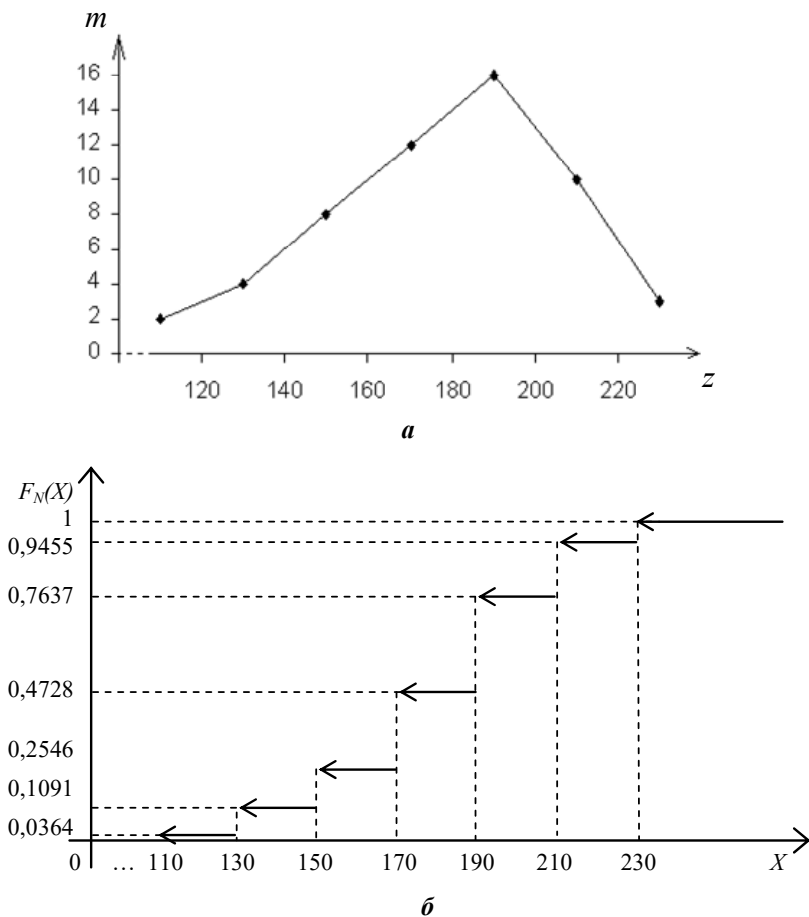


Рис. 1.17. Полігон частот (а), КФРЧ (б)

ЗАДАЧІ

Групування дискретних даних та зображення результатів групування

№ 1

Час, що був витрачений опитуваними на ранкову гімнастику, склав 11, 15, 12, 0, 16, 19, 6, 11, 12, 13, 16, 8, 9, 14, 5, 11, 3 хв. Для цієї вибірки знайти об'єм, розмах, побудувати варіаційний ряд та ряд розподілу частот.

№ 2

За 16 місяців витрати міського бюджету на озеленення склали 17, 18, 16, 16, 17, 18, 19, 17, 15, 17, 19, 18, 16, 16, 18, 18 тис. грн. Знайти об'єм цієї вибірки, її розмах, побудувати варіаційний ряд та ряд розподілу частот.

№ 3

У групі на заняттях із статистики провести експеримент щодо реєстрації місяця народження студентів (опитування виконати, наприклад, за списком групи). Побудувати варіаційний ряд та ряд розподілу частот отриманої вибірки. Зобразити результат у вигляді секторної діаграми.

№ 4

Порахувати кількість хлопців і дівчат у групі на заняттях із статистики. Результати підрахунків зобразити у вигляді секторної, стовпчикової та лінійчастої діаграм.

№ 5

У таблиці множення порахувати кількість усіх цифр від 0 до 9. Результати підрахунків оформити як ряд розподілу частот та зобразити у вигляді стовпчикової і секторної діаграм.

№ 6

Довести, що сума відносних частот у будь-якій вибірці дорівнює одиниці.

№ 7

За секторною діаграмою (рис. 1.18) побудувати відповідний ряд розподілу частот та стовпчикову діаграму:

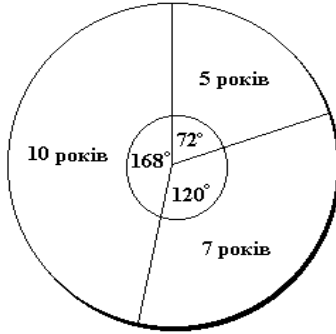


Рис. 1.18. Секторна діаграма

№ 8

Чи можна в попередній задачі знайти об'єм вихідної вибірки, розмах варіаційного ряду? Відповідь пояснити.

№ 9

У першому класі 4 дітей віку шести років, 20 – семи років та 6 – восьми років. Побудувати секторну, стовпчикову та лінійчасту діаграми.

№ 10

Провести опитування студентів на занятті з теорії ймовірностей та записати числа їх днів народження. Знайти розмах вибірки. Згрупувати отримані дані по роках та скласти груповану вибірку. Результати групування зобразити графічно у вигляді гістограми.

№ 11

Для вихідних даних із попередньої задачі провести групування по місяцях та скласти таблицю розподілу частот. Результати групування зобразити графічно у вигляді гістограми та секторної діаграми.

№ 12

Використовуючи умову та розв'язок попередньої задачі, провести групування по порах року (осінь, зима, весна, літо). Результати групування зобразити графічно у вигляді гістограми та секторної діаграми і порівняти їх з відповідниками з попередньої задачі. Зробити висновки.

№ 13

Провести опитування студентів групи на занятті з теорії ймовірностей та записати їх темпераменти (сангвінік, холерик, меланхолік, флегматик). Згрупувати отримані результати за типами темпераменту. Результати групування зобразити у вигляді гістограми та лінійної діаграми.

№ 14

Розв'язати попередню задачу, провівши групування за ознаками "сильний тип" (сангвінік, холерик), "слабий тип" (меланхолік, флегматик).

Групування неперервних даних та зображення результатів групування

№ 1

Час розв'язування задачі учнями 4-го класу склав (у секундах):

38, 60, 41, 51, 33, 42, 45, 21, 53, 60,
68, 52, 47, 46, 49, 49, 14, 57, 54, 59,
77, 47, 28, 48, 58, 32, 42, 58, 61, 30,
61, 35, 47, 72, 41, 45, 44, 55, 30, 40,
67, 65, 39, 48, 43, 60, 54, 42, 59, 50.

Знайти розмах вибірки, число та довжину інтервалів та побудувати групувану вибірку.

№ 2

Час роботи електроламп одного типу склав (у годинах):

13,4 14,7 15,2 15,1 13,0 8,8 14,0 17,9 15,1 16,5 16,6
14,2 16,3 14,6 11,7 16,4 15,1 17,6 14,1 18,8 11,6 13,9
18,0 12,4 17,2 14,5 16,3 13,7 15,5 16,2 8,4 14,7 15,4
11,3 10,7 16,9 15,8 16,1 12,3 14,0 17,7 14,7 16,2 17,1
10,1 15,8 18,3 17,5 12,7 20,7 13,5 14,0 15,7 21,9 14,3
17,7 15,4 10,9 18,2 17,3 15,2 16,7 17,3 12,1 19,2

Знайти розмах вибірки, число та довжину інтервалів і побудувати групувану вибірку.

№ 3

Скласти вибірку росту студентів у групі на занятті зі статистики. Побудувати групувану вибірку для отриманих даних.

№ 4

Скласти вибірку часу, витраченого кожним із студентів у групі на розв'язок попередньої задачі. Побудувати групувану вибірку для отриманих даних.

Кумулятивна функція розподілу частот (КФРЧ) та полігон частот

№ 1

В опитуваних визначався вік, коли вони вперше заробили гроші власною працею. У результаті отримано дані, представлені статистичним рядом:

Вік	15	16	17	18	19
Кількість опитаних	1	4	5	4	2

За даним статистичним рядом побудувати полігон частот та КФРЧ.

№ 2

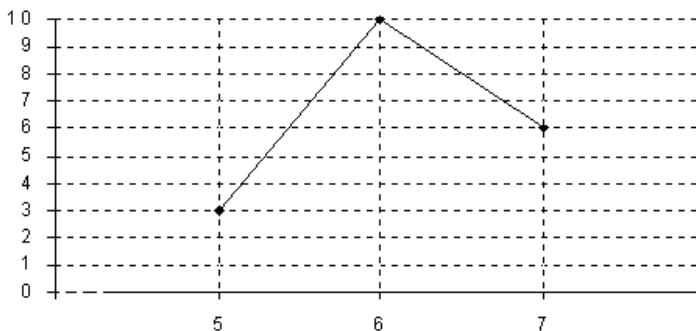
В опитуваних визначалася кількість днів на тиждень, коли вони виконують фізичні вправи. У результаті отримано дані у вигляді статистичного ряду:

Кількість днів на тиждень	0	1	2	3	4	5	6	7
Кількість опитаних	12	16	7	3	2	4	1	1

За даними статистичного ряду побудувати полігон частот та КФРЧ.

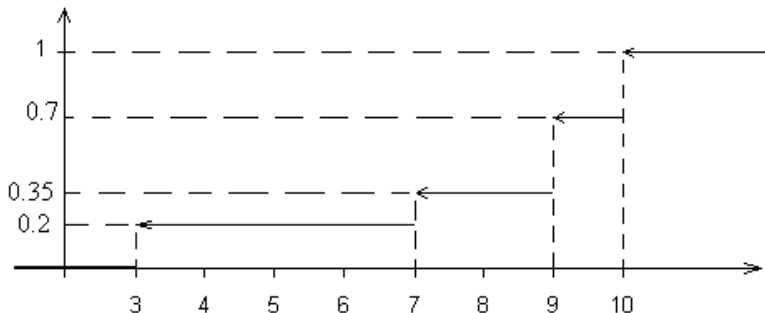
№ 3

За даним полігоном частот побудувати графік відповідної КФРЧ:



№ 4

За КФРЧ побудувати відповідний полігон частот за умови, що об'єм вибірки $n = 100$:



№ 5

У першому класі 4 дітей шести років, 20 – семи років та 6 – восьми років. Побудувати полігон частот та КФРЧ за цими даними.

№ 6

Провести опитування студентів на занятті з теорії ймовірностей та записати їхні дати народження. Згрупувати отримані дані по роках. Побудувати полігон частот та КФРЧ.

№ 7

Для вихідних даних попередньої задачі провести групування по місяцях та побудувати полігон частот і КФРЧ.

№ 8

Використовуючи умову і розв'язок попередньої задачі, провести групування по порах року (осінь, зима, весна, літо) та побудувати полігон частот і КФРЧ.

№ 9

Побудувати гістограму та полігон частот для задачі № 1 із підрозд. 1.3.

№ 10

Побудувати гістограму та полігон частот для задачі № 2 із підрозд. 1.3.

1.3.5. Випадкова величина, розподіли частот випадкових величин

Випадковою величиною називають будь-яку числову характеристику результату спостереження. Це визначення в разі дискретних даних є безумовним, а для неперервного випадку супроводжується технічною вимогою (вимірністю), яка зараз не розглядається.

Позначаються випадкові величини великими латинськими літерами з другої частини абетки (як правило літери X, Y, Z).

Приклад 1.15

1) У результаті тестів характеристик інтелектуального розвитку (коефіцієнт IQ) для можливих темпераментів середніми значеннями є такі:

U	Φ	M	X	C
N	90	95	110	105

Ця таблиця задає значення числової характеристики інтелектуального розвитку, яка є в нашому дослідженні випадковою величиною. Якщо позначити її через Q , то ці значення, як це впливає з таблиці, будуть:

$$Q(\Phi) = 90, Q(M) = 95, Q(X) = 110, Q(C) = 105.$$

2) Випадковими величинами є такі характеристики: стать новонароджених, X – їх зріст, Y – вага новонародженої дитини і т. д.

Числова характеристика перетворюється на випадкову величину, коли вона характеризує результати спостереження випадкового масового явища.

Із кожною випадковою величиною пов'язують розподіл частот \mathcal{C}_X , що описує появу тих чи інших груп B значень цієї випадкової величини за формулою

$$\mathcal{C}_X(B) = \mathcal{C}_N\{X \in B\}.$$

Випадкова величина є зручним засобом опису подій, пов'язаних із значеннями досліджуваної характеристики X (випадкової величини). Набування значень із числової множини B записують як $\{X \in B\}$. Зокрема, подія, що полягає в появі значень досліджуваної характеристики з інтервалу $[a, b)$, записують як $\{X \in [a, b)\}$, а з інтервалу $(-\infty, x)$ – як $\{X \in (-\infty, x)\}$ чи $\{X < x\}$.

Розподіл частот випадкової величини залежно від типу даних (дискретний чи неперервний) множини можливих результатів описують рядом розподілу або КФРЧ, які в цьому випадку називають **рядом розподілу частот випадкової величини** та **кумулятивною функцією розподілу частот випадкової величини**.

Зокрема КФРЧ $F_N^X(x)$ випадкової величини X визначається співвідношенням

$$F_N^X(x) = \mathcal{C}_X((-\infty, x)) = \mathcal{C}_N\{X < x\}.$$

Розподіл частот – це частоти числових груп результатів або подій A з числової множини U .

Якщо при описі числових груп результатів явним чином вживають позначення випадкової величини (досліджуваної числової характеристики), то розподіл частот називають *розподілом частот випадкової величини*. Будь-який розподіл частот може розглядатись як розподіл частот випадкової величини. Наприклад частоти, які пов'язані зі зростом, вагою, часом, об'ємом, масою та іншими фізичними величинами, можна назвати розподілом частот.

1.3.6. Запис умов групування за допомогою випадкової величини

У співвідношенні $\mathcal{C}_X(B) = \mathcal{C}_N\{X \in B\}$, що визначає розподіл частот, є дві різні частоти $\mathcal{C}_X(B)$ та $\mathcal{C}_N\{X \in B\}$, одна з яких знаходиться за допомогою іншої. Запис $\mathcal{C}_N\{X \in B\}$ краще відображає зміст відповідності між множиною B та її відносною частотою. Цей вираз можна прочитати як "відносна частота того, що значення досліджуваної числової характеристики X належать B ".

Принцип "менше ніж". Якщо $B = (-\infty; x)$, то співвідношення $\mathcal{C}_X(B) = \mathcal{C}_N\{X \in B\}$ між розподілом частот випадкової величини та вихідною частотою набуває вигляду

$$\mathcal{C}_X(-\infty; x) = \mathcal{C}_N\{X < x\}.$$

Приклад 1.16

Нехай X – зріст новонародженого у сантиметрах: 58, 71, 72, 71, 73, 72.

Відносні частоти появи зросту, що менші за значення $x = 50, 60, 70, 80$, записують у вигляді

$$\mathcal{C}_N(-\infty; 50) = \mathcal{C}_N\{X < 50\} = 0;$$

$$\mathcal{C}_N(-\infty; 60) = \mathcal{C}_N\{X < 60\} = \frac{m_A}{N} = \frac{1}{6};$$

$$\mathcal{C}_N(-\infty; 70) = \mathcal{C}_N\{X < 70\} = \frac{1}{6};$$

$$\mathcal{C}_N(-\infty; 80) = \mathcal{C}_N\{X < 80\} = 1.$$

Інтервальний принцип групування. Співвідношення щодо частот напівнескінченних ліворуч інтервалів повною мірою відносять і до скінченних інтервалів

$$\mathcal{C}_X[a; b) = \mathcal{C}_N\{a \leq X < b\}.$$

Приклад 1.17

З умови прикладу 1.10 маємо

$$\mathcal{C}_X[50; 80) = \mathcal{C}_N\{50 \leq X < 80\} = \mathcal{C}_N\{X \in [50; 80)\} = 0,1;$$

$$\mathcal{C}_X[80; 110) = \mathcal{C}_N\{80 \leq X < 110\} = \mathcal{C}_N\{X \in [80; 110)\} = 0,3;$$

$$\mathcal{C}_X[110; 140) = \mathcal{C}_N\{110 \leq X < 140\} = \mathcal{C}_N\{X \in [110; 140)\} = 0,5;$$

$$\mathcal{C}_X[140; 170) = \mathcal{C}_N\{140 \leq X < 170\} = \mathcal{C}_N\{X \in [140; 170)\} = 0,1.$$

1.4. Умовна частота.

Формула Байєса для умовних частот

1.4.1. Умовна частота, зв'язок із безумовною

Існує клас задач, в яких потрібно спочатку провести спостереження за фіксованих умов A , а потім з отриманої вибірки додатково відібрати ті спостереження, які задовольняють додаткову вимогу (вимоги) B . Частоту, побудовану за сформованою таким чином вибіркою, називають умовною і позначають $\mathcal{C}_N(A/B)$.

Приклад 1.18

Нехай вихідна вибірка a_1, \dots, a_N об'єму $N = 17$ відповідає спостереженням темпераментів і має вигляд

Ф С С Х М С С М Х С С Ф М М С С Х .

Нехай далі нас цікавлять тільки "помірквані" темпераменти, до яких віднесемо флегматиків та меланхоліків. Позначимо цю умову ("поміркваний" темперамент) через B : $B = \{\Phi, M\}$. Відбір із вихідної вибірки спостережень, що відповідають умові B ("поміркваний" темперамент) дасть вибірку об'єму 6: Φ, M, M, Φ, M, M .

Частоти появи флегматика у вихідній вибірці і у вибірці, сформованій за додатковою вимогою, можуть бути записані у вигляді:

$$\mathcal{C}_N(\Phi) = 2/17 \text{ (2 з 17),}$$

$$\mathcal{C}_N(\{\Phi\}/B) = 2/6 = 1/3 \text{ (2 з 6).}$$

Отже, **умовною частотою події A за умови B** (позначають $\mathcal{C}_N(A/B)$) називають частоту, отриману завдяки додаванню до початкових умов спостереження додаткової вимоги B .

Умовна частота має всі характеристичні властивості частоти і є звичайною частотою.

Термін "умовна" показує на двоступеневий характер її побудови:

1) спостерігається вихідна вибірка (фіксується);

2) із вихідної вибірки вибираються спостереження, що задовольняють додаткову умову (події) B .

В описі прикладної задачі на необхідність застосування умовної частоти може вказувати згадка про умову ("за умови, що...") або наявність виразу "серед ...".

Етапи побудови частот у схемі з умовною частотою мають такий вигляд:

$$\begin{array}{ccccccc} a_1, & \dots, & a_N & \rightarrow & \mathcal{C}_N(A) \\ \downarrow & & \downarrow & & B \\ a_1', & \dots, & a_N' & \rightarrow & \mathcal{C}_N'(A) = \mathcal{C}_N(A/B) \end{array}$$

Якщо в одному контексті одночасно вживаються обидві частоти, то $\mathcal{C}_N(A)$ називають безумовною, а $\mathcal{C}_N(A/B)$ – умовною.

Приклад 1.19

Позначимо

A – захворюваність;

B – належність до групи ризику.

Якщо записати $\mathcal{C}_N(A)$ як частоту захворюваності серед усього населення, то частотою захворювання серед групи ризику буде умовна частота $\mathcal{C}_N(A/B)$.

Хоча визначення умовної частоти передбачає двоступеневу процедуру (первісні спостереження і наступний відбір), підрахувати її можна за безумовною частотою.

Зв'язок умовної та безумовної частот визначається формулою

$$\mathcal{C}_N(A/B) = \frac{\mathcal{C}_N(A \cdot B)}{\mathcal{C}_N(B)}. \quad (1.5)$$

Доведемо цю формулу. Умовна частота, як і будь-яка відносна частота, є відношенням абсолютної частоти m'_A до об'єму вибірки N' . Штрих означає, що це характеристики "прорідженої вибірки", тобто вибірки, до якої застосовано умову B . Абсолютна частота $m'_A = m_{AB}$, а $N' = m_B$. Звідси отримаємо

$$\mathcal{C}_N(A/B) = \frac{m'_A}{N'} = \frac{m_{AB}}{m_B} = \frac{m_{AB} / N}{m_B / N} = \frac{\mathcal{C}_N(A \cdot B)}{\mathcal{C}_N(B)}.$$

Приклад 1.20

Для прикладу 1.18 обчислення умовної частоти безпосередньо за визначенням та обчислення за формулою зв'язку дають однаковий результат. Дійсно, із прикладу 1.18 $\mathcal{C}_N(\{\Phi\}/B) = 2/6 = 1/3$. Позначимо $A = \{\Phi\}$, $B = \{M, \Phi\}$, $A \cdot B = \{\Phi\}$. Тоді

$$\mathcal{C}_N(B) = \mathcal{C}_N(\{M, \Phi\}) = 6/17,$$

$$\mathcal{C}_N(A \cdot B) = \mathcal{C}_N(\{\Phi\}) = 2/17,$$

і за формулою (1.5)

$$\mathcal{C}_N(\{\Phi\}/\{M, \Phi\}) = \frac{\mathcal{C}_N(A \cdot B)}{\mathcal{C}_N(B)} = 2/17 : 6/17 = 2/6 = 1/3.$$

Отже, результати збігаються.

1.4.2. Теорема добутку для частот

Теорема добутку для частот – це формула обчислення частоти добутку двох подій як добутку умовної частоти на частоту умови. Цьому формулюванню відповідають два варіанти запису залежно від того, яку із двох подій віднести до умови:

$$\mathcal{C}_N(A \cdot B) = \mathcal{C}_N(A/B) \cdot \mathcal{C}_N(B),$$

$$\mathcal{C}_N(A \cdot B) = \mathcal{C}_N(B/A) \cdot \mathcal{C}_N(A). \quad (1.6)$$

Це твердження можна отримати з формули зв'язку умовної та безумовної частот (1.5), якщо обидві частини рівності помножити на знаменник правої.

1.4.3. Формула Байєса для частот

Ця формула показує співвідношення між двома ненульовими умовними частотами з оберненим порядком подій:

$$\mathcal{C}_N(A/B) = \frac{\mathcal{C}_N(B/A) \cdot \mathcal{C}_N(A)}{\mathcal{C}_N(B)}. \quad (1.7)$$

Доведення випливає з теореми добутку:

$$\begin{aligned} \mathcal{C}_N(A/B) \cdot \mathcal{C}_N(B) &= \mathcal{C}_N(B/A) \cdot \mathcal{C}_N(A) \Rightarrow \mathcal{C}_N(A/B) = \\ &= \frac{\mathcal{C}_N(B/A) \cdot \mathcal{C}_N(A)}{\mathcal{C}_N(B)}, \text{ якщо } \mathcal{C}_N(B) \neq 0. \end{aligned}$$

Приклад 1.21

Для прикладу 1.18 продемонструємо застосування формули Байєса (1.7).

Позначимо $A = \{\Phi\}$, $B = \{M, \Phi\}$. Тоді $A \cdot B = \{\Phi\}$, а відповідні частоти

$$\mathcal{C}_N(A) = \mathcal{C}_N(\Phi) = 2/17,$$

$$\mathcal{C}_N(B) = \mathcal{C}_N(\{M, \Phi\}) = 6/17,$$

$$\mathcal{C}_N(A \cdot B) = \mathcal{C}_N(\Phi) = 2/17.$$

Умовна частота

$$\mathcal{C}_N(A/B) = \mathcal{C}_N(\{\Phi\}/\{M, \Phi\}) = \frac{\mathcal{C}_N(A \cdot B)}{\mathcal{C}_N(B)} = 2/17 : 6/17 = 2/6 = 1/3,$$

що, власне, є частотою появи флегматиків (подія A) серед осіб із поміркованим темпераментом (подія B).

Застосування формули Байєса дає:

$$\mathcal{C}_N(B/A) = \frac{\mathcal{C}_N(A/B) \cdot \mathcal{C}_N(B)}{\mathcal{C}_N(A)} = \frac{\frac{1}{3} \cdot \frac{6}{17}}{\frac{2}{17}} = 1 \text{ (дійсно, кожен флегматик має поміркований тип темпераменту).}$$

Більш детально та ілюстративно умовні частоти та умовні ймовірності розглядатимуться в розд. 2.

1.5. Інтегральні характеристики вибірки

Інтегральною характеристикою вибірки є функція від вибірки, яка своїм одним значенням характеризує одну з ознак досліджуваного масового явища, за яким побудована вибірка.

Є два основні типи інтегральних характеристик: "типових представників" (або міри центральної тенденції) та "розсіювання" (міри розсіювання).

Вибіркове середнє. До "типових представників" вибірки в першу чергу належать вибіркове середнє (позначається \bar{x}) – середнє арифметичне спостережуваних числових значень:

$$\bar{x} = \frac{a_1 + \dots + a_N}{N}.$$

Аналогічний до вибіркового середнього показник, який обчислюють не для вибірки, а для генеральної сукупності, називають математичним сподіванням випадкової величини X і позначають MX . Його розглянемо нижче, у підрозд. 2.4.

Розмах. Одним із представників інтегральних характеристик вибірки типу "розсіювання" є розмах (R) вибірки – це різниця між максимальним x_{\max} і мінімальним x_{\min} значеннями елементів вибірки:

$$R = x_{\max} - x_{\min}.$$

Приклад 1.22

Для вибірки зросту

170, 189, 182, 184, 182, 172, 181, 178, 171, 180, 183, 173, 181, 173, 189, 176, 179, 191, 186, 192

отримаємо:

$$\bar{x} = 3612 / 20 = 180,6, x_{\min} = 170, x_{\max} = 192, R = 192 - 170 = 22.$$

Отже, вибіркове середнє та розмах характеризують різні боки вибірки зросту: середнє дає "середній зріст", тобто "типового представника" зросту, а розмах характеризує "розсіювання", тобто ступінь розбіжності у зрості осіб, представлених у вибірці.

Вибіркова дисперсія. До інтегральних характеристик типу "розсіювання" належить вибіркова дисперсія, яка позначається $\hat{\sigma}^2$ та S^2 і є інтегральною характеристикою вибірки, котра визначається однією з двох формул:

$$\left\{ \begin{array}{l} \hat{\sigma}^2 = \frac{(a_1 - MX)^2 + \dots + (a_N - MX)^2}{N} \\ S^2 = \frac{(a_1 - \bar{x})^2 + \dots + (a_N - \bar{x})^2}{N-1} \end{array} \right.$$

Вибіркову дисперсію можна записати, як середнє ($\hat{\sigma}^2$) чи "майже середнє" (S^2) квадратичних відхилень елементів вибірки від вибіркового (емпіричного) середнього (принципова відмінність між $\hat{\sigma}^2$ та S^2 розглядається в підрозд. 3.4).

Усі інтегральні характеристики обчислюють або безпосередньо за вибіркою (для дискретного випадку), або за результатами застосування принципу групування ряду розподілу, КФРЧ чи групованої вибірки за інтервальним принципом (для неперервного випадку).

Квартилі, медіана та інтерквартильний розмах

До інтегральних характеристик вибірки належать також три квартилі (позначаються Q_1 , Q_2 , Q_3) та характеристики, побудовані на їх основі.

Квартілями – Q_1 , Q_2 , Q_3 – називають значення досліджуваної характеристики (випадкової величини), які відтинають поспідовно чверть, половину та три чверті кількості менших за величиною значень:

- Q_1 – нижній (перший квартиль),
- Q_2 – медіана Me (другий квартиль),
- Q_3 – верхній (третій квартиль).

Q_3 можна отримати також як значення досліджуваної характеристики, яке відтинає чверть більших за розміром значень.

Медіаною Me називають таке значення досліджуваної характеристики, яке відтинає половину більших (або, що те саме, менших) за величиною значень.

До характеристики, побудованої за квартилями, відносять **інтерквартильний розмах R_1** , який визначають як різницю між верхнім та нижнім квартилями:

$$R_1 = Q_3 - Q_1.$$

R_1 визначає довжину інтервалу, який містить половину значень елементів вибірки, "середніх за величиною", крім чверті менших та чверті більших за величиною значень.

У дискретному випадку кватилі обчислюють за варіаційним рядом вибірки. За Q_1 беруть член варіаційного ряду з порядковим номером $\frac{N}{4}$, якщо N ділиться на 4 і $\left[\frac{N}{4}\right]+1$ у протилежному випадку, тобто той, який відокремлює першу чверть елементів варіаційного ряду за номерами.

$$Q_1 = \begin{cases} a_{\left[\frac{N}{4}\right]}^*, & \text{якщо } \frac{N}{4} \text{ ціле,} \\ a_{\left[\frac{N}{4}\right]+1}^*, & \text{якщо } \frac{N}{4} \text{ не ціле.} \end{cases}$$

Третій кватиль обчислюють за формулою

$$Q_3 = \begin{cases} a_{\left[\frac{3N}{4}\right]}^*, & \text{якщо } \frac{3N}{4} \text{ ціле,} \\ a_{\left[\frac{3N}{4}\right]+1}^*, & \text{якщо } \frac{3N}{4} \text{ не ціле.} \end{cases}$$

Другий кватиль Q_2 (медіана) обчислюють залежно від парності чи непарності об'єму вибірки N .

Якщо N непарне ($N = 2k+1$), то медіаною є член варіаційного ряду з порядковим номером $k+1$:

$$Q_2 = m = a_{k+1}^*.$$

Якщо N парне ($N = 2k$), то медіану обчислюють як середнє двох "середніх за номерами" членів варіаційного ряду:

$$Q_2 = m = \frac{a_k^* + a_{k+1}^*}{2}.$$

Отже,

$$Q_2 = m = \begin{cases} \frac{a_k^* + a_{k+1}^*}{2}, & \text{якщо } N = 2k, \\ a_{k+1}^*, & \text{якщо } N = 2k + 1. \end{cases}$$

Приклад 1.23

Для вибірки зросту

170, 189, 182, 184, 182, 172, 181, 178, 171, 180, 183, 173, 181, 173,
189, 176, 179, 191, 186, 192

маємо варіаційний ряд

170, 171, 172, 173, 173, 176, 178, 179, 180, 181, 181, 182, 182, 183,
184, 186, 189, 189, 191, 192,

$N = 20 = 2k$ (парне, $k = 10$), тому

$$Q_2 = m = \frac{a_k^* + a_{k+1}^*}{2} = \frac{181 + 181}{2} = 181,$$

$$Q_1 = a_{\left[\frac{N}{4}\right]}^* = a_5^* = 173,$$

$$Q_3 = a_{\left[\frac{3N}{4}\right]}^* = a_{15}^* = a_8^* = 184.$$

Квартилі для неперервного випадку обчислюють за групованою вибіркою. У цьому випадку інтервали, до яких потрапили квартилі, називають квартильними. Інтервал, до якого потрапив другий квартиль (медіана) називають медіанним інтервалом. Квартилі обчислюються за формулами

$$Me = Q_2 = x_{Me} + h_{Me} \cdot \frac{\frac{N}{2} - m_{Me}^*}{m_{Me}}, \quad (1.8)$$

де x_{Me} – початок медіанного інтервалу; h_{Me} – його довжина; m_{Me}^* – абсолютна частота; накопичена до початку медіанного інтервалу; m_{Me} – абсолютна частота медіанного інтервалу. Перший та третій квартилі обчислюються як

$$Q_1 = x_{Q_1} + h_{Q_1} \cdot \frac{\frac{N}{4} - m_{Q_1}^*}{m_{Q_1}}, \quad Q_3 = x_{Q_3} + h_{Q_3} \cdot \frac{\frac{3N}{4} - m_{Q_3}^*}{m_{Q_3}}, \quad (1.9)$$

де x_{Q_1} та x_{Q_2} – початки першого та третього квартильних інтервалів; h_{Q_1} та h_{Q_3} – довжини першого та третього інтервалів (у загальному випадку $h_{Q_1} = h_{Me} = h_{Q_3}$, але, якщо інтервал містить недостатню кількість елементів, то його об'єднують із сусіднім,

тому іноді довжина інтервалу може бути кратною h); $m_{Q_1}^*$ та $m_{Q_3}^*$ – абсолютні частоти, накопичені до початку першого та третього квартильних інтервалів; m_{Q_1} та m_{Q_3} – абсолютні частоти першого та третього інтервалів.

Приклад 1.24

Нехай досліджувана характеристика є середньомісячним прибутком осіб у Києві і нехай значення квартилів такі: $Q_1 = 200$, $Q_2 = 1200$, $Q_3 = 19000$.

Це означає, що чверть населення Києва має місячний прибуток менший за 200 грн, половина – 1200 грн і тільки чверть має місячний прибуток більший за 19000 грн. Розмах

$$R_I = Q_3 - Q_1 = 19000 - 200 = 18800.$$

До інтегральних характеристик виду "типовий представник" належить мода *Mo*, яку визначають безпосередньо за вибіркою як значення, що найчастіше зустрічається (повторюється) у вибірці, або за групуваною вибіркою, коли вона є серединою того інтервалу групування, який зустрічається найчастіше (до якого потрапила найбільша кількість спостережуваних значень).

Цей показник найпоширеніший при визначенні ознак нечислової природи: партії у виборчому списку, регіон проживання та ін.

1.5.1. Особливості застосування інтегральних характеристик вибірки

Однією з головних потреб обчислення інтегральних характеристик вибірки є дослідження поведінки вибірки, порівняння вибірок та дослідження тенденцій наповнення вибірки. Часто дослідник досліджує не сам ряд розподілу вибірки, а його певні характеристики. Наприклад, порівнюючи дві вибірки, отримані з неперервної множини значень, неможливо порівнювати два варіаційні ряди перш за все через те, що їх об'єми можуть бути різними.

У соціології, психології та інших галузях науки, де застосовують вибіркові дослідження, ознаки поділяють дещо за іншим принципом, ніж у теорії ймовірностей та математичній статистиці, що розширює методологічну та методичну базу досліджень.

Так, ознаки поділяють на три базові типи: номінальні, порядкові та метричні. Іноді окремим типом виносять дихотомічні ознаки.

Номінальні ознаки є "типовими представниками" дискретних ознак, для яких альтернативи ознаки для зручності просто кодуються числом, тобто варіанти відповідей на питання неможливо логічно впорядкувати за зростанням чи спаданням. До таких ознак належать, наприклад, стать, перелік політичних партій, перелік адміністративних одиниць та ін. Тобто значення 2 та 3 у вибірці не свідчать про те, що другий респондент має більший чи менший рівень якоїсь характеристики, а свідчать про те, що відповіді першого та другого відрізняються.

Порядкові ознаки також відносять до дискретних ознак, але значення, якими кодують альтернативи в ознаці вже мають не формальний, а порядковий характер. Це означає, що самі альтернативи логічно виставляють за зростанням (спаданням). До таких ознак належать, наприклад, рівень задоволеності чимось (1 – зовсім не задоволений; 2 – не задоволений; 3 – як задоволений, так і ні; 4 – задоволений; 5 – повністю задоволений) та ін. Якщо тут у вибірці маємо значення 2 та 3, то це свідчить про те, що рівень задоволеності, наприклад другого респондента, є вищим за рівень задоволеності першого.

Метричні ознаки, що часто також мають дискретний характер, можна вважати аналогом неперервних ознак у теорії ймовірності. До них застосовують ті самі методи аналізу та обчислюють ті самі інтегральні характеристики, що й до неперервних. Це пов'язано з великою кількістю альтернатив. Наприклад ознака "вік", теоретично може змінюватись від 0 до 120 років. До таких ознак часто застосовують методи групування неперервних ознак, описані вище.

Дихотомічний тип є за своїми ознаками номінальним, але відрізняється від нього наявністю всього двох альтернатив "ні" і "так". Тобто за суттю це питання є індикатором наявності або відсутності якоїсь ознаки. Так питання "Чи маєте Ви дітей?" є дихотомічною ознакою, а ознака "Стать респондента" є номінальною, адже на першу з них можна дати відповідь "так" чи "ні", а на другу – таку відповідь дати не можна.

Залежно від задач та типів ознак доцільно застосовувати ту чи іншу інтегральну характеристику, адже деякі з них для певних типів можуть не мати сенсу.

Вибіркове середнє, мода та медіана

Вибіркове середнє застосовують лише для метричних та, іноді, для порядкових ознак і не застосовують для номінальних, хоч іноді його можна інтерпретувати для дихотомічних.

Дійсно, поняття середнього віку та середнього рівня задоволеності пояснити можна, а от пояснити середній номер політичних партій, середній номер гравця у футбол, середній номер телефону або середній регіон України зі списку неможливо. З іншого боку, для дихотомічної ознаки, наприклад статі, що набуває значень 1 (чоловіча) та 2 (жіноча), значення вибіркового середнього більше за 1,5 свідчить про те, що у вибірці жінок більше за чоловіків. Отже, вибіркове середнє використовують для метричних та, іноді, порядкових чи дихотомічних ознак.

Моду, як індикатор відповіді, що зустрічається найчастіше, використовують як головну міру центральної тенденції для номінальних ознак, хоча її легко інтерпретувати і для інших типів ознак. Наприклад мода політичних партій свідчить про те, яку політичну партію вибрала найбільша кількість респондентів, мода номера області говорить про, наприклад, найбільшу кількість респондентів з цієї області. Моду також обчислюють для метричних та порядкових ознак. Наприклад, мода віку респондентів говорить лише про те, якого віку кількість респондентів найбільша у вибірці, однак для метричних ознак є частим випадок багатомодовості або ж виникають проблеми з інтерпретацією моди. А от для групованої ознаки наявність двох чітко виражених мод, що містяться на різних краях гістограми, може свідчити про "двогорбість" розподілу.

Медіана є характеристикою середнього респондента. Таким чином, для номінальних ознак її інтерпретувати неможливо, а от для порядкових та метричних вона часто стає у пригоді.

Дослідження медіани є особливо важливим при розгляді несиметричних розподілів, тобто, коли медіана та середнє не збігаються або принаймні сильно відрізняються. На відміну від

середнього, медіана не чутлива до зростання вибірки та, що навіть більш важливо, викидів. Проілюструємо ці твердження на прикладі.

Остаточно зазначимо, що вибіркове середнє мінімізує вираз $\sum_{i=1}^N (a_i - x)^2$, тобто $\bar{x} = \arg \min_x \left(\sum_{i=1}^N (a_i - x)^2 \right)$ або інакше, мінімум $\sum_{i=1}^N (a_i - x)^2$ досягається на \bar{x} . Медіана мінімізує вираз $\sum_{i=1}^N |a_i - x|$, тобто $Me = \arg \min_x \left(\sum_{i=1}^N |a_i - x| \right)$ або інакше, мінімум $\sum_{i=1}^N |a_i - x|$ досягається на Me .

Приклад 1.25

Нехай розподіл зарплати баскетболістів команди NBA у млн дол. США має вигляд 0.5, 0.6, 0.9, 1, 1, 1.3, 1.9, 2.33, 2.4, 2.4, 2.5, 3, 11.4, 21, 25 (дані чемпіонату США 2010–2011 рр.). Обчислити моду, медіану і середнє та інтерпретувати їх.

Розв'язання

Середньою зарплатою буде вибіркове середнє (середнє арифметичне) зарплат баскетболістів, що становить 5,15 млн дол. США на рік, а моди буде дві – 1 та 2,4 млн дол. США. Медіана, як значення, що міститься посередині варіаційного ряду, дорівнюватиме 2,33 (якщо обчислити медіану за правилами групованої вибірки для неперервних ознак, то вона взагалі дорівнюватиме 2,03 млн дол. США). Для того, щоб пояснити, звідки виникла така розбіжність між середнім та медіаною, побудуємо гістограму за групованою вибіркою (групування за формулою Стерджеса вказує на 3 інтервали, але ми зобразимо гістограму для 3 та для 6 інтервалів).

Із графіка (рис. 1.19) видно, що осіб із меншою зарплатою значно менше, ніж гравців із високою, відповідно відбувається "перекіс" графіка. Про такий розподіл кажуть, що він має правий "скіс", тобто коефіцієнт асиметрії буде додатним¹.

Напрямок асиметрії також легко визначити, порівнюючи середнє та медіану. Якщо медіана менша за середнє, то маємо

¹ Коефіцієнт асиметрії детально розглядатиметься нижче, у розд. 2.

правий "скіс", якщо більша – лівий. У випадку симетричного стосовно середнього розподілу медіана збігається (майже збігається для вибірок) із середнім.

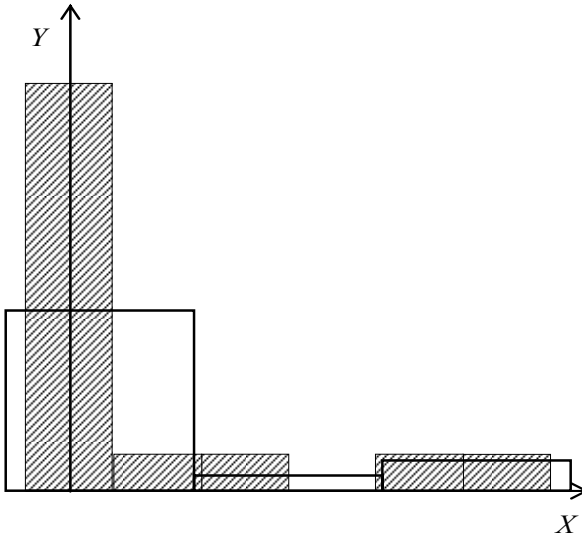


Рис. 1.19. Гістограми за групованою вибіркою (3 та 6 інтервалів):
Y – кількість осіб, **X** – розмір зарплати

Тепер сформулюємо, чим відрізняється середнє від медіани. Середнє свідчить про середні витрати на одного гравця в команді, а медіана – про зарплату середньостатистичного гравця в команді. Тому, коли йдеться про зарплату, то є сенс для аналізу використовувати медіану.

Для пояснення цього твердження розглянемо ситуацію, коли найбільш оплачуваний із гравців отримав контракт на 58 млн дол. США на рік (максимально допустиму у сезоні 2010/11 рр. зарплату у NBA). У цьому випадку медіана не змінить свого значення, а середнє збільшиться до 7,35 млн дол. США на рік. Це означає, що медіана є стійкою до викидів, а середнє – ні. Тобто, наприклад для того, щоб сказати, що у країні трохи зросла середня зарплата, зовсім не обов'язково трохи піднімати її для всіх громадян, досить її дуже сильно збільшити для кількох.

Що ж до мод, то в нашому прикладі вони свідчать лише про те, що серед гравців цієї команди зустрічаються найчастіше ті, що мають зарплату 1 та 2,4 млн дол. США на рік.

Дисперсія

Дисперсія, як міра розсіювання, говорить про середньоквадратичне відхилення елементів вибірки від середнього. Її застосовують для метричних та порядкових ознак. Для номінальних ознак дисперсію обґрунтувати важко або ж узагалі неможливо.

Приклад 1.26

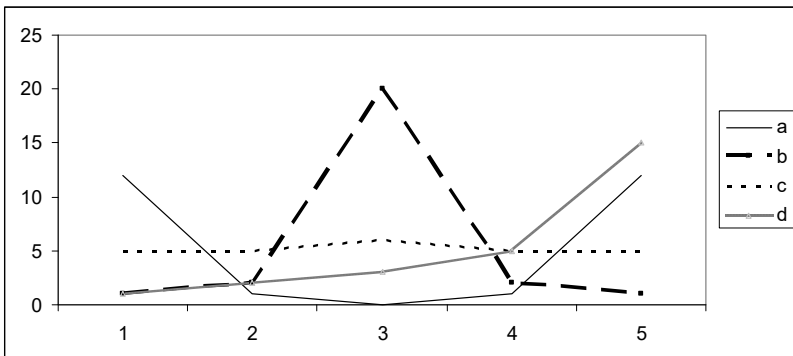
Нехай маємо 4 вибірки з елементами 1, 2, 3, 4 та 5, записані в ряд розподілу:

Вибірка	1	2	3	4	5
a	0,46	0,04	0,00	0,04	0,46
b	0,04	0,08	0,77	0,08	0,04
c	0,19	0,19	0,23	0,19	0,19
d	0,04	0,08	0,12	0,19	0,58

Порівняти дисперсії цих вибірок.

Розв'язання

Спочатку побудуємо лінійчасті діаграми для наведених вибірок:



Дисперсії цих вибірок дорівнюють відповідно $D_a=3,8$, $D_b=0,46$, $D_c=1,92$ і $D_d=1,31$. Дійсно, найбільш щільно навколо

середнього елементи вибірки розташовані у \mathbf{b} , тому й дисперсія $D\mathbf{b}$ – найменша. У вибірці \mathbf{a} ситуація протилежна – елементи вибірки головним чином набувають крайніх значень, тобто, найвіддаленіші від середнього, тому й дисперсія $D\mathbf{a}$ – найбільша. Вибірка \mathbf{c} розпорошена майже рівномірно, тому дисперсія є досить великою. Випадок несиметричної вибірки \mathbf{d} дає порівняно невелику дисперсію через те, що елементи вибірки сконцентровані досить щільно навколо середнього, що дорівнює 4,19.

1.5.2. Обчислення інтегральних характеристик вибірки

До інтегральних характеристик вибірки за шляхом побудови належать інтегральні характеристики, побудовані на базі середнього та не на базі середнього.

Інтегральною характеристикою вибірки, побудованою за функцією $g(x)$, називають середнє перерахованих за допомогою функції $g(x)$ значень елементів вибірки:

$$\overline{g(x)} = \frac{g(a_1) + \dots + g(a_N)}{N}.$$

Прикладами інтегральних характеристик на базі середнього є вибіркове середнє та вибіркова дисперсія:

$$\bar{X} = \frac{g(a_1) + \dots + g(a_N)}{N}, \text{ де } g(x) = X, \quad (1.10)$$

$$\hat{\sigma}^2 = \frac{g(a_1) + \dots + g(a_N)}{N}, \text{ де } g(x) = (X - MX)^2,$$

$$S^2 = \frac{N}{N-1} \frac{g(a_1) + \dots + g(a_N)}{N}, \text{ де } g(x) = (X - \bar{X})^2. \quad (1.11)$$

Загальні формули обчислення інтегральних характеристик на базі середнього

Хоч інтегральні характеристики визначають за вибіркою, їх можна обчислити за рядом розподілу або групуваною вибіркою.

Загальними правилами обчислення інтегральних характеристик на базі середнього є такі: через ряд розподілу, через групувану вибірку.

За рядом розподілу частот

U	u_1	u_2	...	u_n	(...)
f_u	f_1	f_2	...	f_n	(...)

запишемо формулу підрахунку інтегральної характеристики на базі середнього: $\overline{g(x)} = g(u_1)f_1 + g(u_2)f_2 + \dots + g(u_n)f_n (+ \dots)$.

Для результатів групування, представлених групуваною вибіркою

$[x_{i-1}, x_i)$	f_i	x_i^*
$[x_0, x_1)$	f_1	x_1^*
$[x_1, x_2)$	f_2	x_2^*
...
$[x_{M-1}, x_M]$	f_M	x_M^*

формула обчислення має вигляд

$$\overline{g(x)} = g(x_1^*)f_1 + g(x_2^*)f_2 + \dots + g(x_M^*)f_M.$$

Підсумовуючи, маємо

$$\overline{g(x)} = \begin{cases} \text{за рядом розподілу:} \\ g(u_1)f_1 + g(u_2)f_2 + \dots + g(u_N)f_N ; \\ \text{за групуваною вибіркою:} \\ g(x_1^*)f_1 + g(x_2^*)f_2 + \dots + g(x_M^*)f_M . \end{cases}$$

Приклад 1.27

За рядом розподілу частот

U	-1	1	2	3
f_u	0,2	0,3	0,1	0,4

обчислити середнє та дисперсію.

Розв'язання

Для обчислення математичного сподівання (середнього) вважатимемо, що $g(x) = x$ і з (1.10) отримаємо

$$\begin{aligned} \bar{X} = \overline{g(x)} &= g(u_1)f_1 + g(u_2)f_2 + g(u_3)f_3 + g(u_4)f_4 = \\ &= -1 \cdot 0,2 + 1 \cdot 0,3 + 2 \cdot 0,1 + 3 \cdot 0,4 = 1,5. \end{aligned}$$

Для обчислення дисперсії візьмемо за $g(x) = (x - \bar{X})^2$ і з (1.11) одержимо

$$\begin{aligned}\sigma^2 &= \overline{g(x)} = g(u_1)f_1 + g(u_2)f_2 + g(u_3)f_3 + g(u_4)f_4 = \\ &= (-1 - 1,5)^2 \cdot 0,2 + (1 - 1,5)^2 \cdot 0,3 + (2 - 1,5)^2 \cdot 0,1 + (3 - 1,5)^2 \cdot 0,4 = 2,25.\end{aligned}$$

Отже, математичне сподівання (середнє) $\bar{X} = 1,5$, а дисперсія $\sigma^2 = 2,25$.

**Загальна формула обчислення
вибіркового середнього та дисперсії
за рядом розподілу чи групуваною вибіркою**

Запишемо загальну формулу підрахунку вибіркового середнього за рядом розподілу чи групуваною вибіркою:

$$\bar{X} = \begin{cases} u_1 f_1 + \dots + u_n f_n & \text{за рядом розподілу,} \\ x_1^* f_1 + \dots + x_M^* f_M & \text{за групуваною вибіркою.} \end{cases}$$

Для дисперсії формула підрахунку набуває вигляду

$$\hat{\sigma}^2 = \begin{cases} (u_1 - \bar{x})^2 f_1 + \dots + (u_n - \bar{x})^2 f_n, \\ (x_1^* - \bar{x})^2 f_1 + \dots + (x_M^* - \bar{x})^2 f_M. \end{cases}$$

Приклад 1.28 (вибіркове середнє та вибіркова дисперсія за рядом розподілу і групуваною вибіркою)

У процесі опитування фіксувалася зарплата респондентів: 130, 180, 350, 190, 220, 250, 180, 140, 170, 270 умовних грошових одиниць. Обчислити середню зарплату опитаних. Підрахувати середній податок, який сплачують респонденти, якщо відомо, що податок складає 15 % зарплати.

Розв'язання

Вибіркове середнє отриманої вибірки

$$\begin{aligned}\bar{X} &= \frac{1}{10} (130 + 180 + 350 + 190 + 220 + 250 + 180 + 140 + 170 + 270) = \\ &= \frac{1}{10} \cdot 2080 = 208 \text{ умовних грошових одиниць.}\end{aligned}$$

Отримане значення можна інтерпретувати таким чином: якби всі 10 респондентів отримували однакову зарплату, то вона мала би складати саме 208 умовних грошових одиниць. Позначимо як

y_i податок, який сплачує i -й респондент. Оскільки $y_i = g(x_i) = x_i \cdot 0,15$, то середній податок

$$\begin{aligned}\bar{Y} &= \overline{g(x)} = \frac{1}{n} \sum_{i=1}^n g(x_i) = \frac{1}{10} \sum_{i=1}^{10} x_i \cdot 0,15 = \frac{3}{200} \sum_{i=1}^{10} x_i = \frac{3}{200} \cdot 2080 = \\ &= 31,2 \text{ умовних грошових одиниць.}\end{aligned}$$

Отже, середня зарплата дорівнює 208 умовних грошових одиниць, середній податок – 31,2 умовних грошових одиниць.

ЗАДАЧІ

Інтегральні характеристики вибірки на базі середнього

№ 1

Показники лічильника електроенергії за 5 місяців склали 100, 210, 180, 120, 150 кВт-год. Визначити середньомісячну енерговитрату та відповідну вибіркову дисперсію цієї вибірки. Якою в середньому була плата за електроенергію за 1 місяць, якщо 1 кВт-год коштує 15,6 копійки?

№ 2

Микола кожного робочого дня під час зарядки пробігає 1 км, а по вихідних (субота та неділя) – по 2 км. Петро щодня пробігає 1 км 300 м. Порівняти середній денний пробіг Петра та Миколи. Чи зміниться відповідь, якщо за основу розрахунків брати не один, а три тижні? Чому?

№ 3

У класі 15 дівчат і 12 хлопців. Середній зріст дівчат складає 165 см, а середній зріст хлопців складає 180 см. Який середній зріст усіх учнів класу?

№ 4

Чи можна, знаючи середню успішність по класу школи, обчислити середню успішність по всій школі? А якщо відомо, що в кожному класі однакова кількість дітей? Записати формули для пояснення відповідей на запитання задачі.

№ 5

У букіністичному магазині 20 книг за ціною 6 грн, 15 – за 11 грн, 7 – за 4 грн та 18 – за 5 грн. Яка середня ціна однієї кни-

ги? Скільки в магазині книг, ціна яких найближча до середньої?
Найдальша від середньої?

№ 6

Нехай задано ряд розподілу частот для кількості X_i неправильних з'єднань за хвилину на телефонній станції:

X_i	0	1	2	3	4
m_i	75	25	12	5	2

Знайти вибіркочну дисперсію та вибіркоче середнє цієї величини.

№ 7

Нехай задано груповану вибірку для часу обробки тесту (у хв) студентами першого курсу:

X_i	[0; 4)	[4; 8)	[8; 12)	[12; 16)	[16; 20]
m_i	2	7	28	10	3

Знайти вибіркочну дисперсію та вибіркоче середнє цієї величини.

№ 8

Під час дослідження часу обслуговування покупця біля касового апарата, зафіксовано тривалості обслуговування (у хв). Дані наведено у таблиці:

X_i	[0; 0,2)	[0,2; 0,4)	[0,4; 0,6)	[0,6; 0,8)	[0,8; 1,0)	[1,0; 1,2]
m_i	3	6	46	120	20	5

Знайти вибіркочну дисперсію та вибіркоче середнє для тривалості обслуговування.

П'ятиточкова характеристика вибірки та її зображення "ящик з вусами" ("box and whiskey plot")

П'ятиточковою характеристикою вибірки є складена інтегральна характеристика з п'яти чисел: максимальне та мінімальне значення вибірки і три квартилі, що записані в порядку зростання: x_{\min} , Q_1 , Q_2 , Q_3 , x_{\max} .

Приклад 1.29

Для вибірки з прикладу 1.23

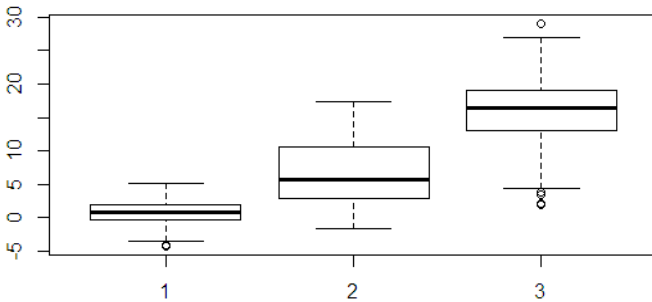
$$x_{\min} = 170, x_{\max} = 192, Q_1 = 173, Q_2 = m = 181, Q_3 = 184.$$

Отже, п'ятиточковою характеристикою цієї вибірки є п'ять значень 170, 173, 181, 184, 192.

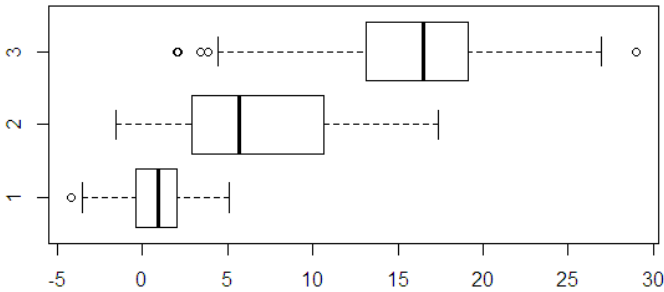
П'ятиточкову характеристику представляють рисунком, який називають "ящик з вусами" від англomовного терміну "box-and-whiskey plot" або "boxplot".

Цей рисунок є прямокутником із перетинкою і сторонами, що паралельні осям координат, та відрізками, паралельними горизонтальній осі, які відходять від середин бiчних сторін прямокутника. Прямокутник із відрізками розташовується так, щоб бiчні сторони проектувались на точки горизонтальної осі, що є квартілями, а кінці горизонтальних відрізків проектуються на точки, що є мінімальним та максимальним елементами вибірки.

Зазначимо, що в зарубіжних дослідженнях цей рисунок розташовують як вертикально (рис. 1.20, а), так і горизонтально (рис. 1.20, б).



а



б

Рис. 1.20. Вертикальний ящик із вусами (а), горизонтальний ящик із вусами (б)

Для п'ятиточкової характеристики прикладу "ящик із вусами" має вигляд як на рис. 1.21.

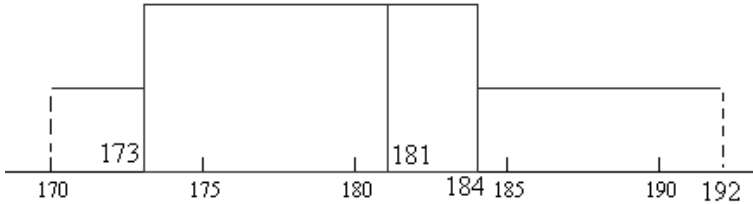


Рис. 1.21. "Ящик із вусами"

Приклад 1.30

Елементами вибірки є вага десяти опитаних (у кг): 74, 72, 73, 74, 75, 73, 73, 73, 74, 75. Знайти для цієї вибірки моду, медіану, перший та третій квартилі. Результати обчислень зобразити у вигляді "ящика з вусами".

Розв'язання

Варіаційний ряд $\{a_i^*\}$ для вихідної вибірки a_i таких: 72, 73, 73, 73, 74, 74, 74, 75, 75. Звідси $a_{\min} = a_1^* = 72$ та $a_{\max} = a_{10}^* = 75$. Об'єм вибірки $N = 10$ є парним числом, тому перший (нижній) квартиль має вигляд

$$Q_1 = a_{\left[\frac{N}{4}\right]+1}^* = a_{\left[\frac{10}{4}\right]+1}^* = a_3^* = 73;$$

а верхній квартиль

$$Q_3 = a_{\left[\frac{3N}{4}\right]+1}^* = a_{\left[\frac{30}{4}\right]+1}^* = a_8^* = 74.$$

Оскільки $N = 10$ є парним, то медіана

$$m = Q_2 = \frac{a_5^* + a_6^*}{2} = \frac{73 + 74}{2} = 73,5.$$

Модюю буде $M_0 = 73$ як значення, що зустрічається у вибірці найчастіше. Шуканий "ящик із вусами" матиме такий вигляд (рис. 1.22):

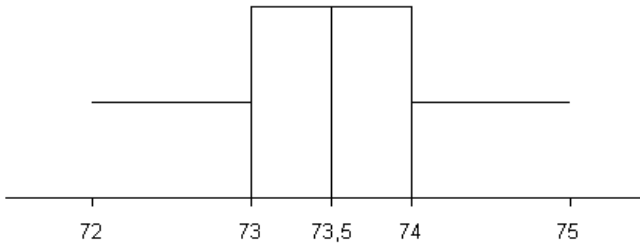


Рис. 1.22. "Ящик із вусами" до прикладу 1.30

Приклад 1.31

Нехай маємо вибірку:

2, 3, 2, 1, 3, 4, 5, 4, 10, 6, 7, 3, 4, 0, 5, 4, 9, 4, 3, 5, 6, 3, 8, 2, 4, 6, 3, 4, 6, 3, 0, 10, 5, 1, 5, 4, 7, 8, 3, 1

За цією вибіркою побудувати її п'ятиточкову характеристику.

Розв'язання

Перш, ніж побудувати п'ятиточкову характеристику, потрібно знайти найменше та найбільше значення вибірки, а також усі 3 кватилі. Для цього сформуємо груповану вибірку, для побудови якої спочатку запишемо варіаційний ряд.

Отримаємо 0, 0, 1, 1, 1, 2, 2, 2, 3, 3, 3, 3, 3, 3, 3, 3, 3, 3, 4, 4, 4, 4, 4, 4, 4, 4, 5, 5, 5, 5, 5, 6, 6, 6, 6, 7, 7, 8, 8, 9, 10, 10.

Об'єм вибірки $N = 40$, $a_{\min} = 0$, $a_{\max} = 10$.

За формулою Стерджеса отримуємо кількість інтервалів групування

$$M = [1 + 3,322 \lg 40] = [6,322] = 6.$$

Знаходимо довжину інтервалів за формулою

$$h = (a_{\max} - a_{\min}) / (M - 1) = (10 - 0) / (6 - 1) = 10 / 5 = 2.$$

Отже, для вибірки об'ємом $N = 40$ найменше значення $a_{\min} = 0$, найбільше значення $a_{\max} = 10$, кількість інтервалів групування $M = 6$, довжина інтервалів групування $h = 2$.

Початок першого інтервалу групування, тобто

$$x_0 = a_{\min} - h/2 = 0 - 2/2 = -1.$$

Для визначення інших границь інтервалів групування використовуємо формулу $x_i = x_{i-1} + h$:

$[x_{i-1}, x_i)$	m_i	m_i^*
$[-1; 1)$		
$[1; 3)$		
$[3; 5)$		
$[5; 7)$		
$[7; 9)$		
$[9; 11)$		

Для кожного інтервалу потрібно визначити кількість елементів вибірки, що потрапляють до нього. Запишемо це так:

$[x_{i-1}, x_i)$	Елементи вибірки	m_i
$[-1; 1)$	0, 0	2
$[1; 3)$	1, 1, 1, 2, 2, 2	6
$[3; 5)$	3, 3, 3, 3, 3, 3, 3, 3, 4, 4, 4, 4, 4, 4, 4	16
$[5; 7)$	5, 5, 5, 5, 5, 6, 6, 6, 6	9
$[7; 9)$	7, 7, 8, 8	4
$[9; 11)$	9, 10, 10	3

Отже, можемо скласти таблицю:

$[x_{i-1}, x_i)$	m_i	m_i^*
$[-1; 1)$	2	0
$[1; 3)$	6	2
$[3; 5)$	16	8
$[5; 7)$	9	24
$[7; 9)$	4	33
$[9; 11)$	3	37

де m_i^* – кумулятивна абсолютна частота, накопичена до i -го інтервалу. Тобто, для першого інтервалу ця частота дорівнює 0 через відсутність будь-яких елементів вибірки, що мають значення, менші за значення початку цього інтервалу. Для другого інтервалу m_i^* дорівнює 2, адже до початку другого інтервалу у вибірці були лише ті елементи, які увійшли до першого інтерва-

лу, а їх $\epsilon 2$. Для третього інтервалу m_i^* дорівнює 8, адже до третього інтервалу у вибірці є лише ті елементи, які потрапили до першого або другого інтервалу, а їх $2+6 = 8$. Відповідно m_i^* для четвертого інтервалу є сумою абсолютних частот, накопичених у 1–3 інтервалах, тобто $2+6+16 = 24$ і т. д.

Наступним кроком є обчислення трьох кватилів для цієї вибірки. Спочатку знайдемо кватильні інтервали, тобто ті інтервали, до яких потрапляють Q_1 , Q_2 та Q_3 відповідно.

За визначенням, перший кватиль відтинає від варіаційного ряду першу чверть елементів, тобто чверть менших за значенням елементів вибірки. Ураховуючи, що $N = 40$, маємо $N/4 = 10$, тобто Q_1 відтинає 10 перших елементів варіаційного ряду. Отже, знаходимо інтервал, до якого потрапляє 10-й елемент варіаційного ряду. Для цього розглянемо стовпчик з m_i^* . Потрібно знайти таке i , щоб виконувалося $m_i^* < 10$, а $m_{i+1}^* \geq 10$. Таким інтервалом є третій інтервал [3; 5). Дійсно, до третього інтервалу було накопичено 8 елементів варіаційного ряду, а в сам 3-й інтервал потрапило 16 елементів, тобто 10-й елемент варіаційного ряду потрапив у 3-й інтервал, отже, 3-й інтервал є кватильним для першого кватіля Q_1 .

Знайдемо тепер саме значення Q_1 . Для цього скористаємось формулою

$$Q_1 = x_{Q_1} + h \cdot \frac{\frac{N}{4} - m_{Q_1}^*}{m_{Q_1}}$$

Початок кватильного інтервалу для першого кватіля $x_{Q_1} = 3$, тому що кватильним інтервалом є інтервал [3; 5), накопичена частота до 3-го інтервалу $m_{Q_1}^* = 8$, абсолютна частота самого інтервалу $m_{Q_1} = 16$, а довжина інтервалу $h = 2$. Отже, отримаємо

$$Q_1 = 3 + 2 \cdot \frac{\frac{40}{4} - 8}{16} = 3 + 2 \cdot \frac{10 - 8}{16} = 3 + \frac{4}{16} = 3 + \frac{1}{4} = 3,25.$$

Зважаючи на те, що кватильним інтервалом для першого кватіля є 3-й інтервал [3; 5), то сам кватиль має потрапляти до нього. Дійсно, $3 \leq Q_1 < 5$, отже, $Q_1 = 3,25$.

Далі знайдемо медіанний інтервал, тобто інтервал для другого кватриля. За визначенням, другий (медіанний) кватиль відтинає від варіаційного ряду другу чверть елементів, тобто половину менших за значенням елементів вибірки. Оскільки $N=40$, то $N/2=20$, тобто Q_2 відтинає 20 перших елементів варіаційного ряду. Отже, потрібно знайти інтервал, до якого потрапляє 20-й елемент варіаційного ряду. Для цього розглянемо стовпчик з m_i^* . Потрібно знайти таке i , щоб справджувалося $m_i^* < 20$, а $m_{i+1}^* \geq 20$. Таким інтервалом є третій інтервал $[3; 5)$. Дійсно, до третього інтервалу було накопичено 8 елементів варіаційного ряду, а в сам 3-й інтервал потрапило 16 елементів, тобто 20-й елемент варіаційного ряду потрапив у 3-й інтервал, отже, 3-й інтервал є кватильним і для другого кватриля Q_2 . Зазначимо, що не обов'язково кожен кватиль має потрапляти у окремий інтервал – є вибірки, в яких навіть усі три кватилі можуть потрапляти до одного інтервалу.

Обчислимо значення медіани:

$$Me = Q_2 = x_{Me} + h \cdot \frac{\frac{N}{2} - m_{Me}^*}{m_{Me}} = 3 + 2 \cdot \frac{20 - 8}{16} = 3 + \frac{24}{16} = 3 + \frac{3}{2} = 4\frac{1}{2},$$

де x_{Me} – початок медіанного інтервалу; h – його довжина; m_{Me}^* – абсолютна частота, накопичена до початку медіанного інтервалу; m_{Me} – абсолютна частота медіанного інтервалу.

Для третього кватриля кватильним є інтервал $Q_3 \in [5; 7)$,

$$Q_3 = x_{Q_3} + h \cdot \frac{\frac{3N}{4} - m_{Q_3}^*}{m_{Q_3}} = 5 + 2 \cdot \frac{30 - 24}{9} = 5 + \frac{12}{9} = 5 + \frac{4}{3} = 6\frac{1}{3} \approx 6,33.$$

Отже, "ящик із вусами" має вигляд (рис. 1.23.)

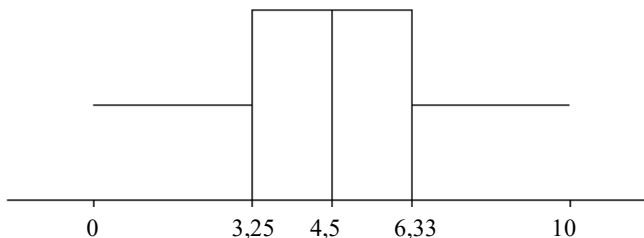


Рис. 1.23. "Ящик із вусами" до прикладу 1.31

ЗАДАЧІ

Інтегральні характеристики вибірки не на базі середнього

У наступних задачах знайти моду і всі необхідні величини для побудови "ящика з вусами" та побудувати його. Дати інтерпретацію модовому (мода), і медіанному (медіана) значенням, нижньому та верхньому квантилям та інтерквартильному розмаху $R = Q_3 - Q_1$.

№ 1

Елементами вибірки є зріст 20 опитаних людей (у см):
172, 175, 173, 172, 177, 182, 179, 175, 176, 173, 169, 174, 173, 178, 180, 179, 176, 172, 173, 176.

№ 2

Елементами вибірки є вік респондентів:
20, 25, 37, 28, 34, 22, 23, 20, 17, 18, 19, 25, 71, 35, 42, 39, 28, 31, 24, 55.

№ 3

Елементами вибірки є вага респондентів (у кг) :
75, 59, 80, 79, 63, 55, 85, 83, 79, 78, 53, 61, 90, 59, 69, 23, 47, 31, 24, 55.

№ 4

Елементами вибірки є кількість дітей у сім'ї респондента :
1, 0, 2, 2, 1, 3, 0, 1, 2, 0, 1, 2, 1, 0, 2, 4, 2, 3, 2, 1.

№ 5

Елементами вибірки є кількість років, які респондент витратив на освіту :
12, 16, 18, 20, 17, 18, 16, 14, 13, 12, 19, 13, 17, 19, 18, 14, 16, 15, 15, 12.

№ 6

Знайти моду та всі необхідні величини для побудови "ящика з вусами" і побудувати його для даних:

X_i	0	1	2	3	4
m_i	75	25	12	5	2

№ 7

Знайти моду та всі необхідні величини для побудови "ящика з вусами" і побудувати його для даних:

X_i	[0; 4)	[4; 8)	[8; 12)	[12; 16)	[16; 20]
m_i	2	7	28	10	3

№ 8

Знайти моду та всі необхідні величини для побудови "ящика з вусами" і побудувати його для даних:

X_i	[0; 0,2)	[0,2; 0,4)	[0,4; 0,6)	[0,6; 0,8)	[0,8; 1,0)	[1,0; 1,2]
m_i	3	6	46	120	20	5

2. ТЕОРІЯ ЙМОВІРНОСТЕЙ

2.1. Коротка історична довідка

Теорія ймовірностей належить до прикладних математичних дисциплін і використовується для розв'язування прикладних задач із застосуванням математичних методів. Вона виникла з чисто практичних міркувань у задачах для випадків, коли зібрати повну інформацію про процес неможливо, а результат дії процесу потрібно хоча б приблизно визначити. Отже, теорія ймовірностей вивчає чисельні закономірності процесів (дослідів), результати яких неможливо передбачити однозначно і точно.

Одним із найпростіших прикладів застосування цієї теорії є передбачення того, якою стороною впаде монета, якщо її підкинути вгору. Знаючи абсолютно точну інформацію про сам процес підкидання, монету та всі фактори, що можуть вплинути на результат підкидання, ми зможемо передбачити його точно. Але зібрати всю цю інформацію неможливо хоча б із тієї причини, що навіть надточні вимірювальні прилади мають похибку вимірювання. Тому для спрощення аналізу та прогнозу результатів таких складних процесів і використовується теорія ймовірностей.

Появу теорії ймовірностей відносять до середини XVII століття і пов'язують з іменами Гюйгенса (1629–1695), Паскаля (1623–1662), Ферма (1601–1665) та Бернуллі (1654–1705). У листуванні Паскаля і Ферма, яке виникло через задачі про азартні ігри, що не вкладались у рамки тогочасної математики, поступово з'являлись такі поняття, як "імовірність" та "математичне сподівання".

Визначна роль у становленні аналітичних методів теорії ймовірностей належить Муавру (1667–1754), Лапласу (1749–1827), Гауссу (1777–1855), Пуассону (1781–1840), Лобачевському (1792–1856).

Із середини XIX і приблизно до 20-х рр. XX століття розвиток теорії ймовірностей був пов'язаний з іменами Буняковського (1804–1889), Чебишова (1821–1894), Маркова (1856–1922), Ляпунова (1857–1918).

Значний внесок у теорію ймовірностей та математичну статистику зробили Бернштейн (1880–1968), Колмогоров (1903–1987), Хінчин (1894–1959), Скороход (1930) та ін.

2.2. Визначення ймовірності. Характеристичні властивості ймовірності

2.2.1. Закон стійкості частот, частотне визначення ймовірності

Закон стійкості частот є емпіричним принципом, який застосовують до масових явищ, дослідження яких можливе методами теорії ймовірностей. Він стосується граничної поведінки частот при нескінченному збільшенні кількості спостережень (об'єму вибірки).

Згідно з цим принципом маємо.

1. Відносна частота $\mathcal{C}_N(A)$ будь-якої події A за нескінченного збільшення об'єму вибірки збігається (прямує) до граничного значення, яке позначають $P(A)$:

$$\lim_{N \rightarrow \infty} \mathcal{C}_N(A) = P(A). \quad (2.1)$$

2. Граничне значення $P(A)$ не залежить від вибірки, за якою отримане, а характеризує масове явище в цілому.

Зміст закону стійкості частот ілюструє схема:

$$\begin{array}{l} \mathbf{a} - a_1, \dots, a_N \rightarrow \mathcal{C}_N(A) \\ \mathbf{б} - b_1, \dots, b_N \rightarrow \mathcal{C}'_N(A) \end{array} \begin{array}{l} \xrightarrow{N \rightarrow \infty} \\ \xrightarrow{N \rightarrow \infty} \end{array} P(A)$$

Закон стійкості частот є основою визначення ймовірності. Відповідно до нього **ймовірністю події A** (позначають $P(A)$) називають граничне значення відносної частоти за нескінченного збільшення об'єму вибірки.

Ймовірність, як і частота, є функцією подій, заданою на просторі елементарних подій U .

2.2.2. Передбачувальна властивість імовірності

Визначення ймовірності через закон стійкості частот дозволяє стверджувати, що для будь-якої послідовності спостережень (вибірки) досліджуваного масового явища відносна частота приблизно дорівнює ймовірності

$$\mathcal{C}_N(A) \approx P(A),$$

а абсолютна, відповідно,

$$m_A = \mathcal{C}_N(A) \cdot N \approx P(A) \cdot N.$$

Ці два співвідношення характеризують передбачувальну властивість імовірності.

Передбачувальна властивість полягає в тому, що значення ймовірності, отримане за однією вибіркою, дозволяє характеризувати (нехай навіть наближено) частоти іншої послідовності спостережень (вибірки) того самого масового явища.

Приклад 2.1

Нехай у пологовому будинку фіксується стать новонароджених вибіркою об'єму $N = 1000$, за якою побудовано ймовірності $P(X) = 0,51$, $P(D) = 0,49$.

Використовуючи передбачувальну властивість імовірності, можемо стверджувати, що для вибірок об'єму 1000, побудованих за спостереженнями статі новонароджених у цьому пологовому будинку, виконуються співвідношення:

$$\mathcal{C}_N(X) \approx P(X) = 0,51, m_N(X) \approx P(X) \times N = 0,51 \times 1000 = 510,$$

$$\mathcal{C}_N(D) \approx P(D) = 0,49, m_N(D) \approx P(D) \times N = 0,49 \times 1000 = 490.$$

2.2.3. Умовні ймовірності

Умовні ймовірності є граничними значеннями умовних частот. **Умовною ймовірністю** події A за умови виконання події B (позначають $P(A/B)$ або $P(A|B)$) називають імовірність, отриману при введенні до початкових умов спостереження додаткової умови B . Умовна ймовірність є граничним значенням умовної відносної частоти:

$$\lim_{N \rightarrow \infty} \mathcal{C}_N(A/B) = P(A/B). \quad (2.2)$$

Умовна й безумовна ймовірності пов'язані співвідношенням, що є граничним значенням частотного відповідника формули (1.5):

$$P(A/B) = \frac{P(AB)}{P(B)}, \text{ якщо } P(B) \neq 0. \quad (2.3)$$

2.2.4. Незалежність події A від події B , еквівалентне визначення незалежності. Незалежні в сукупності події, незалежні випадкові величини

Важливим у використанні ймовірностей є поняття незалежності подій. За своїм змістом незалежність означає "нечутливість" (стійкість) імовірності однієї події до введення додаткової умови в початкові умови спостереження.

Існує два еквівалентні визначення незалежності: вихідне – відбиває сутність цього поняття, еквівалентне – яке є ширшим та кориснішим для застосування.

Вихідне визначення незалежності. Подію A називають незалежною від події B , $P(B) \neq 0$, якщо

$$P(A/B) = P(A). \quad (2.4)$$

Еквівалентне визначення незалежності. Дві події A і B називають незалежними, якщо ймовірність добутку цих подій дорівнює добутку їх імовірностей:

$$P(AB) = P(A)P(B). \quad (2.5)$$

Дійсно, із визначення умовної ймовірності випливає, що $P(AB) = P(A/B)P(B)$. Отже, якщо $P(A/B) = P(A)$, то $P(AB) = P(A)P(B)$.

Навпаки, якщо $P(AB) = P(A)P(B)$, то $P(A) = P(AB)/P(B)$, у якому права частина збігається з умовною ймовірністю (2.3), а отже, $P(A) = P(A/B)$.

Еквівалентне визначення незалежності є ширшим за вихідне, оскільки

а) з нього випливає, що незалежність є властивістю взаємною (симетричною), тобто, якщо подія A не залежить від B , то і B не залежить від A ;

б) не вимагається, щоб імовірності подій були ненульовими.

Парадокс незалежності. Математичне визначення ймовірності, як правило, узгоджене з визначенням незалежності. Наприклад, якщо кидають 2 гральні кубики, то подія "шістка на першому кубіку" та "шістка на другому кубіку" є незалежними як у математичному сенсі, так і у звичайному. Однак ця узгодженість спостерігається не завжди. С. Н. Бернштейн звернув увагу на такий **парадокс**.

Припустимо, що кидають дві правильні (тобто однорідні) монети і нехай визначено події $A = \{\text{на першій монеті випав герб}\}$, $B = \{\text{на другій монеті випав герб}\}$ та $C = \{\text{тільки на одній монеті випав герб}\}$. Очевидно, що події A , B та C є попарно незалежними, але будь-які 2 з них однозначно визначають третю.

Наведемо пояснення парадокса.

Очевидно, що A та B незалежні через те, що результати першого підкидання не залежать від результатів другого. З іншого боку, події A та C (а також пара B та C) на перший погляд видаються залежними. Але з того, що $P(AC) = P(A)P(C) = 1/4$ і $P(BC) = P(B)P(C) = 1/4$, вони насправді є незалежними, що впливає з (2.5). Справедливе також твердження про те, що дві з подій однозначно визначають третю, бо кожна подія (A , B чи C) відбувається тільки тоді, коли відбувається виключно одна з двох інших подій. Цей парадоксальний феномен показав, що **попарна незалежність не означає незалежність у сукупності**. Для незалежності в сукупності необхідно накласти більш жорсткі умови, ніж при попарній незалежності.

Незалежність у сукупності. Якщо поняття незалежності відносять більше, ніж до двох подій, то говорять про незалежність цих подій у сукупності. Змістом незалежності у сукупності, як і у випадку двох подій, є можливість обчислення ймовірності добутку як добутку ймовірностей.

Події A_1, A_2, \dots, A_n називають **незалежними в сукупності**, якщо ймовірність добутку довільної підсукупності (підмножини) цих подій дорівнює добутку їхніх ймовірностей, тобто для довільного $k \leq n$ та для довільних $i_1 < i_2 < \dots < i_k \in \{1, 2, \dots, n\}$ справедливе співвідношення

$$P(A_{i_1} \cdot A_{i_2} \cdots A_{i_k}) = P(A_{i_1}) \cdot P(A_{i_2}) \cdots P(A_{i_k}). \quad (2.6)$$

Із цього визначення випливає, що довільна підмножина незалежних у сукупності подій утворює множину незалежних у сукупності подій.

Незалежність у сукупності є більш сильною, ніж попарна незалежність. Це означає, що з незалежності у сукупності випливає попарна незалежність, а з попарної незалежності не випливає незалежність у сукупності. Підтвердженням цієї тези може виступати парадокс незалежності.

Поняття незалежності визначають також для випадкових величин.

Дві або більше випадкових величин називають **незалежними**, якщо ймовірність добутку подій, пов'язаних зі значеннями цих величин, дорівнює добутку ймовірностей відповідних подій. Для двох незалежних випадкових величин X та Y це означає, що для довільних числових подій A та B справедливе співвідношення

$$P\{X \in A, Y \in B\} = P\{X \in A\}P\{Y \in B\},$$

де комою між двома умовами в лівому виразі для зручності позначають добуток відповідних подій.

Для двох або декількох випадкових величин попарна незалежність та у сукупності визначаються одним терміном, на відміну від незалежності подій.

Із визначення незалежності випадкових величин випливає таке: вибір довільної сукупності з незалежних випадкових величин дає незалежні випадкові величини, а набір подій, пов'язаних із будь-якими незалежними випадковими величинами з цього набору, утворює незалежні в сукупності події.

2.2.5. Характеристичні властивості ймовірності

Характеристичними властивостями ймовірності є відповідники характеристичних властивостей частоти, які отримують із неї граничним переходом:

- 1) невід'ємність: $P(A) \geq 0$;
- 2) адитивність (скінченна адитивність):
якщо A та B – несумісні, то $P(A+B) = P(A) + P(B)$,

якщо A_1, \dots, A_n – попарно несумісні, то

$$P(A_1 + \dots + A_n) = P(A_1) + \dots + P(A_n);$$

3) нормованість: $P(\emptyset) = 0$, $P(U) = 1$.

Адитивністю (скінченна адитивність) ймовірності називають її властивість додавати свої значення при додаванні несумісних чи попарно несумісних подій. Для ймовірності адитивністю вважають **зліченну адитивність**, тобто властивості ймовірності додавати свої значення при додаванні зліченної сукупності попарно несумісних подій.

Характеристичні властивості ймовірності є підставою для **аксіоматичного визначення ймовірності**. В аксіоматичному визначенні **ймовірністю** називають невід'ємну, зліченно-адитивну, нормовану функцію множини, визначену на відповідних подіях простору елементарних подій U .

У такому аксіоматичному визначенні при розв'язанні прикладних задач частота набуває характеру інтерпретації, тобто будь-яку ймовірність інтерпретують як граничну частоту у зв'язку з реальним або віртуальним випадковим масовим явищем.

2.3. Задання ймовірностей

Шлях задання ймовірностей (як і частот) залежить від типу даних (дискретний чи неперервний), а засіб задання отримують із частотних відповідників граничним переходом. Так, у випадку дискретних даних (дискретного простору елементарних подій) ймовірність задають рядом розподілу ймовірностей, а у випадку неперервних даних (простором елементарних подій є відрізок числової осі, півінтервал або вся числова вісь) функцією розподілу ймовірностей (граничним значенням КФРЧ) чи щільністю розподілу ймовірностей (граничним значенням гістограми).

Зазначимо, що, як і для частот, ймовірності, пов'язані з числовими просторами елементарних подій, називаються **розподілами ймовірностей**. За типом простору елементарних подій U (дискретний чи неперервний) розподіли є **дискретні** та **неперервні**.

Якщо розподіл імовірностей записаний з явним позначенням числової характеристики, значення якої досліджуються, то такий розподіл називають **розподілом імовірностей випадкової величини**. Якщо цей розподіл імовірностей є дискретним, то випадкову величину називають дискретно розподіленою (дискретною), а якщо неперервний, то неперервно розподіленою (неперервною).

2.3.1. Задання ймовірностей для дискретних даних

Рядом розподілу ймовірностей (рядом розподілу) для випадку дискретних даних (дискретного простору елементарних подій) називають множину всіх можливих результатів та відповідних їм імовірностей, заданих у вигляді таблиці з двох рядків чи стовпчиків, в одному з яких перераховані всі елементарні події, а у другому – відповідні їм імовірності:

U	u_1	u_2	...	u_n	(...)
P_u	p_1	p_2	...	p_n	(...)

$$p_n = P\{u_n\}.$$

Імовірність будь-якої групи результатів (будь якої складеної події) обчислюють за формулою

$$P(A) = \sum_{u_i \in A} p_i.$$

Характеристичними властивостями ряду розподілу (як граничного переходу ряду розподілу частот) є невід'ємність та сумарне одиничне значення рядка ймовірностей (нормованість).

Якщо ряд розподілу розглядати як розподіл імовірностей випадкової величини X , то з нею відповідні формули мають вигляд

$$p_n = P\{X = u_n\},$$

$$P(X \in A) = \sum_{u_i \in A} p_i = \sum_{u_i \in A} P\{X = u_i\}.$$

2.3.2. Задання розподілів ймовірностей для неперервних даних

Розподіл ймовірностей у випадку неперервних даних задають функцією розподілу ймовірностей або щільністю розподілу ймовірностей.

Функція розподілу ймовірностей

Функцією розподілу ймовірностей (позначають $F(x)$) називають ймовірність напівнескінчених ліворуч інтервалів як функції правого кінця цих інтервалів:

$$F(x) = P(-\infty; x).$$

Якщо розподіл ймовірностей розглядати як розподіл ймовірностей випадкової величини X , то функція розподілу може бути записана у вигляді

$$F(x) = P\{X \in (-\infty; x)\} = P\{X < x\}.$$

Функція розподілу ймовірностей є граничним значенням КФРЧ

$$F(x) = \lim_{N \rightarrow \infty} F_N(x)$$

і має всі її властивості: характеристичні та різницеву, які отримують граничним переходом у властивостях для КФРЧ.

Отже, **характеристичними властивостями функції розподілу ймовірностей є:**

- 1) невід'ємність: $F(x) \geq 0$;
- 2) монотонне неспадання: $x_2 > x_1 \Rightarrow F(x_2) \geq F(x_1)$;
- 3) неперервність зліва,
- 4) нормованість:

$$F(+\infty) = 1, \quad \text{де } F(+\infty) = \lim_{x \rightarrow \infty} F(x),$$

$$F(-\infty) = 0, \quad \text{де } F(-\infty) = \lim_{x \rightarrow -\infty} F(x).$$

Запишемо різницеву властивість функції розподілу ймовірностей. Ймовірність появи значень з інтервалу $[a, b)$ дорівнює різниці значень функції розподілу на кінцях інтервалу:

$$P\{X \in [a, b)\} = F(b) - F(a).$$

Якщо функцію розподілу розглядати як функцію розподілу випадкової величини X , то різницеву властивість записують у вигляді

$$P\{X \in [a, b)\} = P\{a \leq X < b\} = F(b) - F(a).$$

Щільність розподілу ймовірностей

Щільність розподілу є граничним значенням гістограми, що однозначно визначається площинною властивістю гістограми. Згідно з цією властивістю **щільністю розподілу ймовірностей** називають невід'ємну функцію $f(x)$, за якою ймовірність будь-якої події обчислюють як площу під графіком гістограми над відповідною множиною, тобто як визначений інтеграл від щільності за цією множиною

$$P(A) = \int_A f(x) dx.$$

Якщо розподіл розглядають як розподіл ймовірностей випадкової величини X , то визначення набуває вигляду

$$P(X \in A) = \int_A f(x) dx.$$

Характеристичною властивістю щільності $f(x)$ є (як і властивістю її відповідника – гістограми) сумарна одинична площа під графіком функції над віссю OX .

Зв'язок щільності $f(x)$, коли вона існує, із функцією розподілу $F(x)$ здійснюють через похідну

$$f(x) = F'(x).$$

Значення функції розподілу через щільність визначають співвідношенням

$$F(x) = \int_{-\infty}^x f(y) dy. \quad (2.7)$$

2.4. Інтегральні характеристики розподілу ймовірностей

2.4.1. Інтегральні характеристики розподілу ймовірностей на базі середнього

Інтегральною характеристикою розподілу ймовірностей є функція, що побудована за розподілом імовірностей, яка одним числом характеризує той чи інший бік розподілу ймовірностей. Інтегральні характеристики розподілів імовірностей є граничними значеннями частотних відповідників.

Як і для розподілів частот, розрізняють два основні типи інтегральних характеристик:

- "типові представники" або "міри центральної тенденції",
- "розсіювання" або "міри варіації".

За засобами побудови (як і для частот) розрізняють:

- інтегральні характеристики на базі середнього,
- інтегральні характеристики не на базі середнього.

Інтегральною характеристикою випадкової величини (розподілу ймовірностей) на базі середнього, побудованою за допомогою функції $g(x)$ (її позначають $Mg(x)$), називають граничне значення середнього елементів вибірки, перерахованих за допомогою функції $g(x)$:

$$Mg(X) = \lim_{N \rightarrow \infty} \overline{g(X)} = \lim_{N \rightarrow \infty} \frac{g(a_1) + \dots + g(a_N)}{N}.$$

Для $g(x) = x$, $Mg(X) = MX$, а при $g(x) = x^2$ маємо $Mg(X) = MX^2$.

Так само визначають інтегральну характеристику на базі середнього для двох чи більше випадкових величин. У цьому випадку числова функція g повинна мати кількість аргументів, яка відповідає кількості спостережуваних характеристик.

Отже, для двох випадкових величин X та Y спостереженнями є пари (a, b) , а інтегральну характеристику $Mg(X, Y)$ визначають граничним значенням середнього спостережень, перерахованих за допомогою функції g , що складають вибірку:

$$Mg(X, Y) = \lim_{N \rightarrow \infty} \overline{g(X, Y)} = \lim_{N \rightarrow \infty} \frac{g(a_1, b_1) + \dots + g(a_N, b_N)}{N}.$$

Так для $g(x, y) = xy$ маємо $Mg(X, Y) = M(X \cdot Y)$.

До інтегральних характеристик не на базі середнього належать кватилі (серед них медіана), інтеркватильний розмах та мода. Ці характеристики для випадкових величин визначають так само, як для розподілів частот.

Кватильями розподілів імовірностей (Q_1, Q_2, Q_3) – називають значення досліджуваної характеристики, які відтинають за ймовірністю відповідно чверть, половину та три чверті менших за розміром значень. Другий кватиль (як і в частотному випадку) називають медіаною.

Q_3 можна обчислити як значення досліджуваної характеристики, яке відтинає одну чверть більших за ймовірністю значень. Для неперервних розподілів значення кватилів одержують зі співвідношень:

$$Q_1 : j = P(-\infty; Q_1) = P\{X \in (-\infty; Q_1)\} = P\{X < Q_1\} = 0,25;$$

$$Q_2 : S = P\{X < Q_2\} = P\{X \geq Q_2\} = 0,5;$$

$$Q_3 : s = P\{X < Q_3\} = 0,75, P\{X \geq Q_3\} = j.$$

Другий кватиль, або **медіана**, належить до "мір центральної тенденції" ("типові представники").

Інтеркватильним розмахом називають величину $R_I = Q_3 - Q_1$, де півінтервал $[Q_1; Q_3]$ містить 50 % середніх за ймовірністю значень досліджуваної характеристики, тобто за винятком j найбільших та j найменших. Інтеркватильний розмах належить до інтегральних характеристик "розсіювання" ("міри варіації").

Модою (позначають через Mo) розподілу ймовірностей у неперервному випадку називають точку максимуму щільності розподілу, а в дискретному випадку – значення елемента, для якого ймовірність ряду розподілу має найбільше значення (тобто елемент вибірки, який зустрічається найчастіше).

Моду як середню величину використовують для об'єктів нечислової природи: тип населеного пункту, перелік кандидатів у виборчому списку, назва бренду та ін.

2.4.2. Математичне сподівання випадкової величини

Математичним сподіванням випадкової величини X називають інтегральну характеристику на базі середнього, побу-

довану за допомогою функції $g(x)=x$, та позначають MX . Іншими словами, математичним сподіванням є граничне значення середнього елементів вибірки за нескінченного збільшення її об'єму:

$$MX = \lim_{N \rightarrow \infty} \frac{x_1 + \dots + x_N}{N}.$$

Оскільки вибіркове середнє позначають \bar{X} , то формулу визначення можна переписати у вигляді

$$MX = \lim_{N \rightarrow \infty} \bar{X} = \lim_{N \rightarrow \infty} \frac{x_1 + \dots + x_N}{N}.$$

За змістом математичне сподівання належить до інтегральних характеристик, що є "типовими представниками".

Основні властивості математичного сподівання, які безпосередньо випливають із його визначення:

1) $MC = C$, математичне сподівання константи є тією самою константою;

2) $M(X_1+X_2)=MX_1+MX_2$, математичне сподівання суми випадкових величин дорівнює сумі математичних сподівань доданків;

3) $M(aX)=aMX$, константу-співмножник виносять з-під знаку математичного сподівання;

4) математичне сподівання добутку *незалежних* випадкових величин дорівнює добуткові їх математичних сподівань: $M(XY)=MX \cdot MY$, якщо $P(XY)=P(X)P(Y)$.

Властивості математичного сподівання 2) та 3) об'єднують назвою "лінійність".

2.4.3. Дисперсія випадкової величини. Нерівність Чебишова

Дисперсією випадкової величини X (позначають DX) називають інтегральну характеристику на базі середнього, отриману за допомогою функції

$$g(X) = (X - MX)^2,$$

тобто

$$M(X - MX)^2.$$

Таким чином, дисперсія є граничним значенням середнього квадратичних відхилень спостережуваних значень від матема-

тичного сподівання (вибіркового середнього) або середньоквадратичним відхиленням від середнього:

$$DX = M(X - MX)^2 = \lim_{N \rightarrow \infty} \overline{(X - MX)^2} =$$

$$= \lim_{N \rightarrow \infty} \frac{(x_1 - MX)^2 + \dots + (x_N - MX)^2}{N} = \lim_{N \rightarrow \infty} \frac{(x_1 - \bar{x})^2 + \dots + (x_N - \bar{x})^2}{N}.$$

За своїм змістом дисперсія належить до характеристик типу "розсіювання".

Властивості дисперсії, які впливають безпосередньо з визначення:

- 1) $DC = 0$, дисперсія константи дорівнює нулю;
- 2) дисперсія суми або різниці *незалежних* випадкових величин дорівнює сумі дисперсій цих випадкових величин

$$D(X_1 \pm X_2) = DX_1 + DX_2, \text{ якщо } P(X_1 X_2) = P(X_1)P(X_2).$$

Дисперсію суми та різниці двох довільних випадкових величин (не обов'язково незалежних) розглядатимемо у пункті 2.8.3.

- 3) $D(aX) = a^2 DX$, константу-співмножник виносять з-під знаку дисперсії у квадраті.

Використовуючи властивості математичного сподівання та дисперсії, можемо записати простішу формулу обчислення дисперсії через математичні сподівання MX та MX^2 .

За визначенням дисперсії

$$DX = M(X - MX)^2 = / \text{розкриємо дужки} / = M(X^2 - 2X \cdot MX + (MX)^2).$$

За властивістю 2) математичного сподівання маємо

$$M(X^2 - 2X \cdot MX + (MX)^2) = M(X^2) - M(2X \cdot MX) + M(MX)^2.$$

- Обчислене значення MX є константою, тому за властивістю 3) математичного сподівання отримаємо

$$M(X^2) - M(2X \cdot MX) + M(MX)^2 = M(X^2) - 2MX \cdot M(X) + (MX)^2 =$$

$$= M(X^2) - 2(MX)^2 + (MX)^2 = MX^2 - (MX)^2.$$

Отже,

$$DX = MX^2 - (MX)^2. \quad (2.8)$$

Нерівність Чебишова. Для довільного $\varepsilon > 0$ виконується

$$P\{|X - MX| > \varepsilon\} \leq \frac{DX}{\varepsilon^2}.$$

Нерівність Чебишова ілюструє роль математичного сподівання як "типового представника" та дисперсії – як "розсіювання" (варіації) значень досліджуваної характеристики щодо математичного сподівання.

2.4.4. Обчислення математичного сподівання та дисперсії

Загальні формули обчислення математичного сподівання як функції від випадкової величини у дискретному та неперервному випадках. Математичне сподівання у дискретному випадку обчислюють за рядом розподілу, а в неперервному – за щільністю. Формули обчислення мають вигляд

$$Mg(X) = \begin{cases} g(u_1)p_1 + g(u_2)p_2 + \dots + g(u_n)p_n (+\dots) & \text{за рядом розподілу,} \\ \int_{-\infty}^{+\infty} f(x) \cdot g(x) dx & \text{за щільністю розподілу.} \end{cases} \quad (2.9)$$

Формули отримано з частотних відповідників шляхом граничного переходу.

Формули обчислення математичного сподівання за рядом розподілу та щільністю. Для обчислення математичного сподівання загальні формули (2.9), застосовані при $g(x) = x$, дають таке:

$$MX = \begin{cases} u_1 \cdot p_1 + u_2 \cdot p_2 + \dots + u_n \cdot p_n (+\dots) & \text{за рядом розподілу,} \\ \int_{-\infty}^{+\infty} f(x) \cdot x dx & \text{за щільністю розподілу.} \end{cases}$$

Формули обчислення дисперсії за рядом розподілу та щільністю. Для обчислення дисперсії досить застосувати загальні формули (2.9) при $g(x) = (x - MX)^2$. Отримаємо

$$DX = M(X - MX)^2 = \\ = Mg(X) = \begin{cases} g(u_1)p_1 + g(u_2)p_2 + \dots + g(u_n)p_n (+\dots) & \text{за рядом розподілу,} \\ \int_{-\infty}^{+\infty} f(x) \cdot g(x) dx & \text{за щільністю розподілу.} \end{cases}$$

2.5. Розподіли ймовірностей

2.5.1. Задання розподілів імовірностей

Розподіли залежно від типу простору елементарних подій U (дискретний чи неперервний) є дискретні та неперервні, і задаються:

- дискретні – рядом розподілу ймовірностей;
- неперервні – функцією розподілу ймовірностей або щільністю розподілу.

Ми розглядатимемо тільки найуживаніші дискретні та неперервні розподіли, які далі називатимемо основними.

Основні дискретні розподіли:

- 1) p -бернулліївський, $0 \leq p \leq 1$;
- 2) (n, p) -біноміальний, $n = 0, 1, 2, 3, \dots$, $0 \leq p \leq 1$;
- 3) p -геометричний (Паскаля), $0 \leq p \leq 1$;
- 4) λ -пуассонівський, $\lambda > 0$;

основні неперервні розподіли:

- 5) λ -експоненціальний, $\lambda > 0$ (показниковий);
- 6) (a, b) -рівномірний, $a < b$; $a, b \in R$;
- 7) (m, σ^2) -нормальний, $m \in R^1$, $\sigma > 0$ (гауссівський), або $N(m, \sigma^2)$.

Кожен із розподілів визначає цілу сукупність розподілів, залежну від того чи іншого параметра або набору параметрів. Ці параметри вказують у префіксному варіанті, як це показано вище, або у постфіксному (бернулліївський із параметром p , експоненціальний із параметром λ та ін.).

2.5.2. Основні дискретні розподіли

Для опису кожного з перелічених вище дискретних розподілів записують ряд розподілу, тобто простір елементарних подій U та відповідні елементарним подіям імовірності.

Бернулліївський розподіл

У бернулліївському розподілі множина можливих значень U завжди складається з двох елементів: $U = \{0, 1\}$.

p -бернулліївський розподіл, де $0 \leq p \leq 1$, характеризують рядом розподілу:

U	0	1
P_u	q	p

де за 0 позначають наслідок, коли подія A не відбулась, а за 1 – коли відбулась, тобто випадкову величину, розподілену за бернулліївським розподілом можна вважати індикатором того, що подія A відбулась. Тут $p+q = 1$. Отже, характеристична властивість (нормованість) виконується.

Якщо розподіл описують як розподіл імовірностей випадкової величини, наприклад X , то ряд розподілу записують у вигляді

$$X = \begin{cases} 0, & q, \\ 1, & p. \end{cases}$$

Приклад 2.2

1) Розподіл Бернуллі з параметром j задано рядом розподілу

0	1
$1-j$	j

2) розподіл Бернуллі з параметром 0,3 (рис. 2.1, а) задають рядом розподілу

0	1
0,7	0,3

3) розподіл Бернуллі з параметром 1 задають рядом розподілу

0	1
0	1

4) розподіл Бернуллі з параметром 0,8 (рис. 2.1, б) задають рядом розподілу

0	1
0,2	0,8

Розподіл Бернуллі описує випадкову величину X , яка дорівнює "кількості появ події A в одному спостереженні, якщо ймовірність появи цієї події для одного спостереження дорівнює p ".

Для такої бернулліївської випадкової величини математичне сподівання та дисперсія, обчислені за рядом розподілу, мають значення p та pq відповідно:

$$\begin{aligned} MX &= 0 \cdot q + 1 \cdot p = p, \\ DX &= M(X - MX)^2 = \\ &= (0 - p)^2 \cdot q + (1 - p)^2 \cdot p = p^2 \cdot q + p \cdot q^2 = \\ &= p \cdot q \cdot (p + q) = p \cdot q. \end{aligned}$$

Отже,

$MX = p$	$DX = pq$
----------	-----------

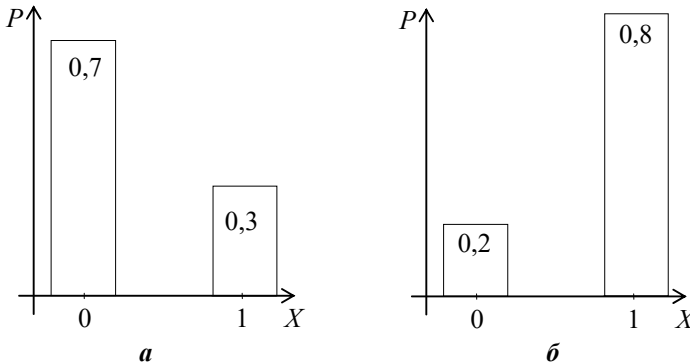


Рис. 2.1. Діаграма розподілу частот бернулліївського розподілу:
а – при $p = 0,3$; б – при $p = 0,8$

Біноміальний розподіл

(n, p) -біноміальний розподіл, де n – натуральне, $0 \leq p \leq 1$, задають рядом розподілу

U	0	1	...	n
P_u	p_0	p_1	...	p_n

Імовірності p_k ряду розподілу визначають співвідношенням

$$p_k = C_n^k p^k \cdot q^{n-k}, \quad k = 0, 1, \dots, n,$$

де C_n^k – біноміальний коефіцієнт, який дорівнює за визначенням кількості k елементних підмножин n -елементної множини. Такі

підмножини називаються **комбінаціями** по k елементів з n , тому C_n^k називають кількістю сполучень з n по k .

Біноміальні коефіцієнти обчислюють за формулою

$$C_n^k = \frac{n!}{k!(n-k)!}, k = 0, \dots, n,$$

де $n!$ (факторіал) є добутком послідовних натуральних чисел від 1 до n включно: $n! = 1 \cdot 2 \cdot \dots \cdot n$. Факторіал від 0 дорівнює 1 тобто $0! = 1$, тому $C_n^0 = 1$.

Виконання характеристичної властивості гарантує біном Ньютона, за яким $(a+b)^n = \sum_{k=0}^n C_n^k a^k b^{n-k}$. Звідси отримуємо справедливість нормованості

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

Біноміально розподіленою випадковою величиною є "кількість появ події A у серії з n незалежних спостережень", якщо ймовірність появи події в одному спостереженні $p = P(A)$. Назву характеристики для скорочення позначають великою літерою латинського алфавіту (наприклад X). Можливими значеннями цієї характеристики є цілі числа від 0 до n , а ймовірності їх появи збігаються з імовірностями відповідних елементарних подій B_k з цієї схеми, які у свою чергу дорівнюють відповідним імовірностям біноміального розподілу:

$$P\{X = k\} = P\{B_k\} = C_n^k \cdot p^k (1-p)^{n-k} = C_n^k \cdot p^k q^{n-k}.$$

Пояснимо появу біноміального коефіцієнта C_n^k у формулі обчислення ймовірності. Для цього розглянемо приклад.

Приклад 2.3

Розглянемо $(3, p)$ -біноміальний розподіл. Ряд розподілу має вигляд

0	1	2	3
p_0	p_1	p_2	p_3

У наведеній серії випробувань усі 3 випробування є незалежними, тому ці події можна записати так:

$p_0 = P\{x_1=0, x_2=0, x_3=0\}$, де $(x_1 \dots x_3)$ – результати в першому, другому та третьому експериментах, а $x_1=0$ означає подію

{у першому випробуванні подія A не відбулась}, $x_2=0$ – {у другому випробуванні подія A не відбулась}, $x_3=0$ – {у третьому випробуванні подія A не відбулась}. Тобто всі результати такої серії випробувань можна умовно записати через подію $p_0=P\{0,0,0\}=(1-p)^3=q^3$;

p_1 є ймовірністю події {у серії з 3-х експериментів подія A відбулась 1 раз}. Результатом такої події може бути наслідок, коли в першому експерименті подія A відбулась, а у другому та третьому – ні, тобто $(1,0,0)$, або ж подія відбулась у другому, а в інших – ні $(0,1,0)$, чи у третьому $(0,0,1)$. Тобто подія {у серії з 3-х експериментів подія A відбулась 1 раз} може бути отримана трьома різними несумісними шляхами $(1,0,0)$, $(0,1,0)$ та $(0,0,1)$, причому ймовірність кожного з них є pq^2 , оскільки подія відбулась 1 раз і 2 рази не відбулась. Остаточно ймовірність p_1 дорівнює сумі ймовірностей подій $(1,0,0)$, $(0,1,0)$ та $(0,0,1)$ за властивістю скінченної адитивності. Тобто для обчислення ймовірності p_1 достатньо обчислити кількість різних шляхів та обчислити ймовірність одного з них. Для обчислення кількості досить розглянути сутність кількості комбінацій C_n^k . За означенням, він дорівнює кількості різних наборів довжиною k , які можна отримати з n різних елементів. У нашому випадку потрібно обрати кількість різних наборів обсягу 1 (1 сприятливий наслідок) із 3-х незалежних експериментів, тобто $C_3^1=3$. Остаточно $p_1=C_3^1 p^1 q^2=3pq^2$;

p_2 є ймовірністю події {у серії з 3-х експериментів подія A відбулась 2 рази}. Результатом такої події можуть бути наслідки, які умовно можна записати як $(0,1,1)$, $(1,0,1)$ чи у $(1,1,0)$. Тобто, міркуючи аналогічно до попереднього випадку, отримаємо $p_2=C_3^2 p^2 q^1=3p^2q$;

$p_3=P\{x_1=1, x_2=1, x_3=1\}$, де $(x_1 \dots x_3)$ – результати в першому, другому та третьому експериментах, тобто подія $\{x_1=1, x_2=1, x_3=1\}$ означає, що у всіх трьох експериментах подія A відбулась і ймовірність цієї події така: $p_3=P\{1,1,1\}=p^3$.

Остаточно ряд розподілу має вигляд

U	0	1	2	3
P	q^3	$3p \cdot q^2$	$3p^2 \cdot q$	p^3

Для побудови діаграми розподілу бернуллівського розподілу розглянемо приклад 2.4.

Приклад 2.4

Якщо розглянути випадкову величину "кількість запізнь на заняття протягом робочого тижня" з імовірністю запізнення кожного дня $1/10$, то ця випадкова величина матиме біноміальний розподіл з $n = 5$ і $p = 1/10$.

Отже, біноміальний розподіл описує розподіл випадкової величини $X = \{\text{кількість появ події } A \text{ у серії з } n \text{ незалежних спостережень}\}$, якщо ймовірність для одного спостереження дорівнює p .

Ряд розподілу частот $p_k = C_n^k p^k q^{n-k}$, $k = 0, 1, \dots, n$, при $n = 5$ і $p = 1/10$ має вигляд

U	0	1	2	3	4	5
P	0,59	0,328	0,073	0,008	0,00045	0,00001

Дійсно,

$$\begin{aligned} p_0 &= P\{X = 0\} = P\{x_1 = 0, x_2 = 0, x_3 = 0, x_4 = 0, x_5 = 0\} = \\ &= C_5^0 p^0 q^5 = (0,9)^5 \approx 0,59; \end{aligned}$$

$$p_1 = P\{X = 1\} = C_5^1 p^1 q^4 = 5 \cdot (0,1) \cdot (0,9)^4 \approx 0,328;$$

$$p_2 = P\{X = 2\} = C_5^2 p^2 q^3 = 10 \cdot (0,1)^2 \cdot (0,9)^3 \approx 0,073;$$

$$p_3 = P\{X = 3\} = C_5^3 p^3 q^2 = 10 \cdot (0,1)^3 \cdot (0,9)^2 \approx 0,008;$$

$$p_4 = P\{X = 4\} = C_5^4 p^4 q^1 = 5 \cdot (0,1)^4 \cdot (0,9)^1 \approx 0,00045;$$

$$p_5 = P\{X = 5\} = C_5^5 p^5 q^0 = 1 \cdot (0,1)^5 \cdot (0,9)^0 \approx 0,00001.$$

Стовпчикову діаграму такого розподілу зображено на рис. 2.2, а.

На рис. 2.2 зображено стовпчикові діаграми бернулліївського розподілу для різних значень параметра p при $n = 5$.

Для обчислення математичного сподівання та дисперсії біноміально розподіленої випадкової величини її записують у вигляді суми n незалежних бернулліївських випадкових величин із параметром p :

$$X = X_1 + X_2 + \dots + X_n.$$

Це представлення впливає з інтерпретації біноміальної випадкової величини як кількості появ події A в n незалежних випробуваннях Бернуллі. Відповідно, кожна з X_i , $i = 1, \dots, n$, представляють як кількість появ події A в i -му спостереженні (X_i

дорівнює 1, якщо подія A відбулась у i -му експерименті; і рівна 0, якщо подія не відбулась). Оскільки кількість X появ події A в n спостережуваних складається із суми кількості X_1 появи події A в першому спостереженні, X_2 – у другому спостереженні і т. д., а незалежність X_i випливає з незалежності спостережень, то

$$MX = M(X_1 + X_2 + \dots + X_n) = MX_1 + MX_2 + \dots + MX_n = np,$$

$$DX = D(X_1 + X_2 + \dots + X_n) = DX_1 + DX_2 + \dots + DX_n = npq.$$

Отже,

$MX = np$	$DX = npq$
-----------	------------

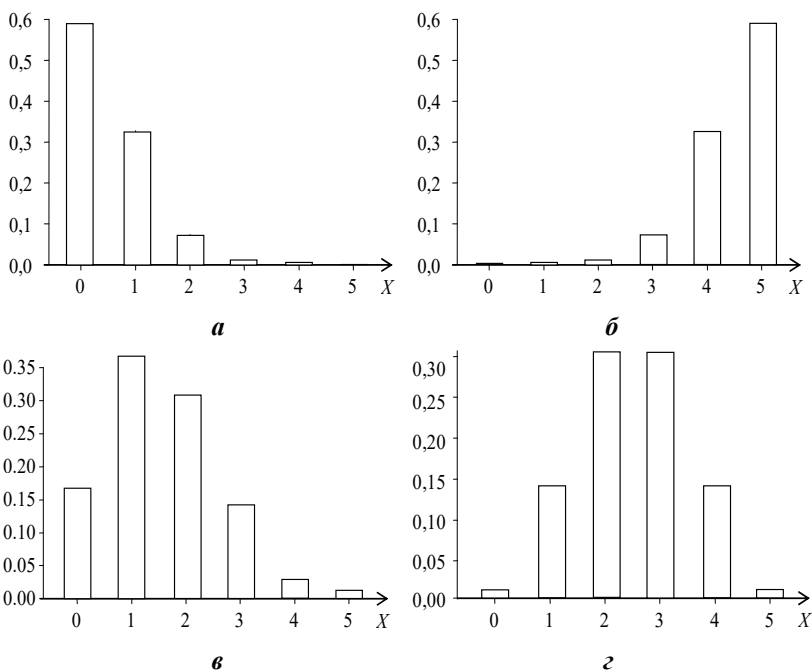


Рис. 2.2. Стовпчикові діаграми біноміального розподілу при $n=5$:
а – при $p=0,1$; **б** – при $p=0,9$; **в** – при $p=0,3$; **г** – при $p=0,5$

Для біноміального розподілу існує ще один спосіб обчислення інтегральних характеристик:

$$DX = M(X - MX)^2 = \sum_{u \in U} (u - MX)^2 p_u = \sum_{k=0}^n (k - MX)^2 C_n^k p^k q^{n-k}.$$

Зауважимо, що за загальними формулами підрахунку цей результат установити значно важче, оскільки

$$MX = \sum_{k=0}^n u_k p_k = \sum_{k=0}^n k C_n^k p^k q^{n-k} .$$

Пуассонівський розподіл

λ -пуассонівський розподіл, $\lambda > 0$. Ряд розподілу запишеться так:

U	0	1	...	n	...
P_u	p_0	p_1	...	p_n	...

$$p_n = P\{n\} = \frac{\lambda^n}{n!} e^{-\lambda} .$$

Якщо розподіл розглядають як розподіл імовірностей випадкової величини, то відповідні ймовірності набувають вигляду

$$p_n = P\{X = n\} = \frac{\lambda^n}{n!} e^{-\lambda} .$$

Для різних значень λ стовпчикову діаграму пуассонівського розподілу зображено на рис. 2.3.

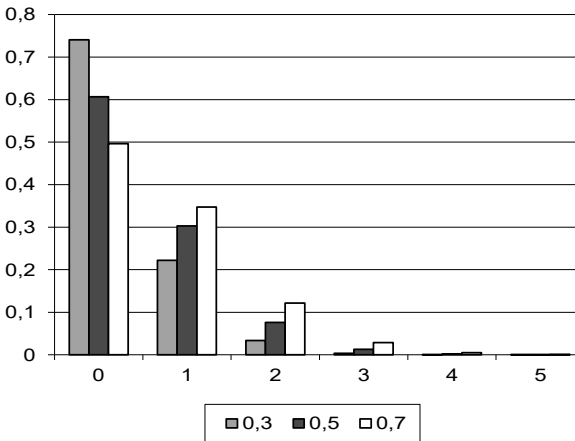


Рис. 2.3. Стовпчикова діаграма розподілу Пуассона при $\lambda=0,3$, $\lambda=0,5$, $\lambda=0,7$

Пуассонівський розподіл справедливий для подій, які мають малу ймовірність чи трапляються нечасто. За його допомогою, наприклад, можна описати ймовірність того, що футболіст заб'є гол у конкретному матчі. Іноді футболіст забиває один гол, рідше два, ще рідше три або більше. На рахунку Пеле (Едісон Арантіс ду Насіменту), кращого у світовій історії футболіста 92 хет-трики (3 голи за 1 матч); 30 матчів, де він забив по 4 голи, та не менше 6 ігор, у яких Пеле забив по 5 голів. У 1964 р. Пеле забив у одному матчі 8 голів у ворота "Ботафого". Найчастіше футболіст не забиває жодного.

Ймовірність забити k голів за гру визначається параметром λ , що є середньою кількістю голів, які забиває футболіст. Якщо λ велике число, то ймовірність має досягати максимуму при якомусь k . У такому випадку йдеться швидше про баскетболіста, який може набирати, наприклад, 22 очки за гру в середньому. Тоді ймовірність набрати 2 очки буде малою. Ймовірність набрати 42 очки теж буде малою, а максимум імовірності становитиме близько 22 очок.

Математичне сподівання та дисперсія випадкової величини, розподіленої за пуассонівським законом, має вигляд

$MX = \lambda$	$DX = \lambda$
----------------	----------------

Геометричний розподіл

Ряд розподілу випадкової величини, розподіленої за p -геометричним розподілом (розподіл Паскаля) має вигляд

U	1	2	...	n	...
P_u	p_1	p_2	...	p_n	...

тут $p_k = P \{k\} = P \{X = k\} = pq^{k-1}$.

Геометричним розподілом описують числову характеристику випадкової величини $X = \{\text{кількість спостережень до першої появи події } A \text{ у схемі незалежних випробувань}\}$. Ця випадкова величина набуває натуральних значень від 1 (під час першого спостереження подія A відбулась) до нескінченності, отже, простір елементарних подій U є зліченим.

На рис. 2.4. зображено стовпчикові діаграми геометричного розподілу для різних значень параметра p .

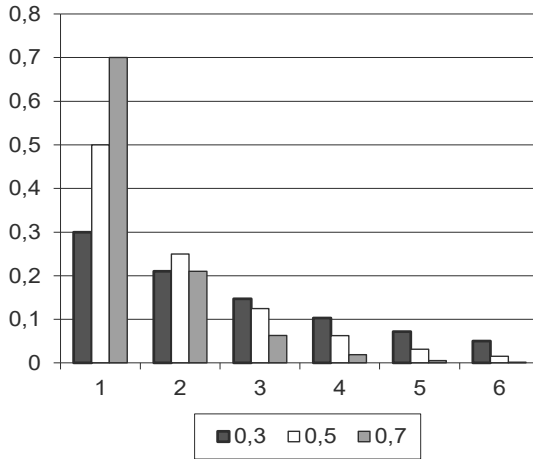


Рис. 2.4. Стовпчикова діаграма геометричного розподілу при $p=0,3$; $p=0,5$; $p=0,7$

Імовірності p_k утворюють геометричну прогресію (звідси пішла назва розподілу) із першим членом p та знаменником q , тобто $p_k = q \cdot p_{k-1}$. Виконання характеристичної властивості гарантує суму нескінченної кількості членів геометричної прогресії

$$\lim_{n \rightarrow \infty} S_n = \lim_{n \rightarrow \infty} p \frac{q^n - 1}{q - 1} = \frac{p}{1 - q} = \frac{p}{p} = 1.$$

Математичне сподівання та дисперсія випадкової величини, розподіленої за геометричним законом, мають вигляд

$MX = \frac{1}{p}$	$DX = \frac{1-p}{p^2} = \frac{q}{p^2}$
--------------------	--

2.5.3. Основні неперервні розподіли

До основних неперервних розподілів належать:

- 1) λ -експоненціальний (показниковий),
- 2) (a, b) -рівномірний,
- 3) (m, σ^2) -нормальний (гауссівський або $N(m, \sigma^2)$ -розподіл).

Неперервні розподіли задають щільністю або функцією розподілу.

Експоненціальний розподіл

Функція λ -експоненціального (показникового) розподілу має вигляд

$$F(x) = P \{ X < x \} = \begin{cases} 0, & x \leq 0, \\ 1 - e^{-\lambda x}, & x > 0. \end{cases}$$

Графік функції розподілу зображено на рис. 2.5.

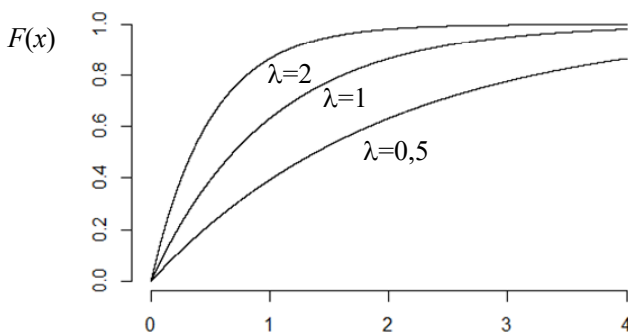


Рис. 2.5. Графік функції експоненціального розподілу при $\lambda = 0,5$, $\lambda = 1$ та $\lambda = 2$

Нескладно переконатись у тому, що всі характеристичні властивості функції розподілу виконуються.

Щільність обчислюють як похідну від функції розподілу (рис. 2.6):

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

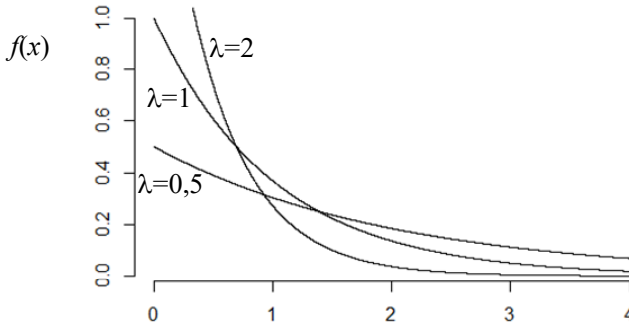


Рис. 2.6. Графік щільності експоненціального розподілу при $\lambda=0,5$, $\lambda=1$ та $\lambda=2$

Для щільності експоненціального розподілу площа під графіком дорівнює одиниці (перевіряють інтегруванням), тобто виконується характеристична властивість щільності.

Так само інтегруванням установлюють, що

$MX = \frac{1}{\lambda}$	$DX = \frac{1}{\lambda^2}$
--------------------------	----------------------------

Типовою ситуацією, де використовують експоненціальний розподіл, є обчислення характеристик X типу "тривалість життя" чи "час безвідмовної роботи". Зазвичай указують середнє значення цієї характеристики (середня тривалість життя чоловіка й жінки, середній час безвідмовної роботи якогось приладу). У всіх цих випадках λ є величиною, оберненою до математичного сподівання, яке інтерпретують як середнє значення відповідної характеристики: $\lambda = 1/(\text{середнє значення})$, тобто

$$\lambda = \frac{1}{MX}.$$

Експоненціальний розподіл зосереджений на додатній півосі. Це означає, що від'ємні значення з'являються з імовірністю, яка дорівнює 0:

$$P(-\infty; 0) = P\{X < 0\} = F(0) = 0.$$

Рівномірний розподіл

Функцією (a, b) -рівномірного розподілу є

$$F(x) = P\{X < x\} = \begin{cases} 0, & x \leq a, \\ \frac{x-a}{b-a}, & a < x \leq b, \\ 1, & x > b. \end{cases}$$

Графік функції розподілу зображено на рис. 2.7.



Рис. 2.7. Графік функції
рівномірного розподілу з параметрами 1, 4

Щільність рівномірного розподілу (рис. 2.8) обчислюють через похідну від функції розподілу

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

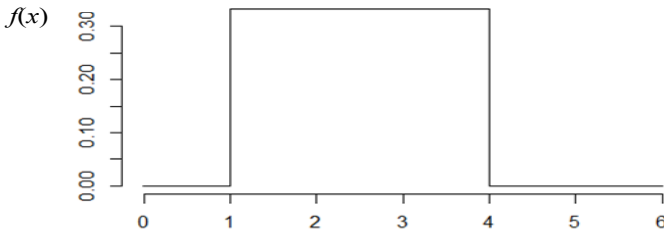


Рис. 2.8. Графік щільності
рівномірного розподілу з параметрами 1, 4

Рівномірний розподіл зосереджено на інтервалі (a, b) . Це означає, що ймовірність появи значень із цього інтервалу дорівнює одиниці:

$$P \{X \in [a, b]\} = P \{a \leq X < b\} = F(b) - F(a) = 1 - 0 = 1.$$

Оскільки за нормованістю та адитивністю ймовірності

$$1 = P \{X \in [a, b]\} + P \{X \notin [a, b]\},$$

то

$$P \{X \notin [a, b]\} = 1 - P \{X \in [a, b]\} = 0.$$

Інтегруванням за відповідними формулами встановлено, що

$MX = \frac{a+b}{2} \text{ (середина } (a, b))$	$DX = \frac{(b-a)^2}{12}$
---	---------------------------

Рівномірний розподіл використовують для неперервних даних, коли кожний із можливих результатів з'являється з однаковою ймовірністю, тобто всі результати є рівноможливими.

Нормальний розподіл

Нормальний $N(m, \sigma^2)$ -розподіл задають щільністю (як правило), тому що для функції розподілу не існує простішого аналітичного виразу, ніж запис через інтеграл від щільності.

Щільність нормального розподілу знаходять за формулою

$$f(x) = \frac{1}{\sqrt{2\pi} \cdot \sigma} \cdot e^{-\frac{(x-m)^2}{2\sigma^2}}, \quad (2.10)$$

де основа натурального логарифма $e \approx 2,718281828459$ (як правило беруть $e \approx 2,72$), $\pi \approx 3,141592654$ (як правило беруть $\pi \approx 3,14$).

Функцію нормального розподілу обчислюють як інтеграл від щільності

$$F(x) = \int_{-\infty}^x f(t) dt. \quad (2.11)$$

Графік щільності симетричний щодо вертикальної прямої у точці з координатами m (математичне сподівання) на осі OX і є тим розлогішим, чим більший параметр σ^2 (дисперсія).

На рис. 2.9. представлено графіки функції розподілу при $m = 0$ (математичне сподівання) для розподілів $N(0, 1)$, $N(0, 2)$ та $N(0, 4)$.

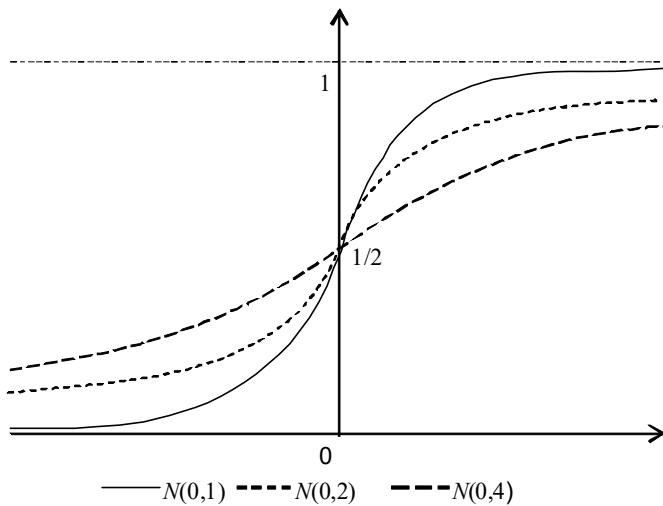


Рис. 2.9. Функція нормального розподілу (нормальна огива)

Щільності нормального розподілу з нульовим математичним сподіванням та різними дисперсіями 1, 2 та 4 відповідно зображено на рис. 2.10.

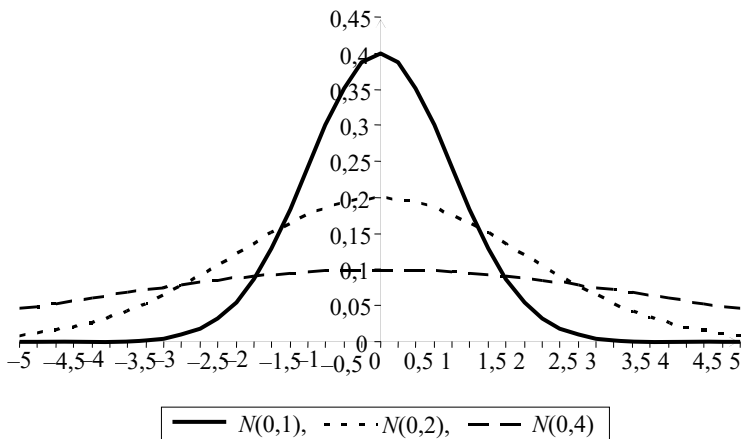


Рис. 2.10. Щільність нормального розподілу

Інтегруванням за щільністю встановлюють, що перший параметр нормального розподілу збігається з математичним сподіванням, а другий – із дисперсією випадкової величини, що має цей розподіл:

$$\begin{aligned} MX &= m, \\ DX &= \sigma^2. \end{aligned}$$

Нормальні розподіли є основними в теорії помилок спостереження:

- m – істинне значення вимірюваної величини;
- σ^2 – розсіювання точності вимірювання приладу, якщо цей прилад не має систематичних помилок (тобто помилок, які повторюються з однаковим значенням, наприклад зсув циферблата на 5°);
- результат вимірювання має $N(m, \sigma^2)$ -розподіл.

Якщо використовується прилад із систематичною помилкою b , то результат має нормальний $N(m+b, \sigma^2)$ -розподіл.

Стандартним нормальним називають нормальний розподіл із параметрами $m=0$ та $\sigma^2=1$, позначається $N(0,1)$. Щільність стандартного нормального розподілу отримують із (2.10):

$$f(x) = \frac{1}{\sqrt{2\pi}} \cdot e^{-\frac{x^2}{2}}.$$

Випадкові величини зі стандартним нормальним розподілом у подальшому позначатимемо через Z . Функцію розподілу $F_{0,1}(x)$ стандартного нормального розподілу позначають $\Phi(x)$ і називають **функцією Лапласа**:

$$F_{0,1}(x) = P \{ Z < x \} = \Phi(x).$$

Значення функції Лапласа наводять у статистичних таблицях (див. наведений далі розд. 4, табл. 4.2). Важливість функції Лапласа полягає в тому, що через неї обчислюють функцію розподілу $F_{m,\sigma^2}(x)$ будь-якого нормального розподілу відповідно до співвідношення

$$F_{m,\sigma^2}(x) = \Phi \left(\frac{x-m}{\sigma} \right). \quad (2.12, a)$$

Із (2.11) функція розподілу стандартного нормального розподілу має вигляд

$$\Phi(x) = \int_{-\infty}^x f(y)dy = S_{(-\infty;x)}^{\frac{1}{\sqrt{2\pi}}e^{-\frac{y^2}{2}}} = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^x e^{-\frac{y^2}{2}} dy,$$

де вираз $S_{(-\infty;x)}^{\frac{1}{\sqrt{2\pi}}e^{-\frac{y^2}{2}}}$ означає площу під графіком щільності над інтервалом $(-\infty, x)$.

Властивості функції Лапласа

1. Функція $\Phi(x)$ визначена для всіх дійсних $x \in (-\infty; \infty)$;

2. Щільність $f(x) = \Phi'(x)$ є симетричною функцією

$$\Phi'(-x) = \Phi'(x), \quad \Phi(-x) = 1 - \Phi(x);$$

3. $\Phi(0) = 1/2$, $\Phi(0) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^0 e^{-\frac{x^2}{2}} dx = \frac{1}{2}$,

оскільки $\frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-\frac{x^2}{2}} dx = \frac{\sqrt{2\pi}}{\sqrt{2\pi}} = 1$ є інтегралом Пуассона;

4. $\Phi(-\infty) = 0$, $\Phi(\infty) = 1$;

5. $\Phi(x)$ є функцією неспадною, оскільки $\Phi'(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} > 0$;

6. $\Phi'(x)$ досягає найбільшого значення в точці 0;

7. $\Phi'(x)$ має перегин у точці 0,

оскільки $\Phi''(0) = 0$, $\Phi''(x)|_{x<0} > 0$, $\Phi''(x)|_{x>0} < 0$.

Лінійна функція від нормально розподіленої випадкової величини є нормально розподіленою. Параметри цієї нової нормальної випадкової величини встановлюють обчисленням її математичного сподівання та дисперсії. Із (2.12, а) впливає ця властивість, її можна переписати так:

$$\frac{X - m}{\sigma} = Z, \quad X = m + \sigma \cdot Z. \quad (2.12, б)$$

Приклад 2.5 (дискретні випадкові величини)

Проводять ряд спроб, щоб завести двигун. Кожна спроба завершується вмиканням двигуна незалежно від інших спроб з імовірністю $p = 0,6$. Кожна спроба займає час t . Знайти розподіл

У загального часу, який буде потрібний для запуску двигуна, а також математичне сподівання та дисперсію.

Розв'язання

Знайдемо спочатку розподіл кількості спроб до першого запуску двигуна. Випадкова величина X , яка дорівнює кількості спроб запуску двигуна до першого запуску, має геометричний розподіл із параметром $p = 0,6$. Звідси, враховуючи характеристики геометричного розподілу, отримаємо

$$P \{X = k\} = pq^{k-1}, \quad M(X) = \frac{1}{p}, \quad D(X) = \frac{1-p}{p^2}$$

або
$$P \{X = k\} = 0,6 \cdot 0,4^{k-1}, \quad M(X) = \frac{1}{0,6} = 10/6,$$

$$D(X) = \frac{1-0,6}{0,6^2} = 40/36.$$

Випадкова величина $Y = X \cdot t$, тому

$$P \{Y = kt\} = 0,6 \cdot 0,4^{k-1}, \quad M(Y) = M(X)t = \frac{1}{0,6} t = 10/6 \cdot t,$$

$$D(Y) = D(X)t^2 = t^2 \frac{1-0,6}{0,6^2} = t^2 40/36.$$

Приклад 2.6

Незалежно кидають n гральних шестигрannих кубиків. Знайти математичне сподівання та дисперсію суми кількості очок, які випадають на всіх кубиках.

Розв'язання

Запишемо ряд розподілу ймовірностей для кількості очок, що випадають на одному кубіку:

U	1	2	3	4	5	6
P_U	1/6	1/6	1/6	1/6	1/6	1/6

На перший погляд, можемо далі для знаходження математичного сподівання та дисперсії записати ряд розподілу суми очок, що випадають на двох, трьох гральних кубиках, але зі збільшенням кількості кубиків зростатиме кількість різних значень, яких може набувати сума очок (від n до $6n$), а отже і

довжина самого ряду розподілу, а вже для довільного n узагалі такий ряд записати не вдасться. Якщо звернутися до умови задачі, то там не йдеться про те, щоб записати ряд розподілу для суми очок, що випадає на n кубиках, а лише про математичне сподівання та дисперсію.

Позначимо як U_i кількість очок, що випадає на i -му кубуку, $i = 1, 2, \dots, n$. Усі U_i є незалежними однаково розподіленими випадковими величинами, тому запишемо як $Y = U_1 + U_2 + \dots + U_n$ випадкову величину, що дорівнює сумі незалежних однаково розподілених випадкових величин.

Звернемось до властивостей математичного сподівання та дисперсії:

1) математичне сподівання суми випадкових величин дорівнює сумі математичних сподівань доданків, тобто $MY = MU_1 + MU_2 + \dots + MU_n$.

2) дисперсія суми незалежних випадкових величин дорівнює сумі їхніх дисперсій, тобто $DY = DU_1 + DU_2 + \dots + DU_n$.

Отже, достатньо знайти MU_i та DU_i для довільного i (тому що всі U_i є однаково розподіленими). Маємо

$$MU_i = 1 \cdot 1/6 + 2 \cdot 1/6 + 3 \cdot 1/6 + 4 \cdot 1/6 + 5 \cdot 1/6 + 6 \cdot 1/6 = 3,5;$$

$$DU_i = (1 - 3,5)^2 \cdot 1/6 + (2 - 3,5)^2 \cdot 1/6 + (3 - 3,5)^2 \cdot 1/6 + (4 - 3,5)^2 \cdot 1/6 + (5 - 3,5)^2 \cdot 1/6 + (6 - 3,5)^2 \cdot 1/6 \approx 2,9.$$

Далі отримаємо

$$MY = n \cdot 3,5 \quad \text{та} \quad DY = n \cdot 2,9.$$

ЗАДАЧІ

Дискретні випадкові величини

№ 1

Двічі підкидають монету. Нехай випадкова величина X – кількість появ герба. Знайти розподіл випадкової величини X , її математичне сподівання MX та дисперсію DX .

№ 2

Двічі підкидають шестигранний гральний кубик. Нехай випадкова величина X – сума отриманих таким чином очок. Знайти розподіл випадкової величини X та її математичне сподівання MX .

№ 3

Монету підкидають доти, доки не випаде герб. Описати простір елементарних подій U . Нехай випадкова величина X – кількість підкидань. Знайти розподіл випадкової величини X .

№ 4

Із партії, що містить 100 виробів, серед яких є 10 дефектних, навмання по черзі виймають 5 виробів, щоразу повертаючи їх назад. Записати закон розподілу випадкової величини X – кількості дефектних виробів, що містяться у вибірці.

№ 5

Випадкова величина X має такий розподіл:

$X: U$	-1	-0.5	-0.1	0	0.1	0.2	0.5	1	1.5	2
P	0.005	0.012	0.074	0.102	0.148	0.231	0.171	0.160	0.081	0.016

Обчислити:

- математичне сподівання та дисперсію випадкової величини X ;
- $P\{X \leq 0,5\}$.

№ 6

Випадкова величина X має такий розподіл:

$X: U$	-1	-0.5	-0.1	0	1	2.0
P	0.05	0.1	0.15	0.2	0.2	0.3

Обчислити:

- математичне сподівання та дисперсію випадкової величини X ;
- $P\{X^2 > 0,5\}$.

2.6. Основна задача теорії ймовірності

2.6.1. Формулювання основної задачі теорії ймовірності (ОЗТЙ)

Вичерпною характеристикою масового явища є ймовірність або розподіл імовірностей із ним пов'язаних, що дає можливість підрахувати появу будь-яких подій у зв'язку з досліджуванним

масовим явищем. Але іноді немає потреби, важко або неможливо визначити всі ймовірності. Ілюстрацією цього може слугувати наступний приклад.

Приклад 2.7

Дві незалежні одна від одної ознаки можуть бути наявними у досліджуваних осіб з ймовірностями 0,2 та 0,6 відповідно. Якою є ймовірність того, що досліджуваний матиме одночасно обидві ознаки.

Характерною відзнакою такого роду задач є необхідність обчислення ймовірностей одних подій за ймовірностями інших, урахуовуючи при цьому зв'язки, що існують між подіями. Таку задачу називають **основною задачею теорії ймовірностей (ОЗТЙ)** в її елементарному прояві.

ОЗТЙ називають задачу визначення ймовірності одних подій за ймовірностями інших на основі зв'язків, що існують між досліджуваними подіями. Ці зв'язки можуть задаватися одним із трьох способів:

- 1) через операції,
- 2) через незалежність,
- 3) через умовні ймовірності.

Спеціальним видом ОЗТЙ є задачі, в яких використовують однакову можливість результатів. У таких задачах на обчислення ймовірностей немає явного посилання на значення ймовірностей, але є посилання на рівну можливість результатів, що передають виразами: "рівноможливі", "навмання", "випадковим чином" і т. д. Зміст задач цього типу зображають такою схемою.

Рівна можливість результатів

("навмання", "випадковим чином", ...) $\Rightarrow P(B)$?

Опис події B

Будь-яку ОЗТЙ розв'язують через розв'язання комбінації найпростіших ОЗТЙ, які називають базовими. Ознакою базової ОЗТЙ є наявність у назві виразів: "формула", "теорема", "схема".

До базових ОЗТЙ належать такі.

- ТЕОРЕМИ (формули):
 - теорема (формула) заперечення,
 - теорема (формула) різниці,
 - теорема (формула) добутку,

- теорема (формула) додавання,
- теорема (формула) Байєса (ФБ),
- формула повної ймовірності (ФПЙ);
- СХЕМИ:
 - незалежних випробувань Бернуллі,
 - незалежних випробувань до першої появи події A .

2.6.2. Постановка, схема, базові задачі ОЗТЙ

Теорема 2.1 (заперечення)

Якщо $B = \bar{A}$, то

$$P(B) = 1 - P(A), P(A) = 1 - P(B), P(A) + P(B) = 1.$$

Доведення випливає з того, що за законом виключення третього маємо $A \cdot B = \emptyset$ (несумісні події) та $A + B = U$. Використовуючи нормованість та адитивність, одержуємо

$$1 = P(U) = P(A + B) = P(A) + P(B).$$

Приклад 2.8

Нехай ймовірність наявності ознаки дорівнює 0,6. Тоді маємо ймовірність її відсутності

$$1 - 0,6 = 0,4.$$

Теорема 2.2 (різниця)

Ця теорема має декілька варіантів формулювання.

1) Якщо $A \supseteq B$, то $P(A \setminus B) = P(A) - P(B)$.

Доведення

З одного боку $(A \setminus B) \cdot B = \emptyset$, а з іншого $A = B + (A \setminus B)$. Використовуючи адитивність, маємо

$$P(A) = P(B + (A \setminus B)) = P(B) + P(A \setminus B),$$

$$P(A) = P(B) + P(A \setminus B),$$

$$P(A \setminus B) = P(A) - P(B).$$

2) Для загального розташування A, B

$$P(A \setminus B) = P(A) - P(A \cdot B).$$

Доведення

Доведення випливає зі співвідношень $A \setminus B = A \setminus (A \cdot B)$ та $A \supseteq A \cdot B$. Останнє дає можливість скористатися попереднім варіантом теореми.

Приклад 2.9

Нехай імовірність захворіти на грип A дорівнює $0,6$, на грип $B - 0,5$, на грип A та B одночасно $- 0,2$.

Тоді ймовірності захворіти на грип A і не захворіти на грип B та навпаки обчислюють за формулами різниці:

$$P(A \setminus B) = P(A) - P(A \cdot B) = 0,6 - 0,2 = 0,4,$$

$$P(B \setminus A) = P(B) - P(A \cdot B) = 0,5 - 0,2 = 0,3.$$

Теорема 2.3 (додавання)

Ця теорема має три варіанти формулювання.

1) Адитивність: у вигляді адитивності, скінченної адитивності чи зліченної адитивності:

- адитивність: якщо $A \cdot B = \emptyset$ (події несумісні), то

$$P(A + B) = P(A) + P(B);$$

- скінченна адитивність: якщо A_1, \dots, A_n такі, що $A_i \cdot A_j = \emptyset$, $i \neq j$ (події попарно несумісні), то

$$P(A_1 + \dots + A_n) = P(A_1) + \dots + P(A_n);$$

- зліченна адитивність: якщо A_1, \dots, A_n, \dots такі, що $A_i \cdot A_j = \emptyset$, $i \neq j$ (події попарно несумісні), то

$$P(A_1 + \dots + A_n + \dots) = P(A_1) + \dots + P(A_n) + \dots$$

2) Загальний варіант для двох подій:

$$P(A + B) = P(A) + P(B) - P(A \cdot B).$$

3) Загальний варіант для незалежних у сукупності подій A_1, \dots, A_n : якщо A_1, \dots, A_n незалежні в сукупності, то

$$\begin{aligned} P(A_1 + \dots + A_n) &= 1 - P(\overline{A_1 + \dots + A_n}) = 1 - P(\overline{A_1} \cdot \dots \cdot \overline{A_n}) = \\ &= 1 - P(\overline{A_1}) \cdot \dots \cdot P(\overline{A_n}) = 1 - (1 - P(A_1)) \cdot \dots \cdot (1 - P(A_n)), \end{aligned}$$

де $B_i = \overline{A_i}$ для $i=1, \dots, n$.

Варіанти адитивності належать до характеристичних властивостей імовірності.

Справедливість загального варіанта для двох подій випливає з того, що $A+B$ можна представити у вигляді суми несумісних подій A та $B \setminus A$, після чого скористатися адитивністю та теоремою різниці.

Останній варіант цієї теореми доводять на основі важливої властивості незалежності, яку можна назвати спадковою властивістю незалежності.

Спадковою властивістю незалежності для незалежних у сукупності подій A_1, \dots, A_n називають здатність незалежності передаватися від цих подій їхнім запереченням у будь-яких комбінаціях.

Зокрема, із незалежності двох подій A та B випливає незалежність подій у трьох парах: \bar{A} та \bar{B} ; A та \bar{B} ; B та \bar{A} .

Так само з незалежності в сукупності A_1, \dots, A_n відповідно зі спадковістю випливає незалежність у сукупності подій $\bar{A}_1, \bar{A}_2, \dots, \bar{A}_n$.

Для доведення третього варіанта теореми додавання користуються спочатку формулою заперечення та законом де Моргана, а потім спадковістю незалежності:

$$\begin{aligned} P(A_1 + \dots + A_n) &= 1 - P(\overline{A_1 + \dots + A_n}) = 1 - P(B_1 \cdots B_n) = \\ &= 1 - P(B_1) \cdots P(B_n) = 1 - (1 - P(A_1)) \cdots (1 - P(A_n)). \end{aligned}$$

Приклад 2.10

В умовах прикладу 2.9 імовірність "захворіти хоча б на один з типів грипу" підраховують за теоремою додавання

$$P(A+B) = P(A) + P(B) - P(A \cdot B) = 0,6 + 0,5 - 0,2 = 0,8.$$

Приклад 2.11

Якщо кожна з трьох незалежних у сукупності ознак може бути присутня з імовірністю відповідно 0,5, 0,3 та 0,4, то ймовірність наявності хоча б однієї з цих ознак підраховують за третім варіантом теореми додавання:

$$\begin{aligned} P(A_1 + A_2 + A_3) &= 1 - (1 - P(B_1)) \cdot (1 - P(B_2)) \cdot (1 - P(B_3)) = \\ &= 1 - 0,5 \cdot 0,7 \cdot 0,6 = 0,79. \end{aligned}$$

Теорема 2.4 (добутку)

Теорема добутку має два основні варіанти: загальний та для незалежних подій.

1) Загальний варіант:

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

тут $P(A) \neq 0$, $P(B) \neq 0$ для тих частин запису, де відповідна подія зустрічається в умові.

2) Для незалежних подій:

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

Для незалежних у сукупності подій:

$$P(A_1 \cdots A_n) = P(A_1) \cdots P(A_n).$$

Доведення першого варіанта теореми впливає зі зв'язку умовної та безумовної ймовірностей:

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

Для отримання твердження теореми треба в останньому співвідношенні помножити обидві частини рівняння на знаменник правої частини.

Щодо варіантів цієї теореми для незалежних подій, то відповідні співвідношення збігаються з визначенням незалежності чи безпосередньо впливають із них.

Приклад 2.12

Нехай ймовірність $A = \{\text{потрапити до групи ризику}\}$ складає 0,1, а $B = \{\text{ймовірність захворіти на грип}\}$ у цій групі складає 0,8. Тоді ймовірність одночасного виконання двох умов (належності до групи ризику та захворіти на грип) обчислюють за формулою

$$P(A \cdot B) = P(A/B) \cdot P(B) = 0,8 \cdot 0,1 = 0,08.$$

Зауважимо, що незалежність тих чи інших об'єктів в умові задачі однозначно тлумачать як незалежність подій, що цим об'єктам відповідають, а незалежність ознак означає незалежність подій.

Спадковістю незалежності загалом називають властивість результатів перетворень успадковувати незалежність подій, що перетворюються. Зокрема, це стосується операцій (застосування операцій зберігає незалежність).

Спадковість незалежності справедлива для всіх результатів застосування всіх операцій. Загалом на операції заперечення подій властивість спадковості ілюструвати найпростіше, тому доведення виконуватиметься саме для цього випадку. Отже, у цьому найпростішому випадку твердження про спадковість має вигляд такої теореми.

Теорема 2.5

Якщо A та B незалежні, то \bar{A} та B незалежні,

A та \bar{B} незалежні,

B та \bar{B} незалежні.

Доведення

Використовуючи представлення різниці подій у вигляді $\bar{A} \cdot B = B \setminus A$, потім незалежність подій A та B і формулу заперечення, отримаємо

$$\begin{aligned}
 P(\bar{A} \cdot B) &= P(B \setminus A) = P(B) - P(B \cdot A) = \\
 &= P(B) - P(A) \cdot P(B) = P(B) \cdot (1 - P(A)) = P(B) \cdot P(\bar{A}),
 \end{aligned}$$

що й доводить теорему.

Другий випадок отримують з першого зміною позначень A на B та навпаки.

Третій випадок теж можна отримати з першого та другого, застосувавши їх послідовно:

A та B є незалежними $\Rightarrow A$ та \bar{B} незалежні $\Rightarrow B, \bar{B}$ незалежні.

2.6.3. Формула повної ймовірності (ФПЙ)

Зміст ФПЙ полягає в можливості обчислення ймовірності події B , коли її (B) можна реалізувати одним із декількох способів, що взаємно виключають один одного. Зміст виразу "способи реалізації" пояснюють через поняття "повна група подій".

Повною групою подій називають такий набір подій $H_1, \dots, H_n, (\dots)$, для яких виконуються такі твердження.

1) Вони є попарно несумісними $H_i \cdot H_j = \emptyset, i \neq j$.

2) Справедливою є одна з двох умов:

- хоча б одна з подій H_k реалізується:

$$H_1 + \dots + H_n (+\dots) = U;$$

- сума ймовірностей цих подій є одиницею:

$$P(H_1) + \dots + P(H_n) (+\dots) = 1.$$

Приклад 2.13

а) При підкиданні монети події $A = \{\text{випаде герб}\}$ та $B = \{\text{випаде номінал}\}$ утворюють повну групу подій із двох елементів:

$$1) A \cdot B = \emptyset,$$

$$2) A + B = U.$$

б) При підкиданні грального шестигранного кубика події $H_i = \{\text{на грані кубика випало число } i\}, i = 1, \dots, 6$, утворюють повну групу подій із шести елементів: усі події H_i є попарно несумісними та їхня сума утворює достовірну подію.

Теорема 2.6 (ФПЙ)

Якщо $H_1, \dots, H_n (\dots)$ утворюють таку повну групу подій, що $P(H_i) \neq 0, i = 1, \dots, n$, то ймовірність будь-якої події B можна об-

числити, як суму добутоків умовних імовірностей на ймовірності умов за всіма подіями повної групи:

$$\begin{aligned}
 P(B) &= P(B/H_1) \cdot P(H_1) + P(B/H_2) \cdot P(H_2) + \\
 &+ \dots + P(B/H_n) \cdot P(H_n) \cdot (+\dots) = \sum_{i=1}^{n(\infty)} P(B/H_i) \cdot P(H_i). \quad (2.13)
 \end{aligned}$$

Зокрема, для $n = 2$ маємо $P(B) = P(B/A) \cdot P(A) + P(B/\bar{A}) \cdot P(\bar{A})$, де $H_1 = A$; $H_2 = \bar{A}$.

Доведення

Доведення проілюструємо для випадку $n = 2$. Ураховуючи закон виключення третього, адитивність та теорему добутку, маємо

$$\begin{aligned}
 P(B) &= P((B \cap \bar{A}) \cup (B \cap A)) = \\
 &= P(B \cap \bar{A}) + P(B \cap A) = P(B | \bar{A}) \cdot P(\bar{A}) + P(B | A) \cdot P(A).
 \end{aligned}$$

Приклад 2.14

У коробці лежать 3 червоні та 5 білих кульок. Навмання виймають одну кульку і відкладають убік. Потім виймають ще одну. Знайти ймовірність того, що вона червоного кольору.

Розв'язання

Коли ми виймаємо першу кульку, то можливими наслідками є два: {вийняли червону кульку} та {вийняли білу кульку}. Ці події утворюють повну групу подій (можна вибрати тільки або червону, або білу), тому для нашої задачі подіями H_1 та H_2 будуть $H_1 = \{\text{вийняли червону кульку}\}$, $H_2 = \{\text{вийняли білу кульку}\}$. Ураховуючи те, що першу кульку вийняли навмання, можемо говорити про однакову ймовірність вийняти кожен з кульок – $1/8$. Таким чином $P(H_1) = 3/8$ та $P(H_2) = 5/8$, тобто вийняти червону кульку ми можемо трьома варіантами з восьми (всього 8 кульок у коробці), а білу – п'ятьма варіантами з восьми.

Тепер сформулюємо подію $B = \{\text{друга кулька є червоною}\}$. Незалежно від того, якого кольору кульку ми вийняли першою, змінюється співвідношення кульок у коробці, тобто якщо першою була біла кулька, то в коробці залишилось 3 червоні та 4 білі кульки, а якщо першою була червона – то 2 червоні та 5 білих. Тобто, якщо записати події $B/H_1 = \{\text{друга кулька червоного кольору, за умови, що перша кулька була червона}\}$ та $B/H_2 = \{\text{друга кулька червоного кольору, за умови, що перша кулька була біла}\}$,

то отримаємо умовні ймовірності $P(B/H_1)=2/7$, $P(B/H_2)=3/7$. Далі, за формулою повної ймовірності маємо

$$\begin{aligned} P(B) &= P(B/H_1) \cdot P(H_1) + P(B/H_2) \cdot P(H_2) = \\ &= 2/7 \cdot 3/8 + 3/7 \cdot 5/8 = 21/56 = 3/8. \end{aligned}$$

2.6.4. Формула Байєса

Формула Байєса встановлює зв'язок між двома умовними ймовірностями зі зворотним порядком подій між $P(A/B)$ та $P(B/A)$. Вона має два варіанти.

1) Основний варіант:

$$P(A/B) = \frac{P(B/A) \cdot P(A)}{P(B)}, \quad (2.14)$$

коли $P(A) \neq 0$, $P(B) \neq 0$.

2) В умовах застосування ФПЙ:

$$P(H_i/B) = \frac{P(B/H_i) \cdot P(H_i)}{P(B/H_1) \cdot P(H_1) + \dots + P(B/H_n) \cdot P(H_n)}. \quad (2.15)$$

Доведення

Доведення основного варіанта впливає із загального варіанта теорема добутку та властивості комутативності подій ($A \cdot B = B \cdot A$), за якими

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

або

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

Якщо в цьому рівнянні поділити обидві частини на $P(A)$, $P(A) \neq 0$, то отримаємо необхідне твердження.

Другий варіант формули Байєса отримують з основного, коли $P(B)$ підраховують згідно із ФПЙ за повною групою подій H_i , $i = 1, \dots, n$, а як A беруть H_i , $A = H_i$.

Приклад 2.15

У коробці лежать 3 червоні та 5 білих кульок. Навмання виймають одну кульку і відкладають убік. Потім виймають ще одну і вона виявляється червоною. Знайти ймовірність того, що відкладена кулька має білий колір.

Розв'язання

Коли ми виймаємо першу кульку, то можливими наслідками є два: {вийняли червону кульку} та {вийняли білу кульку}. Ці події утворюють повну групу подій (можна вибрати тільки або червону, або білу), тому для нашої задачі подіями H_1 та H_2 будуть $H_1 = \{\text{вийняли червону кульку}\}$, $H_2 = \{\text{вийняли білу кульку}\}$. Ураховуючи те, що першу кульку вийняли навмання, можемо говорити про однакову ймовірність вийняти кожен з кульок – $1/8$. Таким чином $P(H_1) = 3/8$ та $P(H_2) = 5/8$, тобто вийняти червону кульку ми можемо трьома варіантами з восьми (усього 8 кульок у коробці), а білу – п'ятьма варіантами з восьми.

На відміну від прикладу 2.14, у нас є ще надлишкова інформація про колір кульки, яку вийняли другою, – ця кулька червоного кольору. Очевидно, що колір першої кульки впливає на ймовірність появи кольору другої, як і навпаки, що відображається формулою Байєса (2.15):

$$P(H_2/B) = \frac{P(B/H_2) \cdot P(H_2)}{P(B/H_1) \cdot P(H_1) + P(B/H_2) \cdot P(H_2)},$$

де подія $B = \{\text{друга кулька є червоною}\}$. Тобто нам потрібно знайти не просто ймовірність події {перша кулька білого кольору}, а {перша кулька білого кольору, за умови, що друга – червоного}. За формулою повної ймовірності отримаємо

$P(B) = P(B/H_1) \cdot P(H_1) + P(B/H_2) \cdot P(H_2) = 2/7 \cdot 3/8 + 3/7 \cdot 5/8 = 21/56$ (див. приклад 2.14). Отже, підставляючи у формулу Байєса всі потрібні значення ймовірностей, дістаємо

$$\begin{aligned} P(H_2/B) &= \frac{P(B/H_2) \cdot P(H_2)}{P(B/H_1) \cdot P(H_1) + P(B/H_2) \cdot P(H_2)} = \\ &= \frac{3/7 \cdot 5/8}{2/7 \cdot 3/8 + 3/7 \cdot 5/8} = \frac{5}{7}. \end{aligned}$$

Приклад 2.16

По мішені стріляють 2 мисливців з ймовірністю попадання 0,6 та 0,7 відповідно. Після пострілу виявилось, що в мішень потрапила тільки одна куля. Знайти ймовірність, що це була куля першого мисливця.

Розв'язання

Спочатку сформулюємо події, ймовірності яких ми шукатимемо. Подія $A = \{\text{перший потрапив у мішень}\}$, подія $\bar{A} = \{\text{перший}$

не потрапив у мішень}, $B = \{\text{у мішень потрапила лише одна куля}\}$. Імовірність події A знаходимо з умови задачі $P(A) = 0,6$. Подія $B/A = \{\text{у мішень потрапила тільки одна куля за умови, що перший потрапив у мішень}\}$ збігається з подією $\{\text{другий не потрапив у мішень}\}$. Уведемо ще одну подію $C = \{\text{другий потрапив у мішень}\}$. З умови маємо $P(C) = 0,7$, а отже $P(\bar{C}) = 1 - P(C) = 1 - 0,7 = 0,3$; $P(B/C) = P\{\text{у мішень потрапила тільки одна куля за умови, що другий потрапив у мішень}\} = P\{\text{перший не потрапив у мішень}\} = P(\bar{A})$, тобто

$$P(B/C) = P(\bar{A}) = 1 - P(A) = 1 - 0,6 = 0,4.$$

Для розв'язку задачі нам потрібно знайти ймовірність події $A/B = \{\text{перший потрапив у мішень за умови, що в мішень потрапила лише одна куля}\}$.

Отже, скористаємося формулою Байєса

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

Для цього знайдемо ймовірність $P(B) = P\{\text{у мішень потрапила тільки одна куля}\}$, яку обчислюємо за формулою повної ймовірності. Подію B можемо сформулювати як $\{\text{перший попав, а другий не попав або перший не попав, а другий попав}\}$. Формально отримаємо

$$\begin{aligned} P(B) &= P(A)P(B/A) + P(\bar{A})P(B/\bar{A}) = \\ &= P(B/C)P(C) + P(B/\bar{C})P(\bar{C}) = P(A)P(\bar{C}) + P(\bar{A})P(C) = \\ &= 0,6 \cdot 0,3 + 0,4 \cdot 0,7 = 0,46. \end{aligned}$$

Підставляючи відповідні ймовірності до ФПЙ, маємо

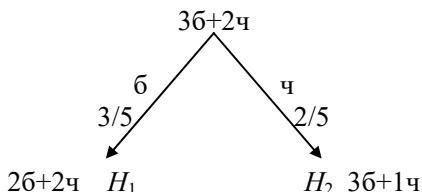
$$P(A/B) = \frac{P(B/A) \cdot P(A)}{P(B)} = \frac{0,3 \cdot 0,6}{0,46} \approx 0,39.$$

2.6.5. Використання графів для розв'язування задач на ФПЙ та формулу Байєса

Одним із методів візуалізації розв'язування певних класів імовірнісних задач є використання графового представлення самих кроків та схеми розв'язування, тобто побудова ймовірнісного графа.

Приклад 2.17

Нехай в урні лежить 5 кульок, із них 3 білі і 2 чорні. Усі можливі наслідки ($2б+2ч$ та $3б+1ч$) того, що ми навмання вийняли 1 кульку та їх імовірності тоді можна записати у вигляді графічного представлення



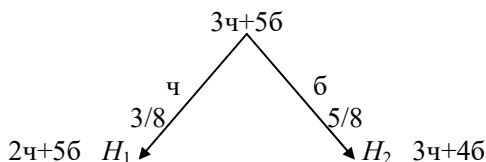
Тут у вершинах вказано подію та стан системи після того, як ця подія відбулась, а на ребрах – імовірність (а також іноді опис події).

Приклад 2.18

У коробці лежить 3 червоні та 5 білих кульок. Навмання виймають одну кульку і відкладають убік. Потім виймають ще одну. Знайти ймовірність того, що вона червоного кольору. Ця умова збігається з умовою прикладу 2.14, тільки розв'язок наводиться з використанням графів.

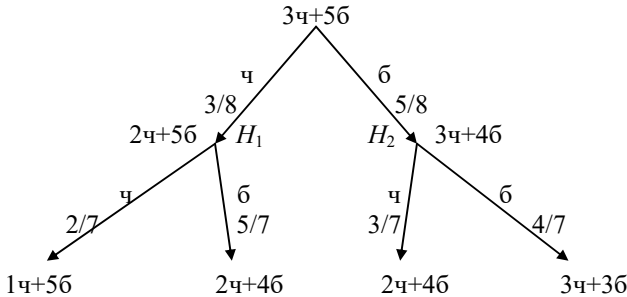
Розв'язання

Розв'яжемо цю задачу з використанням графів. Маємо систему $3ч+5б$. Із неї навмання виймемо 1 кульку, яка або біла з імовірністю $5/8$, або червона з імовірністю $3/8$. Залежно від кольору витягнутої кульки система матиме вигляд $2ч+5б$, якщо вийняли червону кульку, або $3ч+4б$, якщо витягнули білу. На графі це виглядає так:



Тут H_1 та H_2 – події вийняти червону та білу кульку першого разу відповідно. Ці дві події утворюють повну групу подій.

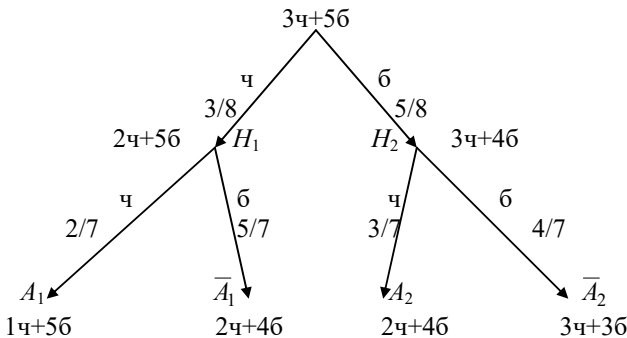
Далі виймають ще одну кульку. Із відповідними ймовірностями це можна зобразити на графі як



Отже, на першому кроці $3/8$ та $5/8$ є ймовірностями вийняти червону та білу кульку відповідно. Дійсно, маючи 3 червоні кульки з 8, отримаємо, за принципом про рівну можливість результатів, таке: імовірність вийняти червону кульку є 3 з 8 тобто $3/8$.

Таким чином, можемо сказати, що ймовірності проміжних подій H_1 та H_2 будуть $3/8$ та $5/8$ відповідно. Кожна з цих подій міняє стан системи H_1 : після того, як із початкової системи видалили одну червону кульку, у системі залишилось 2 червоні та 5 білих; після H_2 – залишилося 3 червоні та 4 білі.

За умовою задачі потрібно знайти ймовірність події $A = \{\text{друга кулька червоного кольору}\}$. Уведемо допоміжні події $A_1 = \{\text{вийняти другу кульку червону після того, як першою вийняли червону}\}$, $\bar{A}_1 = \{\text{не вийняти другу кульку червону після того, як першою вийняли червону, тобто вийняти білу після червоної}\}$, $A_2 = \{\text{вийняти другу кульку червону після того, як першою вийняли білу}\}$ та $\bar{A}_2 = \{\text{вийняти другу кульку білу після того, як першою вийняли білу}\}$ – усі можливі наслідки виймання другої кульки. Зобразимо розкритий імовірнісний граф:

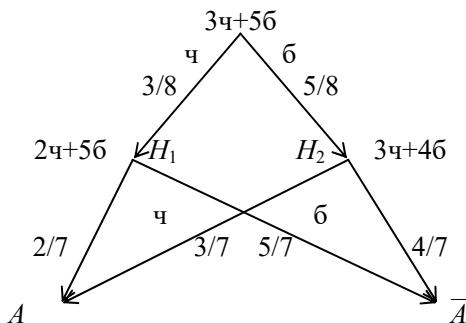


Із визначення цих множин легко бачимо, що події $A_1=A/H_1$, $A_2=A/H_2$, $\bar{A}_1=\bar{A}/H_1$ та $\bar{A}_2=\bar{A}/H_2$. Отже, імовірності дорівнюватимуть $P(A_1)=P(A/H_1)=2/7$, $P(\bar{A}_1)=P(\bar{A}/H_1)=5/7$, $P(A_2)=P(A/H_2)=3/7$ та $P(\bar{A}_2)=P(\bar{A}/H_2)=4/7$.

Для того, щоб досягнути події A , є два ймовірнісні маршрути: $H_1 \rightarrow A$ ($H_1 \rightarrow A_1$) та $H_2 \rightarrow A$ ($H_2 \rightarrow A_2$). Ці два маршрути незалежні, тобто не збігаються, тому ймовірність події A дорівнює сумі ймовірностей кожного з цих маршрутів.

Обчислимо тепер імовірності самих маршрутів. Імовірність того, що в нас відбулася подія A , коли ми рухались через подію H_1 , дорівнює $P(A/H_1)P(H_1)$, тобто, добутку ймовірностей ребер. Аналогічно знаходимо всі інші маршрути.

Далі згрупуємо маршрути графа, які приводять до події A . Одержимо ймовірнісний граф:



Тобто, за теоремою додавання та формулою повної ймовірності отримаємо $P(A) = P(A/H_1)P(H_1) + P(A/H_2)P(H_2) = \frac{2}{7} \cdot \frac{3}{8} + \frac{3}{7} \cdot \frac{5}{8} = \frac{3}{8}$.

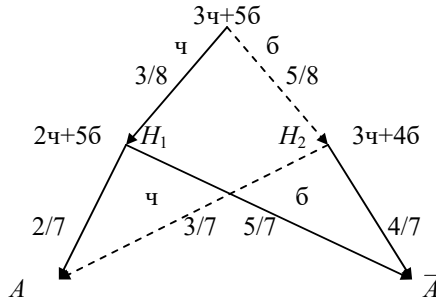
З іншого боку, за формулою повної ймовірності теж маємо $P(A) = P(A/H_1)P(H_1) + P(A/H_2)P(H_2)$. Отже, ймовірнісні задачі можна розв'язувати з використанням графів.

Приклад 2.19

У коробці лежить 3 червоні та 5 білих кульок. Навмання виймають одну кульку і відкладають убік. Потім виймають ще одну і вона виявляється червоного кольору. Знайти ймовірність того, що відкладена кулька має білий колір.

Розв'язання

Із попередньої задачі маємо граф



Тут $A = \{\text{друга куля є червоною}\}$, $H_1 = \{\text{вийняли червону кулю}\}$, $H_2 = \{\text{вийняли білу кулю}\}$, $P(H_1) = 3/8$ та $P(H_2) = 5/8$, тобто вийняти червону кулю ми можемо, використавши три варіанти з восьми (всього 8 кулек у коробці), а білу, застосувавши п'ять варіантів із восьми.

Очевидно, що колір першої кульки впливає на ймовірність появи кольору другої, як і навпаки, що відображається формулою Байєса (2.15):

$$P(H_2 / A) = \frac{P(A / H_2) \cdot P(H_2)}{P(A / H_1) \cdot P(H_1) + P(A / H_2) \cdot P(H_2)},$$

тобто нам потрібно знайти не просто ймовірність події {перша куля білого кольору}, а {перша куля білого кольору, за умови, що друга – червоного}.

Із прикладу 2.18 маємо $P(A) = P(A / H_1)P(H_1) + P(A / H_2)P(H_2) = \frac{2}{7} \cdot \frac{3}{8} + \frac{3}{7} \cdot \frac{5}{8} = \frac{3}{8}$, тому, підставляючи у формулу Байєса всі потрібні значення ймовірностей, отримаємо

$$\begin{aligned} P(H_2 / A) &= \frac{P(A / H_2) \cdot P(H_2)}{P(A / H_1) \cdot P(H_1) + P(A / H_2) \cdot P(H_2)} = \\ &= \frac{3/7 \cdot 5/8}{2/7 \cdot 3/8 + 3/7 \cdot 5/8} = \frac{5}{7}. \end{aligned}$$

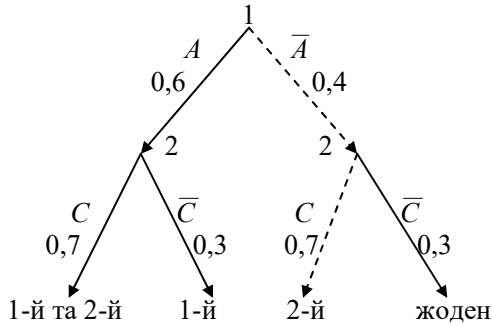
Приклад 2.20

По мішені стріляють 2 мисливців зі ймовірністю попадання 0,6 та 0,7 відповідно. Після пострілу виявилось, що в мішень

потрапила тільки одна куля. Знайти ймовірність, що це була куля першого мисливця.

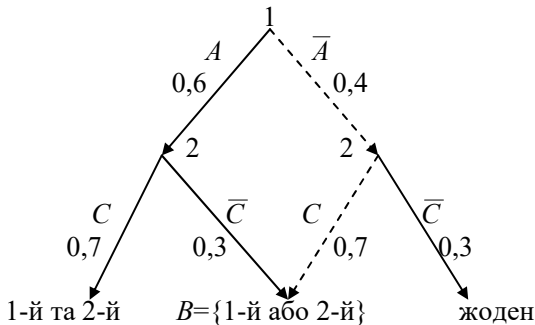
Розв'язання

Визначимо події так, як це робилось у прикладі 2.16. Вважатимемо, що стрілки стрілятимуть один за одним, тоді послідовність подій можна інтерпретувати графом вигляду



тут $A = \{\text{перший потрапив у мішень}\}$, подія $\bar{A} = \{\text{перший не потрапив у мішень}\}$, $C = \{\text{другий потрапив у мішень}\}$ та $\bar{C} = \{\text{другий не потрапив у мішень}\}$. Усі вказані ймовірності отримуємо так само, як у прикладі 2.16. Нам потрібно знайти ймовірність події $\{\text{перший потрапив у мішень при ймовірності, що в мішень потрапила лише одна куля}\}$. Уведемо ще одну подію $B = \{\text{у мішень потрапила тільки одна куля}\}$, тоді шуканою буде ймовірність події A/B .

Перепишемо граф у вигляді



Для використання формули Байєса спочатку потрібно знайти ймовірність події B . Із графа видно, що досягнути подію B можна двома шляхами $A \rightarrow \bar{C}$ або $\bar{A} \rightarrow C$, тому

$$P(B) = 0,6 \cdot 0,3 + 0,4 \cdot 0,7 = 0,46.$$

Ймовірність $P(B/C) = P\{\text{у мішень потрапила тільки одна куля за умови, що другий потрапив у мішень}\} = P\{\text{перший не потрапив у мішень}\} = P(\bar{A})$, відповідно $P(B/A) = \bar{C}$. Отже, за формулою Байєса маємо

$$P(A/B) = \frac{P(B/A) \cdot P(A)}{P(B)} = \frac{0,3 \cdot 0,6}{0,46} \approx 0,39.$$

Значимо, що в задачі не є суттєвим, який стрілок стріляє першим – другий чи перший.

Зауваження

При розв'язуванні задач на використання формули повної ймовірності за допомогою графів шукана ймовірність події A дорівнює сумі ймовірностей усіх різних маршрутів, що приводять до цієї події.

При розв'язуванні задач на використання формули Байєса за допомогою графів шукана ймовірність події B/A дорівнює відношенню суми ймовірностей, що приводять до події A і проходять через подію B , до суми ймовірностей всіх маршрутів, що приводять до події A .

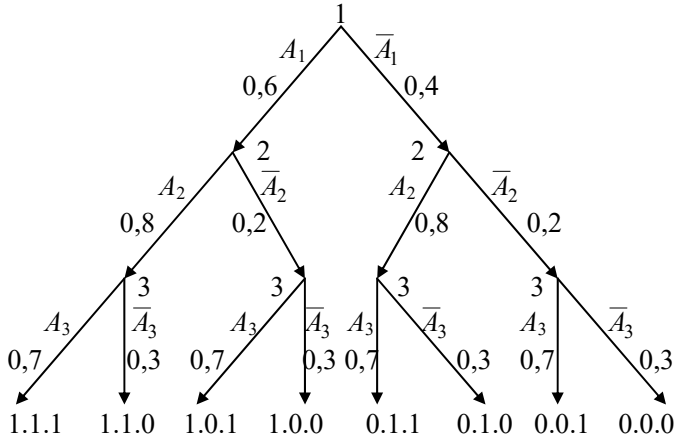
Дійсно, у прикладі 2.20 шукана ймовірність буде відношенням імовірності маршруту $A \rightarrow \bar{C}$ (0,18) до суми ймовірностей усіх маршрутів, що приводять до події B , тобто $A \rightarrow \bar{C}$ (0,18) та $\bar{A} \rightarrow C$ (0,28).

Приклад 2.21

По мішені стріляють 3 мисливців з імовірністю попадання 0,6, 0,8 та 0,7 відповідно. Після пострілу виявилось, що в мішень потрапила тільки одна куля. Знайти ймовірність того, що це була куля першого мисливця.

Розв'язання

Побудуємо граф, який повністю описує задачу, вважаючи, що мисливці стріляють незалежно один за одним (ми вже зазначали, що порядок виконання пострілів не важливий):



Тут запис 1.0.1 означає, що 1-й попав, 2-й не попав, 3-й попав у мішень. Отже, маємо 8 різних маршрутів, з яких 3 приводять до події $B = \{\text{потрапив тільки один}\}$ – це 1.0.0, 0.1.0 та 0.0.1, із них лише один маршрут проходить через подію A_1 – 1.0.0. Запишемо ймовірність події B : $P(B) = 0,6 \cdot 0,2 \cdot 0,3 + 0,4 \cdot 0,8 \cdot 0,3 + 0,4 \cdot 0,2 \cdot 0,7$ як суму ймовірностей маршрутів 1.0.0 ($P(1.0.0) = 0,6 \cdot 0,2 \cdot 0,3$), 0.1.0 ($P(0.1.0) = 0,4 \cdot 0,8 \cdot 0,3$) та 0.0.1 ($P(0.0.1) = 0,4 \cdot 0,2 \cdot 0,7$), а сама шукана умовна ймовірність $P(A_1 / B)$ є така:

$$\frac{0,6 \cdot 0,2 \cdot 0,3}{0,6 \cdot 0,2 \cdot 0,3 + 0,4 \cdot 0,8 \cdot 0,3 + 0,4 \cdot 0,2 \cdot 0,7} \approx 0,19.$$

2.6.6. Дерево прийняття рішень

Наведений у пункті 2.6.5 метод розв'язування ймовірнісних задач можна використати для побудови дерева прийняття рішень. Такі задачі виникають тоді, коли за теоретичними розрахунками ймовірностей потрібно прийняти оптимальне рішення. Класичним прикладом такої задачі є задача про вази (або про урни).

Отже, нехай є 700 ваз 1-го типу і 300 ваз – другого. Випадковим чином обирається ваза і ставиться перед піддослідним. Якщо перед ним стоїть ваза 1-го типу і він вірно вгадає це, то

отримає 350 грн, а якщо не вгадає, то програє 50 грн. Якщо ж перед ним стоїть ваза 2-го типу і він вгадає це, то заробить 500 грн, а якщо не вгадає, то програє 100.

Умову цієї задачі можна записати в табл. 2.1.

Таблиця 2.1

Дерево прийняття рішень

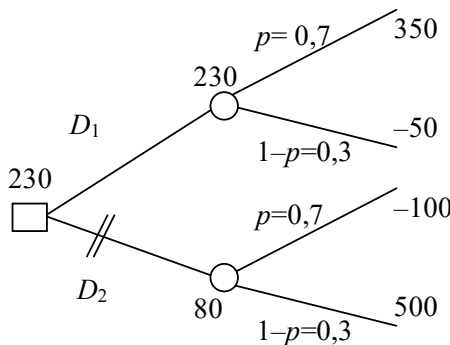
Тип вази	Імовірність вибору вази даного типу	Дії та виграші	
		D_1 (обрана ваза 1-го типу)	D_2 (обрана ваза 2-го типу)
1	0,7	350	-100
2	0,3	-50	500

За цією таблицею обчислюють імовірності стратегій та виграшів (програшів). Позначимо як U функцію видатків, тоді видатки кожної зі стратегій (сказати, що ваза 1-го типу, чи сказати, що ваза 2-го типу) можна записати так:

$$U(D_1) = 0,7 \cdot 350 + 0,3 \cdot (-50) = 230 \text{ грн,}$$

$$U(D_2) = 0,3 \cdot 500 + 0,7 \cdot (-100) = 80 \text{ грн.}$$

Отже, більш вигідною є стратегія говорити, що поряд із підослідним стоїть ваза 1-го типу, тобто, обрати вію (стратегію) D_1 . Така стратегія носить назву стратегії раціонального вибору, її можна записати у вигляді графа-дерева рішень



Тут кружками позначено моменти, коли рішення не приймають, а прямокутником – прийняття рішення. Числа біля кружків

(прямокутників) означають обсяг виграшу або програшу, написи D_1 та D_2 – різні стратегії, значення біля променів – імовірності, подвійне закреслення променя свідчить про відмову від цього варіанта рішення.

Очевидно, що – це найелементарніше дерево прийняття рішень. Розглянемо складнішу задачу про вази, яка дозволить більш детально розглянути сам принцип побудови дерева.

Нехай у вазі 1-го типу лежить 6 червоних та 4 білі кульки, а у вазі 2-го типу – 3 червоні та 7 білих. Нехай піддослідний, отримавши вазу, може за 60 грн вийняти одну кульку з вази і, на основі умовних імовірностей визначити тип, до якого належить ваза. Тобто, по-перше, потрібно визначити чи взагалі варто виймати кульку, витрачаючи на це 60 грн, і по-друге, як, у випадку, коли кульку вийняли, оцінити тип вази.

У цій задачі, для запису дерева прийняття рішень, потрібно обчислити низку умовних імовірностей та скористатись формулою Байєса

$$\begin{aligned} P(D_1) &= 0,7, & P(D_2) &= 0,3, \\ P(\text{ч}/D_1) &= 0,6, & P(\text{б}/D_1) &= 0,4, \\ P(\text{ч}/D_2) &= 0,3, & P(\text{б}/D_2) &= 0,7. \end{aligned}$$

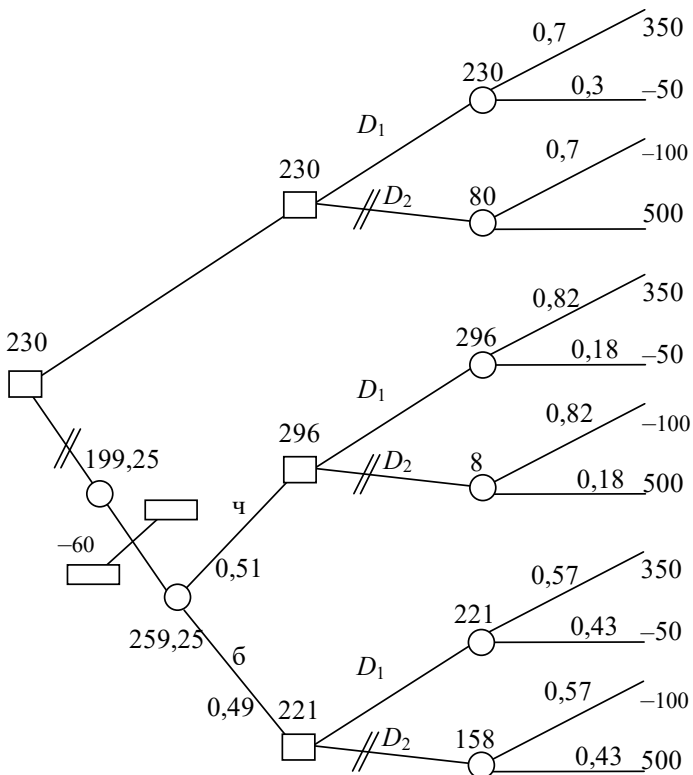
За формулою повної ймовірності

$$\begin{aligned} P(\text{ч}) &= P(\text{ч}/D_1)P(D_1) + P(\text{ч}/D_2)P(D_2) = \\ &= 0,6 \cdot 0,7 + 0,3 \cdot 0,3 = 0,51, \\ P(\text{б}) &= 0,49. \end{aligned}$$

За формулою Байєса

$$\begin{aligned} P(D_1 / \text{ч}) &= \frac{P(\text{ч} / D_1)P(D_1)}{P(\text{ч})} = \frac{0,6 \cdot 0,7}{0,51} \approx 0,82, \\ P(D_2 / \text{ч}) &= \frac{P(\text{ч} / D_2)P(D_2)}{P(\text{ч})} = \frac{0,3 \cdot 0,3}{0,51} \approx 0,18, \\ P(D_1 / \text{б}) &= \frac{P(\text{б} / D_1)P(D_1)}{P(\text{б})} = \frac{0,4 \cdot 0,7}{0,49} \approx 0,57, \\ P(D_2 / \text{б}) &= \frac{P(\text{б} / D_2)P(D_2)}{P(\text{б})} = \frac{0,7 \cdot 0,3}{0,49} \approx 0,43. \end{aligned}$$

Зобразимо дерево прийняття рішень.



Верхня частина (верхня третина графа) є рішенням без виймання кульки і відповідно без утрати 60 грн, решта – 2/3 ілюструють випадки, коли була вийнята червона (ч) та біла (б) кулька відповідно. Як виявилось, найкращим варіантом є такий: не виймати кульку і відповідно не втрачати 60 грн. (230 грн проти $259,25 - 60 = 199,25$.)

2.6.7. Обчислення ряду розподілу функції від незалежних дискретних випадкових величин

Метод обчислення ряду розподілу функції від кількох незалежних випадкових величин розглянемо на прикладі.

Приклад 2.22

Нехай випадкові величини X , Y і Z незалежні та мають такі ряди розподілів імовірностей:

X	-1	0	2		Y	0	1	3		Z	0	1
P	0,2	0,5	0,3		P	0,3	0,4	0,3		P	0,2	0,8

Побудувати ряд розподілу випадкової величини $W = \min\{X^2, YZ, 2\}$ та знайти її математичне сподівання.

Розв'язання

Запишемо всі результати проміжних підрахунків у таблицю, побудовану таким чином: у перших шести стовпчиках випишемо всі можливі комбінації значень випадкових величин X , Y та Z і відповідні ймовірності. Очевидно, що коли X може набувати трьох різних значень, $Y - 3$ та $Z - 2$, то в таблиці має бути $3 \cdot 3 \cdot 2 = 18$ рядків.

X	P_X	Y	P_Y	Z	P_Z							
-1	0,2	0	0,3	0	0,2							
0	0,5	0	0,3	0	0,2							
2	0,3	0	0,3	0	0,2							
-1	0,2	1	0,4	0	0,2							
0	0,5	1	0,4	0	0,2							
2	0,3	1	0,4	0	0,2							
-1	0,2	3	0,3	0	0,2							
0	0,5	3	0,3	0	0,2							
2	0,3	3	0,3	0	0,2							
-1	0,2	0	0,3	1	0,8							
0	0,5	0	0,3	1	0,8							
2	0,3	0	0,3	1	0,8							
-1	0,2	1	0,4	1	0,8							
0	0,5	1	0,4	1	0,8							
2	0,3	1	0,4	1	0,8							
-1	0,2	3	0,3	1	0,8							
0	0,5	3	0,3	1	0,8							
2	0,3	3	0,3	1	0,8							

У наступних двох стовпчиках запишемо значення випадкових величин X^2 та YZ :

X	P_X	Y	P_Y	Z	P_Z	X^2	YZ			
-1	0,2	0	0,3	0	0,2	1	0			
0	0,5	0	0,3	0	0,2	0	0			
2	0,3	0	0,3	0	0,2	4	0			
-1	0,2	1	0,4	0	0,2	1	0			
0	0,5	1	0,4	0	0,2	0	0			
2	0,3	1	0,4	0	0,2	4	0			
-1	0,2	3	0,3	0	0,2	1	0			
0	0,5	3	0,3	0	0,2	0	0			
2	0,3	3	0,3	0	0,2	4	0			
-1	0,2	0	0,3	1	0,8	1	0			
0	0,5	0	0,3	1	0,8	0	0			
2	0,3	0	0,3	1	0,8	4	0			
-1	0,2	1	0,4	1	0,8	1	1			
0	0,5	1	0,4	1	0,8	0	1			
2	0,3	1	0,4	1	0,8	4	1			
-1	0,2	3	0,3	1	0,8	1	3			
0	0,5	3	0,3	1	0,8	0	3			
2	0,3	3	0,3	1	0,8	4	3			

Ураховуючи те, що нам потрібно знайти $\min\{X^2, YZ, 2\}$, для зручності наступним стовпчиком виписуємо числа "2", а далі – найменші значення з X^2 , YZ та числа 2, тобто $W = \min\{X^2, YZ, 2\}$:

X	P_X	Y	P_Y	Z	P_Z	X^2	YZ		W	
-1	0,2	0	0,3	0	0,2	1	0	2	0	
0	0,5	0	0,3	0	0,2	0	0	2	0	
2	0,3	0	0,3	0	0,2	4	0	2	0	
-1	0,2	1	0,4	0	0,2	1	0	2	0	
0	0,5	1	0,4	0	0,2	0	0	2	0	
2	0,3	1	0,4	0	0,2	4	0	2	0	

X	P_X	Y	P_Y	Z	P_Z	X^2	YZ		W	
-1	0,2	3	0,3	0	0,2	1	0	2	0	
0	0,5	3	0,3	0	0,2	0	0	2	0	
2	0,3	3	0,3	0	0,2	4	0	2	0	
-1	0,2	0	0,3	1	0,8	1	0	2	0	
0	0,5	0	0,3	1	0,8	0	0	2	0	
2	0,3	0	0,3	1	0,8	4	0	2	0	
-1	0,2	1	0,4	1	0,8	1	1	2	1	
0	0,5	1	0,4	1	0,8	0	1	2	0	
2	0,3	1	0,4	1	0,8	4	1	2	1	
-1	0,2	3	0,3	1	0,8	1	3	2	1	
0	0,5	3	0,3	1	0,8	0	3	2	0	
2	0,3	3	0,3	1	0,8	4	3	2	2	

Далі потрібно вписати ймовірності відповідних значень випадкової величини W . Виходячи з того, що X , Y та Z є незалежними випадковими величинами, ймовірність того, що випадкова величина набула значення "0" у першому рядку таблиці дорівнює ймовірності того, що одночасно випадкові величини X , Y та Z набувають значень відповідно $X=-1$, $Y=0$ та $Z=0$, тобто $P\{X=-1, Y=0, Z=0\}$ або з використанням теореми добутку

$$\begin{aligned}
 P\{X=-1, Y=0, Z=0\} &= \\
 &= P\{X=-1\}P\{Y=0\}P\{Z=0\}=0,2\cdot0,3\cdot0,2=0,012.
 \end{aligned}$$

Так само обчислюємо ймовірності для всіх значень W :

X	P_X	Y	P_Y	Z	P_Z	X^2	YZ		W	P_W
-1	0,2	0	0,3	0	0,2	1	0	2	0	0,012
0	0,5	0	0,3	0	0,2	0	0	2	0	0,03
2	0,3	0	0,3	0	0,2	4	0	2	0	0,018
-1	0,2	1	0,4	0	0,2	1	0	2	0	0,016
0	0,5	1	0,4	0	0,2	0	0	2	0	0,04
2	0,3	1	0,4	0	0,2	4	0	2	0	0,024
-1	0,2	3	0,3	0	0,2	1	0	2	0	0,012
0	0,5	3	0,3	0	0,2	0	0	2	0	0,03
2	0,3	3	0,3	0	0,2	4	0	2	0	0,018

X	P_X	Y	P_Y	Z	P_Z	X^2	YZ		W	P_W
-1	0,2	0	0,3	1	0,8	1	0	2	0	0,048
0	0,5	0	0,3	1	0,8	0	0	2	0	0,12
2	0,3	0	0,3	1	0,8	4	0	2	0	0,072
-1	0,2	1	0,4	1	0,8	1	1	2	1	0,064
0	0,5	1	0,4	1	0,8	0	1	2	0	0,16
2	0,3	1	0,4	1	0,8	4	1	2	1	0,096
-1	0,2	3	0,3	1	0,8	1	3	2	1	0,048
0	0,5	3	0,3	1	0,8	0	3	2	0	0,12
2	0,3	3	0,3	1	0,8	4	3	2	2	0,072

Отже, для випадкової величини $W = \min\{X^2, YZ, 2\}$ отримано ряд розподілу:

W	0	0	0	0	0	0	0	0	0	0	0
P_W	0,012	0,03	0,018	0,016	0,04	0,024	0,012	0,03	0,018	0,048	0,12

W	0	1	0	1	1	0	2
P_W	0,072	0,064	0,16	0,096	0,048	0,12	0,072

Із цієї таблиці видно, що випадкова величина W може набувати лише трьох різних значень – 0, 1 та 2; причому значення 0 можна отримати 14 шляхами, 1 – трьома шляхами та 2 – тільки одним шляхом. Зведемо цей ряд до стандартного вигляду, об'єднуючи клітинки з однаковими значеннями випадкової величини. Використовуючи теорему додавання, бачимо, що відповідні ймовірності при цьому додаються (через несумісність подій виду $\{W=0\}=\{X=-1, Y=0, Z=0\}$, $\{W=0\}=\{X=0, Y=0, Z=0\}$, $\{W=0\}=\{X=2, Y=0, Z=0\}$ і т. д. для всіх значень W).

Отже, ряд розподілу випадкової величини $W = \min\{X^2, YZ, 2\}$ матиме такий вигляд:

W	0	1	2
P_W	0,72	0,208	0,072

Звідси математичне сподівання

$$MW = 0 \cdot 0,72 + 1 \cdot 0,208 + 2 \cdot 0,072 = 0,352.$$

ЗАДАЧІ

Обчислення ряду розподілу функції від незалежних дискретних випадкових величин

№ 1

Нехай випадкові величини X та Y незалежні та мають такі ряди розподілів імовірностей:

X	1	2	3	4		Y	1	2	3
P	0,2	?	0,3	0,4		P	0,3	?	0,3

Доповнити ці ряди розподілів, замінивши знаки питання потрібними значеннями. Скласти ряд розподілу випадкової величини $X+Y$ та знайти її математичне сподівання.

№ 2

Випадкова величина X має такий ряд розподілу імовірностей:

X	0	2	4	7
P	0,1	0,2	0,3	0,4

Скласти ряд розподілу імовірностей випадкової величини $Y = \max\{X, 3\}$ та знайти її математичне сподівання.

№ 3

Нехай випадкові величини X та Y незалежні та мають такі ряди розподілів імовірностей:

$X:U$	0	1	2	3				$Y:U$	1	3	4
P	0,2	0,3	0,4	0,1				P	0,7	0,2	0,1

Побудувати ряди розподілу випадкових величин $\alpha = \min\{X, Y\}$, $\beta = \max\{X, Y\}$.

ЗАДАЧІ

ОЗТЙ: теореми заперечення, суми, добутку, умовна ймовірність

Приклад 2.23

В урні лежать 5 червоних та 4 білих кульки. З урни виймають послідовно 2 кульки. Знайти ймовірність того, що друга кулька червоного кольору.

Розв'язання

Позначимо подію $A = \{\text{перша кулька – червона}\}$, $B = \{\text{друга кулька червона}\}$. Тоді шукана подія $C = \{\text{друга кулька червоного кольору}\}$ складатиметься з двох несумісних подій $A(B/A)$ та $\bar{A}(B/\bar{A})$. Усього є 9 кульок у коробці, тому, ймовірність вийняти першою червону кульку, очевидно, складає $5/9$, а білу – $4/9$, тобто $P\{A\}=5/9$, а $P\{\bar{A}\}=4/9$. Після того, як вийняли першу кульку, у коробці залишилось 8 кульок, причому якщо ми вийняли першою червону, то в коробці залишилось 4 червоні та 4 білі кульки, а якщо першою виявилась біла кулька, – то 5 червоних та 3 білі, тобто ймовірності набудуть значень $P\{B/A\}=4/8$ та $P\{B/\bar{A}\}=5/8$. Виходячи з теореми суми, отримаємо

$$P\{B\} = P\{A\}P\{B/A\} + P\{\bar{A}\}P\{B/\bar{A}\} = 5/9 \cdot 4/8 + 4/9 \cdot 5/8 = 5/9.$$

№ 1

Для руйнування моста треба, щоб у нього влучили принаймні 2 авіабомби. Незалежно скидають 3 бомби з імовірностями влучення 0,1, 0,3 та 0,4. Яка ймовірність того, що міст буде зруйновано?

№ 2

В урні містяться 3 червоні, 4 сині та 5 жовтих кульок. Навмання беруть три кульки. Обчислити ймовірність таких випадкових подій: A – три кульки виявляться однакового кольору; B – три кульки виявляться різного кольору.

№ 3

В академічній групі 25 студентів. Знайти ймовірність того, що хоча б у двох із них збігаються дні народження.

№ 4

На іспит виносять 60 теоретичних питань. Студент знає відповіді на 30 із них. Екзаменатор задає студенту 6 питань, а для того, щоб скласти іспит, треба відповісти не менш ніж на 3 питання. Яка ймовірність того, що студент складе іспит?

№ 5

Учасник лотереї "Спортлото" із 49 чисел повинен закреслити 6. Повний виграш отримає той, хто правильно вкаже всі 6 чисел.

Виграш одержать і ті, хто вгадає не менше трьох чисел. Обчислити ймовірність повного виграшу у "Спортлото". Яка ймовірність одержати який-небудь виграш?

№ 6

В урні a білих та b чорних куль. Два гравці по черзі беруть кулі з урни, повертаючи взяту кулю в урну. Виграє той, хто першим витягне білу кулю. Знайти ймовірність виграшу для кожного гравця.

№ 7

Залізничний потяг має n вагонів, кожен з яких з ймовірністю P має дефект. Усі вагони перевіряють, незалежно один від одного, два оглядача. Перший виявляє дефект (якщо він є) з імовірністю P_1 , а другий – з імовірністю P_2 . Якщо в жодному з вагонів не виявлено дефекту, потяг вирушає в рейс. Знайти ймовірність того, що в рейс вирушить потяг, в якому є хоча б один дефектний вагон.

№ 8

Із урни, яка містить 3 білі, 5 чорних та 2 червоні кулі, два гравці по черзі виймають по одній кулі без повернення. Виграє той, хто першим витягне білу кулю. Якщо з'явиться червона куля, то оголошують нічию. Знайти ймовірність того, що виграє перший гравець та ймовірність нічиєї.

№ 9

У майстерні працює три верстати. За зміну перший потребує налагодження з імовірністю 0,2; для другого та третього верстатів ці ймовірності відповідно дорівнюють 0,1 і 0,15. Обчислити ймовірності таких подій: A – три верстати потребують налагодження за зміну; B – два верстати; C – один верстат; D – жоден верстат не потребує налагодження за зміну; E – принаймні один верстат потребує налагодження.

№ 10

В урні містяться 3 червоні кульки, 4 синіх та 5 зелених. Кульки з урни виймають по одній без повернення до неї. Таке випробування проводиться до першої появи червоної кульки. Обчислити ймовірність того, що червона кулька з'явиться раніше за зелену.

№ 11

Для психологічного аналізу треба дослідити чотири результати незалежних тестів. Імовірність припуститися помилки при аналізі першого, другого, третього та четвертого тестів відповідно дорівнює 0,01; 0,05; 0,09; 0,08. Яка ймовірність того, що проведений аналіз виявиться вірним?

№ 12

В урні містяться 20 кульок, пронумерованих від 1 до 20. Навмання з неї беруть одну кульку. Яка ймовірність того, що номер, написаний на кульці, буде кратний трьом, якщо відомо, що він є непарне число?

№ 13

Два гравці кидають по одному кубіку. Виграє той, у кого випаде більше число. Яка ймовірність при цьому виграти першому гравцеві?

№ 14

Тричі підкидають монету. Описати простір елементарних подій, події A – двічі випав герб, B – принаймні один раз випав герб. Обчислити $P(A)$, $P(B)$, $P(A \cdot B)$, $P(A/B)$.

№ 15

Кидають три гральні кубики. Обчислити ймовірність того, що принаймні на одному з них випала одиниця, якщо відомо, що на трьох кубиках випали різні грані.

№ 16

В ящику лежить 12 білих, 8 чорних та 10 червоних куль. Навмання беруть дві кулі. Яка ймовірність того, що ці кулі різного кольору, якщо відомо, що червону кулю не вибрано?

№ 17

Підкидають два гральні кубики. Яка ймовірність того, що сума очок більша або дорівнює 9, якщо відомо, що на одному з кубиків випало 4 очки?

№ 18

Відомо, що 5 % чоловіків та 0,25 % усіх жінок – дальтоніки. Навмання вибрана особа – дальтонік. Яка ймовірність того, що це чоловік? (Вважати, що чоловіків та жінок однакова кількість.)

ЗАДАЧІ

ФПІ та формула Байєса

Приклад 2.24

В урні містяться 4 кульки, причому кожна кулька є білою або чорною. Усі можливі припущення про кількість білих кульок рівноймовірні. З урни витягають одночасно дві кульки. Яка ймовірність того, що ці дві кульки різного кольору (тобто, одна – біла, а інша – чорна)?

Розв'язання

Спочатку розглянемо таку допоміжну задачу: нехай у ящику n кульок, серед яких k білих. З ящика навмання витягають m кульок. Знайти ймовірність того, що серед них буде рівно l білих.

Оскільки всього існує $C_n^m = \frac{n!}{m!(n-m)!}$ різних способів витягти m кульок із n , а різних варіантів, коли l кульок у наборі виявляться білими, є $C_k^l C_{n-k}^{m-l}$ (тобто в наборі має бути l із k кульок білого кольору, що містяться в урні, а решта $m-l$ кульок у наборі з $n-k$ кульок у коробці мають бути чорними), то шукана ймовірність складе $\frac{C_k^l C_{n-k}^{m-l}}{C_n^m}$.

Тепер перейдемо до безпосереднього розв'язування нашої задачі. Позначимо події $H_i = \{\text{в урні було рівно } i \text{ білих кульок}\}$. За умовою всі H_i є рівноможливими, тобто $P(H_i) = 1/5$ для всіх $i \in \{0, 1, 2, 3, 4\}$. Через A позначимо подію $A = \{\text{витягнуто дві кульки різного кольору}\}$. Оскільки H_i , $i \in \{0, 1, 2, 3, 4\}$, утворюють повну групу подій, то за формулою повної ймовірності маємо

$$\begin{aligned} P(A) &= \sum_{i=0}^4 P(A | H_i) P(H_i) = \frac{1}{5} \sum_{i=0}^4 P(A | H_i) = \frac{1}{5} \left[0 + \sum_{i=1}^3 \frac{C_i^1 C_{4-i}^1}{C_4^2} + 0 \right] = \\ &= \frac{1}{5} \left[\frac{C_1^1 C_3^1}{C_4^2} + \frac{C_2^1 C_2^1}{C_4^2} + \frac{C_3^1 C_1^1}{C_4^2} \right] = \frac{1}{5} \left[\frac{3}{6} + \frac{4}{6} + \frac{3}{6} \right] = \frac{1}{5} \cdot \frac{10}{6} = \frac{1}{3}. \end{aligned}$$

Приклад 2.25

В урні лежать 5 червоних та 4 білі кульки. З урни виймають одну кульку і відкладають убік. Потім виймають ще одну і вона

виявляється білою. Знайти ймовірність того, що відкладена кулька мала червоний колір.

Розв'язання

У задачах такого типу важливо вірно визначити всі події, які беруть участь у задачі. Позначимо події таким чином: $A = \{\text{перша кулька – червона}\}$, $B = \{\text{друга кулька – біла}\}$ та $B/A = \{\text{друга кулька біла за умови, що перша кулька – червона}\}$. Шуканою є ймовірність події $A/B = \{\text{перша кулька – червона за умови, що друга кулька – біла}\}$.

Далі запишемо в таблицю всі ймовірності, які беруть участь у задачі:

A $P(A)=5/9$		\bar{A} $P(\bar{A})=4/9$	
B/A $P(B/A)=4/8$	\bar{B}/A $P(\bar{B}/A)=4/8$	B/\bar{A} $P(B/\bar{A})=3/8$	\bar{B}/\bar{A} $P(\bar{B}/\bar{A})=5/8$

Для того, щоб застосувати формулу Байєса (2.15), потрібно спочатку знайти ймовірність події B за формулою повної ймовірності (2.13).

Отже, за ФПЙ

$$P(B) = P(B/A)P(A) + P(B/\bar{A})P(\bar{A}) = 4/8 \cdot 5/9 + 3/8 \cdot 4/9 = 4/9.$$

Далі за формулою Байєса отримаємо

$$P(A/B) = \frac{P(B/A)P(A)}{P(B)} = \frac{4/8 \cdot 5/9}{4/9} = \frac{5}{8}.$$

№ 1

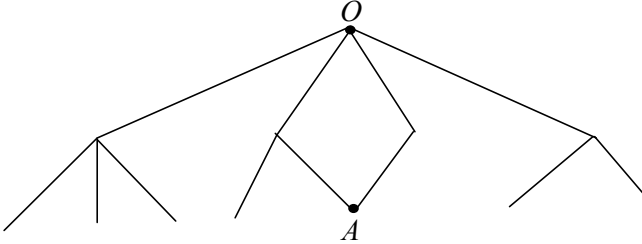
Група студентів, що складає іспит, містить 5 відмінників, 10 із середньою успішністю та 15 слабких. Відмінник завжди отримує оцінку 5, середній студент – 5 та 4 з рівними ймовірностями, слабкий студент – 4, 3 та 2 з рівними ймовірностями. Знайти ймовірність того, що навмання викликаний студент отримає оцінку 4.

№ 2

Серед N екзаменаційних білетів є n "щасливих". Студенти підходять за білетами один за одним. У кого більша ймовірність взяти "щасливий" білет: у того, хто йде першим, чи в того, хто йде другим?

№ 3

Мандрівник виходить з пункту O і на кожному перехресті обирає навмання один із можливих шляхів. Яка ймовірність того, що мандрівник потрапить до пункту A ?



№ 4

В ящику 3 білі та 7 чорних куль. Навмання вийняли одну кулю та відклали вбік. Наступна куля, яку вийняли навмання, виявилася білою. Яка ймовірність того, що відкладена куля була біла?

№ 5

У спеціалізовану лікарню надходять у середньому 40 % хворих на хворобу A , 35 % – на хворобу B , 25 % – на C . Імовірність повного видужання після хвороби A дорівнює 0,75, B – 0,8, C – 0,9. Хворий, прийнятий до лікарні, повністювилікувався. Яка ймовірність того, що він був носієм хвороби A ?

№ 6

При рентгенівському обстеженні ймовірність виявити захворювання на туберкульоз у хворого дорівнює $1-\beta$. Імовірність визнати здорову людину хворою дорівнює α . Нехай частка хворих на туберкульоз щодо всього населення дорівнює γ . Знайти ймовірність того, що людина здорова, якщо вона була визнана хворою при обстеженні.

№ 7

В автобусі їде n пасажирів. На наступній зупинці кожен із них виходить з ймовірністю p . Окрім того, в автобус з ймовірністю p_0 не входить жоден пасажир, а з ймовірністю $1-p_0$ входить один новий пасажир. Знайти ймовірність того, що, коли автобус знову поїде після наступної зупинки, у ньому кількість пасажирів не зміниться.

2.6.8. Схема незалежних випробувань Бернуллі

Задачу обчислення ймовірності появи події A в серії з n незалежних випробувань, при ймовірності появи події A в кожному з випробувань $p = P(A)$, називають **схемою незалежних випробувань Бернуллі** (СНВБ). Поява події A в серії з n випробувань означає реалізацію однієї з $n+1$ подій B_0, B_1, \dots, B_n , де B_k – поява A рівно k разів з n випробувань (тобто подія A відбулась у k випробуваннях і не відбулась у $n - k$ випробуваннях). Отже, зміст СНВБ можна передати схемою

$$p = P(A) \Rightarrow \begin{array}{l} P(B_0) \\ P(B_1) \\ \dots ? \\ P(B_n) \end{array}$$

Теорема 2.7

Розв'язок СНВБ знаходять за загальною формулою для цієї схеми

$$P(B_k) = C_n^k p^k (1-p)^{n-k} = C_n^k p^k q^{n-k}, \quad 0 \leq k \leq n, \quad q = 1 - p. \quad (2.16)$$

Доведення

Позначимо через $C_{i_1 i_2 \dots i_k} = \{ \text{подія } A \text{ відбулась у випробуваннях з послідовними номерами } i_1 < i_2 < \dots < i_k \}, k = 1, \dots, n$. Це означає, що у випробуваннях із рештою номерів подія A не відбувалася. Отже, подія $C_{i_1 i_2 \dots i_k}$ може бути представлена добутком подій $A_i = \{ \text{у } i\text{-му випробуванні подія } A \text{ відбулась} \}$ чи їхніх заперечень: A з номерами i_1, i_2, \dots, i_k входять до добутку без заперечень, решта – із запереченнями. Оскільки події A_1, A_2, \dots, A_n незалежні в сукупності (спадковість незалежності), то $P(C_{i_1 i_2 \dots i_k})$ дорівнюватиме добутку ймовірностей: $P(A) = p$ для компонентів без заперечень та $P(\bar{A}) = 1 - p = q$ для компонентів із запереченнями. Оскільки в записі добутку кожного з $C_{i_1 i_2 \dots i_k}$ рівно k літер A зустрічаються без заперечень і рівно $n - k$ – із запереченнями, то $P(C_{i_1 i_2 \dots i_k}) = p^k \cdot q^{n-k}, \quad k = 1, \dots, n, \quad P(C_0) = P(\bar{A}_1 \bar{A}_2 \dots \bar{A}_n) = P(\bar{A}_1) P(\bar{A}_2) \dots P(\bar{A}_n) = (1-p)^n = q^n$. $C_{i_1 i_2 \dots i_k}$ для різних наборів $\{i_1, \dots, i_k\}$ є несумісними подіями, а виконання події B_k означає

виконання $C_{i_1 i_2 \dots i_k}$ для якогось одного з цих наборів номерів. Оскільки $B_k = \sum_{\{i_1, \dots, i_k\} \subseteq \{1, \dots, n\}} C_{i_1 i_2 \dots i_k}$, $k = 1, \dots, n$, $B_0 = C_0$, то за скінченною адитивністю для $k = 1, \dots, n$ відповідна ймовірність має вигляд $P(B_k) = \sum_{\{i_1, \dots, i_k\} \subseteq \{1, \dots, n\}} P(C_{i_1 i_2 \dots i_k}) = \sum_{\{i_1, \dots, i_k\} \subseteq \{1, \dots, n\}} p^k \cdot q^{n-k}$. В останній сумі всі доданки однакові, а їхня загальна кількість збігається з кількістю варіантів вибору k елементів $\{i_1, \dots, i_k\}$ із загальної множини $\{1, \dots, n\}$. Цю кількість варіантів вибору k елементів з n описують біноміальним коефіцієнтом $C_n^k = \frac{n!}{k!(n-k)!}$.

Отже,

$$P(B_k) = \sum_{\{i_1, \dots, i_k\} \subseteq \{1, \dots, n\}} P(C_{i_1 i_2 \dots i_k}) = \sum_{\{i_1, \dots, i_k\} \subseteq \{1, \dots, n\}} p^k \cdot q^{n-k} = C_n^k p^k \cdot q^{n-k}.$$

Для B_0 , за властивістю спадковості незалежності, маємо

$$P(B_0) = P(C_0) = P(\bar{A}_1 \bar{A}_2 \dots \bar{A}_n) = P(\bar{A}_1) P(\bar{A}_2) \dots P(\bar{A}_n) = q^n.$$

Приклад 2.26

Загальне доведення проілюструємо для випадку $n = 2$. Визначимо ймовірності трьох подій: B_0 – жодної появи A ; B_1 – одна поява A ; B_2 – дві появи A . У цьому випадку маємо таке:

- $C_0 = \bar{A}_1 \cdot \bar{A}_2$ – неоява A в обох спостереженнях,
- $C_1 = A_1 \cdot \bar{A}_2$ – поява A тільки в першому спостереженні, що означає неояву у другому,
- $C_2 = \bar{A}_1 \cdot A_2$ – поява A лише у другому спостереженні, що означає неояву в першому,
- $C_{1,2} = A_1 \cdot A_2$ – поява A в першому та другому (в обох) спостереженнях.

$$B_0 = \{\text{подія } A \text{ не відбулась жодного разу}\},$$

$$B_0 = C_0 = \bar{A}_1 \cdot \bar{A}_2;$$

$B_1 = \{\text{подія } A \text{ відбулась лише 1 раз, тобто, поява } A \text{ в першому чи у другому спостереженні}\},$

$$B_1 = C_1 + C_2 = A_1 \cdot \bar{A}_2 + \bar{A}_1 \cdot A_2;$$

$$B_2 = \{\text{подія } A \text{ відбулась 2 рази}\},$$

$$B_2 = C_{1,2} = A_1 \cdot A_2.$$

$$\begin{aligned}
 P(B_0) &= (1 - P(A_1)) \cdot (1 - P(A_2)) = (1 - p) \cdot (1 - p) = q \cdot q = q^2 \\
 P(B_1) &= P(C_1 + C_2) = P(A_1 \cdot \bar{A}_2) + P(\bar{A}_1 \cdot A_2) = \\
 &= P(A_1) \cdot P(\bar{A}_2) + P(\bar{A}_1) \cdot P(A_2) = P(A_1) \cdot (1 - P(A_2)) + (1 - P(A_1)) \cdot P(A_2) = \\
 &= pq + pq = 2pq ; \\
 P(B_2) &= P(A_1 \cdot A_2) = P(A_1) \cdot P(A_2) = p \cdot p = p^2 .
 \end{aligned}$$

Приклад 2.27

Гральний кубик 4 рази кидають. Знайти ймовірність події $A = \{\text{число "2" випало 2 рази}\}$.

Ймовірність події A в одному підкиданні $P(A) = 1/6 = p$ (кожне з шести чисел має однакову ймовірність), а $q = 5/6$. Ймовірність того, що число "2" випало при 1-му та 2-му підкиданні у серії з 4 незалежних підкидань становить $p^2 q^2 = \left(\frac{1}{6}\right)^2 \left(\frac{5}{6}\right)^2$. Запишемо всі різні варіанти, коли в серії із 4 підкидань подія {випало "2"} відбулась:

ВВНН, ВВНВ, ВВНВ, НВВН, НВВВ, НВВВ.

Легко бачити, що таких варіантів 6 і всі вони є несумісними подіями, тому $P(A) = 6p^2 q^2 = 6\left(\frac{1}{6}\right)^2 \left(\frac{5}{6}\right)^2 = \frac{25}{216} \approx 0,12$.

З іншого боку з формули (2.16) маємо

$$P(A) = C_4^2 p^2 (1-p)^{4-2} = C_4^2 p^2 q^2 = \frac{4!}{2!(4-2)!} \left(\frac{1}{6}\right)^2 \left(\frac{5}{6}\right)^2 = \frac{25}{216} \approx 0,12 .$$

Обчислюючи біноміальні коефіцієнти C_n^k , вважають, що за угодою $C_n^0 = 1, 0! = 1$.

Приклад 2.28

По мішені незалежно здійснюють 10 пострілів. Яка ймовірність семи влучень, якщо ймовірність влучення при кожному пострілі складає 0,5?

Розв'язання

Нехай подія A полягає у влученні в мішень при одному пострілі. Тоді в прикладі йдеться про появу A в декількох незалежних випробуваннях (пострілах) з однією і тією самою ймовірніс-

тю $P(A) = p$. Отже, задачу розв'язуємо за схемою незалежних випробувань Бернуллі з $n = 10$. Згідно з загальною формулою цієї схеми $P(B_k) = C_n^k p^k (1-p)^{n-k}$, де B_k , $k = 0, \dots, n$, – кількість появ A у n випробуваннях:

$$P(B_7) = C_{10}^7 \left(\frac{1}{2}\right)^7 \left(1 - \frac{1}{2}\right)^{10-7} = 120 \cdot \left(\frac{1}{2}\right)^{10} = \frac{120}{1024} = \frac{15}{128} \approx 0,117.$$

Приклад 2.29

Правильний шестигранний гральний кубик кидають 3 рази. Знайти ймовірність події $A = \{\text{число "5" випаде принаймні 1 раз}\}$.

Отже, потрібно знайти ймовірність того, що число "5" випаде принаймні 3 рази, тобто один, два або три рази. Запишемо події $B_1 = \{\text{число "5" випаде тільки один раз}\}$, $B_2 = \{\text{число "5" випаде лише два рази}\}$ та $B_3 = \{\text{число "5" випаде тільки три рази}\}$ і знайдемо їх імовірності. Із (2.16) маємо

$$P(B_1) = C_3^1 \cdot \frac{1}{6} \cdot \left(\frac{5}{6}\right)^2 \approx 0,35, \quad P(B_2) = C_3^2 \cdot \left(\frac{1}{6}\right)^2 \cdot \left(\frac{5}{6}\right) \approx 0,07,$$

$$P(B_3) = C_3^3 \cdot \left(\frac{1}{6}\right)^3 \approx 0,014.$$

Події B_1 , B_2 та B_3 несумісні, тому можемо застосувати теорему додавання ймовірностей. Отже, $P(A) \approx 0,35 + 0,07 + 0,014 = 0,434$.

Узагальнюючим наслідком цього прикладу є теорема.

Теорема 2.8

Нехай k_1 та k_2 – невід'ємні цілі числа такі, що $0 \leq k_1 \leq k_2 \leq n$. Позначимо через $P_n(k_1; k_2)$ імовірність того, що подія A відбудеться не менше, ніж k_1 раз і не більше k_2 разів у n випробуваннях. Тоді

$$P_n(k_1; k_2) = \sum_{k=k_1}^{k_2} C_n^k p^k q^{n-k} \quad (2.17, a)$$

або

$$P_n(k_1; k_2) = 1 - \sum_{k=0}^{k_1-1} C_n^k p^k q^{n-k} - \sum_{k=k_2+1}^n C_n^k p^k q^{n-k} \quad (2.17, б)$$

При застосуванні теореми 2.8 слід використовувати ту з формул, у якій менше доданків або ту, яка має сенс (наприклад при $k_1 = 0$ чи $k_2 = n$ вираз (2.17, б) не має сенсу).

Приклад 2.30

20 разів кидають правильний шестигранний гральний кубик. Знайти ймовірність події $A = \{\text{число "5" випаде не менше за 15 разів і не більше за 17}\}$.

Ймовірністю випадання числа "5" при одноразовому киданні кубика є $p = 1/6$. Із (2.17, *a*) маємо

$$\begin{aligned} P_{20}(15;17) &= \sum_{k=15}^{17} C_{20}^k p^k (1-p)^{20-k} = \\ &= C_{20}^{15} p^{15} (1-p)^5 + C_{20}^{16} p^{16} (1-p)^4 + C_{20}^{17} p^{17} (1-p)^3, \end{aligned}$$

підставляючи значення p , отримуємо

$$\begin{aligned} P_{20}(15;17) &= 15504 \cdot \left(\frac{1}{6}\right)^{15} \cdot \left(\frac{5}{6}\right)^5 + 4845 \cdot \left(\frac{1}{6}\right)^{16} \cdot \left(\frac{5}{6}\right)^4 + \\ &+ 1140 \cdot \left(\frac{1}{6}\right)^{17} \cdot \left(\frac{5}{6}\right)^3 \approx 1,412 \cdot 10^{-8}. \end{aligned}$$

Приклад 2.31

20 разів кидають правильний шестигранний гральний кубик. Знайти ймовірність події $A = \{\text{число "5" випаде не менше за 2 рази і не більше за 17}\}$.

Ймовірність випадання числа "5" при одноразовому киданні кубика є $p = 1/6$. Якщо ми скористаємось (2.17, *a*), то отримаємо суму з 16 доданків, а якщо (2.17, *б*), то всього 5, тому доцільніше скористатись (2.17, *б*). Отже,

$$P_{20}(2;17) = 1 - \sum_{k=0}^1 C_{20}^k \cdot \left(\frac{1}{6}\right) \cdot \left(\frac{5}{6}\right)^{20-k} - \sum_{k=18}^{20} C_{20}^k \cdot \left(\frac{1}{6}\right)^k \cdot \left(\frac{5}{6}\right)^{20-k} \approx 0,9.$$

Приклад 2.32

При проведенні соціологічного опитування виявилось, що ймовірність правильного заповнення анкети респондентом становить $p = 0,9$. Яка ймовірність того, що зі 100 анкет вірно заповнено принаймні 95.

Із (2.17, *a*) маємо

$$\begin{aligned} P_{100}(95) &= \sum_{k=95}^{100} C_{100}^k p^k (1-p)^{100-k} = \\ &= C_{100}^{95} (0,9)^{95} (0,1)^5 + C_{100}^{96} (0,9)^{96} (0,1)^4 + C_{100}^{97} (0,9)^{97} (0,1)^3 + \end{aligned}$$

$$+C_{100}^{98}(0,9)^{98}(0,1)^2 + C_{100}^{99}(0,9)^{99}(0,1)^1 + C_{100}^{100}(0,9)^{100}(0,1)^0 \approx \\ \approx 0,0338+0,01587+0,00589+0,001623+0,000295+0,000026564 \approx 0,05758.$$

Приклад 2.33 (задача Банаха)

Один математик, що палить, носить із собою дві коробки сірників. Щоразу, як він хоче запалити, виймає навмання одну з коробок, дістає звідти сірник і кладе коробку назад. Знайти ймовірність того, що, коли він уперше дістане порожню коробку, то у другій виявиться r сірників, $0 < r \leq n$, де n – початкова кількість сірників у кожній коробці.

Розв'язання

Позначимо через $A = \{\text{виймається сірник із коробки, яка першою стане порожньою}\}$. Якщо коробка, яку витягнули, є порожньою, а в іншій залишилось r сірників, то це означає, що всього з обох коробок взяли $(2n - r)$ сірників. При цьому подія A відбулась рівно n разів, бо одна з коробок виявилась порожньою. Щоразу, коли витягається коробка, імовірність події A складає $1/2$.

За формулою (2.17, а)

$$P_{2n-r}(n) = C_{2n-r}^n \left(\frac{1}{2}\right)^{2n-r} = \frac{(2n-r)!}{n!(2n-r-n)!} \cdot \left(\frac{1}{2}\right)^{2n-r}.$$

Зазначимо, що в задачах на умовні ймовірності та СНВБ дуже важливо вірно сформулювати всі події.

ЗАДАЧІ

Схема незалежних випробувань Бернуллі

№ 1

Якою є ймовірність випадання рівно 5 гербів при підкиданні 10 монет?

№ 2

Несиметричну монету (імовірність появи герба у 3 рази більша ніж імовірність появи решітки) підкидають 6 разів. Яка ймовірність появи герба 4 рази?

№ 3

Імовірність скласти кожний із іспитів складає для студента 0,5. Якою є ймовірність того, що з п'яти іспитів студент складе хоча б два іспити?

№ 4

Симетричну монету підкидають 12 раз. Що ймовірніше:

- а) поява 5 гербів чи поява 6 гербів?
- б) поява 5 гербів чи поява 3 решіток?
- в) поява 5 гербів чи поява 7 решіток?

№ 5

Скільки разів треба кинути гральний кубик, щоб з імовірністю не меншою за 0,5 хоча б один раз випала шістка?

№ 6

Двоє осіб підкидають монету n разів кожна. Знайти ймовірність того, що в них випаде однакова кількість гербів.

№ 7

Симетричну монету підкидають 10 раз. Яке найбільш імовірне число появ герба? Яке найменш імовірне число появ герба?

Формула Стірлінга (Муавра)

При обчисленні C_n^k для великих k та n , а також факторіалів великих чисел часто виникає ситуація, коли порахувати сам факторіал досить важко через його велике значення. Крім того, значення таких факторіалів великих чисел достатньо знати приблизно. Ця проблема була вирішена шотландським математиком Джеймсом Стірлінгом (1692–1770) у 1730 році і, незалежно від Стірлінга, цю проблему вирішив також англійський математик французького походження Абрахам де Муавр (1667–1754).

Теорема 2.9 (формула Стірлінга). У випадку наближеного обчислення значення $n!$ для великих чисел діє формула

$$n! \approx \left(\frac{n}{e}\right)^n \sqrt{2\pi n} \quad (2.18, a)$$

або

$$\ln(n!) \approx \left(n + \frac{1}{2}\right) \ln(n) - n + \frac{1}{2} \ln(2\pi), \quad (2.18, б)$$

де \ln – натуральний логарифм, $e \approx 2,718282$, $\pi \approx 3,141593$.

Приклад 2.34

Обчислити $100!$

Із (2.18, а) отримаємо, що $100! \approx 5,7102107404794E+262$, тобто $5,7102107404794$ помножити на 10 у 262 степені.

Найімовірніша кількість появ випадкової події

При розв'язуванні реальних задач часто виникають випадки, для яких потрібно приблизно вирахувати значення сум $\sum_{k=k_1}^{k_2} P_n(k)$

для заданих k_1 та k_2 і досить великих n .

Для $0 \leq k < n$ виконується співвідношення

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

Дійсно,

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

Звідси випливає таке: якщо $(n-k)p > (k+1)q$, тобто, якщо $np - q > k(p+q)$ або $np - q > k$, то

$$P_n(k+1) > P_n(k).$$

Якщо ж $np - q = k$, то виконується

$$P_n(k+1) = P_n(k).$$

І, нарешті, якщо $np - q < k$, то

$$P_n(k+1) < P_n(k).$$

Звідси видно, що ймовірність $P_n(k)$ зі збільшенням k спочатку зростає, потім досягає максимуму, і, при подальшому зростанні k , спадає (зменшується). При цьому, якщо $np - q$ є цілим числом, то максимального значення ймовірність $P_n(k)$ досягає для двох значень k , $k_0 = np - q$ та $k'_0 = np - q + 1 = np + p$. Якщо ж $np - q$ не є цілим числом, то максимуму $P_n(k)$ досягає при k_0 , що дорівнює найменшому цілому числу, що є більшим за $np - q$.

Найімовірнішою кількістю появ випадкової події A в результаті n незалежних експериментів за схемою Бернуллі називають

вають таке число k_0 , для якого ймовірність $P_n(k)$ є не меншою, або перевищує ймовірність решти наслідків експериментів.

Приклад 2.35

Проводять експеримент із кидання шестигранного правильного грального кубика і досліджують випадання числа "5". Експеримент виконують 6 разів. Знайти найбільш імовірну кількість випадань числа "5".

Очевидно, що ймовірність випадання "5" при одноразовому експерименті складає $1/6$. Запишемо в таблицю наближені значення обчислених ймовірностей:

k	0	1	2	3	4	5	6
$P_6(k)$	0,3342	0,4019	0,2009	0,054	0,0008	0,00064	0,0002

Отже, найбільш імовірним є одноразове випадання "5" при проведенні 6 незалежних експериментів.

З іншого боку, знаючи, що $np - q = 6 \cdot 1/6 - 5/6 = 1/6$ не є цілим числом, маємо $k_0 = 1$.

Приклад 2.36

Проводять експеримент із підкидання правильної монети. Ймовірність випадання герба при одноразовому підкиданні становить $1/2$. Здійснюють серію з 10 підкидань. Знайти найбільш імовірну кількість випадання герба.

Запишемо в таблицю наближені значення ймовірностей:

k	0	1	2	3	4	5	6	7	8	9	10
$P_{10}(k)$	0,001	0,01	0,04	0,1	0,2	0,24	0,2	0,1	0,04	0,01	0,001

Отже, найбільш імовірним є п'ятикратне випадання герба в серії з 10 випробувань.

З іншого боку $np - q = 10 \cdot 0,5 - 0,5 = 4,5$, тобто при $k = 5$ досягається максимальна ймовірність $P_{10}(k)$.

ЗАДАЧІ

№ 1

При проведенні соціологічного опитування потрібно досягти того, щоб 90 анкет були заповнені вірно. Яку найменшу кіль-

кість людей потрібно опитати, якщо ймовірність того, що анкета заповнена вірно, становить 0,9?

№ 2

При психологічному тестуванні ймовірність позитивної відповіді на перше питання становить 0,6. Яка мінімальна кількість людей має пройти тестування, щоб група осіб, які відповіли позитивно на перше питання, складала 30 чоловік?

Формула (теорема) Пуассона

Нехай n – кількість експериментів у схемі Бернуллі є великою, а ймовірність успіху p у кожному з експериментів є малою величиною, причому малий також і добуток $\lambda = np$. Тоді $P_n(m)$ визначають наближено за формулою

$$P_n(m) \approx \frac{\lambda^m}{m!} e^{-\lambda}, \quad m = 0, 1, 2, \dots, n, \quad (2.19, a)$$

або

$$\lim_{\substack{n \rightarrow \infty \\ np = \lambda}} P_n(m) = \frac{\lambda^m}{m!} e^{-\lambda}, \quad m = 0, 1, 2, \dots, n, \quad (2.19, б)$$

яку називають **формулою Пуассона**.

Приклад 2.37

Під час соціологічного дослідження було опитано 3000 осіб. Ймовірність того, що респондент не зазначив в анкеті рівень свого прибутку, становить 0,001. Знайти ймовірність того, що серед 3000 анкет є принаймні 3, де не зазначено рівень прибутку.

Опитування кожного респондента вважатимемо за незалежний експеримент. Позначимо $A = \{\text{у анкетах не зазначено рівень прибутку}\}$. Тоді $p = P(A) = 0,001$, $\lambda = np = 3000 \cdot 0,001 = 3$. Через те, що нам потрібно визначити ймовірність того, що є принаймні 3 (тобто 3 або більше) анкети, де не зазначено рівень прибутку, доцільно обчислити ймовірність протилежної події, тобто наявність 1 або 2 таких анкет серед 3000 і застосувати (2.17, б).

Далі з (2.19, а) маємо

$$P_{3000}(3, 3000) \approx 1 - \frac{3^1}{1!} e^{-3} - \frac{3^2}{2!} e^{-3} = 1 - 0,1493 - 0,224 \approx 0,6266.$$

Локальна теорема Муавра – Лапласа

Якщо у схемі Бернуллі кількість незалежних експериментів n велика і великими є значення pn та qn , то теорему Лапласа застосовувати вже не можна через те, що кількість "успіхів" m росте з ростом кількості експериментів n .

Цю проблему піднято в ряді робіт математиків початку XVIII століття, що були присвячені демографічним дослідженням. Повністю цю задачу розв'язав Абрахам де Муавр (1667–1754), французький математик, що після відміни Нантського едикту переїхав до Англії, де і провів решту свого життя. Там у 1718 році він опублікував свою головну роботу "The Doctrine of Chances" ("Доктрина шансів"). У 1783 році Лаплас узагальнив теорему, доведені Муавром для довільної ймовірності p .

Теорема 2.10 (локальна теорема Муавра – Лапласа). Якщо ймовірність появи випадкової події в кожному з n незалежних експериментів є величина стала і дорівнює p ($0 < p < 1$), то для великих значень m та n ймовірність того, що випадкова подія A відбудеться m разів при n випробуваннях, наближено обчислюють за формулою

$$P_n(m) \approx \frac{\phi(x)}{\sqrt{npq}}, \quad (2.20)$$

де $\phi(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$ називають функцією Гаусса, а $x = \frac{m - np}{\sqrt{npq}}$.

Приклад 2.38

При проведенні соціологічного опитування виявилось, що серед усіх анкет 80 % заповнені вірно. Із пачки анкет навмання беруть 400. Обчислити ймовірність $A = \{\text{серед 400 анкет 300 заповнено вірно}\}$.

Скористаємось локальною теоремою Муавра – Лапласа. Маємо $n = 400$, $p = 0,8$, $m = 300$;

$$x = \frac{m - np}{\sqrt{npq}} = \frac{300 - 320}{8} = -2,5;$$

$$P_{400}(380) \approx \frac{\phi(-2,5)}{8} \approx \frac{0,01752830049}{8} \approx 0,00219103756.$$

Приклад 2.39

Під час психологічного тестування виявилось, що 10 % людей відповіли на десяте питання негативно. Серед усіх опитаних осіб відібрали групу у 500 чоловік. Яка ймовірність того, що серед них 450 відповіли на десяте питання позитивно?

Скористаємось локальною теоремою Муавра – Лапласа. Маємо $n = 500$; $p = 0,9$; $m = 450$;

$$x = \frac{m - np}{\sqrt{npq}} \approx \frac{450 - 450}{6,7} = 0 ;$$

$$P_{500}(450) \approx \frac{\phi(0)}{6,7} \approx \frac{0,3989}{6,7} \approx 0,06 .$$

2.6.9. Схема незалежних випробувань до першої появи події A

Схемою незалежних випробувань до першої появи події A (СНВППА) називають задачу обчислення ймовірностей кількості проведених незалежних спостережень (випробувань) до першої появи події A . Через C_k позначають подію, що полягає в зупинці на k -му кроці, а СНВППА полягає в підрахунку $P(C_k)$ для всіх можливих натуральних значень $k = 1, 2, \dots, n, \dots$, оскільки кількість кроків є необмеженою.

Теорема 2.11

Обчислюють імовірності у СНВППА за формулою

$$P(C_k) = pq^{k-1}, k \geq 1, q = 1 - p.$$

Доведення

Якщо позначити через A_k , $k \geq 1$, появу A на k -му кроці, то для кожного з можливих k

$$C_k = B_1 \cdots B_{k-1} \cdot A_k; C_1 = A_1, C_2 = B_1 \cdot A_2, C_3 = B_1 \cdot B_2 \cdot A_3, \dots$$

Отже, за спадком незалежності

$$\begin{aligned} P(C_k) &= P(B_1 \cdots B_{k-1} \cdot A_k) = P(B_1) \cdots P(B_{k-1}) \cdot P(A_k) = \\ &= (1 - P(A_1)) \cdots (1 - P(A_{k-1})) \cdot P(A_k) = \\ &= (1 - p)^{k-1} \cdot p = pq^{k-1}, k = 1, 2, \dots, n, \dots \end{aligned}$$

Приклад 2.40

Піддослідний може скласти іспит з 1, 2, 3 і більше спроб без обмеження кількості спроб.

1) Яка ймовірність, що іспит буде складено з 1, 2, 3 спроби, якщо ймовірність успішного складання іспиту в кожній зі спроб дорівнює 0,4 і результат однієї спроби не впливає на результат іншої.

2) Якою є ймовірність того, що іспит буде складено не більше ніж із трьох спроб.

Розв'язання

Оскільки в задачі йдеться про незалежні випробування (іспити) до першої появи події $A = \{\text{іспит складено успішно}\}$, то її розв'язують із застосуванням СНВППА з $P(A) = p = 0,4$ і підрахунком у рамках цієї схеми відповідно $P(C_1)$, $P(C_2)$ чи $P(C_3)$ за загальною формулою $P(C_k) = p \cdot q^{k-1}$ для $k = 1, 2, 3$:

$$P(C_1) = p \cdot q^{1-1} = p = 0,4,$$

$$P(C_2) = p \cdot q^{2-1} = p \cdot q = 0,4 \cdot 0,6 = 0,24,$$

$$P(C_3) = p \cdot q^{3-1} = p \cdot q^2 = 0,4 \cdot 0,36 = 0,144.$$

Для обчислення ймовірності скласти іспит не більше ніж із третьої спроби, позначимо цю подію через D і отримаємо

$$D = C_1 + C_2 + C_3.$$

Оскільки C_1 , C_2 , C_3 є попарно несумісними, то за скінченною адитивністю маємо

$$\begin{aligned} P(D) &= P(C_1 + C_2 + C_3) = P(C_1) + P(C_2) + P(C_3) = \\ &= 0,4 + 0,24 + 0,144 = 0,784. \end{aligned}$$

2.6.10. Розв'язування ОЗТЙ

Загальний алгоритм розв'язування ОЗТЙ

Згідно сутності ОЗТЙ загальний алгоритм розв'язування складається з трьох етапів.

1) Уведення явних позначень A_1, \dots, A_n для подій, ймовірності яких відомі, та прив'язка введених позначень до відомих ймовірностей:

$$p_1 = P(A_1), \dots, p_n = P(A_n).$$

2) Уведення позначення, наприклад B , для події, ймовірність якої треба обчислити, та опис її зв'язку з подіями першого кроку.

$$B = B(A_1, \dots, A_n).$$

Перший та другий крок можна об'єднувати.

3) Обчислення ймовірності події B підстановкою виразу цієї події через A_1, \dots, A_n та застосуванням слушної базової ОЗТЙ, як правило, за останнім перетворенням у виразі $B = B(A_1, \dots, A_n)$. Для підрахунку кожної з ймовірностей, що з'явилися після першого кроку в запису $P(B)$, знову шукають слушну базову ОЗТЙ і т. д.:

$$P(B) = P(B(A_1, \dots, A_n)) = \dots$$

Таким чином, формують комбінацію базових ОЗТЙ, що зводить підрахунок $P(B)$ до використання значень

$$p_1 = P(A_1), \dots, p_n = P(A_n).$$

2.6.11. Розв'язування ОЗТЙ через рівну можливість

У деяких задачах на обчислення ймовірності немає явного посилання на значення ймовірностей, але є посилання на рівну можливість результатів, що передаються виразами: "рівноможливі", "навмання", "випадковим чином" і т. д.

Приклад 2.41 (варіант формулювання задачі з посиланням на рівну можливість результатів)

В аудиторії перебуває m осіб, що можуть успішно виконати тест, та n тих, що з ним не справляться. Навмання вибирають: а) одну особу, б) дві особи. Якою є ймовірність того, що а) вибрана особа успішно виконає тест, б) із двох одна виконає, а одна не справиться?

Загальний підхід до розв'язання ОЗТЙ за наявності посилання на рівну можливість результатів. В умовах посилання на рівну можливість, ймовірність будь-якої події з U визначають через "природну" адитивну функцію Z , яка на U має бути скінченною (кількість елементів, довжина, площа, об'єм, ...), за формулою

$$P(A) = \frac{Z(A)}{Z(U)}, \quad Z = \begin{cases} N - \text{кількість,} \\ L - \text{довжина,} \\ S - \text{площа,} \\ V - \text{об'єм,} \\ \dots \end{cases}$$

Визначення $P(A)$ за допомогою N відповідає дискретним даним (дискретному U), а за допомогою довжини L , площі S чи об'єму V – неперервним. Обґрунтуванням такого підходу є реалізація в ньому рівноможливості результатів.

Дійсно, для дискретного $U = \{U_1 \dots U_n\}$ нескінченна кількість елементів виключається, оскільки $Z(U) = N(U)$ – кількість елементів в U має бути скінченною (їх ϵn за визначенням). За загальним підходом у цьому випадку

$$P(A) = \frac{N(A)}{N(U)} = \frac{N(A)}{n}.$$

Зокрема, якщо $A = \{u\}$, то $N(A) = N\{u\} = 1$. Отже,

$$P\{u\} = \frac{N\{u\}}{n} = \frac{1}{n}.$$

Таким чином, визначення ймовірностей через "природну" адитивну функцію для дискретного випадку приводить до рівноможливих результатів, що відповідають змісту задачі. Такий спосіб визначення ймовірності через рівну можливість та скінченну множину U називають **класичним визначенням ймовірності**.

Якщо U – неперервна множина, то вона є частиною прямої (площини, простору) зі скінченною довжиною (площею, об'ємом). У цьому випадку функція Z є довжиною (площею, об'ємом). За такого визначення ймовірностей події A та отримана з A зміщенням на x подія $A+x$ мають однакові ймовірності, оскільки $Z(A+x) = Z(A)$.

Ймовірності в неперервному випадку, визначені за формулою $P(A) = \frac{Z(A)}{Z(U)}$, де $Z(A)$ є довжиною L , площею S , чи об'ємом V ,

називають **геометричними ймовірностями**.

Загальний алгоритм розв'язання ОЗТЙ на рівну можливість результатів. Алгоритм розв'язання в умовах посилення на рівну можливість результатів відрізняють від загального алгоритму розв'язання ОЗТЙ, хоча він є таким самим трикроковим.

1-й крок. На першому кроці описують простір елементарних подій U і визначають його характер: дискретний чи неперервний, залежно від чого визначають "природну" адитивну функцію $Z(A)$ та ймовірність для всіх можливих подій A за формулою:

$$P(A) = \frac{Z(A)}{Z(U)}.$$

Здійснення цього кроку вимагає підрахунку значення $Z(U)$, яке має бути скінченним.

2-й крок. На другому кроці описують подію A як підмножину U та обчислюють значення функції Z для цієї події.

3-й крок. На третьому кроці результати підрахунків першого та другого кроків зводять разом під час знаходження $P(A)$ за формулою

$$P(A) = \frac{Z(A)}{Z(U)}.$$

Приклад 2.42

З урни, де міститься m білих та n чорних куль, "навмання" витягають одну кулю. Якою є ймовірність того, що вона біла?

Розв'язання

1-й крок. Оскільки є посилення на рівну можливість результатів ("навмання"), то відповідно до загального алгоритму розв'язання таких ОЗТЙ, визначають простір елементарних подій U , яким у досліджуваному випадку є множина номерів куль, що містяться в урні:

$$U = \{1, 2, \dots, m + n\}.$$

Оскільки простір елементарних подій U є дискретним, то Z є кількістю елементів у множині A : $Z(A) = N(A)$.

$$Z(U) = N(U) = m + n,$$

$$P(A) = \frac{N(A)}{N(U)} = \frac{N(A)}{m + n}.$$

2-й крок. Подію $A = \{\text{поява білої кулі}\}$ описують множиною номерів цих куль: $A = \{i, \dots, i_m\}$,

$$N(A) = N \{i, \dots, i_m\} = m.$$

3-й крок. Остаточно маємо

$$P(A) = \frac{N(A)}{m + n} = \frac{m}{m + n}.$$

2.6.12. Геометричні ймовірності

Існує низка задач, які розв'язати в загальних термінах теорії ймовірності досить важко (або й зовсім неможливо), але які легко розв'язуються графічним (геометричним) методом. Загальну задачу, що привела до виникнення терміну "геометрична ймовірність", спрощено формулюють так:

нехай маємо квадрат A зі стороною a і всередині нього є ще один квадрат B зі стороною b , $b < a$ (сторони квадрата A не перетинаються зі сторонами квадрата B). Знайти ймовірність події $C = \{\text{навмання кинута у квадрат } A \text{ точка потрапить і у квадрат } B\}$.

"Навмання" означає, що точка може потрапити в довільне місце квадрата A . Тобто, множиною елементарних подій є вся множина точок квадрата A , а множиною сприятливих подій є множина всіх точок квадрата B . Обидві множини є нескінченними й незліченними, тому класичне визначення ймовірності застосувати неможливо. Очевидним виходом із цієї ситуації є введення певної міри (тут функції від множини), за допомогою якої ми можемо ставити множині у відповідність число та порівнювати множини (події). Наприклад, у задачі про два квадрати, описаній вище, такою мірою може бути площа квадратів. Тоді ймовірністю того, що точка потрапить у квадрат B , буде така:

$$P(C) = \frac{a^2}{b^2}.$$

Історична довідка. Жорж Бюфон (1707–1788), видатний французький учений, у своїй роботі, що була написана 1733 року, а опублікована в 1777, заклав новий напрямок у теорії ймовірностей. Розв'язання відомої "задачі про голку", що викладалося в цій статті, вимагало швидше геометричного, а не комбінаторного методу вирішення. У задачах такого типу вважають, що випадкові точки рівномірно розподілені в деякій області. Імовірність потрапити в довільну частину цієї області пропорційна площі цієї частини (довжині, обсягу). Таким чином, для визначення ймовірності достатньо знайти відношення площі "сприятливої" множини до площі (довжини, обсягу) усієї множини.

Приклад 2.43

У кімнаті на підлозі площею 30 м^2 лежить килим площею 2 м^2 . Випадковим чином на підлогу кидають монету. Яка ймовірність того, що вона потрапить на килим? Для знаходження ймовірності можемо скористатись площею як мірою. Тоді шукану ймовірність обчислюють як частку площі килима до площі кімнати, тобто $2/30$.

Приклад 2.44

Нехай визначена множина $U = \{0 \leq x \leq 2, 0 \leq y \leq 2\}$. Знайти ймовірність події $A = \{\text{навмання взята у квадраті точка потрапить у круг із центром у точці } (1;1) \text{ та радіусом } 0,5\}$.

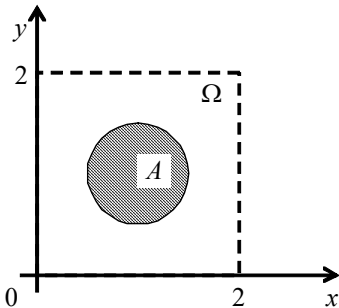


Рис. 2.13. Геометричні ймовірності (до прикладу 2.44)

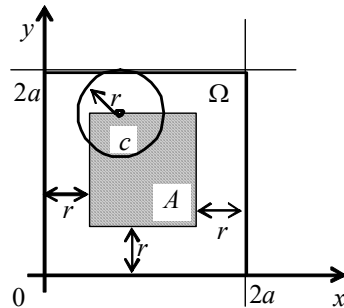


Рис. 2.14. Геометричні ймовірності (до прикладу 2.45)

Скористаємося для порівняння множин площею як мірою. Круг A з центром $(1;1)$ та радіусом $0,5$ лежить усередині квадрата Ω зі стороною 2 (рис. 2.13). Тому відношення площі круга до площі квадрата і буде шуканою ймовірністю $P(A) = \frac{\pi \cdot 0,5^2}{2^2} \approx 0,196$.

Ймовірності, які задають формулою

$$P(A) = \frac{Z(A)}{Z(U)},$$

де A та U – неперервні множини, а функція Z – довжина (площа, об'єм), називають **геометричними ймовірностями**.

Приклад 2.45

Площину розкреслено прямими так, що вони утворюють однакові квадрати зі стороною $2a$. На площину кидають монету радіусом r ($r < a$) випадковим чином (термін "випадковим чином" означає, що центр монети може розташуватись у довільному місці на площині, тобто рівномірно). Знайти ймовірність події $A = \{\text{монета не дотикається і не перетинається з прямими}\}$ (рис. 2.14).

Розглянемо розташування центра c монети у квадраті. Монета не перетинається з прямими, якщо відстань від її центра до прямої більша за r . Тобто, множина всіх сприятливих наслідків складатиме квадрат A (рис. 2.14) зі стороною $2a - 2r$. Отже, шукана ймовірність є відношенням площі квадрата A до площі квадрата U і має вигляд

$$P(A) = \frac{(2a - 2r)^2}{(2a)^2}.$$

Приклад 2.46 (задача Бюфона)

Площина розкреслена паралельними прямими, що віддалені одна від одної на відстань $2a$. На площину навмання кидають тонку голку завдовжки $2l$, $l < a$ (під словом "навмання" мається на увазі, що центр голки може впасти в довільну точку площини, а сама голка під довільним кутом до розкреслених прямих). Знайти ймовірність події $A = \{\text{голка перетне якусь із прямих}\}$.

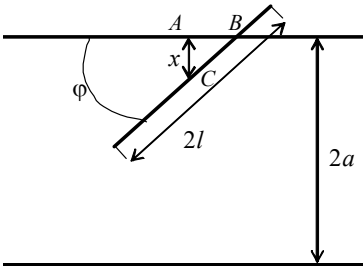


Рис. 2.15. Задача Бюфона.
Розташування голки
стосовно прямих

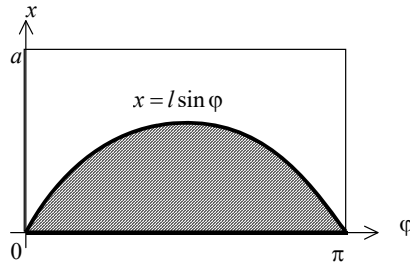


Рис. 2.16. Задача Бюфона.
Множина елементарних подій
та сприятливих наслідків

Позначимо через x відстань від центра голки до найближчої прямої. Маємо $0 \leq x \leq a$, бо x не може перевищувати a – половину відстані між прямими, тому що, коли відстань x від центра голки до якоїсь із прямих більша за a , то до іншої прямої ця відстань буде меншою за a , а ми розглядаємо відстань до найближчої прямої. Це легко бачити з рис. 2.15. Для простоти вважатимемо, що найближча пряма розташована вище за голку. Це означає, що $0 \leq x < a$. Якщо ж $x = a$, то центр голки лежить посередині між прямими. У разі $x > a$ ближчою буде нижня пряма і знову матимемо $0 \leq x < a$. Позначимо як ϕ гострий або прямий кут, який складає голка із прямою ($0 \leq \phi \leq \pi$). Дійсно, із рис. 2.15. видно, що гострий кут голки із прямою утворено ліворуч від точки перетину, тобто маємо $0 \leq \phi < \pi$. Якщо ж $\phi = \frac{\pi}{2}$, то голка складає прямий кут із прямою. Як тільки $\phi > \pi$, то гострий кут голка із прямою утворить уже праворуч від точки перетину, і знову отримаємо $0 \leq \phi < \pi$.

Різні розташування голки визначаються парами точок (x, ϕ) і у декартових координатах є точками прямокутника зі сторонами a та π . Тобто, простір елементарних подій має вигляд $U = \{(x, \phi) \in R^2 : 0 \leq x \leq a, 0 \leq \phi \leq \pi\}$. Із рис. 2.15 видно, що голка не перетинає пряму, якщо виконується нерівність $x \leq l \sin \phi$.

Дійсно, розглянемо трикутник ABC з рис. 2.15. Голка не перетинає пряму, якщо довжина BC не менша за l . З іншого боку, за властивістю прямокутного трикутника, довжина AC дорівнює $|BC| \sin \phi$, тобто $x = |BC| \sin \phi$. Звідси маємо $x \leq l \sin \phi$.

Графічно множину всіх точок, що задовольняють $x \leq l \sin \phi$, зображено на рис. 2.16. Ця нерівність виділяє з прямокутника Ω область, що лежить під синусоїдою (заштрихована область рис. 2.16). Площа заштрихованої області обчислюється як інтеграл $mes A = \int_0^{\pi} l \sin \phi d\phi$.

Отже, шукана ймовірність є відношенням площі заштрихованої області до площі всього прямокутника і становить

$$P(A) = \frac{\int_0^{\pi} l \sin \phi d\phi}{a\pi} = \frac{2l}{a\pi}.$$

Приклад 2.47 (задача про зустріч)

Два чоловіки B та C домовилися про зустріч з 0-ї до 1-ї години дня так, що жоден із них не чекатиме іншого більше за 15 хв. Якщо при цьому другий чоловік не прийде, то перший іде геть. Обчислити ймовірність події $A = \{\text{зустріч відбудеться}\}$.

Позначимо моменти приходу кожного з чоловіків як b та c відповідно. Простір елементарних подій має вигляд $U = \{(b, c) \in \mathbb{R}^2 : 0 \leq b \leq 1, 0 \leq c \leq 1\}$.

B та C зустрінуться тоді, коли $|b - c| \leq \frac{1}{4}$, тобто різниця між моментами їх приходу на зустріч не перевищує 15 хв. Побудуємо простір U і зобразимо його у вигляді квадрата на рис. 2.17.

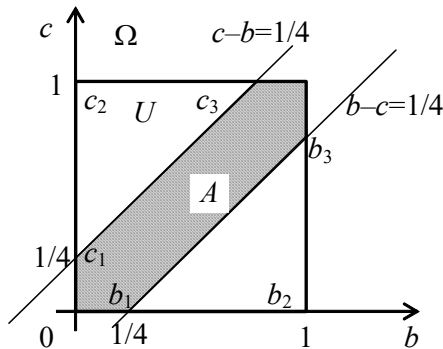


Рис. 2.17. Задача про зустріч

Побудуємо тепер подію A (рис. 2.17). Вона має вигляд $A = \{(b, c) \in \mathbb{R}^2 : 0 \leq c \leq 1, 0 \leq b \leq 1, |b - c| \leq \frac{1}{4}\}$.

Розпишемо більш детально цю подію.

Першим може прийти на зустріч як чоловік B , так і C . Якщо першим прийшов чоловік B , то момент часу b є меншим за момент часу c , і, для виконання умов зустрічі має виконуватись

співвідношення $c - b \leq \frac{1}{4}$. Якщо ж першим прийшов C , то $c < b$ і має виконуватись співвідношення $b - c \leq \frac{1}{4}$. Якщо ці співвідношення об'єднати, то отримаємо систему

$$\begin{cases} b - c \leq \frac{1}{4}, \\ c - b \leq \frac{1}{4}, \end{cases} \quad (2.21, a)$$

яка є ідентичною до $|b - c| \leq \frac{1}{4}$.

Через те, що множина $U = \{(b, c) \in R^2: 0 \leq b \leq 1, 0 \leq c \leq 1\}$ і змінюється від 0 до 1 по осі b та від 0 до 1 по осі c , то на рис. 2.17 вона має вигляд квадрата. Таким чином, усі можливі моменти приходу обох чоловіків містяться в межах квадрата U .

Далі з цього квадрата потрібно виділити область A , яка описана системою (2.21, a) та міститься всередині U . Для цього побудуємо дві прямі, що описуються рівняннями $c - b = \frac{1}{4}$ та

$b - c = \frac{1}{4}$. Шуканою буде область між цими прямими (тобто, нижче за пряму $c - b = \frac{1}{4}$ або ж $c = b + \frac{1}{4}$ та вище за $b - c = \frac{1}{4}$ або ж

$c = b - \frac{1}{4}$), що міститься всередині квадрата U .

Очевидно, що ймовірністю події A буде відношення площі заштрихованої множини до площі всього квадрата (рис. 2.17.)

Площу квадрата U обрахувати легко – вона дорівнює 1 (сторони квадрата дорівнюють 1 кожна).

Для обчислення площі області A логічно обчислювати не саму A , а площу області U , що не потрапляє до A , а потім відняти її від площі U . Для цього обчислимо площі двох трикутників $c_1c_2c_3$ та $b_1b_2b_3$, що входять до U і не входять до A .

Точка c_1 перетину прямої $c - b = \frac{1}{4}$ з віссю OC має координату $(0; 1/4)$, тобто довжина сторони c_1c_2 трикутника $c_1c_2c_3$ така:

$1 - 1/4 = 3/4$. Аналогічно знаходимо $b_1b_2 = 1/4$ (b_1 є точкою перетину прямої $b - c = \frac{1}{4}$ із прямою $c = 0$). Точка c_3 перетину прямої $c - b = \frac{1}{4}$ із прямою $c = 1$ (верхня сторона квадрата) має координату $(3/4; 1)$, тобто сторона c_2c_3 трикутника $c_1c_2c_3$ має довжину $3/4$. Аналогічно обчислюють довжину сторони b_2b_3 трикутника $b_1b_2b_3$ (точка b_3 є точкою перетину прямої $b - c = \frac{1}{4}$ із прямою $b = 1$). Отже, маємо 2 однакові трикутники $b_1b_2b_3$ та $c_1c_2c_3$ з катетами по $3/4$, звідси площу кожного з трикутників можна обчислити як $\frac{1}{2} \cdot \frac{3}{4} \cdot \frac{3}{4} = \frac{9}{32}$. Оскільки трикутників 2, то їх сумарна площа становить $2 \cdot \frac{9}{32} = \frac{9}{16}$, а відповідно площа множини A дорівнює $1 - 2 \cdot \frac{1}{2} \cdot \frac{3}{4} \cdot \frac{3}{4} = 1 - 2 \cdot \frac{9}{32} = 1 - \frac{9}{16} = \frac{7}{16}$.

Отже, остаточно ймовірність події A

$$P(A) = \frac{mesA}{mesU} = \frac{7}{16} = \frac{7}{16},$$

де $mesA$ – це міра множини A , і в нашому випадку мірою є площа.

Приклад 2.48 (задача про зустріч із запізненням)

Два чоловіки B та C домовились про зустріч з 0-ї до 1-ї години дня так, що жоден із них не чекатиме іншого більше 15 хв. Якщо при цьому другий чоловік не прийде, то перший іде геть. Обчислити ймовірність події $A = \{\text{зустріч відбудеться}\}$ у випадку, якщо годинник першого чоловіка запізнюється (відстає) на 10 хв.

Позначимо моменти приходу кожного з чоловіків як b та c відповідно. Оскільки годинник першого чоловіка відстає, то момент, коли в нього на годиннику буде час 0, у реальному часі означатиме, що насправді є 0 годин 10 хвилин, тобто $1/6$. Отже, момент приходу першого чоловіка змінюватиметься від $1/6$ до $7/6$, а момент приходу другого чоловіка становитиме, як у прикладі 2.47, від 0 до 1.

Таким чином, простір елементарних подій має вигляд $U = \{(b, c) \in R^2 : \frac{1}{6} \leq b \leq \frac{7}{6}, 0 \leq c \leq 1\}$. Побудуємо простір U і зобразимо у вигляді квадрата на рис. 2.18.

B та C зустрінуться тоді, коли $|b - c| \leq \frac{1}{4}$, тобто різниця між моментами їх приходу на зустріч не перевищує 15 хв.

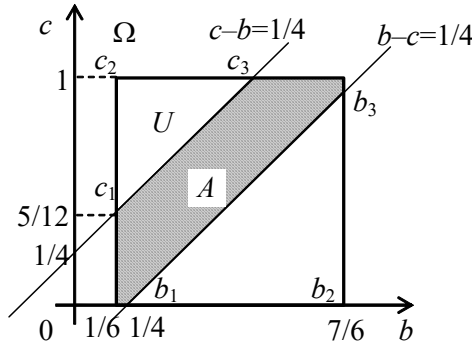


Рис. 2.18. Задача про зустріч із запізненням

Побудуємо тепер на рис. 2.18 подію A . Вона має вигляд $A = \{(b, c) \in R^2 : 0 \leq c \leq 1, \frac{1}{6} \leq b \leq \frac{7}{6}, |b - c| \leq \frac{1}{4}\}$.

Розпишемо більш детально цю подію. Першим може прийти на зустріч як чоловік B , так і C . Якщо першим прийшов чоловік B , то момент часу b є меншим за момент часу c , і, для виконання умов зустрічі має виконуватись співвідношення $c - b \leq \frac{1}{4}$. Якщо ж першим прийшов C , то $c < b$ і має виконуватися співвідношення $b - c \leq \frac{1}{4}$. Якщо ці вирази об'єднати, то отримаємо систему

$$\begin{cases} b - c \leq \frac{1}{4}, \\ c - b \leq \frac{1}{4}, \end{cases} \quad (2.21, б)$$

яка є ідентичною до умови $|b - c| \leq \frac{1}{4}$.

Оскільки множина $U = \{(b, c) \in R^2: \frac{1}{6} \leq b \leq \frac{7}{6}, 0 \leq c \leq 1\}$ і змінюється від $1/6$ до $7/6$ по осі b та від 0 до 1 по осі c , то на рис. 2.18 вона має вигляд квадрата. Таким чином, усі можливі моменти приходу обох чоловіків містяться в межах квадрата U .

Далі з цього квадрата потрібно виділити область A , яка описана системою (2.21, б) та міститься всередині U . Для цього побудуємо 2 прями, що описуються рівняннями $c - b = \frac{1}{4}$ та $b - c = \frac{1}{4}$. Шуканою буде область між цими прямими (тобто, розміщена нижче за пряму $c - b = \frac{1}{4}$ або ж $c = b + \frac{1}{4}$ та вище за $b - c = \frac{1}{4}$ або ж $c = b - \frac{1}{4}$), що міститься всередині квадрата U .

Очевидно, що ймовірність події A буде відношення площі заштрихованої множини до площі всього квадрата (рис. 2.18).

Площу квадрата U обрахувати легко – вона дорівнює 1 (сторони квадрата дорівнюють 1 кожна).

Для обчислення площі області A логічно обчислювати не саму A , а площу області U , що не потрапляє до A , і потім відняти її від площі U . Для цього обчислимо площі двох трикутників $c_1c_2c_3$ та $b_1b_2b_3$, що входять до U і не входять до A .

Точка c_1 перетину прямої $c - b = \frac{1}{4}$ із прямою $b = 1/6$ має координату $(1/6; 1/4 + 1/6)$, тобто $(1/6; 5/12)$, звідки довжина сторони c_1c_2 трикутника $c_1c_2c_3$ дорівнює $1 - 5/12 = 7/12$. Аналогічно знаходимо $b_1b_2 = 7/6 - 1/4 = 11/12$ (b_1 є точкою перетину прямої $b - c = \frac{1}{4}$ із прямою $c = 0$, тобто $b_1 = 1/4$). Точка c_3 перетину прямої $c - b = \frac{1}{4}$ із прямою $c = 1$ (верхня сторона квадрата) має координату $(3/4; 1)$, тобто сторона c_2c_3 трикутника $c_1c_2c_3$ має довжину $3/4 - 1/6 = 7/12$. Аналогічно обчислюємо довжину сторони b_2b_3 трикутника $b_1b_2b_3$ (точка b_3 є точкою перетину прямої $b - c = \frac{1}{4}$ із

прямою $b=7/6$, тобто b_3 має координати $(7/6; 7/6-1/4)=(7/6; 11/12)$. Отже, маємо 2 трикутники $b_1b_2b_3$ з катетами по $11/12$ та $c_1c_2c_3$ з катетами по $7/12$, що визначає площі трикутників відповідно як $\frac{1}{2} \cdot \frac{11}{12} \cdot \frac{11}{12} = \frac{121}{288}$ та $\frac{1}{2} \cdot \frac{7}{12} \cdot \frac{7}{12} = \frac{49}{288}$. Оскільки трикутників 2, то їх сумарна площа становить $\frac{121}{288} + \frac{49}{288} = \frac{170}{288} = \frac{85}{144}$, відповідно площа множини A дорівнює $1 - \frac{85}{144} = \frac{59}{144}$.

Отже, остаточно ймовірність події A така:

$$P(A) = \frac{\text{mes}A}{\text{mes}U} = \frac{\frac{59}{144}}{1} = \frac{59}{144},$$

де $\text{mes}A$ – це міра множини A , і в нашому випадку мірою є площа.

Приклад 2.49

(задача про зустріч із різним часом очікування)

Два чоловіки B та C домовилися про зустріч з 0-ї до 1-ї години дня так, що чоловік B не чекатиме іншого більше 15 хвилин, а чоловік C не чекатиме іншого більше 10 хвилин. Якщо при цьому другий чоловік не прийде, то перший іде з місця зустрічі. Обчислити ймовірність події $A = \{\text{зустріч відбудеться}\}$.

Таким чином, простір елементарних подій має вигляд $U = \{(b, c) \in R^2 : 0 \leq b \leq 1, 0 \leq c \leq 1\}$. Побудуємо простір U і зобразимо у вигляді квадрата на рис. 2.19.

B та C зустрінуться у випадках: 1) якщо першим прийде B (за умовою він чекає не більше 15 хвилин або $1/4$), то $c - b \leq 1/4$, тобто різниця між моментами їх приходу на зустріч не перевищує 15 хвилин або $1/4$, або 2) якщо першим прийде C (за умовою він чекає не більше 10 хвилин або $1/6$), то $b - c \leq 1/6$, тобто різниця між моментами їх приходу не перевищить 10 хвилин або $1/6$.

Побудуємо тепер подію A . Вона матиме вигляд $A = \{(b, c) \in R^2 : 0 \leq c \leq 1, 0 \leq b \leq 1, b - c \leq 1/6, c - b \leq 1/4\}$. Зобразимо її на рис. 2.19.

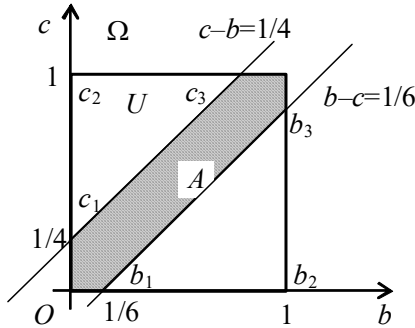


Рис. 2.19. Задача про зустріч із різним часом очікування

Розпишемо більш детально цю подію. Першим може прийти на зустріч як чоловік B , так і C . Якщо першим прийшов чоловік B , то момент часу b є меншим за момент часу c , і для виконання умов зустрічі має бути справедливим співвідношення $c - b \leq \frac{1}{4}$ (B чекає не більше 15 хвилин). Якщо ж першим прийшов C , то $c < b$ і має виконуватись вираз $b - c \leq \frac{1}{6}$ (C чекає не більше 10 хвилин). Якщо ці співвідношення об'єднати, то отримаємо систему

$$\begin{cases} b - c \leq \frac{1}{6}, \\ c - b \leq \frac{1}{4}. \end{cases} \quad (2.21, \text{в})$$

Оскільки множина $U = \{(b, c) \in \mathbb{R}^2 : 0 \leq b \leq 1, 0 \leq c \leq 1\}$ і змінюється від 0 до 1 по осі Ob та від 0 до 1 по осі Oc , то на рис. 2.19 вона має вигляд квадрата. Таким чином, усі можливі моменти приходу обох чоловіків містяться в межах квадрата U .

Далі з цього квадрата потрібно виділити область A , яка описана системою 2.21, в і розміщена всередині U . Для цього побудуємо 2 прямі що описуються рівняннями $c - b = \frac{1}{4}$ та $b - c = \frac{1}{6}$. Шуканою буде область між цими прямими (тобто нижче за прямою $c - b = \frac{1}{4}$ або ж $c = b + \frac{1}{4}$ та вище за $b - c = \frac{1}{6}$ або ж $c = b - \frac{1}{6}$), що перебуває всередині квадрата U .

Очевидно, що ймовірністю події A буде відношення площі заштрихованої множини до площі всього квадрата (рис. 2.19).

Площу квадрата U обрахувати легко – вона дорівнює 1 (сторони квадрата дорівнюють 1 кожна).

Для обчислення площі області A логічно обчислювати не саму A , а площу області U , що не потрапляє до A , і потім відняти її від площі U . Для цього обчислимо площі двох трикутників $c_1c_2c_3$ та $b_1b_2b_3$, що входять до U і не входять до A .

Точка c_1 перетину прямої $c-b = \frac{1}{4}$ з віссю Oc має координату $(0; 1/4)$, звідки довжина сторони c_1c_2 трикутника $c_1c_2c_3$ дорівнює $1-1/4=3/4$. Аналогічно знаходимо $b_1b_2=1-1/6=5/6$ (b_1 є точкою перетину прямої $b-c = \frac{1}{6}$ з віссю Ob , тобто $b_1=1/6$).

Точка c_3 перетину прямої $c-b = \frac{1}{4}$ із прямою $c = 1$ (верхня сторона квадрата) має координату $(3/4; 1)$, тобто, сторона c_2c_3 трикутника $c_1c_2c_3$ має довжину $3/4$. Аналогічно обчислюємо довжину сторони b_2b_3 трикутника $b_1b_2b_3$ (точка b_3 є точкою перетину прямої $b-c = \frac{1}{6}$ із прямою $b=1$, тобто b_3 має координати

$(1; 5/6)$). Отже, маємо 2 трикутники $b_1b_2b_3$ з катетами по $5/6$ та $c_1c_2c_3$ з катетами по $3/4$, що визначають площі трикутників від-

повідно як $\frac{1}{2} \cdot \frac{5}{6} \cdot \frac{5}{6} = \frac{25}{72}$ та $\frac{1}{2} \cdot \frac{3}{4} \cdot \frac{3}{4} = \frac{9}{32}$.

Оскільки трикутників 2, то їх сумарна площа становить $\frac{25}{72} + \frac{9}{32} = \frac{181}{288}$, а відповідно площа множини A дорівнює $1 - \frac{181}{288} = \frac{107}{288}$.

Отже, остаточно ймовірність події A така:

$$P(A) = \frac{mesA}{mesU} = \frac{\frac{107}{288}}{1} = \frac{107}{288},$$

де $mesA$ – це міра множини A , і в нашому випадку мірою є площа.

Парадокс Бертрана. Обчислення ймовірностей подій на незліченних нескінченних множинах породило цілу низку парадоксів. Наприклад, ймовірність потрапити в якусь фіксовану точку такої множини дорівнює 0, тобто ймовірність потрапити точно в центр мішені є нульовою. З іншого боку, потрапити в цю точку можна, і ми повинні розрізняти події, які не можуть відбутись і ті, що відбуваються з ймовірністю 0. Крім того, досить дивно виглядає той факт, що події {поцілити принаймні до однієї точки зі скінченної множини точок} та {поцілити тільки до однієї точки} мають однакову ймовірність 0.

Розглянемо відрізок $[0;1]$ та обчислимо ймовірність того, що навмання вибране число з цього відрізка буде не більше за 0,5. Ця ймовірність дорівнює 0,5. Але якщо всі числа з $[0, 1]$ піднести до квадрата, то ймовірність вибрати число, що не перевищує 0,5, буде приблизно 0,66. Звичайно більш природною у нашому випадку може видатись перша відповідь, хоча для більшості задач не є очевидним вибір "правильності" тієї чи іншої відповіді. Якраз у цьому і полягає суть парадокса Бертрана, що був опублікований у книжці "*Calculus des Probabilities*" у 1889 році Жозефом Луї Бертраном.

Приклад 2.50 (парадокс Бертрана)

Для деякого кола випадковим чином вибирається хорда (частина прямої, що двічі перетинає коло і міститься всередині круга). Знайти ймовірність події $A = \{ \text{довжина цієї хорди більша, за довжину сторони правильного (рівностороннього) трикутника, уписаного в коло} \}$.

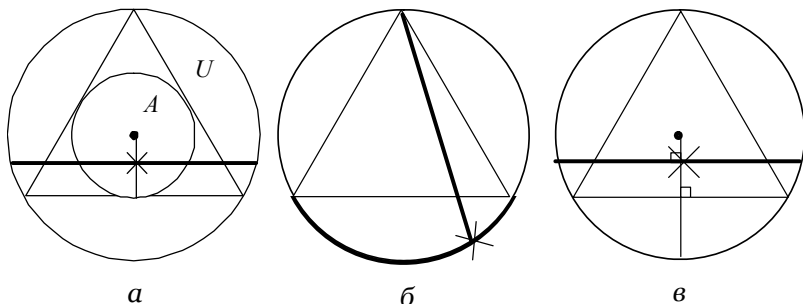


Рис. 2.20. Парадокс Бертрана

Метод 1

Випадковим чином (рівномірно) у цьому крузі вибирається точка (рис. 2.20, *a*). Вона однозначно визначає хорду, центром якої є. Ця хорда буде довшою за сторону вписаного в коло трикутника тоді й тільки тоді, коли вона лежить усередині кола, вписаного у трикутник. Радіус цього кола є половиною радіуса кола, у яке вписаний трикутник (описаного кола), отже, площа вписаного у трикутник круга A у 4 рази менша за площу круга U . Таким чином, імовірність події A становить $P(A) = \frac{1}{4}$.

Метод 2

Виходячи з міркувань симетрії, вважатимемо, що одним кінцем хорди є довільна точка на колі. Візьмемо за цю точку одну з вершин (рис. 2.20, *б*). Виберемо іншу точку на колі для протилежного кінця хорди, вважаючи, що на колі всі точки розташовані рівномірно. Вершини трикутника ділять коло на 3 дуги однакової довжини, тому хорда, яка буде довшою за сторону трикутника, матиме протилежний кінець на дузі, що є протилежною до вершини, з якої починається хорда, тобто хорда має перетинати трикутник. Отже, шуканою ймовірністю буде відношення довжини дуги до периметра кола: $P(A) = \frac{1}{3}$.

Метод 3

Виберемо точку випадковим чином рівномірно на радіусі кола, який проведений перпендикулярно до сторони трикутника, і візьмемо таку хорду, котра перпендикулярна до вибраного радіуса і проходить через вибрану точку (рис. 2.20, *в*). Тоді випадкова хорда буде довшою за сторону трикутника, якщо випадкова точка лежить на тій частині радіуса, що міститься всередині трикутника. Виходячи з міркувань симетрії, не є суттєвим, до якої зі сторін проведено радіус. Тому, ураховуючи, що сторона трикутника ділить радіус кола, проведений перпендикулярно до сторони, на 2 рівні частини, отримаємо шукану ймовірність $P(A) = \frac{1}{2}$.

Отримання різних результатів видається парадоксальним через те, що твердження "рівномірний випадковий вибір" однозначно визначає шукану ймовірність. Але сам парадокс ілюструє, що можливими є різні методи вибору рівномірним чином (на прямій, на колі та у крузі), причому кожен із методів

виглядає досить природно. На думку Пуанкаре (1854–1912) ("Calculus des Probabilités", Париж; 1912 р.), якщо немає ніякої додаткової інформації, то слід скористатися третім методом через те, що в цьому випадку, якщо маємо дві множини хорд, які є геометрично конгруентними (тобто, множини такі, що кожній хорді з однієї множини можна однозначно поставити у відповідність хорду з іншої множини), то з однаковими ймовірностями випадковим чином обрана хорда належатиме будь-якій із цих множин. Вивчення інваріантності (властивості залишатись без змін при певних перетвореннях) такого типу привело до виникнення нового розділу математики, що називається інтегральною геометрією (цей термін уперше ввів Вільгельм Бляшке (1885–1962) у 1934 р.).

ЗАДАЧІ

Класичне визначення ймовірності

№ 1

Симетричний гральний кубик підкидають 6 разів. Обчислити ймовірність події $A = \{\text{випадають усі шість граней}\}$.

№ 2

У коробці містяться 5 однакових занумерованих кубиків. Навмання по одному виймають усі кубики. Яка ймовірність того, що номери кубиків з'являться у зростаючому порядку?

№ 3

Студент прийшов на іспит, знаючи лише 20 із 25 питань програми. Екзаменатор задає студенту 3 питання. Знайти ймовірність того, що студент знає відповіді на всі питання.

№ 4

Гральний кубик підкидають 6 разів. Обчислити ймовірність того, що випадають тільки парні грані.

№ 5

Дев'ять пасажирів навмання сідають у три вагони. Яка ймовірність того, що в кожен вагон сяде по три пасажирі?

№ 6

Обчислити ймовірність того, що дні народження 12 осіб придуться на різні місяці року.

№ 7

Група складається із 20 студентів. Із них 12 хлопців та 8 дівчат. Випадковим чином обирається 5 студентів. Яка ймовірність того, що до цього числа ввійдуть 3 хлопця та 2 дівчини?

№ 8

Монету підкидають п'ять разів. Описати простір елементарних подій U та такі події: A – герб з'явиться один раз; B – герб з'явиться двічі; C – тричі; D – п'ять разів. Обчислити ймовірності подій A, B, C, D .

№ 9

В академічній групі 26 студентів, із них 8 хлопців та 18 дівчат. Групу навмання поділяють на дві рівні частини. Яка ймовірність того, що в кожній частині буде по 4 хлопці?

№ 10

Гральний кубик підкидають 10 разів. Яка ймовірність того, що при цьому цифра 1 з'явиться двічі, 3 – чотири рази, 5 – один раз, 6 – тричі?

№ 11

У ліфт дев'ятиповерхового будинку на першому поверсі зайшло п'ять пасажирів. Кожен із них з однаковою ймовірністю може вийти на будь-якому поверсі, крім першого. Знайти ймовірності таких випадкових подій:

- а) усі пасажери вийдуть на одному поверсі;
- б) усі пасажери вийдуть на п'ятому поверсі;
- в) усі пасажери вийдуть на різних поверхах.

№ 12

Є перетасована колода з 52 карт. Знайти ймовірність того, що перші 4 карти в колоді – тузи.

Геометрична ймовірність

№ 1

На паркетну підлогу кидають монету радіусом r . Паркет складається з прямокутників зі сторонами a і b ($a < b$, $2r < \min\{a, b\}$). Знайти ймовірність того, що монета не перетне жодної зі сторін паркету.

№ 2

Два судна повинні підійти до одного причалу. Появи суден – незалежні випадкові події, рівноможливі протягом доби. Знайти

ймовірність того, що одному із суден доведеться чекати звільнення причалу, якщо час стоянки першого судна (судна номер один) – 1 година, а другого – 2 години.

№ 3

На площині проведено паралельні прямі, відстань між якими дорівнює 8 см. На площину кидають монету радіусом 3 см. Яка ймовірність того, що монета не перетне жодну з прямих?

№ 4

На відрізок $[0,1]$ навмання наносять пару точок. Нехай x – координата однієї точки на відрізку $[0,1]$, y – іншої. Знайти такі ймовірності:

а) $\max\{x,y\} < \frac{1}{2}$;

б) $\max\{x,y\} > \frac{1}{3}$;

в) $\min\{x,y\} < \frac{1}{4}$;

г) $\min\{x,y\} > \frac{1}{2}$;

д) $\max\{x^2,y\} < a$, де $0 < a < 1$; е) $\max\{x,y^2\} > a$, де $0 < a < 1$.

№ 5

У крузі радіусом R навмання наносять точку. Яка ймовірність того, що відстань від цієї точки до центра круга не перевищує r ?

№ 6

У круг вписано квадрат. Точку навмання кидають у круг. Знайти ймовірність того, що вона потрапить у квадрат.

№ 7

У квадрат вписано круг. Точку навмання кидають у квадрат. Знайти ймовірність того, що вона потрапить у круг.

№ 8

У круг вписано правильний трикутник. Точку навмання кидають у круг. Знайти ймовірність того, що вона потрапить у трикутник.

2.7. Багатовимірні розподіли

2.7.1. Дискретні та неперервні багатовимірні розподіли

Розгляд числової характеристики результатів спостережень випадкового масового явища приводить до появи ймовірності, визначеної на підмножинах числової множини, – розподілу ймовірностей. Сама досліджувана числова характеристика називається випадковою величиною. Запис розподілу ймовірностей із використанням позначень для випадкової величини приводить до того, що цей розподіл називають розподілом ймовірностей випадкової величини.

Часто доводиться спостерігати одночасно декілька числових характеристик (випадкових величин). Це приводить до того, що спостерігається не одна випадкова величина, а кілька одночасно, тобто багатовимірний розподіл імовірностей або розподіл імовірностей з R^n , де n кількість досліджуваних числових характеристик. Сам набір досліджуваних числових характеристик (випадкових величин) називається системою випадкових величин (випадковим вектором) або багатовимірною випадковою величиною.

Багатовимірні розподіли

Розподілом імовірностей у R^n називається ймовірність на підмножинах U , що є частиною R^n . Так само, як і в одновимірному випадку, розрізняють дискретні та неперервні багатовимірні розподіли.

Багатовимірний розподіл називають дискретним, якщо U дискретне, та **неперервним**, якщо $U = R^n$ чи є частиною R^n із ненульовим значенням функції множини, яку загалом називають мірою Лебега (Анрі Леон Лебег (1875–1941)) для відповідного R^n . Якщо не обумовлено іншого, то при розгляді неперервного розподілу в R^n йтиметься про ймовірність на підмножинах R^n . При абстрактному (аксіоматичному) математичному заданні ймовірностей областю визначення ймовірності в неперервному випадку вважають усі ті множини, на яких визначена міра Лебега (довжина в R^1 , площа в R^2 , об'єм в R^3 тощо). Усі ці множини в R^n , на яких визначають імовірність у неперервному випадку, називають борелівськими (Еміль Борель (1871–1956)) множинами в R^n .

Як і в одновимірному випадку, задання розподілу ймовірностей в R^n залежить від того, яким він є (дискретним чи неперервним), тобто, дискретний розподіл у R^n задають рядом розподілу, а неперервний – функцією або щільністю розподілу.

2.7.2. Задання багатовимірних розподілів Задання дискретних розподілів у R^n

Дискретні розподіли в R^n задають рядом розподілу. Рядом розподілу в R^n називають сукупність усіх $u \in U$ та відповідних їм ймовірностей $P_u, P_u = P\{u\}$. Ймовірність будь-якої події A обчислюють за формулою

$$P(A) = \sum_{u \in A} p_u,$$

тобто, як суму ймовірностей тих елементарних подій, з яких вона складається.

Характеристичною властивістю ряду розподілу в R^n є невід'ємність та сумарне одиничне значення ймовірностей ряду розподілу.

Приклад 2.51

Розглянемо двовимірну дискретну випадкову величину (X, Y) , де X та Y є одновимірними випадковими величинами, що набувають значень $\{x_1, x_2, x_3\}$ та $\{y_1, y_2\}$ відповідно. Тоді ряд розподілу ймовірностей двовимірної випадкової величини $Z = (X, Y)$ складатиметься зі значень виду (x_i, y_j) та відповідних ймовірностей $P(X = x_i, Y = y_j) = p_{ij}$:

Z	(x_1, y_1)	(x_1, y_2)	(x_2, y_1)	(x_2, y_2)	(x_3, y_1)	(x_3, y_2)
P	p_{11}	p_{12}	p_{21}	p_{22}	p_{31}	p_{32}

Сума всіх ймовірностей $\sum_{i=1, j=1}^{3,2} p_{ij} = 1$, тобто, виконується властивість одиничного сумарного значення ймовірностей.

Задання неперервних розподілів у R^n

Неперервні розподіли у R^n задаються багатовимірною функцією розподілу чи багатовимірною щільністю розподілу.

Багатовимірною функцією розподілу або функцією розподілу в R^n (позначають $F(x_1, \dots, x_n)$) називають функцію, визначену на всьому R^n як імовірність появи значень числових векторів, координати яких одночасно не перевищують відповідних значень x_1, \dots, x_n . Це означає, що багатовимірна функція розподілу визначається як імовірність декартового добутку напівнескінченних ліворуч інтервалів $(-\infty; x_i)$, $i = 1, \dots, n$, і ця ймовірність є функцією від правих границь цих інтервалів:

$$F(x_1, \dots, x_n) = P\{(-\infty; x_1) \times \dots \times (-\infty; x_n)\} = P\{(X_1 < x_1) \times \dots \times (X_n < x_n)\}.$$

Характеристичними властивостями багатовимірної функції розподілу є такі (із відповідними уточненнями ті самі, що й в одновимірному випадку):

- 1) невід'ємність,
- 2) монотонне неспадання,
- 3) неперервність ліворуч,
- 4) нормованість.

Уточнення стосуються монотонного неспадання та нормованості. Монотонне неспадання означає, що векторові, усі координати якого одночасно більші за координати іншого, відповідає не менше значення функції розподілу. Тобто, якщо $x_i > y_i$ для всіх $i = 1, \dots, n$, то $F(x_1, \dots, x_n) \geq F(y_1, \dots, y_n)$.

Нормованість в уточненому варіанті означатиме, що наявність хоча б одного значення $-\infty$ серед значень аргументів дає нульове значення функції розподілу, а одночасне значення $+\infty$ для всіх значень аргументу – її одиничне значення. Тобто, якщо у векторі хоч один елемент (наприклад k -й) $x_k = -\infty$, то матимемо $F(x_1, \dots, x_k, \dots, x_n) = 0$, а якщо $x_i = +\infty$ для всіх $i = 1, \dots, n$, то $F(x_1, \dots, x_n) = 1$.

Багатовимірною щільністю розподілу чи щільністю розподілу в R^n (позначають $f(x_1, \dots, x_n)$) називають невід'ємну функцію, визначену на всьому R^n , за якою ймовірність будь-якої події з R^n обчислюється як визначений кратний інтеграл від цієї функції за цією подією.

Характеристичними властивостями багатовимірної щільності є невід'ємність та сумарне одиничне значення відповідного інтеграла за всім простором R^n .

Приклад 2.52

Двовимірною щільністю розподілу є невід'ємна функція $f(x,y)$, за якою ймовірність множини B обчислюють як двократний (подвійний) визначений інтеграл за цією множиною, тобто, як об'єм фігури над множиною B (рис. 2.21) під поверхнею, що є графіком функції $f(x,y)$:

$$P\{(x,y) \in B\} = \iint_{(x,y) \in B} f(x,y) dx dy = V_B^f.$$

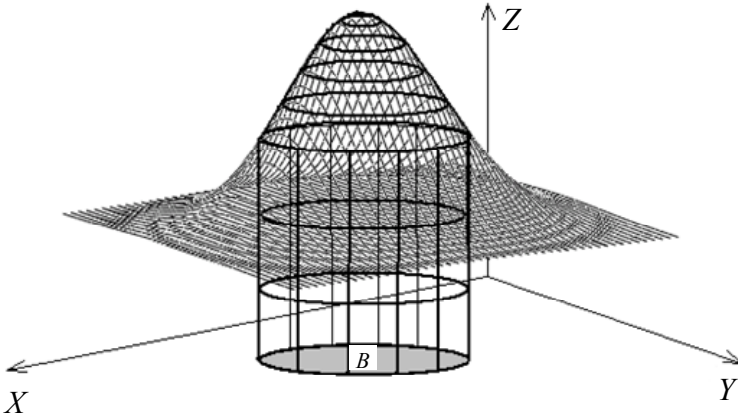


Рис. 2.21. Щільність двовимірного розподілу

2.7.3. Багатовимірні гауссівські розподіли

Двовимірним гауссівським чи нормальним (гауссівським або нормальним у R^2) розподілом називають розподіл імовірностей, який задають щільністю $f(x,y)$, котра визначається п'ятьма параметрами: $m_1, m_2, \sigma_1^2, \sigma_2^2$ та ρ (σ_1, σ_2 – додатні, $|\rho| \leq 1$) і задається співвідношенням

$$f(x,y) = \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}} e^{-\frac{1}{2(1-\rho^2)}\left(\frac{(x-m_1)^2}{\sigma_1^2} - \frac{2\rho(x-m_1)(y-m_2)}{\sigma_1\sigma_2} + \frac{(y-m_2)^2}{\sigma_2^2}\right)} \quad (2.22)$$

Якщо випадкові величини x та y є незалежними, то $\rho = 0$, а сама щільність двовимірного розподілу дорівнює добуткові відповідних щільностей одновимірних випадкових величин x та y :

$$\begin{aligned} f(x, y) &= f(x) \cdot f(y) = \frac{1}{\sqrt{2\pi}\sigma_1} e^{-\frac{1}{2}\left(\frac{(x-m_1)^2}{\sigma_1^2}\right)} \frac{1}{\sqrt{2\pi}\sigma_2} e^{-\frac{1}{2}\left(\frac{(y-m_2)^2}{\sigma_2^2}\right)} = \\ &= \frac{1}{2\pi\sigma_1\sigma_2} e^{-\frac{1}{2}\left(\frac{(x-m_1)^2}{\sigma_1^2} + \frac{(y-m_2)^2}{\sigma_2^2}\right)}. \end{aligned}$$

Змістове наповнення параметра ρ розглядатимемо в підрозд. 2.8.

Багатовимірним гауссівським розподілом (гауссівським розподілом в R^n) називають неперервний розподіл в R^n , який визначають двома параметрами: вектором $m \in R^n$ та симетричною, додатно визначеною матрицею B розмірності $n \times n$, і задають щільністю

$$f(x) = \frac{1}{(\sqrt{2\pi})^n |B|^{\frac{1}{2}}} e^{-\frac{1}{2}(x-m)^T B^{-1}(x-m)}, \quad x \in R^n,$$

де $(x-m)^T$ – транспонований (у цьому випадку повернутий вертикально) вектор-стовпець; B^{-1} – обернена до B матриця (тобто така, що $B \cdot B^{-1} = E$, тут E – одинична матриця

$$\begin{pmatrix} 1 & 0 & \dots & 0 & 0 \\ 0 & 1 & 0 & \dots & 0 \\ \dots & 0 & \dots & 0 & \dots \\ 0 & \dots & 0 & 1 & 0 \\ 0 & 0 & \dots & 0 & 1 \end{pmatrix}; |B| - \text{визначник матриці } B.$$

Для двовимірного випадку маємо

$$m = \begin{pmatrix} m_1 \\ m_2 \end{pmatrix}, \quad B = \begin{pmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{pmatrix}.$$

2.7.4. Системи випадкових величин, розподіли ймовірностей багатовимірних випадкових величин

Системою випадкових величин (багатовимірною випадковою величиною, випадковим вектором) називають декілька випадкових величин X_1, \dots, X_n , що розглядаються одночасно і пов'язані з результатом експерименту.

Наприклад, якщо для осіб, вибраних за тим або іншим принципом, розглядають одночасно дві числові характеристики: X – зріст, Y – вага, то ці характеристики утворюють пару (X, Y) із двох випадкових величин, яку називають системою із двох випадкових величин (двовимірною випадковою величиною чи двовимірним випадковим вектором). Самі випадкові величини, що утворюють систему, називають компонентами (маргіналами) цієї системи (компонентами багатовимірної випадкової величини, випадкового вектора).

Розподілом ймовірностей системи випадкових величин X_1, \dots, X_n називають розподіл ймовірностей в R^n , визначений на подіях з $U_{X_1, \dots, X_n} \subseteq R^n$ співвідношенням

$$P_{X_1, \dots, X_n}(B) = P\{(X_1, \dots, X_n) \in B\}.$$

Це співвідношення ще називають **спільним розподілом випадкових величин**, що складають систему.

Разом зі спільним розподілом існують розподіли ймовірностей кожної випадкової величини (компоненти) X_i , що входять до системи. На відміну від спільного розподілу, ці розподіли називають **маргінальними**.

Якщо компоненти випадкового вектора (X_1, \dots, X_n) – дискретні, то розподіл буде дискретним; якщо неперервні, то неперервним.

У першому випадку множиною можливих значень U_{X_1, \dots, X_n} системи випадкових X_1, \dots, X_n вважають множину всіх можливих комбінацій значень компонентів. За допомогою декартового добутку ця множина може бути представлена у вигляді

$$U_{X_1, \dots, X_n} = U_{X_1} \times \dots \times U_{X_n}.$$

У другому випадку (неперервний) множиною можливих значень є весь простір R^n :

$$U_{X_1, \dots, X_n} = R^n.$$

У цьому разі R^n є комбінацією значень випадкових величин чи системи, кожна з яких набуває значень з R^1 .

Так для системи з двох дискретних випадкових величин: X зі значеннями з $U_X = \{x_1, \dots, x_n\}$ та Y зі значеннями з $U_Y = \{y_1, \dots, y_m\}$, – множиною можливих значень U_{XY} системи (X, Y) будуть усі можливі пари (x_i, y_j) , $i=1, \dots, n$, $j=1, \dots, m$:

$$U_{XY} = U_X \times U_Y = \{(x_i, y_j) : x_i \in U_X, y_j \in U_Y, i=1, \dots, n, j=1, \dots, m\}.$$

Для системи (X, Y) із двох неперервних величин множиною можливих значень системи є множина всіх можливих пар дійсних чисел: $U_{XY} = R^2$.

Задання розподілів імовірностей багатовимірних випадкових величин

Оскільки системи випадкових величин є звичайними багатовимірними розподілами, то вони задаються так само, як останні: рядом розподілу для дискретного випадку та функцією чи щільністю розподілу для неперервного. Щоб відрізнити спільні характеристики від відповідних маргінальних, їх називають спільним рядом, функцією чи щільністю розподілу системи випадкових величин.

Якщо багатовимірний розподіл імовірностей розглядають як розподіл імовірностей системи випадкових величин, то функцію розподілу випадкового вектора записують у вигляді

$$F(x_1, \dots, x_n) = P\{X_1 < x_1, X_2 < x_2, \dots, X_n < x_n\}.$$

Коми у правій частині останньої формули означають добуток відповідних подій.

Імовірності ряду розподілу записують через імовірність того, що відповідні компоненти системи одночасно збігаються з відповідними компонентами $U \in R^n$.

Зв'язок спільного та маргінальних розподілів

Для отримання ряду, функції чи щільності маргінального розподілу, що відповідає компоненті з номером i : $i = 1, \dots, n$, потрібно:

1) щоб одержати ймовірності маргінального ряду розподілу – просумувати ймовірності спільного ряду по значеннях усіх компонент, крім обраної (приклад 2.53) ;

2) для знаходження маргінальної функції розподілу – підставити $+\infty$ замість аргументів усіх компонент, крім обраної, у спільну функцію розподілу ($F_k(x_k) = P\{X_1 < +\infty, X_2 < +\infty, \dots, X_k < x_k, \dots, X_n < +\infty\}$);

3) для отримання маргінальної щільності – зінтегрувати спільну щільність по всіх компонентах, крім обраної.

Приклад 2.53

Дискретна випадкова величина $Z = (X, Y)$ задана сумісним рядом розподілу:

Z	(1;1)	(1;2)	(2;1)	(2;2)	(3;1)	(3;2)
P	0,1	0,2	0,1	0,3	0,2	0,1

Сума всіх імовірностей $\sum_{i=1, j=1}^{3,2} p_{ij} = 1$. Обчислимо маргінальні

ймовірності:

$$P(X=1) = \sum_{j=1}^2 p_{1j} = 0,1 + 0,2 = 0,3,$$

$$P(X=2) = \sum_{j=1}^2 p_{2j} = 0,1 + 0,3 = 0,4,$$

$$P(X=3) = \sum_{j=1}^2 p_{3j} = 0,2 + 0,1 = 0,3,$$

$$P(Y=1) = \sum_{i=1}^3 p_{i1} = 0,1 + 0,1 + 0,2 = 0,4,$$

$$P(Y=2) = \sum_{i=1}^3 p_{i2} = 0,2 + 0,3 + 0,1 = 0,6.$$

Маргінальні ряди розподілу мають вигляд

X	1	2	3	Y	1	2
P	0,3	0,4	0,3	P	0,4	0,6

2.8. Інтегральні характеристики багатовимірного розподілу або системи випадкових величин

2.8.1. Інтегральні характеристики системи випадкових величин на базі середнього

Інтегральні характеристики на базі середнього для системи випадкових величин визначають, як і в одновимірному випадку з тією різницею, що функції, за якими їх визначають, є функціями декількох аргументів. Інтегральна характеристика на базі середнього (для системи X_1, \dots, X_n), яка побудована функцією $G(x_1, \dots, x_n)$, є граничне значення середнього спостережень системи випадкових величин, перерахованих за допомогою функції G . Це граничне значення називають математичним сподіванням функції G від системи X_1, \dots, X_n і позначають $MG(X_1, \dots, X_n)$:

$$MG(X_1, \dots, X_n) = \lim_{N \rightarrow \infty} \frac{G(a_1) + G(a_2) + \dots + G(a_N)}{N},$$

із тим уточненням, що елементи вибірки a_1, \dots, a_N є числовими векторами розмірності n (а не числами, як в одновимірному випадку).

Усі інтегральні характеристики компонентів на основі функції g , визначені раніше для випадкових величин, є інтегральними характеристиками і для системи випадкових величин. Дійсно, беручи для тої чи іншої компоненти з номером i $G(x_1, \dots, x_n) = g(x_i)$, маємо

$$MG(X_1, \dots, X_n) = Mg(X_i).$$

Зокрема, це стосується математичних сподівань та дисперсій компонент:

$$MG(X_1, \dots, X_n) = MX_i$$

для $G(x_1, \dots, x_n) = x_i$,

$$MG(X_1, \dots, X_n) = DX_i = M(X_i - MX_i)^2$$

для $G(x_1, \dots, x_n) = (X_i - MX_i)^2$, $i = 1, \dots, n$.

До інтегральних характеристик системи випадкових величин належать **моменти та центральні моменти системи випадкових величин**:

моменти – $M(X_1^{k_1} \cdot X_2^{k_2} \dots X_n^{k_n})$,

центральні моменти – $M(X_1^{k_1} - MX_1) \cdot (X_2^{k_2} - MX_2) \dots (X_n^{k_n} - MX_n)$

відповідно, які визначають набором невід'ємних цілих чисел k_1, \dots, k_n . Суму степенів k_1, \dots, k_n називають **порядком моменту**.

Першим моментом є маргінальне математичне сподівання, тобто математичне сподівання однієї випадкової величини.

Другими центральними моментами є $M(X_i - MX_i)(X_j - MX_j)$ для всіх $i, j=1, 2, \dots, n$. Кожен із цих центральних моментів, що відповідає номерам i, j , називають **коваріацією i -ї та j -ї компонент** і позначають як $\text{cov}(X_i, X_j)$ або $C_{X_i X_j}$. Для $i = j$ коваріація збігається з дисперсією відповідної компоненти.

Коефіцієнт асиметрії

Якщо третій центральний момент $M(X_i - MX_i)(X_j - MX_j)(X_k - MX_k)$ при $i = j = k$ нормалізувати через маргінальну дисперсію

у степені 3/2 $\gamma_1 = \frac{M(X_i - MX_i)^3}{\sigma_i^3} = \frac{M(X_i - MX_i)^3}{(M(X_i - MX_i)^2)^{3/2}}$, то отримає-

мо числову характеристику симетрії, яку називають **коефіцієнтом асиметрії** (позначають γ_1).

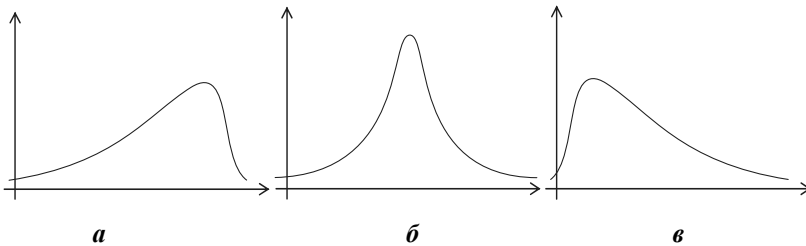


Рис. 2.22. Асиметрія:

a – лівий скіс ($\gamma_1 < 0$); b – симетрія ($\gamma_1 = 0$); v – правий скіс ($\gamma_1 > 0$)

Якщо коефіцієнт асиметрії дорівнює 0, то розподіл є симетричним (рис. 2.22, б), якщо від'ємним, то "довша частина" розподілу міститься ліворуч від математичного сподівання (лівий скіс, рис. 2.22, а), якщо додатна, то праворуч (правий скіс, рис. 2.22, в).

Такий коефіцієнт асиметрії називають коефіцієнтом асиметрії Фішера. Крім того, у теорії ймовірностей та математичній статистиці розглядають ще коефіцієнт асиметрії Пірсона $\beta_1 = (\gamma_1)^2$ та коефіцієнт модальної асиметрії Пірсона.

Ексцес

Якщо четвертий центральний момент

$$M(X_i - MX_i)(X_j - MX_j)(X_k - MX_k)(X_l - MX_l)$$

при $i = j = k = l$

нормалізувати через квадрат маргінальної дисперсії і відняти

3, то матимемо
$$\gamma_2 = \frac{M(X_i - MX_i)^4}{\sigma_i^4} - 3 = \frac{M(X_i - MX_i)^4}{(M(X_i - MX_i)^2)^2} - 3$$
 і

отримаємо **коефіцієнт ексцесу** (позначають γ_2), який указує наскільки яскраво виражена вершина маргінального розподілу в околі маргінального середнього. Коефіцієнт ексцесу набуває значень від -2 до $+\infty$, $\gamma_2 \in [-2; +\infty)$. Чим більший коефіцієнт ексцесу, тим стрімкішим та вищим є пік графіка щільності.

Для стандартного нормального розподілу коефіцієнт ексцесу дорівнює 0.

Коваріація. Властивості коваріації. Коефіцієнт кореляції

Якщо випадкова величина (система випадкових величин) складається з двох компонент, то **коваріацію X та Y** позначають так:

$$C_{XY} = \text{cov}(X, Y) = M(X - MX)(Y - MY).$$

Якщо позначити як X випадковий вектор розмірності n , то матрицю (C_{XX}) , розмірності $n \times n$, складену з коваріацій компонент, називають **матрицею коваріацій системи** випадкових величин: $(C_{XX}) = (\text{cov}(X_i, X_j))$.

Загалом говорять про **матрицю взаємних коваріацій двох систем** випадкових величин $X_1, \dots, X_n, Y_1, \dots, Y_m$, як про $n \times m$ -матрицю C_{XY} , складену з коваріацій кожної компоненти системи X із кожною компонентою системи Y : $(C_{XY}) = (\text{cov}(X_i, Y_j))$. Для C_{XY} виконується співвідношення $|C_{XY}| \leq (DX \cdot DY)^{1/2}$.

Разом із коваріацією розглядають величину, яку називають **коефіцієнтом кореляції Пірсона** (Карл Пірсон (1857–1936)),

іноді нормованим коефіцієнтом кореляції, позначають ρ_{XY} або ρ і визначають за співвідношенням:

$$\rho_{XY} = \frac{C_{XY}}{\sqrt{DX \cdot DY}}. \quad (2.23)$$

Коефіцієнт кореляції за модулем не може перевищувати 1, $|\rho_{XY}| \leq 1$.

Коваріація та коефіцієнт кореляції використовуються для опису зв'язку між випадковими величинами, ці питання розглядатимемо далі в підрозд. 2.8.

Зміст параметрів багатовимірного гауссівського розподілу

Для двовимірного нормального розподілу у (2.22) параметр ρ є коефіцієнтом кореляції (2.23).

Для багатовимірного гауссівського розподілу перший параметр (вектор m) складається з математичних сподівань компонент цього випадкового вектора, а другий параметр (матриця B) є матрицею коваріацій цього випадкового вектора.

Інтегральні характеристики компонент підраховують звичайним чином за маргінальними розподілами (розподілами компонент): за рядом чи щільністю залежно від дискретності чи неперервності відповідного спільного розподілу.

Загалом **формули обчислення інтегральних характеристик на базі середнього** для системи випадкових величин повністю повторюють обчислювальні формули для випадкових величин із тією різницею, що елементарні події у відповідних формулах належать R^n :

$$MG(X_1, \dots, X_n) = \begin{cases} \sum_{u \in U} G(u) p_u, & \text{за спільним рядом розподілу,} \\ \int_{R^n} G(x) f(x) dx, & \text{за спільною щільністю розподілу.} \end{cases}$$

2.8.2. Незалежність випадкових величин. Властивості незалежності

Випадкові чи багатовимірні випадкові величини X та Y називають **незалежними**, якщо будь-які події, пов'язані з їх значен-

нями, є незалежними. Це означає, що для будь-яких "числових" подій B_1, B_2 події $\{X \in B_1\}$ та $\{Y \in B_2\}$ є незалежними:

$$P\{X \in B_1, Y \in B_2\} = P\{X \in B_1\} P\{Y \in B_2\}.$$

Для системи X із випадкових величин X_1, \dots, X_n це означає, що для довільних "числових" подій B_1, \dots, B_n маємо

$$P\{X_1 \in B_1, \dots, X_n \in B_n\} = P\{X_1 \in B_1\} \cdots P\{X_n \in B_n\}.$$

Під **спадковістю незалежності для випадкових величин** (одновимірних чи векторних) розуміють передачу незалежності від них до перерахованих за ними значень. Тобто, якщо X та Y є незалежними між собою випадковими величинами чи випадковими векторами, а G_1, G_2 – будь-якими "припустимими" функціями, то $G_1(X), G_2(Y)$ є незалежними між собою випадковими величинами. "Припустимість" математичною мовою означає виконання тієї властивості, що називається вимірністю щодо числових подій в області визначення та області значення відповідної функції (повний прообраз "числової" події є "числовою" подією). Такі "припустимі" функції у математиці називають борелівськими.

Умови незалежності випадкових величин

Для незалежності двох дискретних випадкових величин чи систем випадкових величин необхідно і достатньо, щоб імовірності ряду розподілу дорівнювали добутку ймовірностей компонент (маргінальних імовірностей):

$$P_{XY}(x_i, y_j) = P_X(x_i) \cdot P_Y(y_j).$$

Необхідною та достатньою умовою незалежності для неперервної системи випадкових величин є збігання спільної функції розподілу з добутком маргінальних:

$$F_{XY}(x, y) = F_X(x) \cdot F_Y(y).$$

У випадку, коли існує спільна щільність розподілу, необхідною і достатньою умовою незалежності є вимога збігання спільної щільності з добутком маргінальних:

$$f_{XY}(x, y) = f_X(x) \cdot f_Y(y).$$

Математичне сподівання добутку незалежних випадкових величин

Незалежність додає до властивостей математичного сподівання ще одну. Згідно з нею математичне сподівання добутку

незалежних випадкових величин X та Y дорівнює добутку математичних сподівань співмножників:

$$M(X \cdot Y) = M X \cdot M Y.$$

Зважаючи на спадковість, можна стверджувати, що для "припустимих" функцій G_1 та G_2 для незалежних систем випадкових величин X та Y

$$M G_1(X) \cdot G_2(Y) = M G_1(X) \cdot M G_2(Y).$$

Незалежність та некорельованість

Із спадковості незалежності та властивостей математичного сподівання для добутку незалежних випадкових величин випливає, що коваріація незалежних випадкових величин X та Y дорівнює нулю:

$$\begin{aligned} C_{XY} = \text{cov}(X, Y) &= M(X - M X)(Y - M Y) = M(X - M X) \cdot M(Y - M Y) = \\ &= (M X - M M X) \cdot (M Y - M M Y) = (M X - M X) \cdot (M Y - M Y) = 0. \end{aligned}$$

Очевидним чином, коваріація та коефіцієнт кореляції дорівнюють нулю одночасно, тому незалежність означає і нульовий коефіцієнт кореляції.

Дві незалежні випадкові величини є прикладом пари некорельованих випадкових величин.

Отже, дві випадкові величини називають **некорельованими**, якщо їх коваріація (коефіцієнт кореляції) дорівнює нулю. Дві випадкові величини називають **корельованими**, якщо вони не є некорельованими.

Зауваження. **Некорельованість** застосовують також для **багатовимірних випадкових величин** X та Y із тією різницею, що нульовою є їхня матриця взаємних коваріацій. Бачимо, що з незалежності випливає некорельованість. Обернене твердження в загальному випадку не є справедливим, але виконується для багатовимірних гауссівських випадкових величин (компонент гауссівського вектора) через те, що для них із некорельованості випливає незалежність. Отже, для компонент гауссівського вектора незалежність є еквівалентною некорельованості.

Таким чином, некорельованість – це властивість, слабша за незалежність, але така, що використовується для характеристики зв'язку між випадковими величинами. Ця можливість впливає з твердження про розклад однієї випадкової величини (Y) за іншою (X).

Теорема 2.12

Для двох випадкових величин X та Y зі скінченними значеннями математичних сподівань MX та MY і дисперсій DX та DY знайдеться випадкова величина Z , некорельована з X , яка має нульове математичне сподівання та дисперсію $DY \cdot (1 - \rho_{XY}^2)$ така, що

$$Y = MY + \frac{C_{XY}}{DX} \cdot (X - MX) + Z. \quad (2.24)$$

Для спільних гауссівських випадкових величин (спільний розподіл гауссівський) у розкладі (2.24) X та гауссівська Z незалежні.

2.8.3. Дисперсія суми та різниці випадкових величин

Дисперсію суми випадкових величин обчислюють за формулою

$$D(X+Y) = DX + DY + 2C_{XY}.$$

Дійсно, із властивостей дисперсії

$$\begin{aligned} D(X+Y) &= M((X+Y) - M(X+Y))^2 = M((X+Y) - MX - MY)^2 = \\ &= M(X - MX + Y - MY)^2 = M((X - MX) + (Y - MY))^2 = \\ &= M((X - MX)^2 + 2(X - MX) \cdot (Y - MY) + (Y - MY)^2) = \\ &= M(X - MX)^2 + 2M(X - MX) \cdot (Y - MY) + M(Y - MY)^2 = \\ &= DX + 2C_{XY} + DY, \end{aligned}$$

тут C_{XY} є коваріацією між X та Y .

Оскільки для незалежних випадкових величин $C_{XY} = 0$, то дисперсія суми незалежних величин дорівнює сумі дисперсій: $D(X+Y) = DX + DY$.

Аналогічно різницю двох випадкових величин обчислюють як

$$D(X-Y) = DX + DY - 2C_{XY}.$$

Дійсно, із властивостей дисперсії

$$\begin{aligned} D(X-Y) &= M((X-Y) - M(X-Y))^2 = M((X-Y) - MX + MY)^2 = \\ &= M(X - MX - Y + MY)^2 = M((X - MX) - (Y - MY))^2 = \\ &= M((X - MX)^2 - 2(X - MX) \cdot (Y - MY) + (Y - MY)^2) = \\ &= M(X - MX)^2 - 2M(X - MX) \cdot (Y - MY) + M(Y - MY)^2 = \\ &= DX - 2C_{XY} + DY. \end{aligned}$$

2.9. Однаково розподілені випадкові величини

2.9.1. Однакові розподіли, однаково розподілені випадкові величини

Розподіли ймовірностей розглядають у зв'язку зі спостереженнями числових характеристик і описують частоту появи тих чи інших результатів або групи результатів. Те саме стосується багатовимірних розподілів тільки для спостережень декількох числових характеристик. Розподіли (одновимірні чи багатовимірні) називають **однаковими**, якщо на однакових подіях вони мають однакові значення ймовірностей. Для збігання (однаковості) розподілів необхідною і достатньою умовою є збігання рядів розподілів для дискретних розподілів, а також функцій розподілів чи щільностей для неперервних розподілів ймовірностей.

Випадкові величини X та Y (одно- чи багатовимірні) називають **однаково розподіленими**, якщо вони мають однакові розподіли, тобто набувають значень з однакових "числових" подій однаково часто. Тобто, випадкові величини X та Y є однаково розподіленими, якщо для довільної "числової" події B

$$P\{X \in B\} = P\{Y \in B\}.$$

Однаково розподілені випадкові величини чи системи випадкових величин мають однакові інтегральні характеристики на базі середнього, оскільки їх обчислюють за рядом чи щільністю розподілу, а вони в однаково розподілених випадкових величинах збігаються.

2.9.2. Незалежні однаково розподілені випадкові величини

Незалежні однаково розподілені випадкові величини X_1, \dots, X_N , зокрема, описують значення числових характеристик, які можуть з'явитися в серії з n незалежних спостережень (дослідів). Власне, вибірка a_1, \dots, a_N є послідовністю спостережуваних значень (результатів спостережень) цих випадкових величин.

Послідовність незалежних однаково розподілених випадкових величин X_1, \dots, X_N, \dots відповідає послідовності незалежних дослідів, в яких спостерігається одна і та сама числова характеристика X (X_1 – те значення, яке може спостерігатися в першому досліді, X_2 – у другому і т. д.).

Приклад 2.54

Послідовність X_i ($i = 1, \dots, n$) спостережень бернуллівських випадкових величин є послідовністю незалежних однаково розподілених випадкових величин, що описують кількість появ події A в окремих спостереженнях СНВБ:

$$X_i = \begin{cases} 0, & \text{коли у } i\text{-му досліді } \bar{A}; \\ 1, & \text{коли у } i\text{-му досліді } A. \end{cases}$$

Ряд розподілу кожної з таких випадкових величин (маргінальний ряд розподілу) має вигляд

$$X_i = \begin{cases} 0, & q = 1 - P(A); \\ 1, & p = P(A). \end{cases}$$

Отже, усі вони є однаково розподіленими, а оскільки відповідають незалежним спостереженням, то й незалежні. Усі X_i , $i = 1, \dots, n$, відповідають спостереженням у різних незалежних дослідях однієї й тієї самої числової характеристики $X = \{\text{кількість появ } A \text{ в одному спостереженні}\}$.

2.10. Граничні теореми

Закон великих чисел (ЗВЧ) та центральну граничну теорему (ЦГТ) розглядають на послідовності незалежних однаково розподілених випадкових величин. Розглядають послідовність незалежних дослідів, в яких спостерігається одна і та сама числова характеристика. Різниця полягає в перетвореннях цих випадкових величин. У ЗВЧ через усереднення отримують одне граничне значення, для ЦГТ через центрування суми випадкових величин її математичним сподіванням та нормування коренем із дисперсії – інше (центрування випадкової величини означає дода-

вання чи віднімання не випадкової константи, а нормування є діленням на не випадкову константу).

Загалом, усереднення незалежних однаково розподілених випадкових величин зберігає значення математичного сподівання, зменшуючи в N разів дисперсію:

$$\begin{aligned} M\left(\frac{X_1 + \dots + X_N}{N}\right) &= \frac{1}{N} M(X_1 + \dots + X_N) = \\ &= \frac{1}{N} (MX_1 + \dots + MX_N) = \frac{1}{N} (m + \dots + m) = \frac{1}{N} Nm = m; \\ D\left(\frac{X_1 + \dots + X_N}{N}\right) &= \frac{1}{N^2} D(X_1 + \dots + X_N) = \\ &= \frac{1}{N^2} (DX_1 + \dots + DX_N) = \frac{1}{N^2} \cdot N \cdot \sigma^2 = \frac{\sigma^2}{N}. \end{aligned}$$

2.10.1. Закон великих чисел

Теорема 2.13 (закон великих чисел)

Нехай $X_1, X_2, \dots, X_N, \dots$ є послідовністю незалежних однаково розподілених випадкових величин зі спільним математичним сподіванням $MX = m$: $MX_i \equiv MX = m, i = 1, 2, \dots$

Тоді середнє за початковими N випадковими величинами за нескінченного збільшення кількості елементів усереднення збігається до спільного математичного сподівання MX :

$$\frac{1}{N} \sum_{i=1}^N X_i \xrightarrow{N \rightarrow \infty} MX = m. \quad (2.25)$$

Наведена збіжність математично може бути такою, що називається "за ймовірністю", чи такою, що визначається як "майже напевно". У другому випадку ЗВЧ називається підсиленням.

Підсилений ЗВЧ за абстрактного визначення ймовірностей може слугувати для обґрунтування визначення інтегральних характеристик на базі середнього у частотній теорії.

Власне у ЗВЧ стверджується, що збіжність середнього до границі (математичного сподівання) відбувається для кожної можливої вибірки при зростанні обсягу вибірки.

2.10.2. Центральна гранична теорема

Зміст центральної граничної теореми полягає в тому, що ймовірності, пов'язані із сумою незалежних однаково розподілених випадкових величин $X_1 + X_2 + \dots + X_N$ тим точніше описуються нормальним розподілом, чим більша кількість доданків у сумі. Параметрами нормального розподілу є сумарне значення математичного сподівання та дисперсії відповідно. Це означає

$$P(X_1 + X_2 + \dots + X_N \in B) \approx P(Y \in B),$$

де випадкова величина Y має $N \left(M \sum_{i=1}^N X_i, D \sum_{i=1}^N X_i \right)$ -розподіл, тобто

нормальний розподіл із параметрами $M \sum_{i=1}^N X_i = N \cdot MX = n \cdot m$ та

$$D \sum_{i=1}^N X_i = N \cdot DX = N \cdot \sigma^2.$$

Точний зміст ЦГТ передається теоремою 2.14.

Теорема 2.14 (ЦГТ)

Нехай $X_1, X_2, \dots, X_N, \dots$ – послідовність незалежних однаково розподілених випадкових величин з однаковим математичним сподіванням $MX = m$ та однаковою дисперсією $DX = \sigma^2$:

$$MX_i \equiv MX = m, DX_i \equiv DX = \sigma^2, i = 1, 2, \dots,$$

Z – випадкова величина зі стандартним гауссівським розподілом ($N(0,1)$ -розподілом).

Тоді для довільної "числової події" B

$$P \left\{ \frac{\sum_{i=1}^N X_i - N \cdot m}{\sigma \sqrt{N}} \in B \right\} \rightarrow P\{Z \in B\}.$$

Твердження ЦГТ записують також у позначеннях, що використовують замість суми випадкових величин $X_1 + X_2 + \dots + X_N$

їхнє середнє \bar{X} : $\bar{X} = \frac{1}{N} \sum_{i=1}^N X_i$, ураховуючи, що

$$\frac{\sum_{i=1}^N X_i - N \cdot m}{\sigma \sqrt{N}} = \frac{N \cdot \left(\frac{1}{N} \sum_{i=1}^N X_i - m \right)}{\sigma \sqrt{N}} = \frac{(\bar{X} - m) \cdot \sqrt{N}}{\sigma}.$$

Отже, справедливою є така теорема.

Теорема 2.15

У припущеннях теореми 2.14 для довільної числової події B

$$P\left\{\frac{(\bar{X} - m) \cdot \sqrt{N}}{\sigma} \in B\right\} \rightarrow P\{Z \in B\}. \quad (2.26)$$

Отже, у ЦГТ стверджують, що з погляду обчислення ймовірностей, розподіл нормованої і центрованої відповідним чином суми випадкових величин приблизно є стандартним нормальним

$$\frac{\sum_{i=1}^N X_i - N \cdot m}{\sigma \sqrt{N}} \approx Z.$$

Із цього випливає, що

$$\sum_{i=1}^N X_i \approx N \cdot m + Z \cdot \sigma \cdot \sqrt{N}.$$

В останньому співвідношенні права частина має гауссівський розподіл із параметрами $N \cdot m$ та $N \cdot \sigma^2$, що дає можливість наближення розподілу суми розподілом гауссівської випадкової величини.

Нормальність розподілу випадкової величини $N \cdot m + Z \cdot \sigma \cdot N^{1/2}$ визначається важливою властивістю гауссівських розподілів, яка полягає в замкненості цих розподілів щодо лінійних перетворень. Це означає, що лінійна функція від випадкової величини з гауссівським розподілом сама розподілена за гауссівським законом. Вираз $N \cdot m + Z \cdot \sigma \cdot \sqrt{N}$ є лінійною функцією від гауссівської випадкової величини Z , а, отже, відповідно до цієї властивості, має гауссівський розподіл. Перший параметр будь-якого гауссівського розподілу є математичним сподіванням а другий – дисперсією відповідної випадкової величини. У досліджуваному випадку, зважаючи на те, що $MZ = 0$, а $DZ = 1$, це $N \cdot m$ та $N \cdot \sigma^2$ відповідно:

$$M(N \cdot m + Z \cdot \sigma \cdot \sqrt{N}) = M(N \cdot m + MZ \cdot \sigma \cdot \sqrt{N}) = N \cdot m.$$

$$D(N \cdot m + Z \cdot \sigma \cdot \sqrt{N}) = D(Z \cdot \sigma \cdot \sqrt{N}) = (\sigma \cdot \sqrt{N})^2 DZ = \sigma^2 \cdot N.$$

2.10.3. Обчислення функції розподілу нормального розподілу через функцію Лапласа

Центральна гранична теорема окреслює важливе місце гаусівських розподілів як універсального засобу наближеного обчислення ймовірностей у зв'язку із сумами випадкових величин. У цьому контексті зазначимо, що обчислення ймовірностей для будь-якого нормального розподілу може бути зведене до обчислення ймовірностей для стандартного нормального розподілу, для якого функція розподілу позначається $\Phi(x)$: $\Phi(x) = P\{Z < x\}$ і називається функцією Лапласа. Значення цієї функції наводять у всіх статистичних таблицях, тому що його не можна записати простішим аналітичним виразом ніж той, що пов'язує її зі щільністю стандартного нормального розподілу:

$$\Phi(x) = \int_{-\infty}^x \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}y^2} dy.$$

Теорема 2.16

Для довільної $N(m, \sigma^2)$ розподіленої випадкової величини Y значення її функції розподілу $F_Y(x)$ пов'язують із функцією Лапласа співвідношенням

$$F_Y(x) = P\{Y < x\} = \Phi\left(\frac{x-m}{\sigma}\right). \quad (2.27)$$

Доведення впливає з двох невеличких лем:

Лема 2.1

Для будь-якої $N(m, \sigma^2)$ розподіленої випадкової величини Y випадкова величина $Z = \frac{Y-m}{\sigma}$ має стандартний нормальний розподіл.

Дійсно, записуючи функцію розподілу Z та підставляючи її вираз через Y , дістаємо

$$\begin{aligned} F_Z(x) &= P\{Z < x\} = P\left\{\frac{Y-m}{\sigma} < x\right\} = P\{Y-m < \sigma \cdot x\} = \\ &= P\{Y < m + \sigma \cdot x\} = F_Y(m + \sigma \cdot x). \end{aligned}$$

Отже, щільність розподілу Z , як похідна для функції розподілу, має вигляд

$$f_Z(x) = F_Z'(x) = (F_Y(m + \sigma x))' = (m + \sigma x)' \cdot F_Y'(m + \sigma x) = \sigma \cdot f_Y(m + \sigma x).$$

Враховуючи, що щільність розподілу $f_Y(x)$ для $N(m, \sigma^2)$ -розподіленої випадкової величини Y має вигляд

$$f_Y(x) = \frac{1}{\sqrt{2\pi}\sigma} \cdot e^{-\frac{1}{2\sigma^2}(x-m)^2},$$

остаточно отримуємо

$$f_Z(x) = \sigma \cdot f_Y(m + \sigma x) = \sigma \cdot \frac{1}{\sqrt{2\pi}\sigma} \cdot e^{-\frac{1}{2\sigma^2}(m+\sigma x-m)^2} = \frac{1}{\sqrt{2\pi}} \cdot e^{-\frac{x^2}{2}}.$$

Останній вираз є щільністю стандартного нормального розподілу. Лему 2.1 доведено.

Лема 2.2

Будь-яку випадкову величину Y , що має $N(m, \sigma^2)$ -розподіл можна подати у вигляді

$$Y = m + \sigma \cdot Z.$$

Доведення леми 2.2 випливає з очевидного представлення для Y та використання леми 2.1:

$$Y = m + \sigma \frac{Y - m}{\sigma} = m + \sigma Z.$$

Доведення теореми 2.16

Дійсно, записуючи вираз функції розподілу для Y та скориставшись лемою 2.2, отримуємо

$$\begin{aligned} F_Y(x) &= P\{Y < x\} = P\{m + \sigma \cdot Z < x\} = P\{\sigma \cdot Z < x - m\} = \\ &= P\left\{Z < \frac{x - m}{\sigma}\right\} = F_Z\left(\frac{x - m}{\sigma}\right) = \Phi\left(\frac{x - m}{\sigma}\right). \end{aligned}$$

2.10.4. Інтегральна теорема Муавра – Лапласа

ЦГТ для випадку, коли послідовність незалежних однаково розподілених величин $X_1, X_2, \dots, X_N, \dots$ утворено бернуллівськими випадковими величинами, називають інтегральною теоремою Муавра – Лапласа. При її формулюванні враховують те, що

сума N таких величин: $X_1 + X_2 + \dots + X_N$ є абсолютною частотою появи одиниці серед значень цих випадкових величин. Цю абсолютну частоту позначають через $v_N = X_1 + X_2 + \dots + X_N$. Беручи до уваги, що для бернулліївських випадкових величин $MX_i = p$, $DX_i = p \cdot q$, $i=1, 2, \dots$, і вибираючи "числову множину" B інтервалом: $B = [a, b)$, отримуємо варіант ЦГТ для бернулліївських випадкових величин, яку інакше називають інтегральною теоремою Муавра – Лапласа.

Теорема 2.17 (інтегральна теорема Муавра – Лапласа)

Нехай $X_1, X_2, \dots, X_N, \dots$ – послідовність незалежних однаково розподілених бернулліївських випадкових величин. Тоді для довільного інтервалу дійсних чисел $[a, b)$ одержимо

$$P \left\{ \frac{v_N - N \cdot p}{\sqrt{N \cdot p \cdot q}} \in [a, b) \right\} \rightarrow P \{Z \in [a, b)\} \quad (2.28)$$

Локальну теорему Муавра – Лапласа розглянуто в підрозд. 2.6 (теорема 2.10).

Зауваження 2.1

Беручи до уваги, що умова належності інтервалу може бути записана через нерівності, а ймовірність приналежності інтервалу – через функцію розподілу, твердження інтегральної теореми Муавра – Лапласа записують у вигляді

$$P \left\{ a \leq \frac{v_N - N \cdot p}{\sqrt{N \cdot p \cdot q}} < b \right\} \rightarrow \Phi(b) - \Phi(a). \quad (2.29)$$

Наслідок 2.1

Для набору незалежних бернулліївських із параметром p випадкових величин X_1, \dots, X_N виконується співвідношення

$$P \{k \leq X_1 + \dots + X_N < l\} \approx \Phi \left(\frac{l - N \cdot p}{\sqrt{N \cdot p \cdot q}} \right) - \Phi \left(\frac{k - N \cdot p}{\sqrt{N \cdot p \cdot q}} \right). \quad (2.30)$$

Дійсно, зважаючи на те, що $X_1 + \dots + X_N = v_N$, а також віднімаючи в обох частинах нерівностей у запису події $\{k \leq X_1 + \dots + X_N < l\}$ величину $N \cdot p$ та ділячи на $\sqrt{N \cdot p \cdot q}$, отримуємо

$$\{k \leq X_1 + \dots + X_N < l\} = \left\{ \frac{k - N \cdot p}{\sqrt{N \cdot p \cdot q}} \leq \frac{v_N - N \cdot p}{\sqrt{N \cdot p \cdot q}} < \frac{l - N \cdot p}{\sqrt{N \cdot p \cdot q}} \right\}.$$

Отже, за інтегральною теоремою Муавра – Лапласа маємо

$$P \{k \leq X_1 + \dots + X_N < l\} = P \left\{ \frac{k - N \cdot p}{\sqrt{N \cdot p \cdot q}} \leq \frac{v_N - N \cdot p}{\sqrt{N \cdot p \cdot q}} < \frac{l - N \cdot p}{\sqrt{N \cdot p \cdot q}} \right\} \approx \\ \approx \Phi \left(\frac{l - N \cdot p}{\sqrt{N \cdot p \cdot q}} \right) - \Phi \left(\frac{k - N \cdot p}{\sqrt{N \cdot p \cdot q}} \right).$$

Зважаючи на те, що v_N має біноміальний розподіл із параметрами N та p , твердження наслідку 2.1 використовують для наближення ймовірностей біноміального розподілу.

Наслідок 2.2

Нехай Y випадкова величина, що має бернулліївський розподіл із параметрами N та p . Тоді для довільних цілих $0 \leq k < l \leq N$ виконується таке:

$$P \{k \leq Y < l\} \approx \Phi \left(\frac{l - N \cdot p}{\sqrt{N \cdot p \cdot q}} \right) - \Phi \left(\frac{k - N \cdot p}{\sqrt{N \cdot p \cdot q}} \right).$$

Приклад 2.55 (парадокс де Муавра)

Згідно схемі незалежних випробувань Бернуллі, при підкиданні правильної монети кількість гербів при збільшенні кількості випробувань прямує до кількості номіналів. З іншого боку ймовірність того, що кількість випадання гербів дорівнює кількості випадань номіналів прямує до 0. Наприклад при шестикратному підкиданні монети ймовірність випадання 3 гербів дорівнює $5/16$, при підкиданні 100 разів ймовірність випадання 50 гербів становить 0,08, а при 1000-кратному підкиданні монети ймовірність випадання 500 гербів становить менше за 0,02. У загальному випадку, коли монету підкидають $2n$ разів, ймовірність випадання n гербів $p = C_{2n}^n 2^{-2n}$ і для достатньо великих n

ймовірність p наближається до $\frac{1}{\sqrt{\pi n}}$ (це легко бачити з доведення

теорема Муавра для $p = 1/2$), що дійсно прямує до 0 з ростом n . Тобто, ймовірність того, що кількість випадання герба приблизно дорівнює кількості випадання номіналу, прямує до 1, а ймовірність того, що кількість випадання герба точно дорівнює кількості випадання номіналу, прямує до 0. Розрив між цими двома фактами був огорнутий "атмосферою парадоксальності" доти, поки де Муавр не побудував між ними математичний міст.

Позначимо через H_n та T_n кількості випадання герба та номіналу відповідно при n підкиданнях монети. Із граничної теореми легко бачити, що ймовірність того, що різниця $|H_n - T_n|$ стає дуже малою порівняно з n , прямує до одиниці. Однак де Муавр зазначив, що величина $|H_n - T_n|$ не є дуже малою порівняно з \sqrt{n} . Він, наприклад, обчислив, що для $n = 3600$ ймовірність того, що $|H_n - T_n|$ не перебільшує 60, дорівнює 0,682688...

Нехай x – довільне додатне число і нехай $A_n(x)$ позначає ймовірність того, що $|H_n - T_n| < x\sqrt{n}$. Згідно з де Муавром, $A_n(x)$ прямує з ростом n до величини $\Phi(x)$, значення якої міститься між 0 та 1, причому при збільшенні x від 0 до $+\infty$ $\Phi(x)$ збільшується від 0 до 1. Ця функція і є тим математичним мостом, про який ішлося вище.

2.10.5. Теорема Пуассона

Можливість наближеного обчислення ймовірностей за бернулліївським розподілом, крім інтегральної теореми Муавра – Лапласа, надає ще теорема Пуассона. Її формулюють для схеми спостережень, яку називають схемою серій.

За цією схемою розглядають послідовно спочатку одну бернулліївську величину X_{11} з параметром p_1 , потім – дві незалежні бернулліївські X_{21} та X_{22} з одним і тим самим параметром p_2 , і т. д. На кроці з номером n маємо n незалежних однаково розподілених із параметром p_n бернулліївських випадкових величин X_{n1}, \dots, X_{nn} . Із кожним кроком пов'язують випадкову величину $Y_n = X_{n1} + \dots + X_{nn}$, що має біноміальний розподіл із параметрами n та p_n . У таких позначеннях у схемі серій справедлива така теорема.

Теорема 2.18 (Пуассона)

Нехай $n \cdot p_n = \lambda$. Тоді для довільного невід'ємного цілого k виконується

$$P\{Y_n = k\} \xrightarrow{n \rightarrow \infty} \frac{\lambda^k}{k!} e^{-\lambda}. \quad (2.31)$$

Оскільки граничні значення ймовірностей відповідають пуассонівському розподілу з параметром $\lambda = n \cdot P_n$, уводячи позначення π для випадкової величини з цим пуассонівським розподілом, твердження теореми Пуассона можна записати у вигляді

$$P\{Y_n = k\} \xrightarrow{n \rightarrow \infty} P\{\pi = k\} = \frac{\lambda^k}{k!} e^{-\lambda}.$$

Простіша версія теореми Пуассона розглядалась у підрозд. 2.6.

2.10.6. Наближене обчислення ймовірностей біноміального розподілу

Теорему Пуассона використовують з одного боку для висновків, щодо пуассонівського розподілу випадкової величини π , що є граничним значенням для послідовності біноміальних випадкових величин Y_n у схемі серій. З іншого – для наближення ймовірностей, пов'язаних з бернулліївським розподілом (див. підрозд. 2.6). Таке наближення полягає в заміні дограничних значень ймовірностей у схемі серій граничними:

$$P\{Y_n = k\} \approx P\{\pi = k\} = \frac{\lambda^k}{k!} e^{-\lambda}.$$

Звісно, таке наближення застосовують в умовах, які відповідають змісту схеми серій у теоремі Пуассона. Власне, схема серій у теоремі Пуассона відповідає мінливості умов спостереження бернулліївських величин від серії до серії. Змінні умови спостереження залишають незмінним математичне сподівання (середнє) для відповідного біноміального розподілу.

Приклад 2.56 (задача про булочку з родзинками)

Використання теореми Пуассона для висновків про розподіл випадкової величини ілюструється задачею про булочку з родзинками.

Середня щільність родзинок у тісті, з якого виготовлена булочка, становить ρ . Вважають, що вони рівномірно розподілені по всьому об'єму тіста. Із цього тіста випікається булочка об'єму V . Яка ймовірність того, що вона містить ту чи іншу кількість родзинок k ?

Розв'язання

Позначимо як π кількість родзинок в об'ємі V . Загалом, виходячи з рівномірної розподіленості в об'ємі тіста, з якого виготовлена булочка, середня кількість родзинок, що припадає на об'єм V , дорівнює $M\pi = \rho \cdot V$.

З іншого боку, якщо розбити весь об'єм булочки на n частин однакового об'єму достатньо малих, щоб можна було вважати, що в кожній з них міститься не більше однієї родзинки, то кількість X_i , $i = 1, \dots, n$, родзинок у кожній із частин є бернуллівською випадковою величиною. Через рівноправність цих частин булочки, вважаємо, що вони однаково розподілені. Це означає однаковість параметра бернуллівського розподілу, який позначимо через p_n , та незалежність випадкових величин X_i , $i = 1, \dots, n$. Якщо позначити як Y_n загальну кількість родзинок у булочці, то, очевидним чином $Y_n = X_1 + \dots + X_n$ має біноміальний розподіл із середнім (математичним сподіванням) $n \cdot p_n$. Зіставляючи два вирази для середньої кількості родзинок: $\rho \cdot V$ та $n \cdot p_n$, доходимо висновку, що $n \cdot p_n = \rho \cdot V$ для всіх можливих розбиттів об'єму V на n частин. Звісно, щоб мати можливість дрібнити об'єм булочки на досить малі частини, припускають, що родзинки мають точковий об'єм (тобто нульовий).

Кількість родзинок π в об'ємі V є "граничною" для біноміальних випадкових величин $Y_n = X_1 + \dots + X_n$ у схемі серій.

Отже, за теоремою Пуассона, робимо висновок про те, що π має пуассонівський розподіл із параметром

$$\lambda = M\pi = n \cdot P_n = \rho \cdot V.$$

Таким чином,

$$P\{\pi = k\} = \frac{\lambda^k}{k!} e^{-\lambda} = \frac{(\rho V)^k}{k!} e^{-\rho V}.$$

Приклад 2.57

(біноміальний розподіл. Теорема Пуассона)

Телефонна станція обслуговує 2000 абонентів. Для кожного з них імовірність того, що протягом 1 години він зателефонує на станцію, становить 0,002. Обчислити ймовірність таких подій: а) протягом години три абоненти зателефонують на станцію; б) не більш як три; в) хоча б один абонент.

Розв'язання

Оскільки кількість незалежних випробувань $n = 2000$ досить велика, а ймовірність "успіху" в одному випробуванні $p = 0,002$ дуже мала, можемо застосувати теорему Пуассона, згідно з якою

$$P\{X_{2000} = k\} \rightarrow \frac{\lambda^k}{k!} e^{-\lambda},$$

де X_{2000} – випадкова величина, що характеризує кількість абонентів, котрі зателефонували протягом години, а за значення параметра береться $\lambda = np$. У цьому випадку $\lambda = 2000 \cdot 0,002 = 4$, а відповідні ймовірності підраховуємо так:

$$\text{а) } P\{X_{2000} = 3\} = \frac{4^3}{3!} e^{-4} = 0,1954;$$

$$\begin{aligned} \text{б) } P\{0 \leq X_{2000} \leq 3\} &= \\ &= P\{X_{2000} = 0\} + P\{X_{2000} = 1\} + P\{X_{2000} = 2\} + P\{X_{2000} = 3\} = \\ &= \frac{4^0}{0!} e^{-4} + \frac{4^1}{1!} e^{-4} + \frac{4^2}{2!} e^{-4} + \frac{4^3}{3!} e^{-4} = \\ &= 0,0183 + 0,0733 + 0,1465 + 0,1954 = 0,4335; \end{aligned}$$

$$\text{в) } P\{1 \leq X_{2000} \leq 2000\} = 1 - P\{X_{2000} = 0\} = 1 - 0,0183 = 0,9817.$$

Приклад 2.58 (теорема Муавра – Лапласа)

Випадковим чином опитують 100 респондентів. Яка ймовірність того, що серед них кількість жінок складатиме від 45 до 55 осіб? (Вважати, що чоловіків та жінок у суспільстві однакова кількість.)

Розв'язання

Згідно з умовою задачі $n = 100$; $p = q = 0,5$; $np = 50$; $npq = 25$. Застосуємо інтегральну теорему Муавра – Лапласа. Тоді

$$\begin{aligned} P\{45 \leq X_{100} \leq 55\} &= \Phi\left(\frac{55 - 50}{\sqrt{25}}\right) - \Phi\left(\frac{45 - 50}{\sqrt{25}}\right) = \\ &= \Phi(1) - \Phi(-1) = 2 \cdot \Phi(1) = 0,6826. \end{aligned}$$

ЗАДАЧІ

Біноміальний розподіл. Теорема Муавра – Лапласа та теорема Пуассона

№ 1

Гральну кістку підкидають 5 разів. Яка ймовірність того, що двічі з'явиться кількість очок, кратна трьом?

№ 2

Яка ймовірність влучити в мішень не менше двох разів, якщо ймовірність влучення дорівнює $1/5$ та проведено 10 незалежних пострілів? Знайти середнє значення кількості влучень у мішень.

№ 3

Ймовірність виграти по одному білету лотереї дорівнює 0,1. Яка ймовірність, що з восьми куплених білетів виграють: а) два білети; б) не більш як два; в) не менш як два; г) лише один білет?

№ 4

На вступних іспитах із математики абітурієнт отримує 11 задач. Імовірність розв'язати задачу для абітурієнта середнього рівня дорівнює 0,6. Щоб дістати оцінку "5", абітурієнт має розв'язати хоча б 10 задач, оцінку "4" – 8–9 задач, "3" – 5–7 задач, "2" – менше 5 задач. Обчислити ймовірність того, що середньо підготовлений студент отримає оцінку: а) "5"; б) "4"; в) "3"; г) "2".

№ 5

Нехай X – випадкова величина, що має біноміальний розподіл із параметрами n і p . Відомо, що $MX = 12$, $DX = 4$. Знайти n і p .

№ 6

Імовірність того, що висіяна зернина ячменю проросте через певний час, становить 0,95. У дослідній лабораторії висіяно 1000 зернин. Обчислити ймовірність таких подій: а) проросте 900 зернин; б) проросте 800–900 зернин.

№ 7

У деякій місцевості кількість хворих на малярію в середньому становить 3 %. Навмання перевіряють 500 осіб. Із якою ймовірністю серед них виявиться $3 \pm 0,5$ % хворих на малярію?

№ 8

У перші класи прийнято 300 дітей. Яка ймовірність того, що серед них 100 дівчаток, коли відомо, що ймовірність вступу до школи хлопчика 0,515?

№ 9

У страховій компанії застраховано 10000 осіб одного віку та однієї соціальної групи. Імовірність смерті протягом року кожної застрахованої особи становить 0,006. Кожна з цих осіб на 1 лютого зробила внесок 12 у. о. страхових. У разі смерті застрахованого зазначена в заповіті особа одержить від компанії 1000 у. о. Знайти ймовірність того, що: а) страхова компанія зазнає збитків; б) одержить прибуток, не менший за 40000 у. о.

№ 10

Імовірність того, що протягом години вийде з ладу один прилад, дорівнює 0,1. Визначити ймовірність того, що за годину зі 100 приладів вийде з ладу: а) не менш як 40 приладів; б) менше 20 приладів; в) від 6 до 18 приладів.

3. МАТЕМАТИЧНА СТАТИСТИКА

3.1. Основна задача математичної статистики

Під **основною задачею математичної статистики (ОЗМС)** розуміють задачу оцінювання основних характеристик (параметрів) розподілів імовірностей на основі емпіричних даних (на основі вибірки a_1, \dots, a_N). Звісно, ця характеристика-параметр розподілу є невідомою. Цей невідомий параметр (числовий чи векторний) далі позначатимемо через θ .

ОЗМС – це задача визначення основних рис-параметрів такого розподілу, який визначає ймовірності появи тих чи інших значень у вибірці. До таких основних характеристик-параметрів можуть належати: математичне сподівання MX , дисперсія DX , чи взагалі $Mg(X)$, функція $F_X(x)$ чи щільність розподілу $f_X(x)$, імовірність $P(A)$ тієї чи іншої події A , квартилі, мода чи медіана розподілу, коваріація та коефіцієнт кореляції. Параметри розподілу не обов'язково визначають розподіл однозначно. До тих, що однозначно визначають розподіли, належать параметри основних розподілів: параметр p бернулліївського, біноміального чи геометричного розподілів, параметр λ для пуассонівського, параметри a та b – для рівномірного, параметр λ – для експоненціального, m та σ^2 – для нормального розподілу.

До параметрів систем випадкових величин належать коваріація C_{XY} та коефіцієнт кореляції ρ_{XY} .

Зміст ОЗМС передають схемою

$$a_1, \dots, a_N \rightarrow \theta = \begin{cases} MX, DX, Mg(X), MG(X, Y), C_{XY}, \rho_{XY}, \\ \text{квартилі,} \\ \text{мода, медіана,} \\ F_X(x), f_X(x), \\ P(A), p \text{ для бернулліївського, біноміального,} \\ \lambda \text{ для пуассонівського, експоненціального,} \\ (a, b) \text{ для рівномірного,} \\ (m, \sigma^2) \text{ для нормального.} \end{cases}$$

Залежно від змісту, який вкладають у розуміння вимоги "визначення", виділяють три основні типи ОЗМС, що відповідають вимогам статистичної обробки даних:

- 1) точкового оцінювання параметра (задача ТО);
- 2) точності точкового оцінювання (задача ТТО);
- 3) перевірки гіпотез (задача ПГ).

3.2. Задача точкового оцінювання (точкова оцінка параметра)

Задача ТО полягає у визначенні функції g від вибірки, значення якої використовують замість невідомого параметра. Цю функцію g від вибірки називають точковою оцінкою параметра. Оцінку параметра θ позначають $\hat{\theta}$. Зміст ТО передають схемою

$$\hat{\theta} = g(a_1 \dots a_N) \rightarrow \theta.$$

Загалом, функцію від вибірки називають **статистикою**, тому точкова оцінка – це статистика, значення якої використовують замість невідомого параметра.

Так, для можливих варіантів параметрів, перелічених вище, точковими оцінками є:

- для ймовірності події A – її відносна частота: $\hat{P}(A) = \mathcal{C}_N(A)$;
- для математичного сподівання – вибіркове середнє: $\hat{M}X = \bar{x}$;
- для дисперсії – вибіркова дисперсія у вигляді S^2 чи $\hat{\sigma}^2$:

$$\hat{D}X = \begin{cases} S^2, & \text{коли математичне сподівання не відоме,} \\ \hat{\sigma}^2, & \text{коли математичне сподівання відоме.} \end{cases}$$

- для функції розподілу – КФРЧ: $\hat{F}(x) = F_N(x)$;
- для щільності розподілу – гістограма: $\hat{f}(x) = f_N(x)$;
- для квантилів – квантілі групованої вибірки;
- для моди та медіани – відповідно мода та медіана розподілу частот, що визначаються за вибіркою;
- для коваріації C_{XY} – вибіркова коваріація \hat{C}_{XY} , що визначається формулою

$$\hat{C}_{XY} = \frac{1}{N-1} \sum_{i=1}^N (x_i - \bar{x})(y_i - \bar{y}),$$

де $a_1 = (x_1, y_1)$ – перше спостереження системи випадкових величин $(X, Y), \dots$, та $a_N = (x_N, y_N)$ – останнє з N спостережень системи випадкових величин (X, Y) ;

▪ для коефіцієнта кореляції r_{XY} – вибіркового коефіцієнта кореляції \hat{r}_{XY} , що обчислюють за формулою:

$$\hat{r}_{XY} = \frac{\hat{C}_{XY}}{\sqrt{S_X^2 \cdot S_Y^2}} = \frac{\sum_{i=1}^N (x_i - \bar{x}) \cdot (y_i - \bar{y})}{\sqrt{\sum_{i=1}^N (x_i - \bar{x})^2 \cdot \sum_{i=1}^N (y_i - \bar{y})^2}};$$

▪ для математичного сподівання функції від випадкової величини чи системи випадкових величин – відповідна інтегральна характеристика на основі середнього розподілу частот:

$$\hat{M}g(X) = \overline{g(X)} = \frac{1}{N} \sum_{i=1}^N g(a_i).$$

Загалом, характеристики розподілів імовірностей, які визначають на основі граничного переходу в частотних відповідниках, оцінюють саме цими частотними відповідниками.

Оцінювання параметрів основних розподілів імовірностей здійснюють за їх зв'язком з інтегральними характеристиками розподілу. Це стосується всіх розподілів із переліку основних, крім рівномірного.

3.2.1. Оцінювання параметрів дискретних розподілів

1) Параметр p бернулліївського розподілу чи випадкової величини X із бернулліївським розподілом оцінюють як імовірність $p = P(A)$, або як математичне сподівання: $p = MX$.

Отже,

$$\hat{p} = \begin{cases} \hat{P}\{X = 1\} = \mathcal{C}_N\{X = 1\}, \\ \hat{M}X = \bar{x}. \end{cases}$$

Оскільки для бернулліївського розподілу вибіркове середнє збігається із частотою появи одиниці $\mathbb{P}\{X=1\} = \hat{MX} = \bar{x}$, то два варіанти оцінювання параметра p збігаються між собою.

2) Параметр p біноміального розподілу або відповідно розподіленої випадкової величини X оцінюють, виходячи з того, що $MX = n \cdot p$, а отже, $p = MX/n$, тобто

$$\hat{p} = \frac{\hat{MX}}{n} = \frac{\bar{x}}{n}.$$

3) Для λ -пуассонівського розподілу $MX = \lambda$, отже, параметр

$$\hat{\lambda} = \hat{MX}.$$

4) Для геометричного розподілу $MX = 1/p$, отже, $p = 1/MX$, тому

$$\hat{p} = \frac{1}{\hat{MX}} = \frac{1}{x}.$$

3.2.2. Оцінювання параметрів неперервних розподілів

1) Параметри a, b рівномірного розподілу, виходячи з міркувань зв'язків цих параметрів із математичним сподіванням, дисперсією та ймовірністю, оцінюють, виходячи зі співвідношень:

$$MX = \frac{a+b}{2}, \quad DX = \frac{(b-a)^2}{12},$$

в які замість математичного сподівання та дисперсії слід підставити їх оцінки: \bar{x} та S^2 відповідно.

За іншим підходом (більш поширеним для великих вибірок, але менш точним) оцінкою для a виступає мінімальний, а для b – максимальний елемент вибірки:

$$\hat{a} = x_{\min}, \quad \hat{b} = x_{\max}.$$

2) Параметри експоненціального розподілу оцінюють, виходячи з того, що $MX = 1/\lambda$ за формулою

$$\hat{\lambda} = \frac{1}{\hat{MX}} = \frac{1}{x}.$$

3) Параметри m та σ^2 нормального розподілу оцінюють, враховуючи те, що $m=MX$, $\sigma^2 = DX$, відповідно до співвідношень:

$$\hat{m} = \hat{MX} = \bar{x}, \text{ тобто вибіркове середнє;}$$

$$\hat{DX} = \hat{\sigma}^2, \text{ вибіркова дисперсія.}$$

Якщо математичне сподівання невідоме і замість нього використовують оцінку математичного сподівання, тобто вибіркове середнє, то за оцінку дисперсії беруть $\hat{DX} = S^2$ (див. далі підрозд. 3.4).

Приклад 3.1 (точкове оцінювання)

Результатом 20 спостережень за реалізаціями випадкової величини X є вибірка 5, -2, 3, -1, 6, 9, 17, -11, 15, 2, 1, 2, -3, 3, 4, 5, 7, -8, 3, 1. Оцінити середнє значення випадкової величини X , її дисперсію та ймовірність того, що X набуде значень, не більших за 4.

Розв'язання

$N=20$. Стандартною оцінкою середнього (математичного сподівання) випадкової величини є вибіркове середнє

$$\hat{MX} = \bar{x} = \frac{1}{N} \sum_{i=1}^N x_i.$$

Оскільки $\sum_{i=1}^N x_i = 5 - 2 + 3 - 1 + 6 + \dots + 3 + 1 = 58$,

то $\bar{x} = \frac{1}{20} \cdot 58 = 2,9$.

Через те, що істинне значення математичного сподівання генеральної сукупності, з якої отримана вибірка, не відоме, дисперсію DX оцінимо вибірковою дисперсією за формулою

$$\hat{DX} = S^2 = \frac{1}{N-1} \sum_{i=1}^N (x_i - \bar{x})^2 = \frac{1}{N-1} \left[\sum_{i=1}^N x_i^2 - \frac{\left(\sum_{i=1}^N x_i \right)^2}{N} \right].$$

Вираз
$$\frac{1}{N-1} \left[\sum_{i=1}^N x_i^2 - \frac{\left(\sum_{i=1}^N x_i \right)^2}{N} \right]$$

можна вважати обчислювальною формулою для підрахунку вибіркової дисперсії.

$$\text{Оскільки } \sum_{i=1}^N x_i^2 = 5^2 + (-2)^2 + 3^2 + (-1)^2 + \dots + 3^2 + 1^2 = 982,$$

$$\text{То } S^2 = \frac{1}{19} \cdot \left(982 - \frac{58^2}{20} \right) = 42,83.$$

Ймовірність того, що значення випадкової величини не перевищить 4 – це $P\{X \leq 4\}$. Оцінкою цієї ймовірності є відповідна відносна частота $\mathcal{C}_N\{X \leq 4\}$. Оскільки у вибірці є 12 значень, що не перевищують 4 ($m_{\{X \leq 4\}} = 12$), а об'єм вибірки $N = 20$, то

$$\mathcal{C}_N\{X \leq 4\} = \frac{m_{\{X \leq 4\}}}{N} = \frac{12}{20} = \frac{3}{5} = 0,6.$$

$$\text{Отже, } \hat{M}X = \bar{x} = 2,9; \hat{D}X = S^2 = 42,83; \\ \hat{P}\{X \leq 4\} = \mathcal{C}_N\{X \leq 4\} = 0,6.$$

Точкове оцінювання коваріації та кореляції здійснюють за вибірковими значеннями відповідних характеристик.

Оцінка коваріації має вигляд

$$\hat{C}_{XY} = \hat{M}(X - \hat{M}X)(Y - \hat{M}Y) = \begin{cases} \frac{1}{N-1} \sum_{i=1}^N (a_i - \bar{x})(b_i - \bar{y}), \\ \frac{1}{N} \sum_{i=1}^N (a_i - MX)(b_i - MY). \end{cases}$$

Коефіцієнт кореляції оцінюють за формулою

$$\hat{\rho} = \frac{\hat{C}_{XY}}{S_X S_Y} = \frac{\frac{1}{N-1} \sum_{i=1}^N (a_i - \bar{x})(b_i - \bar{y})}{\sqrt{\frac{1}{N-1} \sum_{i=1}^N (a_i - \bar{x})^2 \cdot \frac{1}{N-1} \sum_{i=1}^N (b_i - \bar{y})^2}} = \frac{\frac{1}{N} \sum_{i=1}^N (a_i - MX)(b_i - MY)}{\sqrt{\frac{1}{N} \sum_{i=1}^N (a_i - MX)^2 \cdot \frac{1}{N} \sum_{i=1}^N (b_i - MY)^2}}.$$

ЗАДАЧІ

Точкове оцінювання

№ 1

Оцінити невідомий параметр p бернулліївського розподілу та математичне сподівання і дисперсію відповідної випадкової величини, якщо у вибірці 21 нуль та 29 одиниць.

№ 2

Оцінити дисперсію, математичне сподівання та параметр p біноміального розподілу з $n = 10$ за вибіркою 6, 4, 3, 0, 5, 9, 8, 7, 2, 6, 5, 3, 4, 5, 6, 1.

№ 3

Оцінити параметр p геометричного розподілу та відповідне математичне сподівання за вибіркою

1, 2, 1, 3, 2, 1, 3, 2, 4, 1, 3, 2, 1, 2, 2.

№ 4

Оцінити параметр λ пуассонівського розподілу та відповідне математичне сподівання за вибіркою 3, 0, 1, 2, 4, 3, 2, 4, 2, 3, 1, 0, 2, 3, 1.

№ 5

Оцінити параметр λ експоненціального розподілу за вибіркою 3,7; 4,1; 2,4; 1,3; 3,7; 4,2; 1,6; 1,8; 0,9; 2,1; 1,1; 0,7; 1,7; 4,5; 1,7.

№ 6

Оцінити параметри гауссівської випадкової величини за вибіркою 0,1; -0,2; 3,1; 2,6; -0,3; 1,7; 1,9; -0,6; 0,3; -1,1; 1,3; -0,1; 0; -1.

№ 7

Оцінити параметри a , b , MX , DX рівномірного неперервного розподілу за вибіркою 1,15, 1,71, 1,62, 1,03, 1,48, 1,31, 1,56, 1,22, 1,34, 1,91, 1,41, 1,57, 1,47, 1,2, 1,51.

№ 8

Оцінити параметри MX та DX випадкової величини у задачі № 6 із підрозд. 1.6 "Задачі (інтегральні характеристики вибірки на базі середнього)".

№ 9

Оцінити параметри MX та DX випадкової величини у задачі № 7 із підрозд. 1.6 "Задачі (інтегральні характеристики вибірки на базі середнього)".

№ 10

Оцінити параметри MX та DX випадкової величини у задачі № 8 із підрозд. 1.6 "Задачі (інтегральні характеристики вибірки на базі середнього)".

3.3. Стандартні розподіли математичної статистики. Критичні величини

У математичній статистиці, крім розподілів імовірностей із переліку основних імовірнісних, широко вживаними є ще декілька, що пов'язані з нормальним розподілом, тобто будуються на його основі. Цими розподілами є "хі-квадрат з n ступенями вільності", " t -розподіл Стьюдента з n ступенями вільності", " F -розподіл Фішера з m та n ступенями вільності". Випадкові величини за цими розподілами позначають відповідно

- χ_n^2 – для розподілу хі-квадрат;
- t_n – для розподілу Стьюдента;
- F_{mn} – для розподілу Фішера;

Вираз "ступені вільності" у назвах розподілів є еквівалентом виразу "параметр", причому цей параметр чи параметри набувають натуральних значень, тобто значень із множини натуральних чисел.

Усі перелічені вище розподіли неперервні і, як і всі неперервні розподіли, задаються функцією розподілу чи щільністю розподілу. Але, зважаючи на те, що в багатьох статистичних задачах використовують зв'язок цих розподілів із нормальним, їх визначають через нормальний, точніше, стандартний нормальний розподіл.

3.3.1. χ_n^2 -розподіл

χ_n^2 -розподілом (хі-квадратом з n ступенями вільності) називають такий, що є розподілом суми квадратів n незалежних стандартних нормальних випадкових величин:

$$\chi_n^2 = Z_1^2 + \dots + Z_n^2,$$

де $Z_1 \dots Z_n$ – незалежні випадкові величини, розподілені за стандартним нормальним розподілом $N(0,1)$.

У загальному випадку χ_n^2 -розподіл визначають так: нехай $X_1 \dots X_n$ – незалежні випадкові величини, розподілені за однаковим нормальним розподілом $N(a, \sigma^2)$. Тоді функцію розподілу випадкової величини

$$\chi_n^2 = \frac{1}{\sigma^2} \sum_{i=1}^n (X_i - a)^2$$

називають χ_n^2 -розподілом з n ступенями вільності.

Щільність розподілу (рис. 3.1) цієї випадкової величини $f_{\chi_n^2}(x)$ має вигляд

$$f_{\chi_n^2}(x) = \begin{cases} 0, & x < 0, \\ \frac{x^{n/2-1} e^{-x/2}}{2^{n/2} \Gamma(n/2)}, & x \geq 0, \end{cases}$$

де $\Gamma(k) = \int_0^{\infty} e^{-x} x^{k-1} dx$ – гамма-функція.

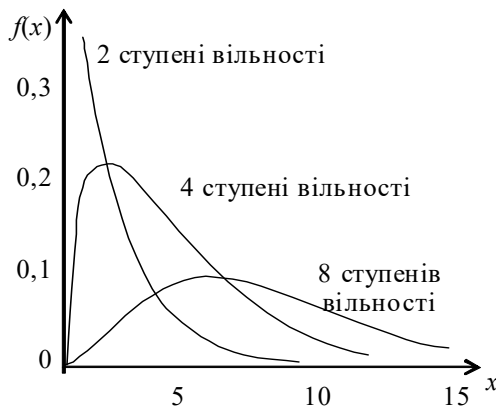


Рис. 3.1. Щільність χ_n^2 -розподілу з 2, 4 та 8 ступенями вільності

Самі числові значення χ_n^2 -розподілу з n ступенями вільності та гамма-функції є табличними (див. далі розділ 4). Зауважимо, що для цілих додатних k виконується $\Gamma(k) = (k-1)!$. Для великих додатних значень k $\Gamma(k) \cong k^k e^{-k} \sqrt{\frac{2\pi}{k}}$.

Математичне сподівання розподілу χ_n^2 з n ступенями вільності дорівнює n , а дисперсія – $2n$, тобто

$M\chi_n^2 = n$	$D\chi_n^2 = 2n$
-----------------	------------------

3.3.2. t_n -розподіл Стьюдента

t_n -розподілом Стьюдента (те-розподілом з n ступенями вільності) називають відношення розподіленої за стандартним нормальним законом випадкової величини Z_0 до незалежного від неї кореня з $\frac{1}{n}\chi_n^2$:

$$t_n = \frac{Z_0}{\sqrt{\frac{1}{n}\chi_n^2}},$$

де $Z_0 \cong N(0;1)$ та χ_n^2 – незалежні випадкові величини.

Щільність розподілу t_n (рис. 3.2) має вигляд

$$f_{t_n}(x) = \frac{\Gamma\left(\frac{1}{2}n + \frac{1}{2}\right)}{n^{1/2}\Gamma\left(\frac{1}{2}\right)\Gamma\left(\frac{1}{2}n\right)} \left(1 + \frac{x^2}{n}\right)^{-\frac{1}{2}(n+1)}, \quad -\infty < x < \infty.$$

Самі числові значення t_n -розподілу з n ступенями вільності та гамма-функції є табличними (див. далі розділ 4).

Зауважимо такє:

- t_n -розподіл є симетричним стосовно середнього, що дорівнює 0, $Mt_n = 0$, а його дисперсія $Dt_n = n/(n-2)$;
- t_n -розподіл набуває важливості у тих випадках, коли розглядають вибіркові середні, а дисперсії генеральних сукупнос-

тей не відомі. У цьому випадку використовують вибіркву дисперсію та t_n -розподіл (див. підрозд. 3.5). Також розподіл Стьюдента використовують у випадках, коли досліджують вибіркві середні й обсяг вибірки є малим.

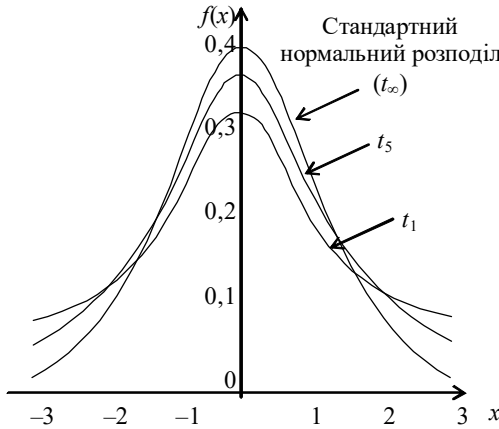


Рис. 3.2. Щільність t_n -розподілу з 1, 2 та ∞ ступенів вільності

Для великих значень n дисперсія випадкової величини t приблизно дорівнює 1, а середнє – 0, тобто t_∞ збігається зі стандартним нормальним розподілом.

$Mt_n = 0$	$Dt_n = n/(n-2)$
------------	------------------

3.3.3. $F_{m;n}$ -розподіл Фішера

$F_{m;n}$ -розподілом Фішера (еф-розподілом з m та n ступенями вільності) називають відношення (частку) двох нормованих незалежних один від одного χ^2 з m та n ступенями вільності відповідно:

$$F_{m;n} = \frac{\frac{1}{m} \chi_m^2}{\frac{1}{n} \chi_n^2}.$$

$F_{m,n}$ -розподіл має функцію щільності ймовірностей (рис. 3.3):

$$f_{F_{m,n}}(x) = \frac{\Gamma\left(\frac{1}{2}m + \frac{1}{2}n\right)}{\Gamma\left(\frac{1}{2}m\right)\Gamma\left(\frac{1}{2}n\right)} \left(\frac{m}{n}\right)^{\frac{1}{2}m} \cdot \frac{x^{\left(\frac{1}{2}m\right)-1}}{\left(1 + \frac{m}{n}x\right)^{\frac{1}{2}(m+n)}}, \quad 0 \leq x < \infty.$$

Самі числові значення $F_{m,n}$ -розподілу Фішера та гамма-функції є табличними (див. далі розд. 4).

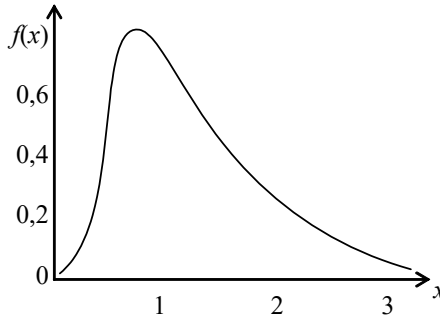


Рис. 3.3. Щільність $F_{m,n}$ -розподілу для $m=10, n=12$

Математичне сподівання $F_{m,n}$ -розподілу запишемо як $MF_{m;n} = n/(n-2)$, а дисперсію як $DF_{m;n} = \frac{2n^2(m+n-2)}{m(n-2)^2(n-4)}, n > 4$.

При $n \rightarrow \infty$ математичне сподівання $F_{m;n}$ прямує до 1, а дисперсія – до $2/m$. Ці величини становлять собою моменти випадкової величини χ_n^2/m . Фактично $mF_{m;\infty}$ має розподіл χ_m^2 .

$F_{m,n}$ -розподіл набуває особливої ваги, коли перевіряють гіпотези про однорідність дисперсій із нормально розподілених сукупностей. Також цей розподіл широко використовують у регресійному та дисперсійному аналізах.

$MF_{m;n} = \frac{n}{n-2}, n > 2$	$DF_{m;n} = \frac{2n^2(m+n-2)}{m(n-2)^2(n-4)}, n > 4$
-----------------------------------	---

3.3.4. Симетричні та несиметричні розподіли ймовірностей

Симетричними розподілами ймовірностей називають такі, для яких імовірності будь-якої пари симетричних подій B та $-B$ (наприклад, $-B = [-2; -1]$, а $B = [1; 2]$) збігаються:

$$P_X\{B\} = P_X\{-B\},$$
$$P\{X \in B\} = P\{X \in -B\}.$$

У цих співвідношеннях X є досліджуваною числовою характеристикою, випадковою величиною.

Розподіли, що не є симетричними, називають несиметричними.

У випадку, якщо розподіл описують щільністю, для симетричності розподілу необхідно і достатньо парності щільності, тобто симетричності її графіка щодо осі OY . Справедливість цього твердження впливає з розгляду елементів рис. 3.4. На ньому ймовірності $P\{X \in B\}$ та $P\{X \in -B\}$ є площами фігур під графіком щільності над відповідними множинами B та $-B$. Якщо вказаний рисунок уявно перегнути по осі OY , то ці фігури збігатимуться, отже, їх площі (на рисунку позначено як A) збігаються і відповідні ймовірності теж.

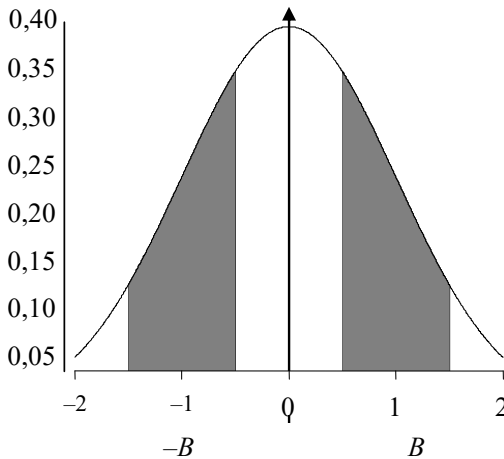


Рис. 3.4. Симетричний розподіл

Симетричні розподіли серед стандартних. Щільності стандартного нормального Z та t_n -розподілу є симетричними. Їхній характер у загальному вигляді показано на рис. 3.5.

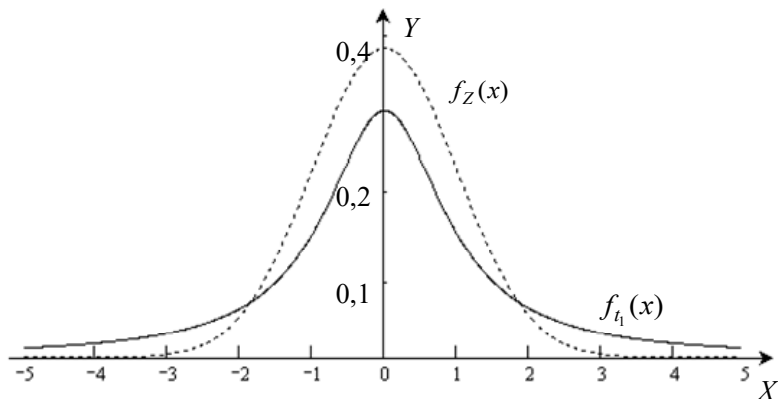


Рис. 3.5. Щільності стандартного нормального розподілу та розподілу Стюдента

Таким чином, Z - та t_n -розподіли (стандартний нормальний та Стюдента) є симетричними.

Несиметричні розподіли серед стандартних. χ_n^2 - та $F_{m,n}$ -розподіли зосереджені на невід'ємній півосі, а отже, несиметричні. Загальний вигляд щільностей цих розподілів ілюструє рис. 3.6.

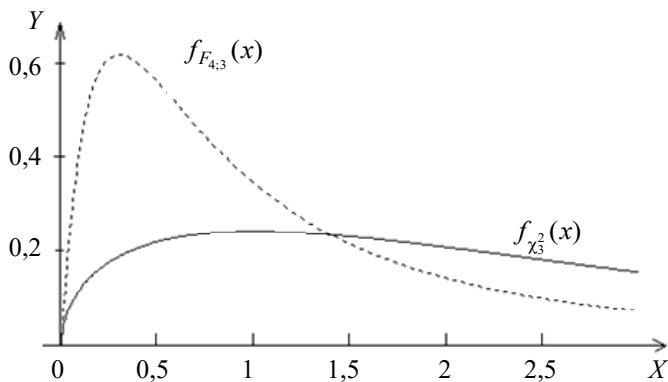


Рис. 3.6. Графіки щільностей хі-квадрат розподілу та розподілу Фішера

3.3.5. Нижні та верхні критичні величини

У практичних задачах часто виникає необхідність отримати значення досліджуваної характеристики, яка відтинає задану "частку" (імовірність) більших чи менших значень. Це властиво ситуаціям, коли в дослідженні не є суттєвим певний відсоток екстремально великих чи екстремально малих значень. Такий випадок виникає, якщо, наприклад, кажуть що 5 % сімей в Україні мають прибуток, нижчий за a , або 5 % – більший за b . Ці 5 % сімей із найбільшим чи 5 % із найменшим прибутком можуть не цікавити дослідника як екстремальні. Значення a та b представляють собою типовий приклад критичних величин (відповідно нижньої та верхньої) рівня $\alpha = 5\%$.

Верхньою (позначають x_α^U) або **нижньою** (позначають x_α^L) **критичною величиною** називають значення досліджуваної характеристики, які відтинають задану ймовірність (частку) α більших чи менших за розміром значень. Величину α називають **рівнем значущості критичної величини**. Величину називають **статистично значущою**, якщо ймовірність її випадкової появи мала.

Для неперервних розподілів критичні величини визначають співвідношеннями:

$$\begin{aligned}x_\alpha^U &: P\{X > x_\alpha^U\} = \alpha, \\x_\alpha^L &: P\{X < x_\alpha^L\} = \alpha.\end{aligned}\tag{3.1}$$

Нижні критичні величини називають також **квантилями**. У такому випадку α може бути довільним. Крім того, коли $\alpha \cdot 100\%$ набуває значень від 1 % до 99 %, то відповідні нижні критичні величини називають **процентилями**.

Верхні критичні величини називають також **процентними або Q-процентними точками**.

Зауважимо, що коли $\alpha = 0,25$, то $x_\alpha^L = Q_1$, тобто відповідна нижня критична величина дорівнює першому квантилю. Якщо $\alpha = 0,5$, то $x_\alpha^L = Q_2$ (другий квантиль або медіана), якщо $\alpha = 0,75$, то $x_\alpha^L = Q_3$, тобто третій квантиль.

Приклад 3.2

Якщо для досліджуваної числової характеристики X для $\alpha = 0,05$ значеннями критичних величин є такі:

$$x_{0,05}^L = 200, \quad x_{0,05}^U = 2000,$$

то це означає, що лише 5 % значень цієї характеристики менші за 200, так само, як тільки 5 % – більші за 2000. Водночас, це свідчить про те, що 90 % значень міститься в інтервалі від 200 до 2000.

3.3.6. Основні співвідношення між критичними величинами

Верхні та нижні критичні величинами пов'язують співвідношеннями, одне з яких має місце для всіх неперервних розподілів, а друге – тільки для симетричних. Наслідком цих співвідношень є те, що у збірниках статистичних таблиць наводяться, як правило, значення лише верхніх критичних величин, а нижні перераховують за верхніми з використанням одного зі співвідношень цього пункту.

Загальний випадок

У загальному співвідношенні стверджують, що нижня критична величина збігається з верхньою з рівнем значущості, доповненим до одиниці:

$$x_{\alpha}^L = x_{1-\alpha}^U \quad (3.2)$$

Дійсно, із співвідношення (3.1), що визначає нижню критичну величину $P\{X < x_{\alpha}^L\} = \alpha$, за теоремою заперечення, зважаючи на неперервність розподілу, маємо

$$\begin{aligned} \alpha &= P\{X < x_{\alpha}^L\} = 1 - P\{X < \overline{x_{\alpha}^L}\} = \\ &= 1 - P\{X \geq x_{\alpha}^L\} = 1 - P\{X > x_{\alpha}^L\}. \end{aligned}$$

Таким чином, отримано співвідношення

$$P\{X > x_{\alpha}^L\} = 1 - \alpha,$$

яке визначає верхню критичну величину рівня $1-\alpha$, що й завершує доведення твердження.

Випадок симетричного розподілу

Для симетричних розподілів верхня та нижня критичні величини розрізняються лише знаком

$$x_{\alpha}^L = -x_{\alpha}^U. \quad (3.3)$$

Твердження випливає з того, що симетричні множини, якими є $(-\infty; -x_{\alpha}^U)$ та $(x_{\alpha}^U; +\infty)$, мають однакові ймовірності, оскільки ймовірності за симетричними розподілами від симетричних множин збігаються. Отже,

$$P\{X \in (-\infty; -x_{\alpha}^U)\} = P\{X \in (x_{\alpha}^U; +\infty)\} = \alpha.$$

Записуючи те саме співвідношення за допомогою нерівностей, отримаємо

$$P\{X < -x_{\alpha}^U\} = P\{X > x_{\alpha}^U\} = \alpha,$$

що й завершує доведення.

Рис. 3.7 ілюструє доведення: на ньому ймовірності напівнескінчених симетричних один до одного інтервалів $(-\infty; -x_{\alpha}^U)$ та $(x_{\alpha}^U; +\infty)$ збігаються як однакові площі під симетричними гілками графіка щільності над цими інтервалами:

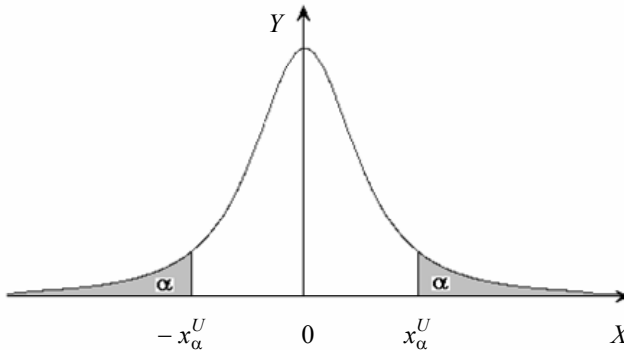


Рис. 3.7. Симетричні розподіли. Критичні точки нормального розподілу

3.3.7. Критичні величини для основних розподілів

Критичні величини стандартних розподілів позначають за позначеннями самих розподілів (табл. 3.1) із додаванням індексу α , що відповідає рівню значущості.

Таблиця 3.1

**Позначення критичних величин
стандартних розподілів**

Розподіл	Верхня критична величина x_{α}^U	Тип розподілу
Z	Z_{α}	Симетричний
χ_n^2	$\chi_{n;\alpha}^2$	Несиметричний
t_n	$t_{n;\alpha}$	Симетричний
$F_{m;n}$	$F_{m;n;\alpha}$	Несиметричний

Нижні критичні величини цих розподілів визначають відповідно до симетричності чи несиметричності стандартного розподілу і записують через верхні критичні величини (табл. 3.2).

Таблиця 3.2

**Позначення верхніх та нижніх критичних величин
стандартних статистичних розподілів**

Розподіл	Верхня критична величина x_{α}^U	Нижня критична величина x_{α}^L
Z	Z_{α}	$-Z_{\alpha}$
χ_n^2	$\chi_{n;\alpha}^2$	$\chi_{n;1-\alpha}^2$
t_n	$t_{n;\alpha}$	$-t_{n;\alpha}$
$F_{m;n}$	$F_{m;n;\alpha}$	$F_{m;n;1-\alpha}$

Прикладом значень критичних величин може слугувати фрагмент таблиці верхніх критичних величин стандартного нормального розподілу:

α	0,1	0,05	0,025	0,01	0,005
Z_{α}	1,282	1,645	1,96	2,326	3,291

Нижні критичні величини цього самого розподілу для тих самих рівнів значущості з урахуванням симетричності розподілу дорівнюють верхнім відповідникам із протилежним знаком:

α	0,1	0,05	0,025	0,01	0,005
$x_{\alpha}^L = -Z_{\alpha}$	-1,282	-1,645	-1,96	-2,326	-3,291

Статистичні таблиці

У розд. 4 наведено таблиці ймовірностей або критичних величин усіх стандартних представлених у цьому посібнику статистичних розподілів. Розглянемо як працюють із такими таблицями.

Таблиця значень функції Лапласа. У табл. 4.2 (розд. 4) наведено значення ймовірностей типу $P(X < x)$ випадкової величини X , розподіленої за стандартним нормальним розподілом. У верхньому рядку таблиці записано соті, у першому стовпчику – цілі та десяті точки x , ймовірність за якою ми шукаємо. Саме значення ймовірності $P(X < x)$ записане на перетині відповідного стовпчика із сотою та рядка з цілою та десятою. Отже, ймовірністю $P(X < x)$ є площа під кривою щільності від $-\infty$ до x .

Для того, щоб знайти верхню критичну величину рівня значущості α x_α^U (тобто точку, яка відтинає праворуч область площею α), звернемось до її визначення (3.1):

$$\alpha = P\{X > x_\alpha^U\} = 1 - P\{X < x_\alpha^U\}.$$

$$\text{Звідси } 1 - \alpha = P\{X < x_\alpha^U\}.$$

Отже, для того, щоб знайти x_α^U усередині таблиці підбирають найближче до $1 - \alpha$ число (ймовірність) і за ним визначають рядок та стовпчик, в якому воно міститься. Сума відповідних цілих, десятих та сотих і є верхньою критичною величиною x_α^U рівня значущості α .

Для відшукування нижньої критичної величини рівня значущості α використовують властивість симетричності стандартного нормального розподілу, для якого справедливе співвідношення (3.3).

У деяких довідниках у таблиці значень функції Лапласа записують ймовірності типу $P(0 \leq X < x)$. Вони за значенням є меншими від наведених у розділі 4 на 0,5 (тобто, на ймовірність $P(X < 0,5)$).

Якщо шукають критичні величини не для стандартного нормального розподілу, а для довільного $N(m; \sigma^2)$, то використовують співвідношення (2.12, б). Спочатку знаходять за таблицями критичні величини рівня значущості α для Z -розподілу (станда-

ртного нормального), а потім за співвідношенням $X = m + \sigma \cdot Z$ перетворюють їх для розподілу $N(m; \sigma^2)$.

Приклад 3.3

Для розподілу $N(1,4)$ знайти верхню та нижню критичні величини рівня значущості $\alpha = 0,01$.

Обчислюючи верхню критичну величину, у табл. 4.2 значень функції Лапласа знаходимо число, яке є найближчим до $1 - 0,01 = 0,99$. Таким числом є 0,9901, яке розміщене на місці перетину стовпчика 0,03 та рядка 2,3. Отже, верхньою критичною величиною рівня значущості 0,01 стандартного нормально розподілу є $2,3 + 0,03 = 2,33$. Відповідно нижньою критичною величиною рівня значущості 0,01 для цього ж розподілу є $-2,33$.

Для того, щоб знайти критичні величини для $N(1,4)$, спочатку запишемо $m = 1$, $\sigma^2 = 4$, тобто $\sigma = 2$. Звідси за (2.12, б) критичними величинами для $N(1,4)$ є $x_{\alpha}^L = 1 + 2 \cdot (-2,33) = -3,66$ та $x_{\alpha}^U = 1 + 2 \cdot 2,33 = 5,66$.

Таблиця значень верхніх критичних величин розподілу χ_n^2 . У табл. 4.3 наведено верхні критичні величини розподілу χ_n^2 залежно від кількості ступенів вільності та рівня значущості α . Значення верхньої критичної величини міститься на місці перетину стовпчика з відповідним рівнем значущості та рядка з кількістю ступенів вільності. Для обчислення нижньої критичної величини використовують співвідношення (3.2).

Приклад 3.4

Для розподілу χ_{29}^2 знайти верхню та нижню критичні величини рівня значущості $\alpha = 0,01$.

У табл. 4.3 шукаємо рядок, який відповідає 29 ступеням вільності та стовпчик для $\alpha = 0,01$, а на їх перетині міститься число 49,5879, яке і є верхньою критичною величиною. Для обчислення нижньої критичної величини з (3.2) знайдемо стовпчик, що відповідає $1 - \alpha = 0,99$, та рядок для $n = 29$. На їхньому перетині розміщена нижня критична величина 14,25641.

Таблиця значень верхніх критичних величин розподілу Стьюдента. У табл. 4.4 записані верхні критичні величини розподілу t_n залежно від кількості ступенів вільності та рівня

значущості α . Верхню критичну величину знаходять на місці перетину стовпчика з відповідним рівнем значущості та рядка з кількістю ступенів вільності. Для обчислення нижньої критичної величини використовують співвідношення (3.3) для симетричних розподілів.

Приклад 3.5

Для розподілу t_{30} знайти верхню та нижню критичні величини рівня значущості $\alpha = 0,005$.

Оскільки $P(t_n > x) = \alpha$, то для обчислення верхньої критичної величини рівня значущості у табл. 4.4 знаходимо рядок, який відповідає 30 ступеням вільності та стовпчик для 0,005. На їх перетині знаходиться число 2,75, яке є верхньою критичною величиною. Для симетричних розподілів із (3.3) нижня критична величина розподілу Стьюдента із 30 ступенями вільності та рівнем значущості 0,005 дорівнює $-2,75$.

Таблиця значень верхніх критичних величин розподілу Фішера. У табл. 4.5 записано верхні критичні величини розподілу F_{mn} залежно від кількості ступенів вільності m та n для різних значень рівня значущості α .

Для критичних величин розподілу Фішера використовують співвідношення, яке не є таким, що зв'яже верхню та нижню критичні величини одного й того самого розподілу, а описує співвідношення між верхніми критичними величинами розподілів, що відповідають двом різним парам параметрів $(m;n)$ та $(n;m)$, і має вигляд

$$F_{m;n;\alpha} = \frac{1}{F_{n;m;1-\alpha}}. \quad (3.4)$$

Як впливає з визначення, обернення випадкової величини з розподілом $F_{m;n}$, дає випадкову величину з розподілом $F_{n;m}$:

$$F_{m;n} = \frac{1}{F_{n;m}}.$$

Використовуючи перехід до обернених величин у нерівності з додатними частинами, маємо

$$\alpha = P\{F_{m;n} > F_{m;n;\alpha}\} = P\left\{\frac{1}{F_{m;n}} < \frac{1}{F_{m;n;\alpha}}\right\} = P\left\{F_{n;m} < \frac{1}{F_{m;n;\alpha}}\right\}.$$

Отже, $\frac{1}{F_{m;n;\alpha}}$ є нижньою критичною величиною рівня α для розподілу $F_{n;m}$, тобто верхньою критичною величиною цього самого розподілу рівня $1-\alpha$, що й доводить твердження.

Приклад 3.6

Для розподілу $F_{25;10}$ знайти верхню та нижню критичні величини рівня значущості $\alpha = 0,01$.

Для обчислення верхньої критичної величини у табл. 4.5 для $\alpha = 0,01$ знаходимо стовпчик, який відповідає $m = 25$ ступеням вільності та рядок для $n = 10$ ступенів вільності. Оскільки стовпчика з $m = 25$ немає у таблиці, але є $m = 24$, то братимемо наближене значення для $m = 24$. На перетині міститься число 4,33, яке є верхньою критичною величиною. Для розподілу Фішера нижня критична величина обчислюється за формулою

$$F_{m;n;1-\alpha} = \frac{1}{F_{n;m;\alpha}} : F_{25;10;1-0,01} = \frac{1}{F_{10;25;0,01}} = \frac{1}{3,13} \approx 0,319489.$$

3.3.8. Довірчі області

Із критичною величиною пов'язують поняття довірчої області. Якщо критичні величини відтинають "непотрібне", то довірча область "зберігає" потрібне із заданою часткою (імовірністю) $1-\alpha$, як правило близькою до одиниці. Залежно від того, що відтинається як "непотрібне", розрізняють три типи областей.

Довірчою областю рівня або довірчої ймовірності $1-\alpha$ (позначають як $D_{1-\alpha}$) називають інтервал можливих значень досліджуваної числової характеристики (випадкової величини), який містить задану частку $1-\alpha$ значень цієї характеристики за винятком:

- а) частки α менших за величиною значень, або
- б) частки α більших за величиною значень, або
- в) частки α менших та більших значень рівними частинами (тоб-

то, відповідно по $\frac{\alpha}{2}$ частки найбільших та найменших значень).

Перші дві області називають односторонніми, третю – двосторонньою.

Довірчі області записують за допомогою верхніх та нижніх критичних величин

$$D_{1-\alpha} = \begin{cases} (x_{\alpha}^L; +\infty), \\ (-\infty; x_{\alpha}^U), \\ (x_{\alpha/2}^L; x_{\alpha/2}^U). \end{cases}$$

Довірчі області симетричних стандартних розподілів Z, t_n

До симетричних розподілів належать стандартний гауссівський (нормальний) розподіл та розподіл Стюдента. Беручи до уваги те, що *верхні критичні величини позначають так само як самі розподіли з додаванням відповідного індексу*, та те, що для симетричних розподілів нижня критична величина дорівнює верхній із протилежним знаком, отримаємо записи довірчих областей:

- довірча область стандартного нормального розподілу

$$D_{1-\alpha} = \begin{cases} (-Z_{\alpha}; +\infty), \\ (-\infty; Z_{\alpha}), \\ \left(-Z_{\frac{\alpha}{2}}; Z_{\frac{\alpha}{2}}\right); \end{cases}$$

- довірча область розподілу Стюдента

$$D_{1-\alpha} = \begin{cases} (-t_{n;\alpha}; +\infty), \\ (-\infty; t_{n;\alpha}), \\ \left(-t_{n;\frac{\alpha}{2}}; t_{n;\frac{\alpha}{2}}\right). \end{cases}$$

Приклад 3.7

Для довірчої ймовірності 0,95 довірчі області для стандартного нормального розподілу мають вигляд

$$D_{0,95} = \begin{cases} (-Z_{1-0,95}; +\infty), \\ (-\infty; Z_{1-0,95}), \\ \left(-Z_{\frac{1-0,95}{2}}; Z_{\frac{1-0,95}{2}} \right) \end{cases} = \begin{cases} (-Z_{0,05}; +\infty), \\ (-\infty; Z_{0,05}), \\ (-Z_{0,025}; Z_{0,025}). \end{cases}$$

За табл. 4.2 маємо $D_{0,95} = \begin{cases} (-1,64; +\infty), \\ (-\infty; 1,64), \\ (-1,96; 1,96). \end{cases}$

Так само, для довірчої ймовірності 0,99 отримаємо

$$D_{0,99} = \begin{cases} (-Z_{1-0,99}; +\infty), \\ (-\infty; Z_{1-0,99}), \\ \left(-Z_{\frac{1-0,99}{2}}; Z_{\frac{1-0,99}{2}} \right) \end{cases} = \begin{cases} (-Z_{0,01}; +\infty), \\ (-\infty; Z_{0,01}), \\ (-Z_{0,005}; Z_{0,005}). \end{cases}$$

Довірчі області стандартних несиметричних розподілів: χ_n^2 , $F_{m;n}$

Довірчі області цих розподілів мають вигляд

▪ χ_n^2 -розподіл: $D_{1-\alpha} = \begin{cases} (\chi_{n;1-\alpha}^2; +\infty), \\ (-\infty; \chi_{n;\alpha}^2), \\ \left(\chi_{n;1-\frac{\alpha}{2}}^2; \chi_{n;\frac{\alpha}{2}}^2 \right); \end{cases}$

$$\blacksquare F_{m;n}\text{-розподіл: } D_{1-\alpha} = \begin{cases} (F_{m;n;1-\alpha}; +\infty), \\ (-\infty; F_{m;n;\alpha}), \\ \left(F_{m;n;1-\frac{\alpha}{2}}; F_{m;n;\frac{\alpha}{2}} \right). \end{cases}$$

Ці співвідношення отримують із міркувань несиметричності стандартних розподілів хі-квадрат та Фішера, тобто, з того, що

$$x_{\alpha}^L = x_{1-\alpha}^U,$$

а отже, $x_{\frac{\alpha}{2}}^L = x_{1-\frac{\alpha}{2}}^U$.

Зважаючи на те, що досліджувані розподіли зосереджені на невід'ємній півосі, довірчі області звужують природним чином, відкинувши від'ємні значення.

Отже, остаточно маємо

$$\blacksquare \chi_n^2\text{-розподіл: } D_{1-\alpha} = \begin{cases} (\chi_{n;1-\alpha}^2; +\infty), \\ (0; \chi_{n;\alpha}^2), \\ \left(\chi_{n;1-\frac{\alpha}{2}}^2; \chi_{n;\frac{\alpha}{2}}^2 \right); \end{cases}$$

$$\blacksquare F_{m;n}\text{-розподіл: } D_{1-\alpha} = \begin{cases} (F_{m;n;1-\alpha}; +\infty), \\ (0; F_{m;n;\alpha}), \\ \left(F_{m;n;1-\frac{\alpha}{2}}; F_{m;n;\frac{\alpha}{2}} \right). \end{cases}$$

Двосторонні довірчі області стандартних розподілів мають вигляд

$$D_{1-\alpha} = \left(x_{\frac{\alpha}{2}}^L; x_{\frac{\alpha}{2}}^U \right) = \left[\begin{array}{c} \left(-Z_{\frac{\alpha}{2}}; Z_{\frac{\alpha}{2}} \right), \\ \left(-t_{n; \frac{\alpha}{2}}; t_{n; \frac{\alpha}{2}} \right), \\ \left(\chi_{n; 1-\frac{\alpha}{2}}^2; \chi_{n; \frac{\alpha}{2}}^2 \right), \\ \left(F_{m; n; 1-\frac{\alpha}{2}}; F_{m; n; \frac{\alpha}{2}} \right). \end{array} \right.$$

Тобто, за межами довірчого інтервалу ймовірності $1-\alpha$ залишаються з кожного боку по інтервалу ймовірності $\frac{\alpha}{2}$ кожен.

3.4. Задача точності точкового оцінювання (інтервальне оцінювання параметра)

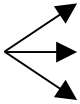
3.4.1. Другий тип ОЗМС (задача ТТО)

Задачею точності точкового оцінювання (задача ТТО) є визначення ступеня близькості між точковою оцінкою параметра $\hat{\theta}$ та його істинним, але невідомим значенням θ . Тобто, результатом цієї задачі є інтервал, в якому із заданою ймовірністю може перебувати статистика.

Як правило, другий тип ОЗМС (задачу ТТО) розглядають разом із першою (задачею ТО).

Загальний підхід до розв'язання задачі ТТО

Мірою ступеня близькості між параметром θ та його оцінкою $\hat{\theta}$ вибирають $|\theta - \hat{\theta}|$. Оскільки $\hat{\theta}$, як точкова оцінка, є функцією від вибірки, то вона змінює своє значення разом із вибіркою, а отже, є випадковою величиною щодо умов, за яких отримують серію спостережень (вибірку):

"Умови" 

$$a_1, \dots, a_N \rightarrow \hat{\theta}_1 = g(a_1 \dots a_N),$$

$$b_1, \dots, b_N \rightarrow \hat{\theta}_2 = g(b_1 \dots b_N),$$

$$c_1, \dots, c_N \rightarrow \hat{\theta}_3 = g(c_1 \dots c_N).$$

Таким чином $|\theta - \hat{\theta}|$ є випадковою величиною, яку досліджують за принципом групування, тобто, обчислюють частоту чи ймовірність появи тих чи інших значень цього відхилення. Для різних ε досліджують імовірність $P_\theta\{|\theta - \hat{\theta}| < \varepsilon\}$. Індекс θ у ймовірності означає, що розглядають імовірність, яка відповідає параметру θ . Надалі для спрощення записів цей індекс у записі ймовірностей тих чи інших подій буде опускатись. Так само вчинимо для математичних сподівань за відповідними ймовірностями. У статистичній практиці задовольняються значенням граничного допустимого відхилення оцінки від істинного значення параметра ε , що залежить від імовірності $1 - \alpha$, яку називають довірчою. Це граничне допустиме відхилення позначають $\varepsilon_{1-\alpha}^\theta$ і називають **точністю точкового оцінювання**.

Точністю точкового оцінювання невідомого параметра θ його точковою оцінкою $\hat{\theta}$ називають граничне допустиме відхилення $\varepsilon_{1-\alpha}^\theta$, яке гарантують з імовірністю $1 - \alpha$, тобто, для якого виконується співвідношення:

$$P_\theta\{|\theta - \hat{\theta}| < \varepsilon_{1-\alpha}^\theta\} = 1 - \alpha$$



Імовірність $1 - \alpha$, яку називають довірчою, вибирають близькою до одиниці: 0,9, 0,95, 0,99 та ін.

Виконання умови $\{|\theta - \hat{\theta}| < \varepsilon\}$ є еквівалентним умові $\{\theta \in (\hat{\theta} - \varepsilon_{1-\alpha}; \hat{\theta} + \varepsilon_{1-\alpha})\}$:

$$|\theta - \hat{\theta}| < \varepsilon \Leftrightarrow -\varepsilon < \theta - \hat{\theta} < \varepsilon \Leftrightarrow$$

$$\Leftrightarrow -\varepsilon + \hat{\theta} < \theta < \varepsilon + \hat{\theta} \Leftrightarrow \theta \in (\hat{\theta} - \varepsilon_{1-\alpha}; \hat{\theta} + \varepsilon_{1-\alpha}).$$

Це означає, що з імовірністю $1-\alpha$ істинне невідоме значення параметра θ "накривають" інтервалом $(\hat{\theta}-\varepsilon_{1-\alpha}; \hat{\theta}+\varepsilon_{1-\alpha})$ із випадковими границями

$$P\{|\theta-\hat{\theta}|<\varepsilon_{1-\alpha}\}=1-\alpha \Leftrightarrow P\{\theta\in(\hat{\theta}-\varepsilon_{1-\alpha}; \hat{\theta}+\varepsilon_{1-\alpha})\}=1-\alpha.$$

Цей інтервал $(\hat{\theta}-\varepsilon_{1-\alpha}; \hat{\theta}+\varepsilon_{1-\alpha})$ позначають $I_{1-\alpha}^0$ і називають **довірчим інтервалом рівня $1-\alpha$ для невідомого параметра θ** .

Таким чином, задача ТТО (обчислення $\varepsilon_{1-\alpha}$ за заданим $1-\alpha$) є еквівалентною до задачі побудови інтервалу $I_{1-\alpha}^0$ (довірчого інтервалу), який "накриває" істинне значення параметра із заданою ймовірністю.

Задачу побудови інтервалу називають ще **задачею інтервального оцінювання**.

Загалом, **довірчим інтервалом для параметра θ ($I_{1-\alpha}^0$)** називають інтервал із випадковими границями $\hat{\theta}_1$ та $\hat{\theta}_2$ (що залежать від вибірки), який "накриває" невідоме істинне значення параметра з довірчою ймовірністю $1-\alpha$:

$$P\{\theta\in I_{1-\alpha}^0\}=P\{\hat{\theta}_1<\theta<\hat{\theta}_2\}=1-\alpha:$$

$$I_{1-\alpha}^0=(\hat{\theta}_1;\hat{\theta}_2).$$

За загальним алгоритмом розв'язування задачі точності ТО для деяких видів параметрів задача ТТО одночасно має розв'язок як у вигляді $\varepsilon_{1-\alpha}$, так і у вигляді $I_{1-\alpha}^0$, а для деяких – тільки у вигляді $I_{1-\alpha}^0$.

Загальний алгоритм розв'язування задачі ТТО включає три кроки.

1-й крок. *Розгляд статистики задачі ТТО.* На цьому кроці вибирають статистику $G(\theta, \hat{\theta})$ як функцію від вибірки та невідомого параметра таким чином, щоб її розподіл не залежав від невідомого параметра і збігався з одним із стандартних розподілів: стандартним нормальним, t -розподілом Стьюдента, χ^2 -розподілом, чи F -розподілом Фішера:

$G(\theta, \hat{\theta}) \cong$ стандартний:

$$\left[\begin{array}{l} Z - \text{стандартний нормальний,} \\ t_n - \text{розподіл Стьюдента з } n \text{ ступенями вільності,} \\ \chi_n^2 - \chi^2 \text{ із } n \text{ ступенями вільності,} \\ F_{mn} - F - \text{розподіл.} \end{array} \right.$$

Значення функцій розподілу чи верхніх критичних величин стандартних розподілів наведено у розд. 4 "Статистичні таблиці".

2-й крок. Побудова довірчої області. На другому етапі за стандартним розподілом першого етапу будують довірчу область $D_{1-\alpha}$ ймовірності $1-\alpha$ (як правило, двосторонню):

$$\text{Стандартний розподіл} \xrightarrow{1-\alpha} D_{1-\alpha} = \left[\begin{array}{l} \left(x_{\frac{\alpha}{2}}^L; x_{\frac{\alpha}{2}}^U \right), \\ \left(x_{\alpha}^L; +\infty \right), \\ \left(-\infty; x_{\alpha}^U \right). \end{array} \right.$$

3-й крок. Отримання виразу-розв'язку задачі ТТО. На цьому етапі в записі $P\{G(\theta, \hat{\theta}) \in D_{1-\alpha}\} = 1-\alpha$ умову $G(\theta, \hat{\theta}) \in D_{1-\alpha}$ розкривають щодо θ , після чого отримують довірчий інтервал $I_{1-\alpha}^{\theta}$ для параметра θ .

3.4.2. Точність точкового оцінювання параметрів нормального $N(m; \sigma^2)$ -розподілу

Точкове оцінювання параметрів нормального розподілу в теорії помилок спостереження за відсутності систематичних складових помилок інтерпретують як оцінювання істинного значення вимірюваної величини та дисперсії випадкової складової помилки. Отже, точність точкового оцінювання – це точність оцінювання істинного значення виміру та дисперсії помилки спостереження.

Розглядають випадки великої та невеликої генеральних сукупностей, тобто, коли генеральна сукупність значно більша за вибірку, і тоді розміром генеральної сукупності вважають нескінченність, і коли обсяг генеральної сукупності співмірний з обсягом вибірки – у цьому випадку враховують крім об'єму вибірки також об'єм генеральної сукупності. Якщо не вказують інше, то йдеться про велику генеральну сукупність.

Точність точкового оцінювання m

Точністю оцінювання першого параметра нормального розподілу в теорії помилок спостереження є точність оцінювання істинного значення вимірюваної величини. Розв'язання задачі ТТО в цьому випадку залежить від того, відоме чи невідоме значення параметра σ^2 – дисперсії "випадкової" складової помилки спостережень. Відповідно до цього статистика $G(\theta; \hat{\theta})$ має стандартний нормальний розподіл або розподіл Стюдента.

Варіант 1, коли σ^2 відоме. У цьому випадку

$$\varepsilon_{1-\alpha}^m = Z_{\frac{\alpha}{2}} \cdot \frac{\sigma}{\sqrt{N}}. \quad (3.5)$$

де $Z_{\frac{\alpha}{2}}$ є верхньою критичною величиною рівня $\frac{\alpha}{2}$ для стандартного нормального розподілу з довірчим інтервалом для параметра m : $I_{1-\alpha}^m = \left(\bar{x} - Z_{\frac{\alpha}{2}} \cdot \frac{\sigma}{\sqrt{N}}; \bar{x} + Z_{\frac{\alpha}{2}} \cdot \frac{\sigma}{\sqrt{N}} \right)$.

Дійсно, за стандартним алгоритмом розв'язування задачі ТТО:

1. $G(m, \bar{x}) = \frac{(\bar{x} - m)\sqrt{N}}{\sigma} \cong Z$ – вибірка, що складається з незалежних гауссівських спостережень, має спільний гауссівський розподіл, тому \bar{x} – нормально розподілене з параметрами, що визначають як $M\bar{x} = m$, $D\bar{x} = \frac{\sigma^2}{N}$. Так само, за замкненістю гауссівських розподілів щодо лінійних перетворень, гауссівською буде випадкова величина $Z = \frac{\bar{x} - M\bar{x}}{\sqrt{D\bar{x}}}$. Нескладно переконатись, що

$$Z = \frac{\bar{x} - M\bar{x}}{\sqrt{D\bar{x}}} = \frac{(\bar{x} - m)}{\sqrt{\frac{\sigma^2}{N}}} = \frac{(\bar{x} - m)\sqrt{N}}{\sigma} \quad \text{і}$$

$MZ=0, DZ=1$, а отже, $Z \cong N(0,1)$.

2. $N(0;1) \xrightarrow{1-\alpha} D_{1-\alpha} = \left(x_{\frac{\alpha}{2}}^L; x_{\frac{\alpha}{2}}^U \right) = \left(-Z_{\frac{\alpha}{2}}; Z_{\frac{\alpha}{2}} \right)$ – оскільки вибірка

із симетричного розподілу, яким є стандартний нормальний, то верхня та нижня критичні величини розрізняються тільки знаком.

3. У співвідношенні $P\{G \in D_{1-\alpha}\} = 1 - \alpha$ подію $\{G \in D_{1-\alpha}\}$ подають у вигляді

$$\begin{aligned} \{G \in D_{1-\alpha}\} &\sim \left\{ \frac{(\bar{x} - m)\sqrt{N}}{\sigma} \in \left(-Z_{\frac{\alpha}{2}}; Z_{\frac{\alpha}{2}} \right) \right\} \sim \\ &\sim \left\{ -Z_{\frac{\alpha}{2}} < \frac{(\bar{x} - m)\sqrt{N}}{\sigma} < Z_{\frac{\alpha}{2}} \right\} \sim \\ &\sim \left\{ \left| \frac{(\bar{x} - m)\sqrt{N}}{\sigma} \right| < Z_{\frac{\alpha}{2}} \right\} \sim \left\{ |\bar{x} - m| < Z_{\frac{\alpha}{2}} \cdot \frac{\sigma}{\sqrt{N}} \right\}. \end{aligned}$$

Таким чином, $P\{G \in D_{1-\alpha}\} = 1 - \alpha \Leftrightarrow P\left\{ |\bar{x} - m| < Z_{\frac{\alpha}{2}} \cdot \frac{\sigma}{\sqrt{N}} \right\} = 1 - \alpha$,

а отже, $\varepsilon_{1-\alpha}^m = Z_{\frac{\alpha}{2}} \cdot \frac{\sigma}{\sqrt{N}}$.

За інших рівних умов, точність оцінювання вища (відповідно величина $\varepsilon_{1-\alpha}^m$ граничного допустимого відхилення менша) у випадку:

- 1) коли менша дисперсія σ^2 ;
- 2) коли більший об'єм вибірки.

Розв'язання задачі ТТО оцінювання m за допомогою \bar{x} при відомій дисперсії представляють довірчим інтервалом, який має

вигляд $I_{1-\alpha}^m = \left(\bar{x} - Z_{\frac{\alpha}{2}} \cdot \frac{\sigma}{\sqrt{N}}; \bar{x} + Z_{\frac{\alpha}{2}} \cdot \frac{\sigma}{\sqrt{N}} \right)$ або самим гранично допу-

стимим відхиленням $\varepsilon_{1-\alpha}^m = Z_{\frac{\alpha}{2}} \cdot \frac{\sigma}{\sqrt{N}}$.

При малих об'ємах вибірки ($N < 30$) навіть за відомої дисперсії використовують для статистики розподіл Стьюдента, як у наступному випадку при невідомій дисперсії.

Варіант 2, коли σ^2 невідоме. У цьому випадку

$$\varepsilon_{1-\alpha}^m = t_{N-1; \frac{\alpha}{2}} \cdot \frac{S}{\sqrt{N}}, \quad (3.6)$$

де $t_{N-1; \frac{\alpha}{2}}$ (за прийнятими позначеннями) є верхньою критичною

величиною рівня $\frac{\alpha}{2}$ для стандартного розподілу t_{N-1} , а

$$S^2 = \frac{(a_1 - \bar{x})^2 + \dots + (a_N - \bar{x})^2}{N-1}.$$

Для оцінювання точності виконують три кроки.

1-й крок. Розглядають статистику

$$G(m, \bar{x}) = \frac{(\bar{x} - m)\sqrt{N}}{S},$$

$$G(m, \bar{x}) = \frac{(\bar{x} - m)\sqrt{N}}{S} \cong t_{N-1}.$$

При доведенні цього твердження використовують той факт, що $\frac{(\bar{x} - m)\sqrt{N}}{\sigma}$ має стандартний нормальний розподіл, а також те, що

оцінка дисперсії S^2 має $\frac{\sigma^2}{N-1} \chi_{N-1}^2$ -розподіл і не залежить від \bar{x} :

$$\begin{aligned} G(m, \bar{x}) &= \frac{(\bar{x} - m)\sqrt{N}}{S} = \\ &= \frac{(\bar{x} - m)\sqrt{N}}{\sqrt{\frac{\sigma^2}{N-1} \chi_{N-1}^2}} = \frac{(\bar{x} - m)\sqrt{N}}{\sigma \sqrt{\frac{1}{N-1} \chi_{N-1}^2}} = \\ &= \frac{(\bar{x} - m)\sqrt{N}}{\sigma} = \frac{Z}{\sqrt{\frac{1}{N-1} \chi_{N-1}^2}} = t_{N-1}. \end{aligned}$$

В останньому співвідношенні чисельник і знаменник є незалежними за спадком, оскільки \bar{x} та S^2 незалежні, а отже, за визначенням $\frac{Z}{\sqrt{\frac{1}{N-1}\chi_{N-1}^2}}$ має t_{N-1} -розподіл Стьюдента.

2-й крок. На другому кроці знаходять довірчу область $D_{1-\alpha} = \left(x_{\frac{\alpha}{2}}^L; x_{\frac{\alpha}{2}}^U \right) = \left(-t_{N-1; \frac{\alpha}{2}}; t_{N-1; \frac{\alpha}{2}} \right)$, яка є симетричною, зважаючи на симетричність розподілу Стьюдента.

3-й крок. Повторення відповідника випадку відомої дисперсії з очевидною заміною σ на S та $Z_{\frac{\alpha}{2}}$ на $t_{N-1; \frac{\alpha}{2}}$, після чого отримують розв'язок задачі ТТО.

Розв'язання задачі ТТО оцінювання m за допомогою \bar{x} при невідомій дисперсії представляють довірчим інтервалом, який

має вигляд $I_{1-\alpha}^m = \left(\bar{x} - t_{N-1; \frac{\alpha}{2}} \cdot \frac{S}{\sqrt{N}}; \bar{x} + t_{N-1; \frac{\alpha}{2}} \cdot \frac{S}{\sqrt{N}} \right)$ або самим значенням

гранично допустимого відхилення $\varepsilon_{1-\alpha}^m = t_{N-1; \frac{\alpha}{2}} \cdot \frac{S}{\sqrt{N}}$.

Точність точкового оцінювання σ^2

Дисперсія помилок спостережень визначає точність вимірювання величини m : точність точкового оцінювання цієї величини. Тому в теорії помилок спостережень термін "точність приладу" чи просто "точність" вживають як синонім дисперсії помилки спостережень. Щоб уникнути колізій зі вживанням терміну "точність" у ширшому значенні, тому що він фігурує у другому типі ОЗМС (задача ТО), для позначення дисперсії цей термін запишемо в лапках.

Отже, ТО параметра σ^2 (дисперсії) нормального розподілу інтерпретують як визначення "точності" вимірювального приладу, а задача ТТО для цього параметра означає дослідження точності визначення "точності" приладу.

Можливі два варіанти цієї задачі відповідно до умов вимірювання у теорії помилок спостережень:

- m – відоме, що відповідає калібруванню приладу;
- m – невідоме, що відповідає вимірюванню приладом невідомої "точності".

В обох цих варіантах розв'язок задачі ТТО представляють тільки довірчим інтервалом.

Варіант 1, коли m відоме. Розв'язання задачі ТТО оцінювання σ^2 за допомогою $\hat{\sigma}^2$ подають довірчим інтервалом, який має вигляд

$$I_{1-\alpha}^{\sigma^2} = \left(\frac{N \cdot \hat{\sigma}^2}{\chi_{N; \frac{\alpha}{2}}^2}; \frac{N \cdot \hat{\sigma}^2}{\chi_{N; 1-\frac{\alpha}{2}}^2} \right). \quad (3.7)$$

Як і зазвичай, для розв'язування задачі ТТО необхідно виконати три стандартні кроки.

Для проходження першого потрібна лема.

Лема

Статистика $\hat{\sigma}^2 = \frac{1}{N} \cdot \sum_{i=1}^N (a_i - m)^2$ має $\frac{\sigma^2}{N} \cdot \chi_N^2$ -розподіл.

Бачимо, що статистика $\hat{\sigma}^2 = \frac{1}{N} \cdot \sum_{i=1}^N (a_i - m)^2$ і виступає оцінкою дисперсії у досліджуваному випадку.

Доведення

Дійсно, з визначення вибіркової дисперсії

$$\begin{aligned} \hat{\sigma}^2 &= \frac{1}{N} \cdot \sum_{i=1}^N (a_i - m)^2 = \frac{1}{N} \cdot \frac{\sigma^2}{\sigma^2} \cdot \sum_{i=1}^N (a_i - m)^2 = \\ &= \frac{1}{N} \cdot \sigma^2 \cdot \sum_{i=1}^N \frac{(a_i - m)^2}{\sigma^2} = \frac{1}{N} \cdot \sigma^2 \cdot \sum_{i=1}^N \left(\frac{a_i - m}{\sigma} \right)^2. \end{aligned}$$

Випадкові величини $\frac{a_i - m}{\sigma}$, $i = 1, \dots, N$, є незалежними за спадком та гауссівськими за замкненістю гауссівських розподілів щодо лінійних перетворень. Параметрами цих випадкових величин є відповідно 0 та 1 (див. 1-й крок алгоритму ТТО для оцінювання середнього при відомій дисперсії). Отже, $\frac{a_i - m}{\sigma} = Z_i$, $i = 1, \dots, N$, тобто, є набором незалежних стандартно розподіле-

них гауссівських випадкових величин. Таким чином, за означенням хі-квадрат розподілу

$$\hat{\sigma}^2 = \frac{1}{N} \sigma^2 \cdot \sum_{i=1}^N Z_i^2 = \frac{1}{N} \sigma^2 \cdot \chi_N^2.$$

Розв'язування ТТО-задачі відбувається у три кроки.

1-й крок. Розгляд статистики задачі ТТО.

Це статистика $G(\sigma^2, \hat{\sigma}^2) = \frac{N \cdot \hat{\sigma}^2}{\sigma^2}$, яка має χ_N^2 -розподіл

$$G(\sigma^2, \hat{\sigma}^2) = \frac{N \cdot \hat{\sigma}^2}{\sigma^2} \cong \chi_N^2, \text{ що є безпосереднім наслідком наведе-}$$

ної вище леми.

2-й крок. Побудова довірчої області.

Другий крок дає довірчу область $D_{1-\alpha} = \left(\chi_{N; 1-\frac{\alpha}{2}}^2; \chi_{N; \frac{\alpha}{2}}^2 \right)$:

$$D_{1-\alpha} = \left(x_{\frac{\alpha}{2}}^L; x_{\frac{\alpha}{2}}^U \right) = \left(x_{1-\frac{\alpha}{2}}^U; x_{\frac{\alpha}{2}}^U \right) = \left(\chi_{N; 1-\frac{\alpha}{2}}^2; \chi_{N; \frac{\alpha}{2}}^2 \right).$$

3-й крок. Отримання виразу-розв'язку задачі ТТО.

Перетворимо запис події $\{G(\sigma^2, \hat{\sigma}^2) \in D_{1-\alpha}\}$, записавши умову належності інтервалу через нерівності та перейшовши до обернених величин у нерівностях із додатними частинами:

$$\begin{aligned} \{G(\sigma^2, \hat{\sigma}^2) \in D_{1-\alpha}\} &\sim \left\{ \frac{N \cdot \hat{\sigma}^2}{\sigma^2} \in \left(\chi_{N; 1-\frac{\alpha}{2}}^2; \chi_{N; \frac{\alpha}{2}}^2 \right) \right\} \sim \\ &\sim \left\{ \chi_{N; 1-\frac{\alpha}{2}}^2 < \frac{N \cdot \hat{\sigma}^2}{\sigma^2} < \chi_{N; \frac{\alpha}{2}}^2 \right\} \sim \left\{ \frac{1}{\chi_{N; \frac{\alpha}{2}}^2} < \frac{\sigma^2}{N \cdot \hat{\sigma}^2} < \frac{1}{\chi_{N; 1-\frac{\alpha}{2}}^2} \right\}. \end{aligned}$$

Домножаючи обидві частини нерівностей в останньому записі події на $N \hat{\sigma}^2$, остаточно отримуємо

$$\{G(\sigma^2, \hat{\sigma}^2) \in D_{1-\alpha}\} \sim \left\{ \frac{N \cdot \hat{\sigma}^2}{\chi_{N; \frac{\alpha}{2}}^2} < \sigma^2 < \frac{N \cdot \hat{\sigma}^2}{\chi_{N; 1-\frac{\alpha}{2}}^2} \right\} \sim$$

$$\sim \left\{ \sigma^2 \in \left(\frac{N \cdot \hat{\sigma}^2}{\chi^2_{N; \frac{\alpha}{2}}}; \frac{N \cdot \hat{\sigma}^2}{\chi^2_{N; 1 - \frac{\alpha}{2}}} \right) \right\},$$

що й завершує вирішення питання про вигляд довірчого інтервалу для досліджуваного випадку.

Розв'язання задачі ТТО оцінювання σ^2 за допомогою $\hat{\sigma}^2$ при відомому m представляють довірчим інтервалом, який має вигляд

$$I_{1-\alpha}^{\sigma^2} = \left(\frac{N \cdot \hat{\sigma}^2}{\chi^2_{N; \frac{\alpha}{2}}}; \frac{N \cdot \hat{\sigma}^2}{\chi^2_{N; 1 - \frac{\alpha}{2}}} \right).$$

Варіант 2, коли m невідоме. Точковою оцінкою дисперсії σ^2 у цьому випадку є статистика S^2 :

$$S^2 = \frac{1}{N-1} \cdot \sum_{i=1}^N (a_i - \bar{x})^2,$$

яка, має $\frac{1}{N-1} \cdot \sigma^2 \cdot \chi^2_{N-1}$ -розподіл.

Розв'язок задачі ТТО оцінювання σ^2 за допомогою S^2 подають довірчим інтервалом, який має вигляд

$$I_{1-\alpha}^{\sigma^2} = \left(\frac{(N-1) \cdot S^2}{\chi^2_{N-1; \frac{\alpha}{2}}}; \frac{(N-1) \cdot S^2}{\chi^2_{N-1; 1 - \frac{\alpha}{2}}} \right). \quad (3.8)$$

Розгляд цього варіанта задачі ТТО для дисперсії повністю повторює попередній із тією різницею, що статистику задачі визначають за формулою

$$G(\sigma^2, S^2) = \frac{(N-1) \cdot S^2}{\sigma^2}, \text{ яка має } \chi^2_{N-1} \text{-розподіл:}$$

$$G(\sigma^2, S^2) = \frac{(N-1) \cdot S^2}{\sigma^2} \cong \chi^2_{N-1}.$$

Отже, довірчу область $D_{1-\alpha}$ на другому кроці будують за χ^2_{N-1} -розподілом і записують через верхні критичні величини так:

$$D_{1-\alpha} = (\chi_{N-1; 1-\frac{\alpha}{2}}^2; \chi_{N-1; \frac{\alpha}{2}}^2).$$

Проходження третього кроку з очевидними змінами повторює відповідник попереднього варіанта.

Розв'язання задачі ТТО оцінювання σ^2 за допомогою S^2 при невідомому m представляють довірчим інтервалом, який має

$$\text{вигляд } I_{1-\alpha}^{\sigma^2} = \left(\frac{(N-1) \cdot S^2}{\chi_{N-1; \frac{\alpha}{2}}^2}; \frac{(N-1) \cdot S^2}{\chi_{N-1; 1-\frac{\alpha}{2}}^2} \right).$$

**Точність точкового оцінювання
параметра m нормального $N(m; \sigma^2)$ -розподілу
при співрозмірній генеральній сукупності**

Позначимо обсяг генеральної сукупності як N_G . Результати подамо в конспективному вигляді через те, що ідеологічних розбіжностей у критеріях для випадку великої та невеликої генеральних сукупностей немає. Співрозмірність обсягу генеральної сукупності з обсягом вибірки впливає на оцінку вибіркової дисперсії, тому є сенс розглядати випадок оцінювання параметра за невідомої дисперсії. Позначимо через $S_{N_G}^2$ вибірковою дисперсією при співрозмірній генеральній сукупності та $S_{\bar{x}}^2$ – дисперсією вибіркового середнього при співрозмірній генеральній сукупності.

У цьому випадку оцінка вибіркової дисперсії

$$S_{N_G}^2 = S^2 \left(\frac{N_G - N}{N_G - 1} \right), \tag{3.9}$$

а дисперсія параметра $\hat{m} = \bar{x}$ така: $\hat{D}\hat{m} = S_{\bar{x}}^2 = \frac{S_{N_G}^2}{N}$.

Дійсно, якщо розглянути дуже велике значення N_G порівняно з невеликим N , то при граничному переході отримаємо

$$\lim_{N_G \rightarrow \infty} S_{N_G}^2 = \lim_{N_G \rightarrow \infty} S^2 \left(\frac{N_G - N}{N_G - 1} \right) = S^2,$$

тобто випадок, розглянутий вище для нескінченного (великого) обсягу генеральної сукупності.

Далі записуємо

$$\varepsilon_{1-\alpha}^m = t_{N-1; \frac{\alpha}{2}} \cdot \frac{S_{N_G}}{\sqrt{N}} = t_{N-1; \frac{\alpha}{2}} \cdot S_{\bar{x}}.$$

Звідси довірчий інтервал для параметра m

$$I_{1-\alpha}^m = \left(\bar{x} - t_{N-1; \frac{\alpha}{2}} \cdot \frac{S}{\sqrt{N}} \sqrt{\frac{N_G - N}{N_G - 1}}; \bar{x} + t_{N-1; \frac{\alpha}{2}} \cdot \frac{S}{\sqrt{N}} \sqrt{\frac{N_G - N}{N_G - 1}} \right).$$

3.4.3. Нормальне наближення в обчисленні ТТО інтегральних характеристик на базі середнього

Результати розв'язку задачі ТТО $m = MX$ нормального розподілу за відомої чи невідомої дисперсії подають у вигляді

$$\varepsilon_{1-\alpha}^m = \begin{cases} Z_{\frac{\alpha}{2}} \cdot \frac{\sigma}{\sqrt{N}}, & \sigma^2 = DX \text{ відома,} \\ t_{N-1; \frac{\alpha}{2}} \cdot \frac{S}{\sqrt{N}}, & \sigma^2 = DX \text{ не відома.} \end{cases}$$

Наближена формула ТТО для математичного сподівання нормального розподілу

Для вибірок із великим об'ємом, зважаючи на збіжність t -розподілу Стьюдента до стандартного нормального зі збільшенням кількості ступенів вільності, критичні величини t -розподілу з відповідним ступенем наближення замінюють критичними величинами стандартного нормального розподілу: $t_{N-1; \frac{\alpha}{2}} \approx Z_{\frac{\alpha}{2}}$. Отже, для невідомої дисперсії задача ТТО для пара-

метра m набуває наближеного вигляду $\varepsilon_{1-\alpha}^m \approx Z_{\frac{\alpha}{2}} \cdot \frac{S}{\sqrt{N}}$.

Тоді ТТО для параметра m нормального розподілу у випадку великих N можна записати так:

$$\varepsilon_{1-\alpha}^m = Z_{\frac{\alpha}{2}} \cdot \begin{cases} \frac{\sigma}{\sqrt{N}}, \sigma^2 = DX \text{ відома,} \\ \frac{S}{\sqrt{N}}, \sigma^2 = DX \text{ не відома.} \end{cases} \quad (3.10)$$

Тут для невідомої дисперсії рівність має наближений характер.

Для оцінювання на базі середнього задачу ТТО розв'язують наближено з **використанням нормального наближення**, формули якого мають вигляд

$$\varepsilon_{1-\alpha}^0 \approx Z_{\frac{\alpha}{2}} \cdot \begin{cases} \sqrt{D\hat{\theta}}, \\ \sqrt{\hat{D}\hat{\theta}}. \end{cases}$$

Це впливає з того, що формули ТТО для параметра m нормального розподілу з наближенням в одній із них записують як

$$\varepsilon_{1-\alpha}^m = Z_{\frac{\alpha}{2}} \cdot \begin{cases} \sqrt{D\hat{m}}, \\ \sqrt{\hat{D}\hat{m}}. \end{cases}$$

Дійсно, $D\hat{m} = D\bar{x} = \frac{\sigma^2}{N}$, звідки $\sqrt{D\hat{m}} = \sqrt{D\bar{x}} = \sqrt{\frac{\sigma^2}{N}} = \frac{\sigma}{\sqrt{N}}$. Тоді

оцінку дисперсії можна використати у записі виразу ТТО для m за відомої дисперсії σ^2 .

Для варіанта, коли дисперсія $DX = \sigma^2$ невідома, її оцінюють за допомогою S^2 , тому дисперсію середнього $\frac{\sigma^2}{N}$ можна оці-

нити за допомогою статистики $\frac{S^2}{N}$ $\hat{D}\bar{x} = \hat{D}\hat{m} = \frac{S^2}{N}$. Таким чи-

ном, $\frac{S}{\sqrt{N}} = \sqrt{\hat{D}\hat{m}}$ використовують у записі ТТО для m за невідомої дисперсії.

ТТО для параметра p бернулліївського розподілу за нормального наближення¹

За нормальним наближенням задачі ТТО для параметра p бернулліївського розподілу розв'язують за формулами

$$\varepsilon_{1-\alpha}^0 \approx Z_{\frac{\alpha}{2}} \cdot \sqrt{\frac{\hat{p}(1-\hat{p})}{N}}. \quad (3.11)$$

Вираз (3.11) аналогічний формулі нормального наближення, якщо взяти до уваги, що $\hat{p} = \mathbb{P}_N\{X = 1\} = \bar{x}$, а отже, $D\hat{p} = D\bar{x} = \frac{DX}{N}$. Таким чином, записуємо оцінку дисперсії точкової оцінки p у вигляді

$$\hat{D}\hat{p} = \hat{D}\bar{x} = \frac{\hat{D}X}{N}.$$

Оскільки для бернулліївського розподілу $DX = p(1-p)$, то як $\hat{D}\hat{p}$ обирають $\hat{D}\hat{p} = \hat{D}\bar{x} = \frac{\hat{p}(1-\hat{p})}{N}$.

Відповідно статистика

$$G(p, \hat{p}) = \frac{\hat{p} - p}{\sqrt{\frac{\hat{p}(1-\hat{p})}{N}}} \cong Z$$

розподілена за стандартним нормальним законом.

На рис. 3.8 з'явлено графіки щільності нормального розподілу (точки позначено символом +) та відповідні значення біноміального розподілу (позначено символом ○). Як видно, графіки є дуже близькими.

Існує кілька шляхів обчислення довірчого інтервалу для параметра бернулліївського розподілу.

¹ У літературі часто ці оцінки називають оцінками параметра p біноміального розподілу. Оскільки параметр p біноміального розподілу є бернулліївським, то це не є помилкою чи відмінністю. З погляду біноміального розподілу параметр \hat{p} є відношенням кількості сприятливих наслідків до загальної кількості експериментів у вибірці. Суми елементів вибірок бернулліївського з параметром p розподілу довжини n є біноміально розподіленими випадковими величинами з параметрами $(n; p)$.

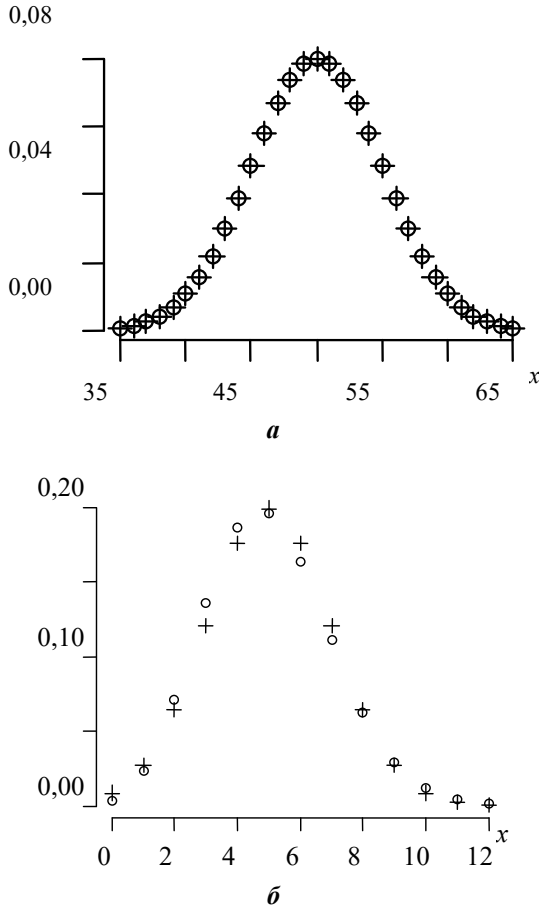


Рис. 3.8. Графіки щільності розподілу: на a позначено – як "o" біноміальний розподіл ($n=100, p=0,5$), "+" нормальний $N(50; 25)$; на b позначено – "o" біноміальний ($n=25, p=0,2$), "+" нормальний $N(5; 4)$

Довірчий інтервал для параметра p бернулліївського розподілу за нормального наближення будується найпростіше. Ця побудова базується на оцінці \hat{p} параметра p та похибки

$$\varepsilon_{1-\alpha}^{\theta} \approx Z_{\frac{\alpha}{2}} \cdot \sqrt{\frac{\hat{p}(1-\hat{p})}{N}}$$

і з урахуванням ЦГТ має вигляд

$$I_{1-\alpha}^p = \left(\hat{p} - Z_{\frac{\alpha}{2}} \cdot \sqrt{\frac{\hat{p}(1-\hat{p})}{N}}; \hat{p} + Z_{\frac{\alpha}{2}} \cdot \sqrt{\frac{\hat{p}(1-\hat{p})}{N}} \right). \quad (3.12)$$

Справді,

$$\begin{aligned} 1 - \alpha &= P \left\{ -Z_{\frac{\alpha}{2}} < G(p, \hat{p}) < Z_{\frac{\alpha}{2}} \right\} = \\ &= P \left\{ -Z_{\frac{\alpha}{2}} < \frac{\hat{p} - p}{\sqrt{\frac{\hat{p}(1-\hat{p})}{N}}} < Z_{\frac{\alpha}{2}} \right\} = \\ &= P \left\{ -Z_{\frac{\alpha}{2}} \sqrt{\frac{\hat{p}(1-\hat{p})}{N}} < \hat{p} - p < Z_{\frac{\alpha}{2}} \sqrt{\frac{\hat{p}(1-\hat{p})}{N}} \right\} = \\ &= P \left\{ \hat{p} - Z_{\frac{\alpha}{2}} \cdot \sqrt{\frac{\hat{p}(1-\hat{p})}{N}}; \hat{p} + Z_{\frac{\alpha}{2}} \cdot \sqrt{\frac{\hat{p}(1-\hat{p})}{N}} \right\}, \end{aligned}$$

тобто справджується (3.12).

Такий підхід називають методом Уальда.

Але ЦГТ є не зовсім придатною, якщо обсяг вибірки менший за 30 або ж параметр p бернулліївського розподілу близький до 0 або 1. Вважають, що застосування нормального наближення для параметра p бернулліївського розподілу є легітимним, якщо $Np > 5$ та $N(1-p) > 5$.

Довірчий інтервал Вілсона для параметра p бернулліївського розподілу є "покращеною" версією оцінки інтервалу за нормального наближення. Однією з кардинальних відмінностей є те, що його можна застосовувати для малих вибірок та для екстремальних імовірностей, тобто, коли p близький до 0 або 1, а сам розподіл має сильний скіс – має коефіцієнт симетрії, який значно відмінний від 0. Запишемо цей інтервал у вигляді

$$\left(\frac{1}{1 + \frac{1}{N} Z_{\frac{\alpha}{2}}^2} \left[\hat{p} + \frac{1}{2N} Z_{\frac{\alpha}{2}}^2 \pm Z_{\frac{\alpha}{2}} \sqrt{\frac{1}{N} \hat{p}(1-\hat{p}) + \frac{1}{4N^2} Z_{\frac{\alpha}{2}}^2} \right] \right), \quad (3.13)$$

або

$$I_{1-\alpha}^p = \left(\begin{array}{c} \frac{1}{1 + \frac{1}{N} Z_{\frac{\alpha}{2}}^2} \left[\hat{p} + \frac{1}{2N} Z_{\frac{\alpha}{2}}^2 - Z_{\frac{\alpha}{2}} \sqrt{\frac{1}{N} \hat{p}(1-\hat{p}) + \frac{1}{4N^2} Z_{\frac{\alpha}{2}}^2} \right]; \\ \frac{1}{1 + \frac{1}{N} Z_{\frac{\alpha}{2}}^2} \left[\hat{p} + \frac{1}{2N} Z_{\frac{\alpha}{2}}^2 + Z_{\frac{\alpha}{2}} \sqrt{\frac{1}{N} \hat{p}(1-\hat{p}) + \frac{1}{4N^2} Z_{\frac{\alpha}{2}}^2} \right] \end{array} \right),$$

де для обчислення нижньої границі інтервалу беруть знак $-$, а для

верхньої $+$. Центр інтервалу Вілсона² міститься в точці $\frac{\hat{p} + \frac{1}{2N} Z_{\frac{\alpha}{2}}^2}{1 + \frac{1}{N} Z_{\frac{\alpha}{2}}^2}$.

Довірчий інтервал Агресті – Коула є ще одним підходом щодо побудови довірчого інтервалу для параметра p бернулліївського розподілу при $\hat{p} \rightarrow 0$ або $\hat{p} \rightarrow 1$. У цьому методі дуже просто нівелюється проблема скосу щільності розподілу.

Нехай k – кількість сприятливих наслідків у вибірці з N елементів. Тоді, якщо позначити оцінку p як

$$\tilde{p} = \frac{k + \frac{1}{2} Z_{\frac{\alpha}{2}}^2}{N + Z_{\frac{\alpha}{2}}^2}, \quad (3.14)$$

а не як \hat{p} ³, то довірчий інтервал матиме вигляд

$$I_{1-\alpha}^p = \left(\tilde{p} - Z_{\frac{\alpha}{2}} \sqrt{\frac{\tilde{p}(1-\tilde{p})}{N}}; \tilde{p} + Z_{\frac{\alpha}{2}} \sqrt{\frac{\tilde{p}(1-\tilde{p})}{N}} \right). \quad (3.15)$$

Довірчий інтервал для параметра p бернулліївського розподілу за нормального наближення при малій генеральній сукупності будують інакше, ніж для випадку "нескінченної"

² Уперше цей інтервал отримано Едвіном Бідвелом Вілсоном у 1927 році.

³ Вище ми будували оцінку параметра як $\hat{p} = \frac{k}{N}$.

генеральної сукупності, – за допомогою оцінки дисперсії, яка враховує обсяг генеральної сукупності N_G . Оцінкою дисперсії в цьому випадку є не $S^2 = \hat{p}(1 - \hat{p})$, а

$$S_{\hat{p}}^2 = \frac{\hat{p}(1 - \hat{p})}{N} \frac{N_G - N}{N_G - 1}. \quad (3.16)$$

Похибка та довірчий інтервал для цього випадку мають вигляд відповідно

$$\varepsilon_{1-\alpha}^{\theta} \approx Z_{\frac{\alpha}{2}} \cdot \sqrt{\frac{\hat{p}(1 - \hat{p})}{N} \left(\frac{N_G - N}{N_G - 1} \right)} \quad (3.17)$$

та

$$\begin{aligned} I_{1-\alpha}^P &= \left(\hat{p} - Z_{\frac{\alpha}{2}} \cdot \sqrt{\frac{\hat{p}(1 - \hat{p})}{N} \left(\frac{N_G - N}{N_G - 1} \right)}; \hat{p} + Z_{\frac{\alpha}{2}} \cdot \sqrt{\frac{\hat{p}(1 - \hat{p})}{N} \left(\frac{N_G - N}{N_G - 1} \right)} \right) = \\ &= \left(\hat{p} - Z_{\frac{\alpha}{2}} \sqrt{S_{\hat{p}}^2}; \hat{p} + Z_{\frac{\alpha}{2}} \sqrt{S_{\hat{p}}^2} \right). \end{aligned} \quad (3.18)$$

Приклад 3.8

У вибірці обсягу 400 коледжів з 1395 виявлено, що математичну статистику викладають протягом двох семестрів у 141 коледжі. Оцінити частку серед усіх коледжів, в яких цей курс викладають протягом двох семестрів, та знайти 95%-й довірчий інтервал для цієї частки.

Розв'язання

Запишемо спочатку відомі параметри: $\hat{p} = \frac{141}{400} \approx 0,3525$; $N = 400$;

$N_G = 1395$.

Це означає, що близько у 35,25 % усіх коледжів курс викладають протягом 2 семестрів.

Маємо оцінку дисперсії

$$S_{N_G}^2 = \hat{p}(1 - \hat{p}) \frac{N_G - N}{N_G - 1} = \frac{141}{400} \left(1 - \frac{141}{400} \right) \left(\frac{1395 - 400}{1395 - 1} \right) \approx 0,162914.$$

Звідси

$$S_{\hat{p}}^2 = \frac{S_{N_G}^2}{N} \approx 0,000407, \text{ звідки } S_{\hat{p}} \approx 0,0202.$$

Для 95%-го довірчого інтервалу $Z_{\frac{\alpha}{2}} \approx 1,95996 \approx 1,96$, тому до-

вірчий інтервал $I_{1-\alpha}^p = \left(\hat{p} - Z_{\frac{\alpha}{2}} \sqrt{S_{\hat{p}}^2}; \hat{p} + Z_{\frac{\alpha}{2}} \sqrt{S_{\hat{p}}^2} \right)$ має вигляд

$$I_{1-\alpha}^p = \left(0,3525 - 1,96 \cdot \sqrt{0,000407}; 0,3525 + 1,96 \cdot \sqrt{0,000407} \right) \approx \\ \approx (0,3129; 0,3921).$$

ТТО для ймовірності $P(A)$ події A за нормальним наближенням

Розв'язком задачі ТТО в цьому випадку є

$$\varepsilon_{1-\alpha}^{p(A)} \approx Z_{\frac{\alpha}{2}} \sqrt{\frac{\mathcal{C}_N(A)(1 - \mathcal{C}_N(A))}{N}}.$$

Розв'язок задачі ТТО для ймовірності збігається з розв'язком задачі ТТО для параметра p бернулліївського розподілу, якщо бернулліївська випадкова величина визначена так, що набуває одиничного значення, коли подія A відбувається, і 0 – коли A не відбувається. Очевидним чином

$$P = P\{X=1\} = P(A).$$

Отже,

$$\hat{p} = \hat{P}(A) = \hat{P}\{X = 1\} = \mathcal{C}_N\{X = 1\}.$$

Звідси

$$\varepsilon_{1-\alpha}^{p(A)} = \varepsilon_{1-\alpha}^p \approx Z_{\frac{\alpha}{2}} \cdot \sqrt{\hat{D}\hat{p}} = Z_{\frac{\alpha}{2}} \sqrt{\frac{\hat{p} \cdot (1 - \hat{p})}{N}} = Z_{\frac{\alpha}{2}} \sqrt{\frac{\hat{P}(A) \cdot (1 - \hat{P}(A))}{N}} = \\ = Z_{\frac{\alpha}{2}} \sqrt{\frac{\mathcal{C}_N(A) \cdot (1 - \mathcal{C}_N(A))}{N}}$$

та

$$I_{1-\alpha}^p = \left(\mathcal{C}_N(A) - Z_{\frac{\alpha}{2}} \sqrt{\frac{\mathcal{C}_N(A)(1 - \mathcal{C}_N(A))}{N}}; \mathcal{C}_N(A) + Z_{\frac{\alpha}{2}} \sqrt{\frac{\mathcal{C}_N(A)(1 - \mathcal{C}_N(A))}{N}} \right).$$

ТТО для значення функції розподілу за нормальним наближенням

Оскільки значення функції розподілу для будь-якого аргументу є ймовірністю $F(x) = P\{X < x\}$ і її оцінюють у кожній точці значенням частоти емпіричної функції розподілу $F_N(x) = \frac{1}{N} \sum_{i=1}^n I_{\{X_i < x\}}$ ($\hat{F}(x) = F_N(x)$), то задача ТТО для неї може бути розв'язана за формулами ТТО для ймовірності

$$\varepsilon_{1-\alpha}^{F(x)} \approx Z_{\frac{\alpha}{2}} \sqrt{\frac{F_N(x)(1-F_N(x))}{N}}.$$

Приклад 3.9

Вибіркове середнє значення ємності шістнадцяти ($N = 16$) досліджених конденсаторів склало 20 мкФ (мікрофарада – одиниця ємності конденсатора), а відоме середньоквадратичне відхилення $\sigma = 4$ мкФ. Визначити ТТО ємності вибіркоким середнім для довірчої ймовірності 0,9. Якою мусить бути мінімальна кількість вимірів, щоб із тією самою довірчою ймовірністю точність була не меншою за 0,5 мкФ.

Розв'язання

Основні параметри задачі: $N = 16$, $\sigma = 4$, довірча ймовірність 0,9. У рамках теорії помилок спостережень використовують розв'язки задачі ТТО для параметрів нормального розподілу (у нашому випадку – для ТТО математичного сподівання за відомої дисперсії).

Оскільки дисперсія спостереження відома, використовують перший варіант розв'язку ТТО для математичного сподівання

нормального розподілу: $\varepsilon_{1-\alpha}^m = Z_{\frac{\alpha}{2}} \cdot \sqrt{D\hat{m}} = Z_{\frac{\alpha}{2}} \cdot \frac{\sigma}{\sqrt{N}}$.

За табл. 4.2 верхніх критичних величин стандартного нормального розподілу (таблиця значень функції Лапласа) знаходимо необхідне значення: $Z_{\frac{\alpha}{2}} = Z_{\frac{1-0,9}{2}} = Z_{0,05} = 1,64$.

$$\text{Отже, } \varepsilon_{0,9}^m = Z_{\frac{\alpha}{2}} \cdot \frac{\sigma}{\sqrt{N}} = Z_{0,05} \cdot \frac{4}{\sqrt{16}} = Z_{0,05} = 1,64.$$

Еквівалентним представленням розв'язку задачі ТТО є довірчий інтервал

$$I_{0,9}^m = (\bar{x} - \varepsilon_{0,9}^m; \bar{x} + \varepsilon_{0,9}^m) = (20 - 1,64; 20 + 1,64) = (18,36; 21,64).$$

Для відшукання об'єму вибірки, що забезпечує задану точність (позначимо її через Δ), зазначимо, що з виразу для точності

$$\varepsilon_{1-\alpha}^m = Z_{\frac{\alpha}{2}} \cdot \frac{\sigma}{\sqrt{N}} \text{ маємо } \Delta = Z_{\frac{\alpha}{2}} \cdot \frac{\sigma}{\sqrt{N}}. \text{ Звідси } N = Z_{\frac{\alpha}{2}}^2 \cdot \frac{\sigma^2}{\Delta^2}.$$

$$\text{Отже, } N = \left[1,64^2 \cdot \frac{4^2}{0,5^2} \right] + 1 = [172,1344] + 1 \approx 173.$$

Істинне значення ємності конденсатора оцінюють вибірко-вим середнім із точністю 1,64 для довірчої ймовірності 0,9, тобто, воно належить довірчому інтервалу (18,36; 21,65) з ймовірністю 0,9.

Мінімальна кількість вимірів, що забезпечує задану точність, дорівнює 173.

ЗАДАЧІ

№ 1

Розв'язати приклад 3.9, якщо σ невідоме, вибіркоче середньо-квадратичне відхилення склало $S = 4$ мкФ. Зробити порівняльний аналіз результатів обох задач.

№ 2

Для часу x (у днях) безвідмовної роботи ліфта маємо вибіркоче середнє $\bar{x} = 500$ днів при $N = 100$ дослідженнях. Відомо, що $\sigma = 10$ днів. Знайти довірчий інтервал для MX при $\alpha = 0,01$.

№ 3

Розв'язати попередню задачу для $S = 13$ днів. Зробити порівняльний аналіз результатів обох задач.

№ 4

Елементами вибірки є три числа 1, 2 та 3. Побудувати довірчі інтервали для MX та DX при $\alpha = 0,1$.

№ 5

Оцінка величини опору для партії резисторів, підрахована за результатами 100 тестів, склала $\bar{x} = 10$ кОм.

а) Вважаючи, що $\sigma^2 = 1 \text{ кОм}^2$, знайти ймовірність того, що для всієї партії опір становить 10 ± 1 кОм.

б) Скільки вимірів (тестів) треба зробити, щоб зі ймовірністю, не меншою за 0,95, стверджувати, що опір складає 10 ± 1 кОм.

№ 6

Результати 10 вимірів ємності конденсаторів приладом, що не має систематичної помилки ($m = 0$), склали такі відхилення від номіналу (пкФ): 5.4; -13.9; -11; 7.2; -15.6; 29.2; 1.4; -0.3; 6.6; -9.9. Знайти 90%-й довірчий інтервал для дисперсії та середньоквадратичного відхилення.

3.5. Перевірка статистичних гіпотез

3.5.1. Третій тип ОЗМС: перевірка гіпотез типу H_0 та типу альтернативного вибору

До третього типу задач математичної статистики належать задачі перевірки істинності тверджень про значення параметра розподілу, якому відповідає досліджувана вибірка. Задачами цього типу ОЗМС є перевірка наскільки апіорі інформація чи припущення узгоджуються з тією інформацією, що міститься у спостереженнях. Наприклад, піддаючи тестуванню людину з метою визначення придатності до виконання тих або інших обов'язків, припускають що вона придатна (чи непридатна) і підтверджують чи заперечують припущення на основі емпіричної інформації (спостережень), тобто вибірки. Такою емпіричною інформацією є, наприклад, результати тестування або інших однотипних випробувань. Так, якщо за однотипних тестувань отримано результати 118, 120, 117, 121, 122, 116 балів, то постає питання, визнавати чи не визнавати людину здатною виконувати пропоновані обов'язки, якщо для їх виконання достатньо, щоб рівень умінь оцінювався не нижче за 120 балів.

Звісно, якщо вимоги до того, хто проходить тестування, жорсткі, то можна відхиляти кандидатуру, коли хоча б в одному тесті він набрав менше за 120 балів. Іншим може бути підхід, за якого відхилення за нижню границю вимог викликане випадковістю. Отже, за результатами тестувань треба визначитися з тим, чи дійсно відхилення є випадковими.

До тверджень, які підлягають перевірці, коли йдеться про числові значення параметра θ , належать такі:

$$\begin{aligned}\theta &= \theta_0; \\ \theta &\geq \theta_0; \\ \theta &< \theta_0.\end{aligned}\tag{3.19}$$

Твердження, яке підлягає перевірці, називають **основною гіпотезою** і позначають як H_0 . Такими основними гіпотезами у (3.19) є кожне з наведених тверджень.

Загалом, крім основного твердження, обов'язково присутня (явно чи неявно) альтернатива основній гіпотезі. Це альтернативне твердження називають **альтернативною гіпотезою** і позначають через H_1 . Так, альтернативною гіпотезою в першому виразі (3.19) може бути $H_1: \theta = \theta_1$ чи $H_1: \theta \neq \theta_0$. В останньому випадку альтернативною гіпотезою є заперечення основної: $H_1 = \overline{H_0}$.

Якщо основна гіпотеза має вигляд $H_0: \theta = \theta_0$, її називають **простою**.

У разі, коли альтернатива простій основній гіпотезі має той самий вигляд, що й основна: $H_1: \theta = \theta_1$, то таку задачу перевірки гіпотез називають **задачею перевірки гіпотез типу альтернативного вибору**.

Задачу перевірки гіпотез, в якій альтернативна гіпотеза є запереченням основної: $H_1 = \overline{H_0}$, називають **задачею перевірки гіпотез типу H_0** . У цьому випадку альтернативу явно не згадують і розуміють за замовчуванням як протилежну до нульової.

Загалом у визначенні альтернативи через основну гіпотезу показують "нерівноправність" основної та альтернативної гіпотез. Ця нерівноправність є характерною для третьої ОЗМС і відображається в загальній формалізації прикладу 3.11, наведеного далі у пункті 3.5.4.

3.5.2. Помилки першого та другого роду. Формалізація третього типу ОЗМС

Приймаючи чи відхиляючи гіпотези (основну або альтернативну) можна припуститись помилки, яка може виникнути через відхилення (відкидання) твердження у випадку, коли воно є

істинним. Відхилення основної гіпотези, коли вона є істинною (прийняття хибної альтернативної) називають **помилкою першого роду**, прийняття основної гіпотези, коли істинною є альтернативна (відхилення істинної альтернативної), називають **помилкою другого роду**. Ті самі терміни "помилка першого роду", "помилка другого роду" вживають також тоді, коли говорять про ймовірності відповідних помилок. Загалом із контексту твердження мусить бути зрозуміло, в якому сенсі вживають відповідний вираз – щодо самої помилки чи її ймовірності.

Для формалізації третього типу ОЗМС історично склався підхід, згідно з яким при заданому наперед рівні помилок першого роду вибирають правило приймання чи відхилення основної гіпотези, яке забезпечує мінімальний рівень помилки другого роду. **Імовірність помилки першого роду** називають рівнем значущості і позначають через α . **Імовірність похибки другого роду** позначають через β .

У запропонованій формалізації основна та альтернативна гіпотези є нерівноправними: рівень помилок першого роду визначають верхньою границею, а помилки другого роду не мають певного значення – їх просто мінімізують.

При формулюванні вимог до помилок першого роду задають не рівень значущості, а величину, що доповнює його до одиниці. Цю ймовірність називають довірчою (зазвичай позначають $1-\alpha$) і вона має зміст імовірності правильного приймання основної гіпотези (тобто приймання гіпотези H_0 , коли вона вірна). Довірчу ймовірність вибирають близькою до одиниці.

Отже, **третім типом ОЗМС є задача перевірки справедливості того чи іншого твердження про істинне значення параметра θ** . Правило визнання істинним твердження (прийняття гіпотези H_0) має забезпечувати приймання, коли воно істинне в більшості випадків, що характеризується довірчою ймовірністю $1-\alpha$, забезпечуючи мінімум помилки другого роду. Це означає, що твердження основної гіпотези, коли воно істинне, має прийматися приблизно у $(1-\alpha) \cdot 100$ випадків із 100.

Приклад 3.10 (формулювання третьої ОЗМС)

Учасника дослідів визнають здатним виконувати певну роботу, коли за результатами однотипних тестувань він набирає

більше 120 очок. Зважаючи на те, що методику оцінювання характеризують стандартним відхиленням σ у 1,5 бали, перевірити гіпотезу про відповідність учасника досліджуваній умові для довірчої ймовірності 0,95, якщо результати тестування такі: 118, 120, 117, 121, 122, 116.

Розв'язання

У сформульованій задачі вибірка характеризує можливість учасника досліджуваній умові, з очевидними випадковими збуреннями від певного рівня, визначеного можливостями особи, яку перевіряють.

Основною гіпотезою виберемо твердження про те, що істинні можливості m перевищують пороговий рівень вимог, $H_0: m \geq 120$.

За умовою випробувань довірна ймовірність $1-\alpha$ дорівнює 0,95, а рівень значущості відповідно $\alpha = 0,05$.

Оскільки альтернативне твердження не наведено явно, то за замовчуванням воно полягає в запереченні основного, $H_1: m < 120$. Таким чином, запропоновану задачу відносять до третьої ОЗМС у варіанті перевірки гіпотез типу H_0 .

Відповідно до формулювання третьої ОЗМС за вибіркою роблять вибір або на користь H_0 , або на користь альтернативного твердження. При цьому правило вибору одного чи іншого твердження має забезпечувати приймання основного твердження (твердження про відповідність вимогам) приблизно у 95 випадках зі 100, коли тестована особа відповідає вимогам, а відхилення альтернативної гіпотези, коли вона вірна (особа дійсно не відповідає вимогам), відбувалося з мінімальною ймовірністю. Не знаючи ймовірності похибки другого роду, можна говорити не про прийняття нульової гіпотези, а **про неможливість її відхилення** (*fail to reject*).

Зауваження. Твердження гіпотез можна поміняти місцями, тобто, за основну прийняти гіпотезу про "невідповідність". У цьому варіанті задачі за тої самої довірчої ймовірності треба забезпечити приймання гіпотези про "невідповідність" у 95 випадках зі 100, забезпечуючи мінімальну ймовірність відхилення твердження про "відповідність".

Очевидним чином у двох наведених варіантах фігурують різні ОЗМС. Звернемо увагу, що розглянуті варіанти ОЗМС отримані один з одного переміною місць основного та альтернативного

тверджень. Відмінністю у двох задачах, про які йдеться, є прояв несиметричності основного та альтернативного тверджень.

Приклад обчислення похибки другого роду розглянуто у пункті 3.5.5.

3.5.3. Загальний алгоритм розв'язання задачі перевірки статистичних гіпотез типу H_0 : критерій згоди чи критерій значущості

Третя ОЗМС як у варіанті перевірки гіпотез типу H_0 , так і альтернативного вибору, має стандартні розв'язки. Ці розв'язки для двох варіантів формулювання ОЗМС є різними і вимагають різної додаткової інформації. Розв'язання задачі у першому варіанті (гіпотеза типу H_0) дає правило, яке називають критерієм згоди чи критерієм значущості, у другому варіанті – правило, яке називають тестом.

Таке строге розрізнення за назвами розв'язків третьої ОЗМС є характерним для вітчизняних чи російськомовних джерел. В англomовній літературі термін "тест" вживають і для обох типів гіпотез (правил прийняття чи відхилення гіпотез типу H_0 і для альтернативного вибору).

Далі, зважаючи на більшу поширеність, детально розглядатиметься загальний вигляд і конкретні приклади розв'язування задачі перевірки гіпотез типу H_0 .

Існує два підходи до перевірки гіпотез типу H_0 : 1) перевірку гіпотези здійснюють побудовою довірчого інтервалу через критичні значення відповідного розподілу; 2) перевірка гіпотези за допомогою обчислення *p-рівня* (*p-value*).

Перевірка гіпотези через довірчий інтервал

Критерієм згоди (критерієм значущості) є загальне правило прийняття чи відхилення гіпотези H_0 , яке передбачає виконання таких трьох кроків.

1-й крок. Визначення функції $G(x_1, \dots, x_N, \theta_0)$ від вибірки та гіпотетичного значення параметра, яка у випадку справедливості гіпотези H_0 , має один зі стандартних розподілів. До стандартних належать:

- Z – стандартний нормальний розподіл,
- t_n – розподіл Стюдента з n ступенями вільності,
- χ^2_n – розподіл хі-квадрат з n ступенями вільності,
- F_{mn} – розподіл Фішера – Снедекора з m та n ступенями вільності.

2-й крок. За стандартним розподілом і довірчою ймовірністю $1-\alpha$ будують довірчу область $D_{1-\alpha}$ рівня $1-\alpha$. Ця область, як правило, є двосторонньою для рівностей і односторонньою для нерівностей:

$$D_{1-\alpha} = \begin{cases} (-\infty; x_\alpha^U) - \text{права одностороння область,} \\ (x_\alpha^L; +\infty) - \text{ліва одностороння область,} \\ (x_{\alpha/2}^L; x_{\alpha/2}^U) - \text{двостороння область.} \end{cases}$$

3-й крок. Обчислюють G за наявною вибіркою a_1, \dots, a_N та гіпотетичним значенням θ_0 . а потім перевіряють отримане значення на належність довірчій області $D_{1-\alpha}$. У випадку, коли обчислене значення належить області $D_{1-\alpha}$, гіпотезу приймають, у протилежному – відхиляють:

$$G \in D_{1-\alpha} \Rightarrow H_0 - \text{приймають,}$$

$$G \notin D_{1-\alpha} \Rightarrow H_0 - \text{відхиляють.}$$

Цей підхід використовують тоді, коли потрібно знати допустимий діапазон значень характеристики (довірчий інтервал). Якщо значення статистики потрапляє до цього інтервалу, то гіпотезу приймають, в іншому разі – відхиляють. Сама ж статистика є показником відмінності між оцінкою та гіпотетичним значенням у критерії. Крім того, цей підхід дозволяє визначити довірчий інтервал $I_{1-\alpha}^\theta$ для самого параметра θ , що дає можливість проаналізувати діапазон можливих допустимих значень цього параметра.

Перевірка гіпотези через р-рівень

Будуючи довірчі інтервали, легко бачити, що довірчий інтервал рівня 95 % вузьчий за довірчий інтервал 99 %, тобто цілком можливо є ситуація, коли гіпотезу для рівня значущості 0,05 відхиляють, а для рівня 0,01 приймають. Тобто, знижуючи рі-

вень значущості, знижують імовірність відхилити справедливую нульову гіпотезу. Таким чином, це викликає підозру стосовно правильності відкидання нульової гіпотези.

Крім перевірки гіпотези при заданому рівні значущості, визначають також найменший рівень значущості, при якому нульова гіпотеза може бути відхилена, виходячи з інформації, отриманої з даної вибірки. Такий найменший рівень значущості називають p -рівень.

P -значення (p -value, probability value, p -рівень) – це ймовірність отримати значення випадкової величини, розподіленої за тим самим законом, що і статистика, таке або більш екстремальне порівняно зі значенням статистики, отриманим при справедливій нульовій гіпотезі. Таким чином, p -рівень – це мінімальна величина рівня значущості, при якому нульова гіпотеза може бути відхилена, виходячи з інформації, отриманої з даної вибірки. Це означає таке: якщо p -рівень для даної гіпотези є більшим за заданий у задачі рівень значущості, то нульову гіпотезу приймають, в іншому разі – відхиляють.

Алгоритм прийняття (відхилення) гіпотези має такий вигляд:

1. Визначають статистику $G(x_1, \dots, x_N, \theta_0)$ як функцію від вибірки та гіпотетичного значення параметра, яка у випадку справедливості гіпотези H_0 , має один зі стандартних розподілів. До стандартних належать:

- Z – стандартний нормальний розподіл,
- t_n – розподіл Стьюдента з n ступенями вільності,
- χ_n^2 – хі-квадрат з n ступенями вільності,
- F_{mn} – розподіл Фішера – Снедекора з m та n ступенями вільності.

2. Обчислюють значення G_0 статистики G для даної вибірки та знаходять інтервал D_{G_0} залежно від типу нульової гіпотези:

$$D_{G_0} = \begin{cases} (-\infty; G_0) - \text{права одностороння область при } H_0 : \theta < \theta_0, \\ (G_0; +\infty) - \text{ліва одностороння область при } H_0 : \theta > \theta_0. \end{cases}$$

Зазначимо, що для гіпотези $H_0 : \theta = \theta_0$ використовують обидва інтервали.

3. Обчислюють значення p -рівня, позначаючи через X – випадкову величину, розподілену за тим самим розподілом, що і статистика G :

$$p(G) = \min_{\alpha} \{G \notin D_{1-\alpha}\} = P(X \notin D_{G_0}). \quad (3.20)$$

Для гіпотези $H_0 : \theta = \theta_0$ знаходять обидва значення $p_1 = P(X > G_0)$ та $p_2 = P(X < G_0)$, а p -рівень обчислюють за формулою

$$p\text{-value} = 2 \min(p_1, p_2) \\ \text{або } p\text{-value} = 2 \min(p_1, 1 - p_1), \quad (3.21)$$

оскільки $p_1 = P(X > G_0)$ та $p_2 = P(X < G_0)$, через те, що $P(X = G_0) = 0$ у сумі дорівнюють 1 як імовірність достовірної події.

Далі p -рівень порівнюють із рівнем значущості α . Якщо $\alpha < p\text{-value}$, гіпотезу приймають, в іншому випадку – відхиляють.

Існує ще одна важлива перевага цього методу: p -рівень дає більш точну інформацію про "силу" відкидання нульової гіпотези.

3.5.4. Критерій згоди для параметра m нормального розподілу

У теорії помилок спостережень вважають, що істинне значення вимірюваної величини m при вимірюванні збурене помилкою спостережень ε , яка має нормальний розподіл $N(b; \sigma^2)$, отже, результатом вимірювання, згідно з властивостями нормального розподілу, є розподіл $N(m+b; \sigma^2)$:

$$m + \varepsilon \approx N(m + b; \sigma^2).$$

Параметр розподілу b помилки спостережень характеризує систематичну складову помилки, σ^2 – несистематичну, яку часто називають випадковою складовою.

Якщо систематична складова відсутня ($b = 0$, що є досить поширеним при вимірюваннях), то результат вимірювання має розподіл $N(m; \sigma^2)$: не має систематичних складових помилки і характеризується розсіюванням σ^2 .

Таким чином, задача оцінювання гіпотез про параметр m для нормального розподілу є задачею перевірки гіпотез про істинне значення вимірюваної величини.

Якщо розсіювання σ^2 відоме, то кажуть, що вимірювання проводиться з **відомою точністю**, в іншому випадку – з невідомою. Вигляд обговорюваного критерію згоди залежить від того, чи відома відповідна точність, чи ні.

**Перевірка через довірчий інтервал гіпотези
про значення параметра t нормального розподілу
у випадку, коли точність відома**

Відповідно до загальної структури критерію згоди, його побудову здійснюють за три кроки.

1-й крок. Вибір статистики зі стандартним розподілом для виконання гіпотези. Така статистика визначається співвідношенням

$$G(x_1, \dots, x_N, \theta_0) = G(x_1, \dots, x_N, m_0) = \frac{(\bar{x} - m_0)}{\sqrt{\frac{\sigma^2}{N}}} = \frac{(\bar{x} - m_0)\sqrt{N}}{\sigma}, \quad (3.22)$$

а її розподіл, якщо задовольняються твердження гіпотези, є стандартним нормальним

$$G(x_1, \dots, x_N, \theta_0) = G(x_1, \dots, x_N, m_0) \cong Z.$$

2-й крок. Побудова довірчої області.

Здійснюють за стандартним нормальним розподілом, область є двосторонньою. Зважаючи на позначення для верхніх критичних величин (у даному випадку для стандартного нормального розподілу) та приймаючи до уваги симетричність цього розподілу, маємо

$$Z \xrightarrow{\alpha} D_{1-\alpha} = \left(x_{\alpha}^L; x_{\alpha}^U \right) = \left(-Z_{\alpha}; Z_{\alpha} \right). \quad (3.23)$$

3-й крок. Останній крок полягає в обчисленні статистики

$$G(a_1, \dots, a_N, m_0) = \frac{(\bar{x} - m_0)\sqrt{N}}{\sigma}$$

та перевірці отриманого значення на належність до інтервалу попереднього кроку. У разі належності гіпотезу приймають, у протилежному випадку – відхиляють.

При перевірці гіпотез явно чи неявно разом з основним твердженням розглядають альтернативне, що є запереченням основного. Для прикладу, якщо основним твердженням є одне з

трьох: $[m = m_0]$; $[m \geq m_0]$; $[m \leq m_0]$, то альтернативним є таке: $[m \neq m_0]$; $[m < m_0]$; $[m > m_0]$ відповідно.

При перевірці гіпотез із самого початку маємо ймовірність відхилення тверджень, коли вони істинні:

- а) основного твердження;
- б) альтернативного твердження.

Відхилення істинних тверджень, коли вони є істинними, – це помилки:

першого роду, коли відхиляють основне твердження; *другого роду*, коли відхиляють альтернативне твердження.

Із самого початку у формулювання задачі перевірки гіпотез типу H_0 включають вимогу, щоб помилка першого роду не перевищувала α .

Критерій згоди вибирають таким чином, щоб мінімізувати ймовірність помилки другого роду. Причому цю вимогу потрібно включати до формулювання задачі перевірки гіпотез типу H_0 .

Наведемо зауваження до критерію згоди гіпотез про параметр m нормального розподілу з нерівностями: $H_0: m \geq m_0$ чи $H_0: m \leq m_0$. Перевірку гіпотез із нерівностями здійснюють залежно від того, відома дисперсія чи ні. Відповідно до кроків перевірки гіпотез про m при відомій та невідомій дисперсіях, замість двосторонніх беруть відповідні односторонні області.

$$H_0: m = m_0 - \left(\frac{x_\alpha^L}{2}; \frac{x_\alpha^U}{2} \right),$$

$$H_0: m \geq m_0 - \left(x_\alpha^L; +\infty \right),$$

$$H_0: m \leq m_0 - \left(-\infty; x_\alpha^U \right).$$

Приклад 3.11

При перевірці побічних дій препарату на серцево-судинну систему потрібно визначити, чи не підвищується тиск у пацієнта після вживання препарату. Для пацієнтів із цим захворюванням систолічний тиск має складати 129 мм рт. ст. із дисперсією 18 мм^2 . У піддослідній групі обсягом $N = 20$ показники тиску мають за середнє 130 мм рт. ст. Препарат не відмінитимуть до-ти, поки з довірчою ймовірністю 0,95 не виявиться, що його дія приводить до підвищення тиску у пацієнтів.

Розв'язання

Препарат відмінять тільки у випадку, коли будуть переконливі докази того, що він дійсно піднімає тиск у пацієнтів, тобто нульовою гіпотезою в нашому прикладі виступає $H_0: m \leq 129$. Оскільки дисперсія відома, то за статистику беруть

$$G(x_1, \dots, x_N, \theta_0) = \frac{(\bar{x} - m_0)\sqrt{N}}{\sigma}, \quad \text{тобто,} \quad G_0 = \frac{(130 - 129)\sqrt{20}}{\sqrt{18}} \approx 1,05.$$

Розподіл статистики – стандартний нормальний. Довірчий інтервал, виходячи з нульової гіпотези, є одностороннім вигляду $(-\infty; x_\alpha^U)$, тобто для цього випадку $D_{1-\alpha} = (-\infty; 1,645)$. Оскільки значення статистики G_0 потрапляє до довірчого інтервалу, то, маючи довірчу ймовірність 0,95, не маємо підстав відхилити нульову гіпотезу.

Отже, вважаємо, що за даними вибірки з довірчою ймовірністю 0,95 побічні дії препарату не підвищили тиск пацієнтів.

Якби середнє значення піддослідної групи було вищим, наприклад, за 131 мм рт. ст., то значення $G_0 = \frac{(131 - 129)\sqrt{20}}{\sqrt{18}} \approx 2,1$ і нульову гіпотезу з імовірністю 0,95 було б відхилено.

Перевірка гіпотези про значення параметра m нормального розподілу у випадку альтернативного вибору

Попередньо нами розглянуто задачі перевірки гіпотез типу H_0 , тобто випадки, коли альтернативною була гіпотеза, протилежна до нульової. У разі, коли альтернативна гіпотеза не є протилежною, дещо інакше будують довірчий інтервал.

Нехай нульовою гіпотезою є $H_0: m = m_0$ проти альтернативних $H_1: m < m_0$ та $H_2: m > m_0$. Якщо за статистику взяти

$$G(x_1, \dots, x_N, \theta_0) = \frac{(\bar{x} - m_0)\sqrt{N}}{\sigma}, \quad \text{то нульову гіпотезу відхиляють у}$$

разі $G_0 \leq -Z_{1-\alpha}$ при альтернативній гіпотезі $H_1: m < m_0$ і якщо $G_0 \geq Z_{1-\alpha}$ при альтернативній гіпотезі $H_2: m > m_0$. Тобто на довірчий інтервал, який будують для статистики, впливає альтернативна гіпотеза.

Аналогічні кроки роблять у випадку невідомої дисперсії, тільки за розподіл статистики беруть розподіл Стьюдента.

Приклад 3.12

Харчова цінність склянки кефіру складала 100 ккал. У результаті технологічних змін харчова цінність мала зменшитись. Для перевірки цього проведено 50 незалежних експертиз, за якими вибіркова середня харчова цінність склянки кефіру складала $\bar{x} = 92$ ккал. Припустимо, що вибірка експертиз отримана з нормальної генеральної сукупності із невідомим середнім m та дисперсією $\sigma^2 = 100$ ккал². Перевірити гіпотезу про те, що технологічні зміни не змінили харчової цінності продукту, обравши рівень значущості $\alpha = 0,05$.

Розв'язання

Визначимо основну гіпотезу $H_0: m = 100$. За статистику критерію візьмемо оцінку математичного сподівання (вибіркове середнє \bar{x}). Оскільки вибірка одержана з нормально розподіленої генеральної сукупності, то вибіркове середнє теж має нормальний розподіл, але з дисперсією $\frac{\sigma^2}{N} = \frac{100}{50} = 2$. Якщо вірна гіпотеза H_0 , то математичне сподівання цього розподілу є 100. Нормована статистика критерію (3.22) $G = \frac{\bar{x} - 100}{\sqrt{2}}$ має нормальний $N(0;1)$ розподіл. Вибіркове значення нормованої статистики становить

$$G_0 = \frac{92 - 100}{\sqrt{2}} = -\frac{8}{\sqrt{2}} = -4\sqrt{2} = -5,656.$$

Оскільки альтернативною гіпотезою є $H_1: m < 100$, то для спростування нульової гіпотези застосовують правосторонній критерій, і критичною (із (3.22)) є область $U < U_\alpha$, де $U_\alpha = U_{0,05} = -U_{0,95} = 1,645$, тобто довірчий інтервал має вигляд $(-\infty; 1,645)$. Ураховуючи, що значення статистики G потрапляє до довірчого інтервалу $(-\infty; 1,645)$, то гіпотезу $H_1: m < 100$ приймаємо з довірчою ймовірністю 0,95, а нульову відхиляємо.

Отже, вважаємо, що технологічні зміни призвели до зменшення харчової цінності кефіру.

**Перевірка через p -рівень гіпотези
про значення параметра t нормального розподілу
у випадку, коли точність відома**

Згідно із загальною структурою критерію згоди, його побудову здійснюють у три кроки.

1-й крок. Вибір статистики зі стандартним розподілом для виконання гіпотези. Такою є статистика (3.22):

$G(x_1, \dots, x_N, \theta_0) = \frac{(\bar{x} - m_0)\sqrt{N}}{\sigma}$, а її розподіл, якщо задовольняються твердження гіпотези, є стандартним нормальним $G(x_1, \dots, x_N, \theta_0) = G(x_1, \dots, x_N, m_0) \cong Z$.

2-й крок. Обчислення статистики

$$G(x_1 \dots x_N, m_0) = \frac{(\bar{x} - m_0)\sqrt{N}}{\sigma} \cong Z.$$

3-й крок. Підраховують значення p -рівня за формулами (3.20)–(3.21) та порівнюють p -рівень із рівнем значущості. Якщо рівень значущості менший за p -значення, гіпотезу приймають, у противному випадку – відхиляють.

Приклад 3.13

Перевірити гіпотезу про те, що технологічні зміни не змінили харчової цінності продукту, обравши рівень значущості $\alpha = 0,05$ за умовою прикладу 3.12 за допомогою p -рівня.

Розв'язання

З прикладу 3.12 $G = \frac{92 - 100}{\sqrt{2}} = -\frac{8}{\sqrt{2}} = -4\sqrt{2} = -5,656$.

Тепер обчислимо p -рівень для гіпотези $H_0: m = 100$. Для цього за формулами (3.20)–(3.21) маємо обчислити p -рівень для гіпотез $H_1: m < 100$ та $H_2: m > 100$.

Оскільки G за наших умов розподілено за Z -розподілом, то шуканою ймовірністю для $H_1: m < 100$ буде

$$p_1 = P(X_Z > -5,656) = 1 - \Phi(-5,656) = \Phi(5,656) = 0,999999992,$$

а для $H_2: m > 100$ імовірність дорівнює

$$p_2 = P(X_Z < -5,656) = \Phi(-5,656) = 1 - \Phi(5,656) = 0,000000008.$$

Таким чином, p -рівень для гіпотези $H_0: m = 100$ має значення $p\text{-value} = 2\min(0,999999992, 0,000000008) = 0,000000016$.

Оскільки p -рівень менший за рівень значущості, то нульову гіпотезу відкидають.

Тут можна сказати, що для цієї вибірки нульову гіпотезу можна прийняти лише з рівнем значущості, меншим за 0,000000016.

Відповідно гіпотезу H_1 приймають тому, що її p -рівень більший за рівень значущості 0,05, а H_2 відхиляють з аналогічних міркувань.

Отже, вважаємо, що технологічні зміни привели до зменшення харчової цінності кефіру.

ЗАДАЧІ

№ 1

Розв'язати приклад 3.12 за умови, що $\sigma = 25$ ккал. Виконати порівняльний аналіз результатів цієї та вихідної задач.

№ 2

Розв'язати приклад 3.12 при $N = 10$. Порівняти результати цієї задачі та прикладу 3.12.

№ 3

В умовах вихідної задачі знайти критичну область для статистики \bar{x} .

Перевірка через довірчий інтервал гіпотези про значення параметра m нормального розподілу у випадку, коли точність невідома

Для перевірки гіпотези виконують три стандартні кроки побудови критерію.

1-й крок. Статистикою критерію в досліджуваному випадку є статистика, визначена співвідношенням

$$G(x_1, \dots, x_N, \theta_0) = G(x_1, \dots, x_N, m_0) = \frac{(\bar{x} - m_0)}{\sqrt{\frac{S^2}{N}}} = \frac{(\bar{x} - m_0)\sqrt{N}}{S}, \quad (3.24)$$

котра, якщо виконуються твердження гіпотези, має розподіл Стьюдента з $N-1$ ступенем вільності:

$$\begin{aligned} G(x_1, \dots, x_N, \theta_0) &= \\ &= G(x_1, \dots, x_N, m_0) \cong t_{N-1}. \end{aligned}$$

2-й крок. Побудову довірчої області здійснюють за стандартним розподілом, яким у нашому випадку є розподіл Стьюдента з відповідною кількістю ступенів вільності, довірна область є двосторонньою. Вживаючи позначення для критичних величин стандартних розподілів та враховуючи симетричність розподілу Стьюдента, маємо

$$t_{N-1} \xrightarrow{1-\alpha} D_{1-\alpha} = \left(x_{\frac{\alpha}{2}}^L; x_{\frac{\alpha}{2}}^U \right) = \left(-t_{N-1; \frac{\alpha}{2}}; t_{N-1; \frac{\alpha}{2}} \right). \quad (3.25)$$

3-й крок. Обчислення статистики $G(a_1 \dots a_N, m_0) = \frac{(\bar{x} - m_0) \sqrt{N}}{S}$ і перевірка отриманого значення на належність до інтервалу попереднього кроку. У разі належності гіпотезу приймають, у протилежному випадку – відхиляють.

Приклад 3.14

Нехай 6 вимірювань невідомої величини m дали значення 118, 120, 117, 121, 122, 116. Для довірчої ймовірності $1-\alpha = 0,95$ перевірити гіпотезу про те, що істинне значення вимірюваної величини не менше за 120.

Розв'язання

За умовами задачі об'єм вибірки $N = 6$.

Обчислимо спочатку вибіркове середнє та вибіркочу дисперсію.

Відповідь на питання задачі є еквівалентною перевірці гіпотези вигляду $H_0: m \geq 120$.

Відповідно до стандарту критерію згоди в досліджуваному випадку для математичного сподівання нормального розподілу за невідомої дисперсії розглядаємо статистику з розподілом Стьюдента (3.24):

$$\begin{aligned} G(x_1, \dots, x_N, \theta_0) &= G(x_1, \dots, x_N, m_0) = \\ &= \frac{(\bar{x} - m_0)}{\sqrt{\frac{S^2}{N}}} = \frac{(\bar{x} - m_0) \sqrt{N}}{S} \cong t_{N-1}. \end{aligned}$$

На другому кроці з (3.25) за розподілом Стьюдента з $N-1$ ступенем вільності будемо односторонню довірчу область $D_{0,95}$:

$$Z \xrightarrow{0,95} D_{0,95} = (-x_{1-0,95}^U; +\infty) = (-t_{N-1;0,05}; +\infty) = (-t_{5;0,05}; +\infty) = (-2,015; +\infty).$$

На третьому етапі підраховуємо статистику G :

- $\hat{M}X = \bar{x} = \frac{1}{N} \sum_{i=1}^N x_i = 119$,
- $\hat{D}X = S^2 = \frac{1}{N-1} \sum_{i=1}^N (x_i - \bar{x})^2 \approx 5,6$,
- $m_0 = 120$.

Отже,

$$\begin{aligned} G &= G(x_1, \dots, x_N, m_0) = \frac{(\bar{x} - m_0)\sqrt{N}}{S} = \\ &= \frac{(119 - 120)\sqrt{6}}{\sqrt{5,6}} = -\frac{\sqrt{6}}{\sqrt{5,6}} \approx -1,035. \end{aligned}$$

Оскільки обчислене значення $G \approx -1,035$ потрапляє до довірчої області $D_{0,95} = (-2,015; +\infty)$, доходимо висновку про можливість прийняття нульової гіпотези. Таким чином, із довірчою ймовірністю 0,95 спостережувані дані є узгодженими з висунутою гіпотезою.

Перевірка через p -рівень гіпотези про значення параметра t нормального розподілу у випадку, коли точність невідома

Для перевірки гіпотези виконують три стандартні кроки побудови критерію.

1-й крок. Статистикою критерію в досліджуваному випадку є статистика, визначена співвідношенням (3.24):

$$G(a_1, \dots, a_N, \theta_0) = \frac{(\bar{x} - m_0)\sqrt{N}}{S},$$

із (якщо виконуються твердження гіпотези) розподілом Стьюдента з $N-1$ ступенем вільності $G(a_1, \dots, a_N, \theta_0) = G(a_1, \dots, a_N, m_0) \cong t_{N-1}$.

2-й крок. Обчислення статистики

$$G(a_1, \dots, a_N, \theta_0) = \frac{(\bar{x} - m_0)\sqrt{N}}{S} \cong t_{N-1}.$$

3-й крок. Пошук значення p -рівня за формулами (3.20)–(3.21) та порівняння p -рівня з рівнем значущості. Якщо рівень значущості є меншим за p -рівень, то гіпотезу приймають, у протилежному випадку – відхиляють.

Приклад 3.15

Нехай 6 вимірювань невідомої величини m дали значення 118, 120, 117, 121, 122, 116. Для довірчої ймовірності $1-\alpha = 0,95$ перевірити гіпотезу про те, що істинне значення вимірюваної величини не менше за 120.

Розв'язання

Статистикою критерію виступає

$$G(x_1, \dots, x_N, \theta_0) = \frac{(\bar{x} - m_0)\sqrt{N}}{S} \cong t_{N-1},$$

розподілена за розподілом Стьюдента з $N-1$ ступенем вільності.

Обчислимо тепер значення цієї статистики: за умовами задачі об'єм вибірки $N = 6$.

Знайдемо вибірккові середнє та дисперсію:

$$\hat{M}X = \bar{x} = \frac{1}{N} \sum_{i=1}^N x_i = 119, \quad \hat{D}X = S^2 = \frac{1}{N-1} \sum_{i=1}^N (x_i - \bar{x})^2 \approx 5,6, \quad m_0 = 120.$$

Отже, маємо значення статистики

$$G = G(x_1, \dots, x_N, m_0) = \frac{(\bar{x} - m_0)\sqrt{N}}{S} = \frac{(119 - 120)\sqrt{6}}{\sqrt{5,6}} = -\frac{\sqrt{6}}{\sqrt{5,6}} \approx -1,035.$$

Відповідь на питання задачі є еквівалентною перевірці гіпотези вигляду $H_0: m \geq 120$, тому з (3.20) отримуємо p -рівень $P(X_{t_{N-1}} < G_0) = P(X_{t_{N-1}} < -1,035) = 0,174$.

Оскільки p -рівень більший за рівень значущості, то з довірчою ймовірністю 0,95 спостережувані дані є узгодженими з висунутою нульовою гіпотезою, а отже, нема підстав відхилити нульову гіпотезу.

3.5.5. Інтерпретація похибок першого та другого роду. Потужність критерію

Як видно з означень похибок першого та другого роду для гіпотез типу H_0 є взаємно симетричними, тобто, якщо поміняти місцями гіпотези нульову та альтернативну (протилежну до нульової), то похибка першого роду перетвориться на похибку другого, а похибка другого роду – на першого.

На практиці, такої плутанини, як правило, не спостерігають через те, що за нульову гіпотезу беруть найочікуванішу, відповідно альтернативна гіпотеза є менш імовірною, неординарною, такою, що вимагає якоїсь реакції. Через це похибку першого роду називають ще хибною тривоگوю – наприклад аналіз на ВІЛ показав наявність захворювання, хоча насправді людина здорова, або металодетектор спрацював на пряжку пояса чи металеву ручку, або ми забракували фрукт, який насправді стиглий. Інакше цю похибку ще називають хибно-позитивною.

Відповідно похибку другого роду називають пропуском події або хибно-негативною. Наприклад, коли людина хвора, а тест не показав захворювання, або ж ми відібрали фрукт, який виявився гнилим.

Як правило, похибки першого та другого роду корелюють між собою. Наприклад, у медичному тесті, чим чутливіша система (менша похибка другого роду), тим більше хворих пацієнтів (небезпек) вона виявляє, але водночас кількість хибних спрацювань теж зростає, тобто зростає похибка першого роду. "Параноїдально" настроєний критерій може викликати ефект, коли з одного боку, усі випадки захворювань будуть виявлені, але і всі пацієнти за цим критерієм будуть визнані хворими.

Крім імовірностей похибок першого α та другого β роду, у статистиці використовують ще термін "потужність критерію", який обчислюють як $1 - \beta$.

Різні значення α мають свої переваги та недоліки. Менші α дають більшу впевненість у тому, що вже встановлена альтернативна гіпотеза значуща, але при цьому є більший ризик не відкинути помилкову нульову гіпотезу (похибка другого роду) і таким чином отримати меншу статистичну потужність.

Вибір α неминує вимагає компромісу між значущістю і потужністю, а, отже, між імовірністю похибок першого і другого роду.

Приклади використання похибок першого та другого роду

Боротьба з комп'ютерними вірусами. Якщо антивірусна програма класифікує файл, де вірус, при тому, що файл таким не є, то похибку відносять до похибки першого роду. Якщо ж антивірусна програма "пропускає" заражений файл – це похибка другого роду. Як правило таке відбувається, коли невірно працює евристичний аналізатор, який за "поведінкою" файла класифікує в ньому шкідливий код або ж не помічає його. Така сама ситуація із троянськими та шпигунськими програмами.

Оптичне розпізнавання тексту. Під час розпізнавання відсканованого тексту OCR-програми діють також за ймовірнісним принципом. Якщо програма сприймає літеру як сміття на сторінці і відкидає її, то маємо похибку першого роду, якщо ж програма "дофантазує" текст із забрудненої частини сторінки, то маємо похибку другого роду.

Фільтрування спаму. Якщо антиспамова програма помилково вважає звичайний лист спамом, то це класифікується як похибка першого роду. Якщо ж спам-лист "пропускається" системою, то отримуємо похибку другого роду.

Але в антиспамових програмах, крім необхідності позбуватись спаму та блокувати його розповсюдження, не менш важливим є вірно визначати звичайні листи (і не відносити їх до спаму), уникати хибно-позитивних випадків, тобто, мінімізувати похибку першого роду. Низький рівень похибки другого роду говорить про високу точність алгоритму.

Поки ще антиспам-системи не можуть на 100 % вірно фільтрувати листи, як не вдалося позбутися залежності між похибкою першого та другого роду. Так, у сучасних антиспамових системах похибка першого роду коливається у межах 0,001–3 %, а другого – від 1 до 30 %. Тому користувачеві самому потрібно вирішити, що для нього важливіше – не отримувати спаму, чи не втратити жодного звичайного листа, і відповідно налаштувати систему.

Обчислення похибки другого роду та потужності критерію

Як зазначалося вище, імовірністю похибки другого роду є ймовірність прийняти хибну нульову гіпотезу. Потужністю називають імовірність відхилити хибну нульову гіпотезу, тобто $P_{OW} = 1 - \beta$.

Розглянемо гіпотезу $H_0: m = m_0$ при відомій дисперсії σ^2 з альтернативною гіпотезою $H_1: m > m_0$. Оскільки в нашому випадку маємо гіпотезу типу альтернативного вибору, то для неї нульову гіпотезу відхилимо при $G_0 \geq Z_{1-\alpha}$, тобто $\frac{\bar{x} - m_0}{\sigma/\sqrt{N}} \geq Z_{1-\alpha}$,

звідки $\bar{x} \geq \bar{x}_0 = m_0 + Z_{1-\alpha} \frac{\sigma}{\sqrt{N}}$. Звідси можемо записати, що похи-

бку другого роду обчислюють як

$$\beta = P(\bar{x} < \bar{x}_0 / m = m_0) = P\left(Z \leq \frac{\bar{x}_0 - m_0}{\sigma/\sqrt{N}}\right), \quad (3.26)$$

а відповідно потужність дорівнює

$$1 - \beta = 1 - P(\bar{x} < \bar{x}_0 / m = m_0) = 1 - P\left(Z \leq \frac{\bar{x}_0 - m_0}{\sigma/\sqrt{N}}\right). \quad (3.27)$$

Приклад 3.16

Перевіряється гіпотеза про те, що середній обсяг ампули препарату, що виробляється на фармацевтичному заводі, складає 5 мл проти альтернативної гіпотези, що її обсяг є більшим за 5 мл. Проведено перевірку 16 ампул, випадковим чином відібраних із партії препарату, і виявилось, що обсяг ампул має нормальний розподіл зі стандартним відхиленням $\sigma = 0,1$ мл. Обравши рівень значущості $\alpha = 0,05$, обчислити похибку другого роду та потужність критерію.

Розв'язання

Визначимо основну гіпотезу $H_0: m = m_0 = 5$ при альтернативній

$H_1: m > m_0 = 5$. Оскільки вибірка одержана з нормально розподіленої генеральної сукупності, та відоме стандартне відхилення, а отже й дисперсія, тому статистикою критерію є

$G_0 = \frac{(\bar{x} - m_0)\sqrt{N}}{\sigma} \cong Z$. Через те, що альтернативною гіпотезою є

$H_1: m > 5$ при $H_0: m = 5$, то маємо перевірку гіпотези типу альте-

рнативного вибору і довірчий інтервал для статистики будемо за альтернативною гіпотезою, тобто односторонній $(-\infty; Z_{0,05}) = (-\infty; 1,645)$. Звідси випливає, що нульова гіпотеза

відхиляється при $G_0 = \frac{(\bar{x} - m_0)\sqrt{N}}{\sigma} > 1,645$. Це спостерігаємо,

підставляючи значення, при $\bar{x} > 5 + 1,645 \frac{0,1}{\sqrt{16}} = 5,041$ (на лівому

графіку рис. 3.9 виділено критичну область). Якщо ж вибіркове середнє $\bar{x} \leq 5,041$, то ми не матимемо підстав відхилити нульову гіпотезу. Отже, $\bar{x}_0 = 5,041$.

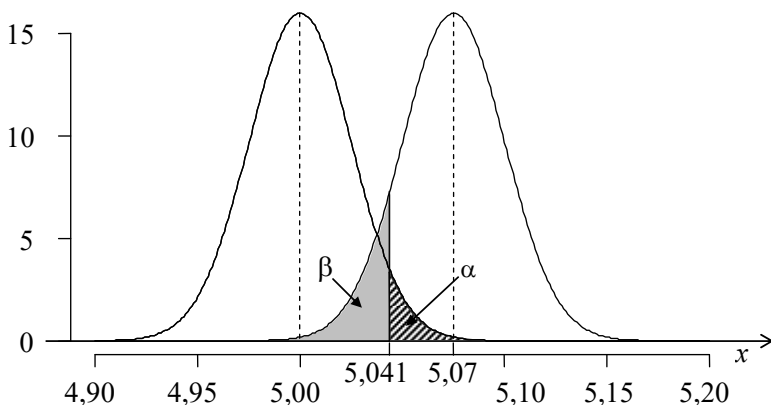


Рис. 3.9. Похибки першого та другого роду

Припустимо, що потрібно визначити ймовірність того, що в нас нема підстав відхилити нульову гіпотезу, якщо істинне значення середнього обсягу ампули препарату, що виробляється заводом (істинне значення математичного сподівання), дорівнює 5,07 мл. Це припущення підтверджує альтернативну гіпотезу і ми хочемо визначити ймовірність того, що все ж нема підстав відхилити й нульову гіпотезу, тобто отримати похибку другого роду. Іншими словами, нам потрібно обчислити ймовірність того, що вибіркове середнє є $\bar{x} \leq 5,041$, якщо середній обсяг ампул препарату на заводі (тобто істинне значення математичного сподівання) дорівнює 5,07.

Для цього скористаємося формулою (3.24), звідки отримаємо

$$\beta = P(\bar{x} < \bar{x}_0 / m = m_0) = P\left(Z \leq \frac{5,041 - 5,07}{\sigma / \sqrt{N}}\right) = P\left(Z \leq \frac{5,041 - 5,07}{0,1 / \sqrt{16}}\right) =$$

$$= P(Z \leq -1,16) = 0,123.$$

Тобто похибка другого роду при середньому обсязі ампул препарату на заводі (істинному значенні математичного сподівання) 5,07 мл дорівнює 0,123 (рис. 3.9, правий графік).

Відповідно, маємо потужність критерію для 5,07 із виразу (3.17): $1 - \beta = 1 - 0,123 = 0,877$.

3.5.6. Перевірка гіпотези про значення параметра σ^2 нормального розподілу

Перевірка гіпотез цього типу для нормального розподілу здійснюють за одним із двох критеріїв залежно від того, відоме чи невідоме математичне сподівання (тобто значення параметра m розподілу або генеральної сукупності).

Отже, маємо вибірку, отриману з нормальної генеральної сукупності. За вибірковою дисперсією потрібно перевірити гіпотезу про значення дисперсії генеральної сукупності.

Перевірка через довірчий інтервал гіпотези про значення параметра σ^2 нормального розподілу, коли m відоме

1-й крок. Статистикою критерію є функція G , яку визначають співвідношенням

$$G = \frac{N \cdot \hat{\sigma}^2}{\sigma^2} = \frac{N}{N} \sum_{i=1}^N (x_i - m)^2 / \sigma^2 = \sum_{i=1}^N (x_i - m)^2 / \sigma^2, \quad (3.28)$$

а стандартним розподілом критерію – χ^2 -квадрат з N ступенями вільності:

$$G \cong \chi_N^2.$$

2-й крок. Довірчою областю критерію $D_{1-\alpha}$ є двостороння довірча область за χ^2 -квадрат розподілом, яку, з урахуванням

загального зв'язку між нижніми та верхніми критичними величинами, представляють у вигляді

$$D_{1-\alpha} = \left(\chi_{N; 1-\frac{\alpha}{2}}^2; \chi_{N; \frac{\alpha}{2}}^2 \right) \quad (3.29)$$

3-й крок. Застосування критерію є стандартним – обчислення статистики критерію та перевірка обчисленого значення на належність довірчій області.

***Перевірка через р-рівень
гіпотези про значення параметра σ^2
нормального розподілу, коли t відоме***

1-й крок. Статистикою критерію є функція G , яку визначають співвідношенням (3.28): $G = \sum_{i=1}^N (x_i - m)^2 / \sigma^2$, розподілена за хі-квадрат розподілом з N ступенями вільності, тобто $G \cong \chi_N^2$.

2-й крок. Обчислення значення статистики критерію.

3-й крок. Підрахунок за (3.20) та (3.21) значення p -рівня відповідно до нульової гіпотези. Якщо значення p -рівня більше за рівень значущості, то нульову гіпотезу приймають (не мають підстав відхилити), в іншому випадку – відхиляють.

***Перевірка через довірчий інтервал
гіпотези про значення параметра σ^2
нормального розподілу, коли t невідоме***

1-й крок. Статистику критерію G у цьому варіанті перевірки гіпотези про значення дисперсії визначають співвідношенням

$$G = \frac{(N-1) \cdot S^2}{\sigma^2} = \frac{N-1}{N-1} \sum_{i=1}^N (a_i - \bar{x})^2 / \sigma^2 = \sum_{i=1}^N (a_i - \bar{x})^2 / \sigma^2 \quad (3.30)$$

стандартним розподілом для критерію залишають хі-квадрат, зменшуючи кількість ступенів вільності на одиницю

$$G \cong \chi_{N-1}^2.$$

2-й крок. Довірчою областю критерію $D_{1-\alpha}$ є двостороння довірча область за хі-квадратом, але з відповідно зміненою кількістю ступенів вільності:

$$D_{1-\alpha} = \left(\chi_{N-1; 1-\frac{\alpha}{2}}^2; \chi_{N-1; \frac{\alpha}{2}}^2 \right) \quad (3.31)$$

3-й крок. Застосування критерію є стандартним – обчислення статистики критерію та перевірка підрахованого значення на належність до довірчої області.

Приклад 3.17

Із нормальної генеральної сукупності отримано вибірку об'єму $N = 21$ і по ній знайдено вибіркочну дисперсію $S^2 = 16,2$. Потрібно при рівні значущості $\alpha = 0,01$ перевірити нульову гіпотезу $H_0: \sigma^2 = \sigma_0^2 = 15$.

Розв'язання

Знайдемо спочатку з (3.30) значення статистики критерію

$$G = \frac{(N-1) \cdot S^2}{\sigma^2} = \frac{(21-1) \cdot 16,2}{15} = 21,6.$$

За табл. 4.3 верхніх критичних значень, наведеною далі, із (3.31) маємо довірчий інтервал

$$D_{1-\alpha} = \left(\chi_{20; \frac{1,99}{2}}^2; \chi_{20; \frac{0,01}{2}}^2 \right) = \left(\chi_{20; 0,995}^2; \chi_{20; 0,005}^2 \right) = (7,4338; 39,9968).$$

Значення статистики G потрапляє до цього інтервалу, тому приймаємо нульову гіпотезу про те, що дисперсія генеральної сукупності дорівнює 15.

Перевіримо тепер, чи можемо ми прийняти альтернативні гіпотези $H_1: \sigma^2 > 15$ та $H_2: \sigma^2 < 15$. Довірчий інтервал для альтернативної гіпотези H_1 має вигляд $D_{1-\alpha}^{H_1} = \left(\chi_{20; 0,99}^2; +\infty \right) = (8,26; +\infty)$ та $D_{1-\alpha}^{H_2} = \left(+\infty; \chi_{20; 0,01}^2 \right) = (-\infty; 37,566)$.

Тобто, ми з довірчою ймовірністю 0,99 (рівнем значущості 0,01) приймаємо всі гіпотези. Це відбувається через те, що всі запропоновані гіпотези мають область перетину довірчих областей для статистики G , до якої потрапляє саме обчислене значення статистики $G_0 = 21,6$. Відповідно, довірчі інтервали для параметра S^2 у кожній із гіпотез матимуть спільну область. Справді, довірчими інтервалами при альтернативних гіпотезах

$H_A: \sigma^2 \neq 15$, $H_1: \sigma^2 > 15$ та $H_2: \sigma^2 < 15$ будуть $(7,43;39,997)$, $(-\infty;37,57)$ та $(8,26;+\infty)$ відповідно (рис. 3.10).

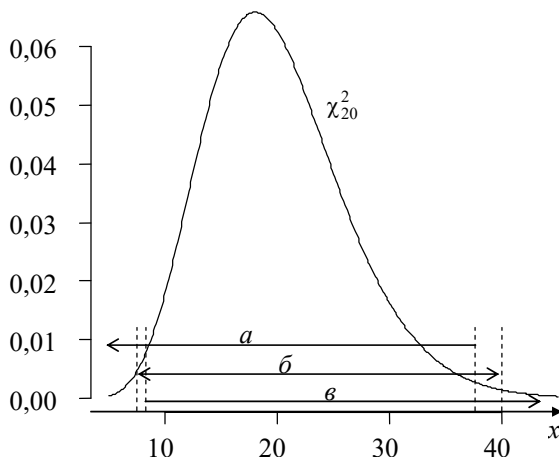


Рис. 3.10. Довірчі інтервали для статистики G при гіпотезі

$H_0: \sigma^2 = \sigma_0^2 = 15$ та різних альтернативних гіпотезах:

$a - H_1: \sigma^2 > 15$ $(-\infty;37,57)$; $b - H_A: \sigma^2 \neq 15$ $(7,43;39,997)$;

$c - H_2: \sigma^2 < 15$ $(8,26;+\infty)$

Перевірка через p -рівень гіпотези про значення параметра σ^2 нормального розподілу, коли t відоме

1-й крок. Статистика критерію – функція G , яку визначають співвідношенням (3.30): $G = \frac{(N-1) \cdot S^2}{\sigma^2} = \sum_{i=1}^N (a_i - \bar{x})^2 / \sigma^2$, розподілена за хі-квадрат розподілом з $N-1$ ступенями вільності, тобто $G \cong \chi_{N-1}^2$.

2-й крок. Обчислення значення G_0 статистики критерію G .

3-й крок. Підрахунок за (3.20) та (3.21) значення p -рівня відповідно до нульової гіпотези. Якщо значення p -рівня більше за рівень значущості, то нульову гіпотезу приймають, у протилежному випадку – відхиляють.

Приклад 3.18

Із нормальної генеральної сукупності отримано вибірку об'єму $N = 21$ і по ній знайдено вибіркочну дисперсію $S^2 = 16,2$. Потрібно при рівні значущості $\alpha = 0,01$ перевірити нульову гіпотезу $H_0: \sigma^2 = \sigma_0^2 = 15$.

Розв'язання

Знайдемо спочатку з (3.30) значення статистики критерію G :

$$G_0 = \frac{(N-1) \cdot S^2}{\sigma^2} = \frac{(21-1) \cdot 16,2}{15} = 21,6.$$

Для обчислення p -значення для нульової гіпотези з (3.20)–(3.21) потрібно спочатку отримати його для гіпотез $H_1: \sigma^2 > 15$ та $H_2: \sigma^2 < 15$. Отже, $p_1 = P(X_{\chi_{N-1}^2} < G_0) = 1 - 0,3626 = 0,6374$, відповідно

$$p_2 = P(X_{\chi_{N-1}^2} > G_0) = 0,3626.$$

Звідси p -рівень для гіпотези $H_0: \sigma^2 = 15$ запишемо як $p\text{-value} = 2\min(0,3626; 0,6374) = 0,725$. Отже, нема підстав відхилити нульову гіпотезу. Таке високе значення p -рівня говорить про велику надійність нульової гіпотези щодо поширення її на всю генеральну сукупність.

3.5.7. Довірчі інтервали для параметрів нормального розподілу

При перевірці гіпотез, як правило, у першу чергу, будують довірчий інтервал для відповідних статистик, необхідний для перевірки самої гіпотези. Але часто виникає потреба в отриманні довірчого інтервалу для самого параметра θ . Цього можна досягнути для кожної з гіпотез, виходячи з вигляду статистики та її довірчого інтервалу.

Довірчий інтервал для MX нормального розподілу

При відомій дисперсії DX за гіпотези $H_0: m = m_0$ використовують статистику $G = \frac{(\bar{x} - m_0)}{\sqrt{\frac{\sigma^2}{N}}} = \frac{(\bar{x} - m_0)\sqrt{N}}{\sigma} \cong Z$ із довірчим ін-

тервалом $D_{1-\alpha} = \left(-Z_{\frac{\alpha}{2}}; Z_{\frac{\alpha}{2}} \right)$. Запишемо вигляд довірчого інтерва-

лу $I_{1-\alpha}^m$ для параметра m із довірчою ймовірністю $1-\alpha$, тобто інтервал, "розв'язуваний" щодо m_0 .

$$\text{Маємо} \quad -Z_{\frac{\alpha}{2}} < \frac{(\bar{x} - m_0)\sqrt{N}}{\sigma} < Z_{\frac{\alpha}{2}}.$$

Звідси

$$\begin{aligned} -Z_{\frac{\alpha}{2}}\sigma < (\bar{x} - m_0)\sqrt{N} < Z_{\frac{\alpha}{2}}\sigma &\Rightarrow -Z_{\frac{\alpha}{2}}\frac{\sigma}{\sqrt{N}} < (\bar{x} - m_0) < Z_{\frac{\alpha}{2}}\frac{\sigma}{\sqrt{N}} \Rightarrow \\ \Rightarrow -Z_{\frac{\alpha}{2}}\frac{\sigma}{\sqrt{N}} - \bar{x} < -m_0 < Z_{\frac{\alpha}{2}}\frac{\sigma}{\sqrt{N}} - \bar{x} &\Rightarrow \\ \Rightarrow \bar{x} - Z_{\frac{\alpha}{2}}\frac{\sigma}{\sqrt{N}} < m_0 < \bar{x} + Z_{\frac{\alpha}{2}}\frac{\sigma}{\sqrt{N}}. \end{aligned}$$

Отже, $m_0 \in \left(\bar{x} - Z_{\frac{\alpha}{2}}\frac{\sigma}{\sqrt{N}}; \bar{x} + Z_{\frac{\alpha}{2}}\frac{\sigma}{\sqrt{N}} \right)$, а це означає, що довірчий інтервал $I_{1-\alpha}^m$ для нульової гіпотези $H_0: m = m_0$ з довірчою ймовірністю $1-\alpha$ має вигляд $I_{1-\alpha}^m = \left(\bar{x} - Z_{\frac{\alpha}{2}}\frac{\sigma}{\sqrt{N}}; \bar{x} + Z_{\frac{\alpha}{2}}\frac{\sigma}{\sqrt{N}} \right)$.

Аналогічно щодо гіпотез:

- для $H_0: m \geq m_0$ довірчий інтервал має вигляд

$$I_{1-\alpha}^m = \left(-\infty; \bar{x} + Z_{\alpha}\frac{\sigma}{\sqrt{N}} \right);$$

- для $H_0: m < m_0$ довірчий інтервал має вигляд

$$I_{1-\alpha}^m = \left(\bar{x} - Z_{\alpha}\frac{\sigma}{\sqrt{N}}; +\infty \right).$$

Таким самим шляхом визначають довірчі інтервали у випадку невідомої дисперсії:

- для $H_0: m = m_0$ маємо довірчий інтервал

$$I_{1-\alpha}^m = \left(\bar{x} - t_{N-1; \frac{\alpha}{2}}\frac{\sigma}{\sqrt{N}}; \bar{x} + t_{N-1; \frac{\alpha}{2}}\frac{\sigma}{\sqrt{N}} \right);$$

- для $H_0 : m \geq m_0$ маємо $I_{1-\alpha}^m = \left(-\infty; \bar{x} + t_{N-1; \alpha} \cdot \frac{\sigma}{\sqrt{N}} \right)$;
- для $H_0 : m < m_0$ маємо $I_{1-\alpha}^m = \left(\bar{x} - t_{N-1; \alpha} \cdot \frac{\sigma}{\sqrt{N}}; +\infty \right)$.

Довірчий інтервал для ДХ нормального розподілу

При відомому математичному сподіванні $MX = m$ за гіпотези

$$H_0 : \sigma^2 = \sigma_0^2 \text{ використовують статистику } G = \frac{N \cdot \hat{\sigma}^2}{\sigma_0^2} = \frac{\sum_{i=1}^N (a_i - m)^2}{\sigma_0^2} \cong \chi_N^2$$

з довірчим інтервалом $\left(\chi_{N; 1-\frac{\alpha}{2}}^2; \chi_{N; \frac{\alpha}{2}}^2 \right)$. Запишемо вигляд довірчого

інтервалу $I_{1-\alpha}^{\sigma^2}$ з довірчою ймовірністю $1-\alpha$, тобто інтервал, "розв'язуваний" стосовно σ^2 .

$$\text{Маємо } \chi_{N; 1-\frac{\alpha}{2}}^2 < \frac{N \cdot \hat{\sigma}^2}{\sigma_0^2} < \chi_{N; \frac{\alpha}{2}}^2.$$

Якщо розв'язати цю нерівність стосовно σ^2 , то отримаємо

$$\begin{aligned} \chi_{N; 1-\frac{\alpha}{2}}^2 < \frac{N \cdot \hat{\sigma}^2}{\sigma_0^2} < \chi_{N; \frac{\alpha}{2}}^2 &\Rightarrow \frac{1}{\chi_{N; \frac{\alpha}{2}}^2} < \frac{\sigma_0^2}{N \cdot \hat{\sigma}^2} < \frac{1}{\chi_{N; 1-\frac{\alpha}{2}}^2} \Rightarrow \\ &\Rightarrow \frac{N \cdot \hat{\sigma}^2}{\chi_{N; \frac{\alpha}{2}}^2} < \sigma_0^2 < \frac{N \cdot \hat{\sigma}^2}{\chi_{N; 1-\frac{\alpha}{2}}^2}. \end{aligned}$$

Тобто, істинне значення σ^2 генеральної сукупності з довірчою ймовірністю $1-\alpha$ має перебувати в інтервалі

$$I_{1-\alpha}^{\sigma^2} = \left(\frac{N \cdot \hat{\sigma}^2}{\chi_{N; \frac{\alpha}{2}}^2}; \frac{N \cdot \hat{\sigma}^2}{\chi_{N; 1-\frac{\alpha}{2}}^2} \right).$$

Аналогічно щодо гіпотез:

- для $H_0 : \sigma^2 \geq \sigma_0^2$ довірчий інтервал має вигляд

$$I_{1-\alpha}^{\sigma^2} = \left(-\infty; \frac{N \cdot \hat{\sigma}^2}{\chi_{N;1-\alpha}^2} \right);$$

- для $H_0 : \sigma^2 < \sigma_0^2$ довірчий інтервал має вигляд

$$I_{1-\alpha}^{\sigma^2} = \left(\frac{N \cdot \hat{\sigma}^2}{\chi_{N;\alpha}^2}; +\infty \right).$$

Таким самим шляхом визначають довірчі інтервали у випадку невідомого математичного сподівання:

- для $H_0 : \sigma^2 = \sigma_0^2$ маємо довірчий інтервал

$$I_{1-\alpha}^{\sigma^2} = \left(\frac{N \cdot S^2}{\chi_{N-1; \frac{\alpha}{2}}^2}; \frac{N \cdot S^2}{\chi_{N-1; 1-\frac{\alpha}{2}}^2} \right);$$

- для $H_0 : \sigma^2 \geq \sigma_0^2$ – $I_{1-\alpha}^{\sigma^2} = \left(-\infty; \frac{N \cdot S^2}{\chi_{N-1; 1-\alpha}^2} \right);$

- для $H_0 : \sigma^2 < \sigma_0^2$ – $I_{1-\alpha}^{\sigma^2} = \left(\frac{N \cdot S^2}{\chi_{N-1; \alpha}^2}; +\infty \right).$

3.5.8. Перевірка гіпотези про значення параметрів (таблиця критеріїв)

Запишемо сформульовані вище критерії у вигляді таблиць.

Критерій згоди для МХ нормального розподілу

Нульовою гіпотезою є $H_0 : m = m_0 \begin{pmatrix} m \geq m_0 \\ m < m_0 \end{pmatrix}$.

<i>DX</i> – відома	<i>DX</i> – невідома
1. $G(a_1, \dots, a_N, \theta_0) =$ $= G(a_1, \dots, a_N, m_0) =$ $= \frac{(\bar{x} - m_0)}{\sqrt{\frac{\sigma^2}{N}}} = \frac{(\bar{x} - m_0)\sqrt{N}}{\sigma} \cong Z;$	1. $G(a_1, \dots, a_N, \theta_0) =$ $= G(a_1, \dots, a_N, m_0) =$ $= \frac{(\bar{x} - m_0)}{\sqrt{\frac{S^2}{N}}} = \frac{(\bar{x} - m_0)\sqrt{N}}{S} \cong t_{N-1};$

<i>DX</i> – відома	<i>DX</i> – невідома
2. $D_{1-\alpha}$: $\left(x_{\alpha}^L; x_{\alpha}^U \right) [H_0 : m = m_0] = \left(-Z_{\frac{\alpha}{2}}; Z_{\frac{\alpha}{2}} \right),$ $\left(x_{\alpha}^L; +\infty \right) [H_0 : m \geq m_0] = (-Z_{\alpha}; +\infty),$ $\left(-\infty; x_{\alpha}^U \right) [H_0 : m < m_0] = (-\infty; Z_{\alpha});$	2. $D_{1-\alpha}$: $\left(x_{\alpha}^L; x_{\alpha}^U \right) [H_0 : m = m_0] = \left(-t_{N-1; \frac{\alpha}{2}}; t_{N-1; \frac{\alpha}{2}} \right),$ $\left(x_{\alpha}^L; +\infty \right) [H_0 : m \geq m_0] = (-t_{N-1; \alpha}; +\infty),$ $\left(-\infty; x_{\alpha}^U \right) [H_0 : m < m_0] = (-\infty; t_{N-1; \alpha});$
3. $G(a_1, \dots, a_N, m_0) \in D_{1-\alpha} = (x_{\alpha}^L; x_{\alpha}^U)$;	3. $G(a_1, \dots, a_N, m_0) \in D_{1-\alpha} = (x_{\alpha}^L; x_{\alpha}^U)$;
4. $I_{1-\alpha}^m : H_0 : m \geq m_0 -$ $\left(\bar{x} - Z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{N}}; \bar{x} + Z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{N}} \right),$ $H_0 : m \geq m_0 - \left(-\infty; \bar{x} + Z_{\alpha} \frac{\sigma}{\sqrt{N}} \right),$ $H_0 : m < m_0 - \left(\bar{x} - Z_{\alpha} \frac{\sigma}{\sqrt{N}}; +\infty \right);$	4. $I_{1-\alpha}^m : H_0 : m = m_0 -$ $\left(\bar{x} - t_{N-1; \frac{\alpha}{2}} \frac{\sigma}{\sqrt{N}}; \bar{x} + t_{N-1; \frac{\alpha}{2}} \frac{\sigma}{\sqrt{N}} \right),$ $H_0 : m \geq m_0 - \left(-\infty; \bar{x} + t_{N-1; \alpha} \frac{\sigma}{\sqrt{N}} \right),$ $H_0 : m < m_0 - \left(\bar{x} - t_{N-1; \alpha} \frac{\sigma}{\sqrt{N}}; +\infty \right).$

Критерій згоди для *DX* нормального розподілу

Нульовою гіпотезою є $H_0 : \sigma^2 = \sigma_0^2 \left(\begin{matrix} \sigma^2 \geq \sigma_0^2 \\ \sigma^2 < \sigma_0^2 \end{matrix} \right)$.

<i>MX</i> – відоме	<i>MX</i> – невідоме
1. $G(a_1, \dots, a_N, \theta_0) = G(a_1, \dots, a_N, \sigma_0^2) =$ $= \frac{N \cdot \hat{\sigma}^2}{\sigma^2} = \frac{\sum_{i=1}^N (a_i - m)^2}{\sigma^2} \cong \chi_{N-1}^2;$	1. $G(a_1, \dots, a_N, \theta_0) = G(a_1, \dots, a_N, \sigma_0^2) =$ $= \frac{(N-1) \cdot S^2}{\sigma^2} = \frac{\sum_{i=1}^N (a_i - \bar{x})^2}{\sigma^2} \cong \chi_{N-1}^2;$
2. $D_{1-\alpha}$: $\left(x_{\alpha}^L; x_{\alpha}^U \right) [H_0 : \sigma^2 = \sigma_0^2] = \left(\chi_{N-1; 1-\frac{\alpha}{2}}^2; \chi_{N-1; \frac{\alpha}{2}}^2 \right)$ $\left(x_{\alpha}^L; +\infty \right) [H_0 : \sigma^2 \geq \sigma_0^2] = \left(\chi_{N-1; 1-\alpha}^2; +\infty \right),$ $\left(-\infty; x_{\alpha}^L \right) [H_0 : \sigma^2 < \sigma_0^2] = \left(-\infty; \chi_{N-1; \alpha}^2 \right);$	2. $D_{1-\alpha}$: $\left(x_{\alpha}^L; x_{\alpha}^U \right) [H_0 : \sigma^2 = \sigma_0^2] = \left(\chi_{N-1; 1-\frac{\alpha}{2}}^2; \chi_{N-1; \frac{\alpha}{2}}^2 \right)$ $\left(x_{\alpha}^L; +\infty \right) [H_0 : \sigma^2 \geq \sigma_0^2] = \left(\chi_{N-1; 1-\alpha}^2; +\infty \right),$ $\left(-\infty; x_{\alpha}^L \right) [H_0 : \sigma^2 < \sigma_0^2] = \left(-\infty; \chi_{N-1; \alpha}^2 \right);$

<i>MX</i> – відоме	<i>MX</i> – невідоме
3. $G(a_1, \dots, a_N, \sigma_0^2) \in D_{1-\alpha} =$ $= (x_\alpha^L; x_\alpha^U);$	3. $G(a_1, \dots, a_N, \sigma_0^2) \in D_{1-\alpha} =$ $= (x_\alpha^L; x_\alpha^U);$
4. $I_{1-\alpha}^{\sigma^2} :$ $H_0 : \sigma^2 = \sigma_0^2 - \left(\frac{N \cdot \hat{\sigma}^2}{\chi_{N; \frac{\alpha}{2}}^2}; \frac{N \cdot \hat{\sigma}^2}{\chi_{N; 1-\frac{\alpha}{2}}^2} \right),$ $H_0 : \sigma^2 \geq \sigma_0^2 - \left(-\infty; \frac{N \cdot \hat{\sigma}^2}{\chi_{N; 1-\alpha}^2} \right),$ $H_0 : \sigma^2 < \sigma_0^2 - \left(\frac{N \cdot \hat{\sigma}^2}{\chi_{N; \alpha}^2}; +\infty \right);$	4. $I_{1-\alpha}^{\sigma^2} :$ $H_0 : \sigma^2 = \sigma_0^2 - \left(\frac{N \cdot S^2}{\chi_{N-1; \frac{\alpha}{2}}^2}; \frac{N \cdot S^2}{\chi_{N-1; 1-\frac{\alpha}{2}}^2} \right),$ $H_0 : \sigma^2 \geq \sigma_0^2 - \left(-\infty; \frac{N \cdot S^2}{\chi_{N-1; 1-\alpha}^2} \right),$ $H_0 : \sigma^2 < \sigma_0^2 - \left(\frac{N \cdot S^2}{\chi_{N-1; \alpha}^2}; +\infty \right).$

ЗАДАЧІ

№ 1

Для вибірки

-4,6141; -3,01; -2,44; -2,07; -1,78; -1,53; -1,33; -1,14; -0,97;
-0,81; -0,67; -0,53; -0,4; -0,27; -0,15; -0,03; 0,08; 0,19; 0,3; 0,4;
0,51; 0,61; 0,71; 0,81; 0,91; 1,01; 1,11; 1,21; 1,3; 1,42; 1,52; 1,62;
1,73; 1,84; 1,95; 2,06; 2,18; 2,3; 2,43; 2,56; 2,7; 2,85; 3,01; 3,18;
3,38; 3,59; 3,84; 4,15; 4,56; 5,22

а) перевірити гіпотезу $H_0: m_0=1$ при відомій дисперсії 4 з довірчою ймовірністю 0,9;

б) перевірити гіпотезу $H_0: m_0=1$ при невідомій дисперсії з довірчою ймовірністю 0,9;

в) перевірити гіпотезу $H_0: \sigma_0^2=4$ при відомому середньому 1 із довірчою ймовірністю 0,9;

г) перевірити гіпотезу $H_0: \sigma_0^2=4$ при невідомому середньому з довірчою ймовірністю 0,9;

д) знайти p -рівень для гіпотези $H_0: m_0=1$ при невідомій дисперсії;

е) знайти p -рівень для гіпотези $H_0: \sigma_0^2=4$ при невідомому середньому.

№ 2

Для вибірки

-7,4211; -5,01; -4,17; -3,6; -3,17; -2,8; -2,49; -2,21; -1,95; -1,72;
-1,5; -1,29; -1,09; -0,91; -0,73; -0,55; -0,38; -0,22; -0,06; 0,1;
0,26; 0,41; 0,57; 0,72; 0,87; 1,02; 1,17; 1,32; 1,5; 1,62; 1,78; 1,94;
2,1; 2,26; 2,42; 2,59; 2,77; 2,95; 3,14; 3,34; 3,55; 3,77; 4,01; 4,27;
4,56; 4,89; 5,27; 5,73; 6,34; 7,33

а) перевірити гіпотезу $H_0: m_0=1$ при відомій дисперсії 9 із довірчою ймовірністю 0,8;

б) перевірити гіпотезу $H_0: m_0=1$ при невідомій дисперсії з довірчою ймовірністю 0,8;

в) перевірити гіпотезу $H_0: \sigma_0^2=9$ при відомому середньому 1 із довірчою ймовірністю 0,8;

г) перевірити гіпотезу $H_0: \sigma_0^2=9$ при невідомому середньому з довірчою ймовірністю 0,8;

д) знайти p -рівень для гіпотези $H_0: m_0=1$ при невідомій дисперсії;

е) знайти p -рівень для гіпотези $H_0: \sigma_0^2=9$ при невідомому середньому.

№ 3

Для вибірки

-3,6141; -2,01; -1,44; -1,07; -0,78; -0,53; -0,33; -0,14; 0,03; 0,19;
0,33; 0,47; 0,6; 0,73; 0,85; 0,97; 1,08; 1,19; 1,3; 1,4; 1,51; 1,61;
1,71; 1,81; 1,91; 2,01; 2,11; 2,21; 2,3; 2,42; 2,52; 2,62; 2,73; 2,84;
2,95; 3,06; 3,18; 3,3; 3,43; 3,56; 3,7; 3,85; 4,01; 4,18; 4,38; 4,59;
4,84; 5,15; 5,56; 6,22

а) перевірити гіпотезу $H_0: m_0=2$ при відомій дисперсії 4 з довірчою ймовірністю 0,8;

б) перевірити гіпотезу $H_0: m_0=2$ при невідомій дисперсії з довірчою ймовірністю 0,8;

в) перевірити гіпотезу $H_0: \sigma_0^2=4$ при відомому середньому 2 з довірчою ймовірністю 0,8;

г) перевірити гіпотезу $H_0: \sigma_0^2=4$ при невідомому середньому з довірчою ймовірністю 0,8;

д) знайти p -рівень для гіпотези $H_0: m_0=2$ при невідомій дисперсії;

е) знайти p -рівень для гіпотези $H_0: \sigma_0^2=4$ при невідомому середньому.

№ 4

Для вибірки

-6,4211; -4,01; -3,17; -2,6; -2,17; -1,8; -1,49; -1,21; -0,95; -0,72; -0,5; -0,29; -0,09; 0,09; 0,27; 0,45; 0,62; 0,78; 0,94; 1,1; 1,26; 1,41; 1,57; 1,72; 1,87; 2,02; 2,17; 2,32; 2,5; 2,62; 2,78; 2,94; 3,1; 3,26; 3,42; 3,59; 3,77; 3,95; 4,14; 4,34; 4,55; 4,77; 5,01; 5,27; 5,56; 5,89; 6,27; 6,73; 7,34; 8,33

а) перевірити гіпотезу $H_0: m_0=2$ при відомій дисперсії 9 із довірчою ймовірністю 0,9;

б) перевірити гіпотезу $H_0: m_0=2$ при невідомій дисперсії з довірчою ймовірністю 0,9;

в) перевірити гіпотезу $H_0: \sigma_0^2=9$ при відомому середньому 2 з довірчою ймовірністю 0,9;

г) перевірити гіпотезу $H_0: \sigma_0^2=9$ при невідомому середньому з довірчою ймовірністю 0,9;

д) знайти p -рівень для гіпотези $H_0: m_0=2$ при невідомій дисперсії;

е) знайти p -рівень для гіпотези $H_0: \sigma_0^2=9$ при невідомому середньому.

3.5.9. Критерій Пірсона χ^2

Цей критерій застосовують для перевірки гіпотез про значення ймовірностей групованої вибірки (див. підрозд. 1.3) та перевірки гіпотези про незалежність ознак.

У обох випадках ідея застосування цього критерію така:

- для перевірки гіпотези про значення ймовірностей групованої вибірки за існуючими даними будують груповану вибірку та обчислюють її ймовірності. Потім на основі гіпотези будують гіпотетичну вибірку. Далі обчислюють відмінність між ймовірностями групованих вибірок за формулою Пірсона (3.32) і перевіряють, чи не перевищує значення статистики допустиме критичне значення.

- у разі перевірки гіпотези про незалежність двох категоріальних ознак або однієї категоріальної та однієї інтервальної, будують таблицю спряженості за існуючими даними. Далі за маргінальними розподілами будують гіпотетичну таблицю

спряженості, якою вона мала би бути у випадку незалежності двох ознак (див. пункт 2.8.2) за формулою $P\{X \in B_1, Y \in B_2\} = P\{X \in B_1\}P\{Y \in B_2\}$. Після цього також обчислюють відмінність між імовірностями групованих вибірок за формулою Пірсона (3.32) і перевіряють, чи не перевищує значення статистики допустиме критичне значення.

Перевірка гіпотези про значення ймовірностей групованої вибірки

Ця гіпотеза стосується ймовірностей належності кожного з можливих спостережень до цих груп. Належність до відповідних груп визначають інтервалом можливих значень досліджуваної характеристики для неперервних даних або окремими значеннями для дискретних даних. Отже, загалом гіпотеза критерію χ^2 має вигляд

$$H_0: p_1 = p_1^{(0)}, p_2 = p_2^{(0)}, \dots, p_M = p_M^{(0)},$$

де p_i , $i = \overline{1, M}$, – позначення для істинних, але невідомих значень імовірностей щодо належності до відповідних груп, а $p_i^{(0)}$, $i = \overline{1, M}$, – припущення про значення відповідних імовірностей (одночасно про всі). Вочевидь, сума гіпотетичних імовірностей $p_i^{(0)}$, $i = \overline{1, M}$, дорівнює 1, оскільки групи разом утворюють повну групу подій.

Статистикою критерію (тобто функцією G першого кроку стандартного критерію згоди у критерії Пірсона) є функція, яку визначають співвідношенням

$$\begin{aligned} G(a_1, \dots, a_N; p_1^{(0)}, \dots, p_M^{(0)}) &= \sum_{k=1}^M \frac{(m_k - Np_k^{(0)})^2}{Np_k^{(0)}} = \\ &= \frac{(m_1 - Np_1^{(0)})^2}{Np_1^{(0)}} + \dots + \frac{(m_M - Np_M^{(0)})^2}{Np_M^{(0)}}, \end{aligned} \quad (3.32)$$

де m_k – абсолютна частота k -ї групи за вибіркою; $p_k^{(0)}$ – істинне (гіпотетичне) значення ймовірності k -ї групи групованої вибірки; $Np_k^{(0)}$ – істинне (гіпотетичне) значення абсолютної частоти k -ї групи групованої вибірки.

Статистика цього критерію має асимптотичний (тобто при нескінченному збільшенні кількості груп) розподіл, яким є хі-квадрат із $M-1$ ступенем вільності. Таким чином, застосування цього розподілу для статистики критерію в конкретній вибірці має наближений характер

$$G \approx \chi_{M-1}^2.$$

Ця наближеність ніяк не впливає на стандартний алгоритм застосування критерію (обчислення значення статистики, побудова довірчої області та перевірка обчисленого значення на належність до довірчої області).

Довірчою областю критерію $D_{1-\alpha}$ є одностороння довірча область за розподілом хі-квадрат, але з відповідно зміненою кількістю ступенів вільності:

$$D_{1-\alpha} = \left(0; \chi_{M-1; \alpha}^2\right). \quad (3.33)$$

Дійсно, хоч гіпотеза і сформульована як $H_0: p_1 = p_1^{(0)}, p_2 = p_2^{(0)}, \dots, p_M = p_M^{(0)}$ (тобто слід використовувати двосторонній інтервал), статистикою є значення відмінності існуючих та гіпотетичних даних, яке не повинно перевищувати допустиме критичне значення, яке формується гіпотезою, рівнем значущості та вибірковими характеристиками. Оскільки статистика за своєю побудовою є невід'ємною, а саме значення статистики $G(a_1, \dots, a_N; p_1^{(0)}, \dots, p_M^{(0)}) \approx \chi_{M-1}^2$ дорівнюватиме 0 лише у випадку, коли $m_i = Np_i^{(0)}$ для всіх $i = \overline{1, M}$, тобто, коли всі наявні ймовірності групованої вибірки збігатимуться з гіпотетичними, то гіпотезу можна додатково сформулювати як $H_0: G(a_1, \dots, a_N; p_1^{(0)}, \dots, p_M^{(0)}) \leq \chi_{M-1; \alpha}^2$.

Отже, чим більше відхилення між абсолютними частотами існуючої та гіпотетичної вибірки, тим більшим буде значення χ_{M-1}^2 . Як тільки воно перевищить критичне значення $\chi_{M-1; \alpha}^2$, то це свідчатиме, що ми не можемо з імовірністю $1-\alpha$ стверджувати, що немає підстав відхилити гіпотезу H_0 , тобто, зі ймовірністю $1-\alpha$ гіпотезу H_0 буде відхилено.

Приклад 3.19

У групі зі 100 осіб, що проходили навчання, виявилося, що вчаться на добре та відмінно 25 % осіб групи, задовільні показ-

ники характерні для 55 %, а вчать незадовільно – 20 %. Загалом групам такого роду властиві показники, відповідно 20 %, 60 %, 20 %. Для довірчої ймовірності 0,95 перевірити, чи не є відхилення від стандартних показників такими, що визначаються випадковими факторами.

У наведеному прикладі присутня групована вибірка: елементи вихідної вибірки загальним об'ємом $N=100$ представлені у вигляді трьох груп із відносними частотами належності до груп (у відсотках) відповідно 25, 55 та 20 %. Припущення про можливі істинні значення частот групованої вибірки представлені частотами "нормативної" успішності: 20 %, 60 %, 20 %. Саме ці значення і будуть складати ймовірності гіпотези про істинні значення ймовірностей групованої вибірки, і відповідно позначатимуться $p_1^{(0)} = 0,2$, $p_2^{(0)} = 0,6$ та $p_3^{(0)} = 0,2$.

Таким чином гіпотеза в цілому має вигляд

$$H_0: p_1 = 0,2; p_2 = 0,6; p_3 = 0,2.$$

Значення статистики критерію з (3.22) представлено виразом

$$\begin{aligned} G(a_1, \dots, a_{100}; p_1^{(0)}, p_2^{(0)}, p_3^{(0)}) &= \\ &= \frac{(m_1 - 100p_1^{(0)})^2}{100p_1^{(0)}} + \frac{(m_2 - 100p_2^{(0)})^2}{100p_2^{(0)}} + \frac{(m_3 - 100p_3^{(0)})^2}{100p_3^{(0)}} = \\ &= \frac{(25 - 100 \cdot 0,2)^2}{100 \cdot 0,2} + \frac{(55 - 100 \cdot 0,6)^2}{100 \cdot 0,6} + \frac{(20 - 100 \cdot 0,2)^2}{100 \cdot 0,2} = 1,66. \end{aligned}$$

Для довірчої ймовірності 0,95 та кількості груп $M = 3$ використовують критичну величину $\chi_{M-1; \alpha}^2 = \chi_{3-1; 1-0,95}^2 = \chi_{2; 0,05}^2$, значення якої береться з таблиці критичних величин для відповідного χ^2 -розподілу (див. далі табл. 4.3) $\chi_{2; 0,05}^2 \approx 5,99$. Оскільки розрахункове значення статистики розподілу становить 1,6 і це значення менше за табличне 5,99, то немає підстави відхилити нульову гіпотезу, тобто з довірчою ймовірністю 0,95 гіпотезу про те, що відхилення від стандартних показників є такими, що визначаються випадковими факторами, приймають.

Приклад 3.20 (застосування критерію Пірсона для групованої вибірки)

Вибірку об'єму 50 представлено таблицею :

0,632	2,233	-0,526	1,587	7,7
0,149	4,334	-0,230	4,549	1,835
-0,029	2,386	3,259	1,383	3,970
3,830	-1,221	2,851	0,902	1,735
5,457	3,358	1,682	3,517	3,754
3,819	1,347	1,463	5,663	-2,3
0,929	3,701	-0,631	2,753	0,709
3,180	5,479	4,061	6,340	2,235
2,406	1,922	2,180	-0,353	2,276
0,439	1,687	1,411	3,600	1,576

Для довірчої ймовірності 0,9 перевірити гіпотезу про те, що ці дані є вибіркою з гауссівського $N(2;3)$ розподілу.

Гіпотези про узгодженість вибірових значень із тим чи іншим розподілом є типовим класом гіпотез, які перевіряють через застосування критерію Пірсона. Таке застосування вимагає перетворення вихідної вибірки у груповану. Кількість інтервалів групування визначають за формулою Стерджеса (1.2).

Абсолютні частоти $m_i, i = \overline{1, M}$, для кожної з груп підраховують як кількість елементів вихідної вибірки, що належать відповідному інтервалу.

У нашому прикладі $M = [1+3,322\lg 50] = [6,64] = 6$.

Найменший з елементів вибірки дорівнює $-2,3$, найбільший $7,7$. Обчислимо розмах

$$R = 7,7 - (-2,3) = 10$$

та довжину інтервалу за формулою (1.3)

$$h = \frac{10}{6-1} = 2.$$

Розіб'ємо інтервал $(-2,3-h/2; 7,7+h/2)$ на 6 однакових за довжиною інтервалів

$[-3,3; -1,3), [-1,3; 0,7), [0,7; 2,7), [2,7; 4,7), [4,7; 6,7), [6,7; 8,7)$.

До цих інтервалів із вибірки потрапило відповідно 1, 9, 20, 15, 4, 1 елементів, тобто, абсолютними частотами для відповідних інтервалів будуть 1, 9, 20, 15, 4, 1.

Імовірності гіпотези про групувану вибірку формують як імовірності гіпотетичного розподілу для інтервалів розбиття. Для нашого прикладу ці ймовірності з (2.12, а) визначають за формулами

$$p_1^{(0)} = P\{\xi \in (-3, 3; -1, 3)\} = F_\xi(-1, 3) - F_\xi(-3, 3) = \\ = \Phi\left(\frac{-1, 3 - 2}{\sqrt{3}}\right) - \Phi\left(\frac{-3, 3 - 2}{\sqrt{3}}\right) \approx \Phi(-1, 9) - \Phi(-3, 06) \approx 0, 027'$$

де $F_\xi(-1, 3)$ – значення функції розподілу в точці $-1, 3$ ймовірностей для нормального (у нашому випадку $N(2; 3)$ розподілу); $F_\xi(-3, 3)$ – для точки $-3, 3$. Табличні значення функції Лапласа наведено для стандартного нормального розподілу, тому ми вимушені скористатися співвідношенням $F_{m, \sigma^2}(x) = \Phi\left(\frac{x - m}{\sigma}\right)$, тобто, зважаючи на те, що за умовою задачі $m = 2$, а $\sigma^2 = 3$, отримаємо

$$F_\xi(-1, 3) = \Phi\left(\frac{-1, 3 - 2}{\sqrt{3}}\right) \text{ і відповідно } F_\xi(-3, 3) = \Phi\left(\frac{-3, 3 - 2}{\sqrt{3}}\right).$$

Аналогічно

$$p_2^{(0)} = P\{\xi \in (-1, 3; 0, 7)\} = F_\xi(0, 7) - F_\xi(-1, 3) = \\ = \Phi\left(\frac{0, 7 - 2}{\sqrt{3}}\right) - \Phi\left(\frac{-1, 3 - 2}{\sqrt{3}}\right) \approx 0, 2;$$

$$p_3^{(0)} = P\{\xi \in (0, 7; 2, 7)\} = F_\xi(2, 7) - F_\xi(0, 7) = \\ = \Phi\left(\frac{2, 7 - 2}{\sqrt{3}}\right) - \Phi\left(\frac{0, 7 - 2}{\sqrt{3}}\right) \approx 0, 43;$$

$$p_4^{(0)} = P\{\xi \in (2, 7; 4, 7)\} = F_\xi(4, 7) - F_\xi(2, 7) = \\ = \Phi\left(\frac{4, 7 - 2}{\sqrt{3}}\right) - \Phi\left(\frac{2, 7 - 2}{\sqrt{3}}\right) \approx 0, 284;$$

$$p_5^{(0)} = P\{\xi \in (4, 7; 6, 7)\} = F_\xi(6, 7) - F_\xi(4, 7) = \\ = \Phi\left(\frac{6, 7 - 2}{\sqrt{3}}\right) - \Phi\left(\frac{4, 7 - 2}{\sqrt{3}}\right) \approx 0, 056;$$

$$\begin{aligned}
 p_6^{(0)} &= P\{\xi \in (6, 7; 8, 7)\} = F_\xi(8, 7) - F_\xi(6, 7) = \\
 &= \Phi\left(\frac{8, 7 - 2}{\sqrt{3}}\right) - \Phi\left(\frac{6, 7 - 2}{\sqrt{3}}\right) \approx 0,003.
 \end{aligned}$$

Тоді значення статистики критерію з (3.32) має вигляд

$$\begin{aligned}
 G_0 &= \frac{(1 - 50 \cdot 0,027)^2}{50 \cdot 0,027} + \frac{(9 - 50 \cdot 0,2)^2}{50 \cdot 0,2} + \frac{(20 - 50 \cdot 0,43)^2}{50 \cdot 0,43} + \frac{(15 - 50 \cdot 0,284)^2}{50 \cdot 0,284} + \\
 &+ \frac{(4 - 50 \cdot 0,056)^2}{50 \cdot 0,056} + \frac{(1 - 50 \cdot 0,003)^2}{50 \cdot 0,003} \approx 5,65.
 \end{aligned}$$

Оскільки довірна ймовірність за умовою задачі дорівнює 0,9, то з (3.33) за табл. 4.3 знаходять критичний рівень значущості 0,1, а оскільки груп 6, то кількість ступенів вільності $6 - 1 = 5$. Цим табличним значенням є 9,2364: $\chi_{5;0,1}^2 \approx 9,2364$. Оскільки $G = 5,65 < 9,2364 = \chi_{5;0,1}^2$ (тобто потрапляє до довірчого інтервалу), то гіпотезу про узгодженість вибірки з розподілом $N(2,3)$ приймають.

Зауважимо, що для нашого прикладу навіть для довірчої ймовірності 0,7 одержимо $\chi_{5;0,3}^2 \approx 6,06443$, тобто маємо $G = 5,65 < 6,06443$ і гіпотезу теж приймаємо.

Більш складним і звичним випадком є перевірка гіпотези про узгодженість вибірки розподілу, коли параметри розподілу відсутні, а відомий лише його тип. Розв'язання задачі в цьому випадку полягає у попередньому оцінюванні цих параметрів розподілу, а решту дій виконують, як у прикладі 3.20.

Обчислення параметрів розподілу зменшує кількість ступенів вільності в розподілі статистики критерію на кількість оцінених за вихідною вибіркою параметрів (у прикладі 3.21 на два – невідомі математичне сподівання та дисперсію як параметри нормального розподілу).

Приклад 3.21

Для вибірки об'єму 100, що представлена таблицею, для довірчої ймовірності 0,999 перевірити гіпотезу про те, що вибірка отримана з гауссівського розподілу:

3,082	0,767	0,597	1,634	1,494	2,217	1,316	0,744	-0,508	0,901
0,885	1,517	0,878	0,23	2,976	-0,788	0,471	0,404	-0,174	1,068
0,218	0,464	1,727	1,036	3,407	0,486	1,816	0,444	-0,03	0,618
1,88	0,641	0,447	0,249	1,023	0,554	-0,548	0,665	2,296	-0,349
3,139	0,843	0,28	1,574	2,626	0,806	-0,322	1,369	1,019	0,237
2,755	-0,222	2,154	2,868	0,754	0,77	-0,199	-0,183	0,8	2,44
0,83	0,200	3,244	-1,108	1,585	0,246	1,935	0,239	0,267	0,916
0,448	2,008	0,928	2,38	2,944	-0,201	1,459	2,35	1,456	0,739
0,31	1,060	2,35	-1,993	0,904	1,306	0,949	1,038	1,524	2,284
1,010	-0,102	1,114	1,502	1,873	0,343	0,24	0,734	2,438	2,977

Спочатку оцінимо параметри гіпотетичного гауссівського розподілу: математичного сподівання m та дисперсії σ^2 відповідно вибіркоvim середнім та вибірковою дисперсією (див. підрозд. 3.2):

$$\hat{m} = \bar{x} = \frac{1}{100} \sum_{i=1}^{100} a_i \approx 1,05, \quad S^2 = \frac{1}{99} \sum_{k=1}^{100} (a_i - \bar{x})^2 \approx 1,09.$$

Отже, задача, як у прикладі 3.20, звелася до перевірки гіпотези про нормальність вибірки, коли параметри розподілу відомі, тобто про узгодженість спостережень із припущенням про $N(1,05; 1,09)$ розподіл.

Із формули Стерджеса (1.2) маємо кількість інтервалів $M=7$. Обчислимо елементи розбиття можливих значень та відповідні їм абсолютні частоти. Оскільки найменше спостереження $-1,993$, а найбільше $3,407$, то довжина інтервалу $h = (3,407+1,993)/6 = 0,9$, а розбиття інтервалу $(-1,993-h/2; 3,5+h/2)$ на 7 частин має вигляд $[-2,443; -1,543)$, $[-1,543; -0,643)$, $[-0,643; 0,257)$, $[0,257; 1,157)$, $[1,157; 2,057)$, $[2,057; 2,957)$, $[2,957; 3,857)$ із відповідними абсолютними частотами 1, 2, 19, 41, 18, 13, 6.

На практиці інтервали, до яких потрапило менше трьох спостережень, об'єднують, що і зробимо, об'єднавши перші два інтервали: $[-2,443; -0,643)$, $[-0,643; 0,257)$, $[0,257; 1,157)$, $[1,157; 2,057)$, $[2,057; 2,957)$, $[2,957; 3,857)$. Відповідні цим інтервалам абсолютні частоти мають такі значення: 3, 19, 41, 18, 13, 6.

Імовірності гіпотези про групувану вибірку формують як імовірності гіпотетичного розподілу для інтервалів розбиття:

$$\begin{aligned}
 p_1^{(0)} &= P\{\xi \in (-2,443; -0,643)\} = F_\xi(-0,643) - F_\xi(-2,443) = \\
 &= \Phi\left(\frac{-0,643 - 1,05}{\sqrt{1,09}}\right) - \Phi\left(\frac{-2,443 - 1,05}{\sqrt{1,09}}\right) \approx 0,052; \\
 p_2^{(0)} &= P\{\xi \in (-0,643; 0,257)\} \approx 0,17; \\
 p_3^{(0)} &= P\{\xi \in (0,257; 1,157)\} \approx 0,317; \\
 p_4^{(0)} &= P\{\xi \in (1,157; 2,057)\} \approx 0,292; \\
 p_5^{(0)} &= P\{\xi \in (2,057; 2,957)\} \approx 0,134; \\
 p_6^{(0)} &= P\{\xi \in (2,957; 3,857)\} \approx 0,03.
 \end{aligned}$$

Тоді значення статистики критерію з (3.32) має вигляд

$$\begin{aligned}
 G_0 &= \frac{(3 - 100 \cdot 0,052)^2}{100 \cdot 0,052} + \frac{(19 - 100 \cdot 0,17)^2}{100 \cdot 0,17} + \\
 &+ \frac{(41 - 100 \cdot 0,317)^2}{100 \cdot 0,317} + \frac{(18 - 100 \cdot 0,292)^2}{100 \cdot 0,292} + \\
 &+ \frac{(13 - 100 \cdot 0,134)^2}{100 \cdot 0,134} + \frac{(6 - 100 \cdot 0,03)^2}{100 \cdot 0,03} \approx 11,07.
 \end{aligned}$$

Оскільки довірча ймовірність за умовою задачі дорівнює 0,999, то з (3.33) визначенню за табл. 4.3 підлягає критична величина рівня 0,001, а оскільки груп 6, то кількість ступенів вільності $6 - 2 - 1 = 3$. Цим табличним значенням є 16,27: $\chi_{3,0,001}^2 \approx 16,27$. Оскільки $G = 11,07 < \chi_{3,0,001}^2 \approx 16,27$, то гіпотезу про узгодженість вибірки з гауссівським розподілом $N(1,05; 1,09)$ приймають, тобто можемо вважати, що вибірка здійснена з гауссівського розподілу.

Специфікою статистики критерію χ^2 є інформація про різницю між наявною абсолютною частотою групи m_i та передбачуваною абсолютною частотою $N \cdot p_i^{(0)}$ за передбачуваною властивістю ймовірностей.

Перевірка гіпотези про незалежність двох ознак

Критерій хі-квадрат також застосовують для перевірки гіпотези про незалежність нечислових (номінальних, дихотомічних) ознак.

Для застосування критерію хі-квадрат у цьому випадку потрібно виконати алгоритм, що складається з 5 кроків.

1-й крок. Позначимо ознаки, стосовно яких перевіряємо гіпотезу незалежності X та Y . Спочатку за двовимірною вибіркою (дві ознаки та N респондентів, із цього маємо двовимірну вибірку $2 \times N$) будують таблицю спряженості m_{ij} , кожним елементом якої є абсолютна частота спільної появи наслідків кожної з двох подій. Розмір цієї таблиці дорівнює $k \times n$, де k – кількість альтернатив першої ознаки, n – кількість альтернатив другої ознаки. Елементи таблиці дорівнюють абсолютним частотам $m_{ij} = m_{\{X=x_i, Y=y_j\}}$, де $x_i = x_1, \dots, x_k$ та $y_j = y_1, \dots, y_n$ альтернативи ознак X та Y відповідно.

$Y \backslash X$	x_1	x_2	...	x_k
y_1	m_{11}	m_{12}	...	m_{1k}
y_2	m_{21}	m_{22}	...	m_{2k}
...
y_n	m_{n1}	m_{n2}	...	m_{nk}

2-й крок. Будують маргінальні розподіли кожної з ознак X та Y . Для цього знаходять абсолютні частоти кожної з ознак: для ознаки X додають частоти у стовпчиках, для Y – у рядках.

$Y \backslash X$	x_1	x_2	...	x_k	m_Y
y_1	m_{11}	m_{12}	...	m_{1k}	$m_{11} + m_{12} + \dots + m_{1k}$
y_2	m_{21}	m_{22}	...	m_{2k}	$m_{21} + m_{22} + \dots + m_{2k}$
...
y_n	m_{n1}	m_{n2}	...	m_{nk}	$m_{n1} + m_{n2} + \dots + m_{nk}$
m_X	$m_{11} + m_{21} + \dots + m_{n1}$	$m_{12} + m_{22} + \dots + m_{n2}$...	$m_{1k} + m_{2k} + \dots + m_{nk}$	N

Оскільки кількість респондентів дорівнює N , то маргінальні ряди розподілу мають вигляд

X	x_1	x_2	...	x_k
$\mathcal{C}_N(X)$	$\frac{\sum_{i=1}^n m_{i1}}{N}$	$\frac{\sum_{i=1}^n m_{i2}}{N}$...	$\frac{\sum_{i=1}^n m_{ik}}{N}$

та

Y	x_1	x_2	...	x_k
$\mathcal{C}_N(Y)$	$\frac{\sum_{j=1}^k m_{1j}}{N}$	$\frac{\sum_{j=1}^k m_{2j}}{N}$...	$\frac{\sum_{j=1}^k m_{nj}}{N}$

3-й крок. За маргінальними розподілами будують гіпотетичну таблицю спряженості, використовуючи властивість незалежності двох ознак (2.5), за якою сумісна ймовірність (відносна частота) двох незалежних подій дорівнює добутку маргінальних імовірностей (відносних частот) кожної з подій, тобто для незалежних X та Y виконується $\mathcal{C}_N\{X = a, Y = b\} = \mathcal{C}_N\{X = a\}\mathcal{C}_N\{Y = b\}$ та властивості (1.1) абсолютної частоти $m_A = N \cdot \mathcal{C}_N(A)$:

$Y' \backslash X'$	x_1	x_2	...	x_k
y_1	$N \frac{\sum_{j=1}^k m_{1j}}{N} \cdot \frac{\sum_{i=1}^n m_{i1}}{N}$	$N \frac{\sum_{j=1}^k m_{1j}}{N} \cdot \frac{\sum_{i=1}^n m_{i2}}{N}$...	$N \frac{\sum_{j=1}^k m_{1j}}{N} \cdot \frac{\sum_{i=1}^n m_{ik}}{N}$
y_2	$N \frac{\sum_{j=1}^k m_{2j}}{N} \cdot \frac{\sum_{i=1}^n m_{i1}}{N}$	$N \frac{\sum_{j=1}^k m_{2j}}{N} \cdot \frac{\sum_{i=1}^n m_{i2}}{N}$...	$N \frac{\sum_{j=1}^k m_{2j}}{N} \cdot \frac{\sum_{i=1}^n m_{ik}}{N}$
...
y_n	$N \frac{\sum_{j=1}^k m_{nj}}{N} \cdot \frac{\sum_{i=1}^n m_{i1}}{N}$	$N \frac{\sum_{j=1}^k m_{nj}}{N} \cdot \frac{\sum_{i=1}^n m_{i2}}{N}$...	$N \frac{\sum_{j=1}^k m_{nj}}{N} \cdot \frac{\sum_{i=1}^n m_{ik}}{N}$

4-й крок. Позначимо $p_{ij}^{(0)} = \frac{\sum_{a=1}^k m_{ia}}{N} \cdot \frac{\sum_{b=1}^n m_{bj}}{N}$ – відносна частота⁴

одночасної появи наслідків $X' = j$ та $Y' = i$ для незалежних X' та Y' , тобто $p_{ij}^{(0)} = \mathbb{P}_N(X' = j, Y' = i)$. Тоді статистика критерію має вигляд

$$G(X_1, \dots, X_N; Y_1, \dots, Y_N) = \sum_{j=1}^k \sum_{i=1}^n \frac{(m_{ij} - Np_{ij}^{(0)})^2}{Np_{ij}^{(0)}} \cong \chi_{(k-1)(n-1)}^2.$$

5-й крок. Для рівня значущості α знаходять верхнє критичне значення $\chi_{(k-1)(n-1); \alpha}^2$ та порівнюють його зі значенням статистики. У випадку, коли статистика менша за це критичне значення, тобто рівень відхилення між частотами даної таблиці спряженості та гіпотетичної (якими вони мали би бути у випадку незалежних випадкових величин) є меншим за максимально допустимий, то гіпотезу про незалежність випадкових величин (ознак) приймають, у протилежному випадку – відхиляють. Довірчою областю критерію є $D_{1-\alpha} = (0; \chi_{(k-1)(n-1); \alpha}^2)$.

Проілюструємо роботу алгоритму перевірки гіпотези про незалежність двох ознак на прикладі.

Приклад 3.22

Розглянемо дві номінальні ознаки: "Де Ви плануєте проживати після закінчення університету?" та "Ваша стать". Перша з ознак має 5 альтернатив: "1. У Києві", "2. Повернись до рідного міста / села", "3. Переїду до іншого міста в межах України", "4. Планую виїхати за кордон" та "5. Інше". Було опитано 1227 студентів, за результатами опитування побудовано таблицю спряженості:

Місце проживання	Стать		
	Ч	Ж	Всього
1. У Києві	301	499	800
2. Повернись до рідного міста / села	21	34	55

⁴ Тут ідеться саме про відносну частоту, а не про ймовірність, оскільки ми оперуємо поняттями, які стосуються вибірки, а не генеральної сукупності.

3. Переїду до іншого міста в межах України	9	26	35
4. Планую виїхати за кордон	87	156	243
5. Інше	28	66	94
Всього	446	781	1227

Для рівня значущості 0,05 перевірити гіпотезу про незалежність цих ознак.

Розв'язання

Отже, крок 1 алгоритму перевірки гіпотези про незалежність двох ознак виконано в умові – побудована таблиця спряженості.

За таблицею спряженості ознак можна побудувати маргінальні розподіли для кожної з цих ознак за кроком 2.

Для ознаки "Стать":

Стать	Ч	Ж
Відносна частота	446/1227 \approx 0,363	781/1227 \approx 0,637

Для ознаки "Місце проживання":

Місце проживання	1	2	3	4	5
Відносна частота	800/1227 \approx \approx 0,652	55/1227 \approx \approx 0,045	35/1227 \approx \approx 0,029	243/1227 \approx \approx 0,198	94/1227 \approx \approx 0,077

Наведені маргінальні ряди розподілу означають, що при розгляді окремо двох ознак: "Де Ви плануєте проживати після закінчення університету?" та "Ваша стать", їх розподіли окремо один від одного матимуть саме такий вигляд.

Наступним кроком є побудова гіпотетичної таблиці спряженості у випадку незалежності даних ознак. На кроці 3 треба побудувати таку таблицю спряженості, для якої маргінальні ймовірності кожної з ознак збереглися б, а сумісні ймовірності були побудовані для незалежних ознак, тобто, для яких виконувалося б співвідношення $\mathcal{P}_N\{X = a, Y = b\} = \mathcal{P}_N\{X = a\}\mathcal{P}_N\{Y = b\}$.

Якщо позначити ознаку "Де Ви плануєте проживати після закінчення університету?" через X , а ознаку "Ваша стать" через

Y, то з кроку 3 алгоритму, сумісна відносна частота $P(X=1, Y=1)$ означає ймовірність того, що респондент планує після закінчення університету залишитись у Києві і є чоловіком дорівнює добутку таких маргінальних імовірностей: респондент планує залишитися у Києві та респондент є чоловіком $P(X=1, Y=1) = 0,652 \times 0,363 \approx 0,237$, а абсолютною частотою буде $0,237 \times 1227 \approx 290,8$. Тобто, якби ці дві ознаки насправді були б незалежними, то у клітинці на перетині стовпчика "чоловік" та рядка "У Києві" мала би стояти абсолютна частота 290,8. Зазначимо, що при підрахунку гіпотетичних абсолютних частот допускають наявність дробових значень.

Аналогічно будують усі інші значення таблиці спряженості, отримуючи в результаті гіпотетичну таблицю спряженості:

Місце проживання \ Стать	Ч	Ж	Всього
1. У Києві	290,8	509,2	800
2. Повернусь до рідного міста / села	20,0	35,0	55
3. Переїду до іншого міста в межах України	12,7	22,3	35
4. Планую виїхати за кордон	88,3	154,7	243
5. Інше	34,2	59,8	94
Всього	446	781	1227

На 4-му кроці будується функція хі-квадрат критерію за формулою (3.32). У нашому випадку вона має вигляд, який умовно позначають $\sum \frac{(O-E)^2}{E}$, тобто, як суму квадратів відхилень спостережуваних (O – *observed*) та гіпотетичних (E – *estimated*) абсолютних частот, поділених на гіпотетичні абсолютні частоти.

Отже,

$$G_0 = \frac{(301-290,8)^2}{290,8} + \frac{(499-509,2)^2}{509,2} + \frac{(21-20)^2}{20} + \frac{(34-35)^2}{35} + \frac{(9-12,7)^2}{12,7} + \frac{(26-22,3)^2}{22,3} + \frac{(87-88,3)^2}{88,3} + \frac{(156-154,7)^2}{154,7}$$

$$\frac{(28-34,2)^2}{34,2} + \frac{(66-59,8)^2}{59,8} \approx 0,358+0,204+0,050+0,029+$$

$$+1,078+0,614+0,019+0,011+1,124+0,643=4,129$$

(якщо не заокруглювати значення абсолютних частот на етапі формування гіпотетичної таблиці спряженості, то $G_0 = 4,134$).

Статистика цього критерію має розподіл хі-квадрат із $(5-1) \times (2-1) = 4$ ступенями вільності. На 5-му кроці визначаємо, що для рівня значущості 0,05 верхньою критичною величиною є $\chi_{4;0,05}^2 = 9,488$. Оскільки значення статистики менше за критичну величину $G_0 = 4,129 < \chi_{4;0,05}^2 = 9,488$, то з довірчою ймовірністю 0,95 нема підстав відхилити гіпотезу про незалежність спостережуваних випадкових величин (ознак).

ЗАДАЧІ

№ 1

Для вибірки

-4,6141; -3,01; -2,44; -2,07; -1,78; -1,53; -1,33; -1,14; -0,97; -0,81; -0,67; -0,53; -0,4; -0,27; -0,15; -0,03; 0,08; 0,19; 0,3; 0,4; 0,51; 0,61; 0,71; 0,81; 0,91; 1,01; 1,11; 1,21; 1,3; 1,42; 1,52; 1,62; 1,73; 1,84; 1,95; 2,06; 2,18; 2,3; 2,43; 2,56; 2,7; 2,85; 3,01; 3,18; 3,38; 3,59; 3,84; 4,15; 4,56; 5,22 перевірити з довірчою ймовірністю 0,9 гіпотезу про те, що ці дані є:

- а) вибіркою з гауссівського $N(1;4)$ розподілу;
- б) вибіркою з гауссівського розподілу.

№ 2

Для вибірки

-7,4211; -5,01; -4,17; -3,6; -3,17; -2,8; -2,49; -2,21; -1,95; -1,72; -1,5; -1,29; -1,09; -0,91; -0,73; -0,55; -0,38; -0,22; -0,06; 0,1; 0,26; 0,41; 0,57; 0,72; 0,87; 1,02; 1,17; 1,32; 1,5; 1,62; 1,78; 1,94; 2,1; 2,26; 2,42; 2,59; 2,77; 2,95; 3,14; 3,34; 3,55; 3,77; 4,01; 4,27; 4,56; 4,89; 5,27; 5,73; 6,34; 7,33 перевірити з довірчою ймовірністю 0,8 гіпотезу про те, що ці дані є:

- а) вибіркою з гауссівського $N(1;9)$ розподілу;
- б) вибіркою з гауссівського розподілу.

№ 3

Для вибірки

-3,6141; -2,01; -1,44; -1,07; -0,78; -0,53; -0,33; -0,14; 0,03; 0,19;
0,33; 0,47; 0,6; 0,73; 0,85; 0,97; 1,08; 1,19; 1,3; 1,4; 1,51; 1,61;
1,71; 1,81; 1,91; 2,01; 2,11; 2,21; 2,3; 2,42; 2,52; 2,62; 2,73; 2,84;
2,95; 3,06; 3,18; 3,3; 3,43; 3,56; 3,7; 3,85; 4,01; 4,18; 4,38; 4,59;
4,84; 5,15; 5,56; 6,22 вказати найменшу (із точністю до десятих)
довірчу ймовірність, при якій можемо стверджувати, що ці дані є:

- а) вибіркою з гауссівського $N(2,4)$ розподілу;
- б) вибіркою з гауссівського розподілу.

3.5.10. Критерій згоди Колмогорова

Критерій згоди застосовується для перевірки гіпотез щодо значення функцій розподілу $H_0: F(x) = F_0(x)$ і є непараметричним критерієм згоди.

На відміну від χ^2 -критерію Пірсона, де оцінюється ступінь відхилення щільності гіпотетичного розподілу від гістограми емпіричного, у критерії Колмогорова оцінюють ступінь відхилення емпіричної кумулятивної функції розподілу частот від гіпотетичної функції розподілу.

Теорема Колмогорова

Нехай

$$K(\lambda) = \begin{cases} \sum_{i=-\infty}^{\infty} (-1)^i e^{(-2i^2\lambda^2)} & \text{при } \lambda > 0; \\ 0 & \text{при } \lambda \leq 0, \end{cases} \quad \text{— функція розподілу Колмогорова.}$$

Тоді, якщо функція $F(x)$ є неперервною, то для довільного λ

$$P(\sqrt{N} \max |F(x) - F_N(x)| < \lambda) \rightarrow K(\lambda) \quad \text{при } N \rightarrow \infty.$$

Статистикою критерію G_N є добуток \sqrt{N} і максимальної різниці між значенням гіпотетичної функції розподілу та КФРЧ, тобто

$$G_N = \sqrt{N} \max |F(x) - F_N(x)|. \quad (3.34)$$

Довірча область є односторонньою, будується за функцією Колмогорова $K(\lambda)$ і має вигляд

$$D_{1-\alpha} = (0; x_\alpha^L) = (0; K_\alpha^L). \quad (3.35)$$

Значення x_α^L шукаємо таким чином: усередині табл. 4.7 (розд. 4) знаходимо рівень значущості α (у таблиці подано значення, помножені на 100 000), а в лівому стовпчику та верхньому рядку знаходимо відповідне значення λ , що й виступатиме нижньою критичною величиною. Зауважимо, що в табл. 4.7 для функції Колмогорова наведено значення **нижньої критичної величини**.

Схема застосування критерію Колмогорова є досить простою: будують статистичну функцію розподілу $F_N(x)$ та гіпотетичну функцію $F(x)$, визначають максимум модуля різниці між ними (рис. 3.11).

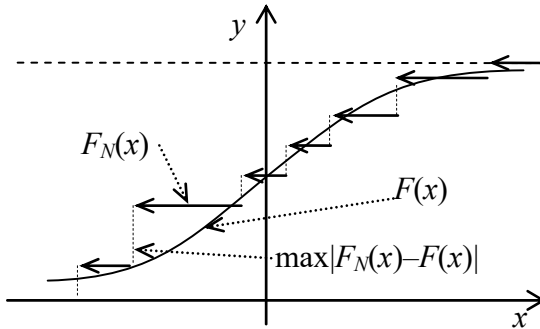


Рис. 3.11. Графічне зображення дії критерію Колмогорова

Далі обчислюють величину $G_N = \sqrt{N} \max |F(x) - F_N(x)|$. За статистичною табл. 4.7 знаходять відповідне значення λ , якому відповідає ймовірність α (рівень значущості). Якщо значення функції G_N потрапляє в довірчий інтервал, то гіпотезу про збігання розподілів приймають, у протилежному випадку – відхиляють.

Критерій Колмогорова своєю простотою вигідно відрізняється від описаного вище χ^2 -критерію, тому його досить часто застосовують на практиці. Однак зазначимо, що цей критерій можна використовувати лише у випадках, коли відомий сам гіпотетичний розподіл $F(x)$ та всі його параметри. На практиці такі

випадки досить рідкі – як правило, нам відомий сам вигляд функції розподілу, а не його параметри. При застосуванні критерію Пірсона цю проблему вирішують зменшенням кількості ступенів вільності в розподілі χ^2 . Критерій Колмогорова такого узгодження не дає, тому, якщо застосувати цей критерій у випадку, коли не відомі всі параметри гіпотетичного розподілу, а вони обчислені за статистичними даними (вибіркою), то критерій дасть занижене значення статистики G_N і ми ризикуємо прийняти гіпотезу, яка насправді погано узгоджується з дослідними даними.

Приклад 3.23

Для вибірки $-6,4211; -4,01; -3,17; -2,6; -2,17; -1,8; -1,49; -1,21; -0,95; -0,72; -0,5; -0,29; -0,09; 0,09; 0,27; 0,45; 0,62; 0,78; 0,94; 1,1; 1,26; 1,41; 1,57; 1,72; 1,87; 2,02; 2,17; 2,32; 2,5; 2,62; 2,78; 2,94; 3,1; 3,26; 3,42; 3,59; 3,77; 3,95; 4,14; 4,34; 4,55; 4,77; 5,01; 5,27; 5,56; 5,89; 6,27; 6,73; 7,34; 8,33$ з довірчою ймовірністю 0,9 перевірити, чи є вона вибіркою з гауссівського розподілу $N(2,9)$.

Занесемо саму вибірку та відповідні ймовірності в таблицю:

x	$F_N(x)$	$F(x)$	$ F(x) - F_N(x) $
-6,4211	0,02	0,00250	0,01750
-4,01	0,04	0,02257	0,01743
-3,17	0,06	0,04241	0,01759
-2,6	0,08	0,06260	0,01740
-2,17	0,1	0,08226	0,01774
-1,8	0,12	0,10264	0,01736
-1,49	0,14	0,12235	0,01765
-1,21	0,16	0,14231	0,01769
-0,95	0,18	0,16272	0,01728
-0,72	0,2	0,18229	0,01771
-0,5	0,22	0,20233	0,01767
-0,29	0,24	0,22263	0,01737
-0,09	0,26	0,24301	0,01699
0,09	0,28	0,26217	0,01783
0,27	0,3	0,28208	0,01792

x	$F_M(x)$	$F(x)$	$ F(x) - F_N(x) $
0,45	0,32	0,30269	0,01731
0,62	0,34	0,32276	0,01724
0,78	0,36	0,34213	0,01787
0,94	0,38	0,36192	0,01808
1,1	0,4	0,38209	0,01791
1,26	0,42	0,40258	0,01742
1,41	0,44	0,42204	0,01796
1,57	0,46	0,44301	0,01699
1,72	0,48	0,46282	0,01718
1,87	0,5	0,48272	0,01728
2,02	0,52	0,50266	0,01734
2,17	0,54	0,52259	0,01741
2,32	0,56	0,54247	0,01753
2,5	0,58	0,56618	0,01382
2,62	0,6	0,58186	0,01814
2,78	0,62	0,60257	0,01743
2,94	0,64	0,62299	0,01701
3,1	0,66	0,64307	0,01693
3,26	0,68	0,66276	0,01724
3,42	0,7	0,68201	0,01799
3,59	0,72	0,70194	0,01806
3,77	0,74	0,72240	0,01760
3,95	0,76	0,74215	0,01785
4,14	0,78	0,76218	0,01782
4,34	0,8	0,78230	0,01770
4,55	0,82	0,80234	0,01766
4,77	0,84	0,82208	0,01792
5,01	0,86	0,84215	0,01785
5,27	0,88	0,86214	0,01786
5,56	0,9	0,88232	0,01768
5,89	0,92	0,90263	0,01737

x	$F_N(x)$	$F(x)$	$ F(x) - F_N(x) $
6,27	0,94	0,92268	0,01732
6,73	0,96	0,94256	0,01744
7,34	0,98	0,96246	0,01754
8,33	1	0,98257	0,01743

Найбільшим модулем різниці є число 0,01814. Обчислимо тепер $G_N = \sqrt{N} \max |F(x) - F_N(x)| = \sqrt{50} \cdot 0,01814 \approx 0,128269$. За табл. 4.7 значень функції Колмогорова маємо, що для рівня значущості $\alpha = 1 - 0,9 = 0,1$ нижня критична величина приблизно дорівнює 0,575. Дійсно, усередині табл. 4.7 шукаємо значення, яке дорівнює $0,1 \times 10^5 = 10000$ (у таблиці всі значення ймовірностей представлені як число, помножене на 10^5). Найближчими значеннями будуть 9865,64 та 11039,44, що відповідають $\lambda = 0,57$ та $0,58$ відповідно. Отже, беремо середнє значення $\lambda = 0,575$ і приймаємо нульову гіпотезу, оскільки значення $G_N \approx 0,128269$ не перевищує $0,575$ і потрапляє в довірчий інтервал $(0; 0,575)$.

Посилений критерій Колмогорова

На відміну від звичайного критерію Колмогорова в посиленому розрізняють "праві" та "ліві" відхилення КФРЧ від функції розподілу (рис. 3.12.). Для цього спочатку вводять позначення:

$$G_N^+ = \sqrt{N} \max_{i=1, N} |F(x_i) - F_N(x_i)|, \quad (3.36)$$

$$G_N^- = \sqrt{N} \max_{i=1, N} |F(x_i) - F_N(x_{i-1})|,$$

де $F_N(x_0) = 0$. Тоді справедливе співвідношення

$$G_N = \max(G_N^-, G_N^+). \quad (3.37)$$

Потім обчислюють величини G_N^+ та G_N^- , знаходять більшу з них $G_N = \max(G_N^-, G_N^+)$.

Далі за статистичною табл. 4.7 знаходять значення λ , якому відповідає рівень значущості α . Якщо значення функції G_N потрапляє в довірчий інтервал, то гіпотезу про збігання розподілів приймають, у протилежному випадку – відхиляють.

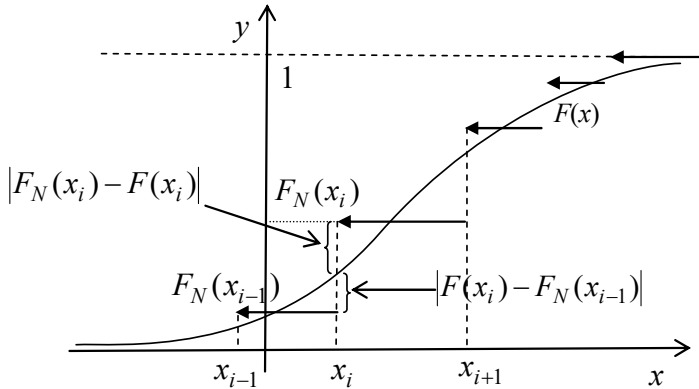


Рис. 3.12. Графічне зображення дії посиленого критерію Колмогорова

Приклад 3.23

Для вибірки

-6,421; -4,01; -3,17; -2,6; -2,17; -1,8; -1,49; -1,21; -0,95; -0,72; -0,5; -0,29; -0,09; 0,09; 0,27; 0,45; 0,62; 0,78; 0,94; 1,1; 1,26; 1,41; 1,57; 1,72; 1,87; 2,02; 2,17; 2,32; 2,5; 2,62; 2,78; 2,94; 3,1; 3,26; 3,42; 3,59; 3,77; 3,95; 4,14; 4,34; 4,55; 4,77; 5,01; 5,27; 5,56; 5,89; 6,27; 6,73; 7,34; 8,33 з довірчою ймовірністю 0,9 перевірити, чи є вона вибіркою з гауссівського розподілу $N(2;9)$.

Спочатку вибірку потрібно перетворити на варіаційний ряд. У нашому випадку вона записана у порядку зростання, що й утворює варіаційний ряд.

Далі потрібно записати дані в таблицю:

x	$F_N(x_i)$	$F(x_i)$	$ F(x_i) - F_N(x_{i-1}) $	$ F(x_i) - F_N(x_i) $
-----	------------	----------	---------------------------	-----------------------

Для цього спочатку формують стовпчик імовірностей $F(x_i)$. Проблема виникає через те, що таблиці розподілу для $N(2;9)$ немає, тому для побудови цього стовпчика використовують співвідношення (2.27) теореми 2.16, за яким $F(x_i) = P\{X_{N(2;9)} < x_i\} = \Phi\left(\frac{x_i - 2}{\sqrt{9}}\right)$. Наприклад, для значення

$x = -6,421 \quad F(-6,421) = \Phi\left(\frac{-6,421-2}{3}\right) = \Phi(-2,807)$. У таблиці розподілу стандартного нормального розподілу (функції Лапласа) для $x = -2,807$ значення ймовірності не вписано (у таблиці наведено значення лише для додатних x). За властивістю симетричних розподілів $P\{X < -x\} = P\{X > x\}$, з іншого боку $P\{X > x\} = 1 - P\{X < x\}$, тому $\Phi(-2,807) = 1 - \Phi(2,807)$. Значення ймовірності для $x = 2,807$ у таблиці відсутнє, але можна його наблизити значенням $x = 2,81$, тобто $\Phi(-2,807) = 1 - \Phi(2,807) \approx 1 - \Phi(2,81) = 1 - 0,9975 = 0,0025$. Таку процедуру здійснюють для всіх значень варіаційного ряду.

Оскільки кожне значення з варіаційного ряду об'ємом N дає приріст КФРЧ у $\frac{1}{N}$, то значення самої кумулятивної функції роз-

поділу частот для елемента варіаційного ряду x_i дорівнює $\frac{i}{N}$.

Занесемо сам варіаційний ряд та відповідні ймовірності в таблицю:

x	$F_N(x_i)$	$F(x_i)$	$ F(x_i) - F_N(x_{i-1}) $	$ F(x_i) - F_N(x_i) $
-6,421	0,02	0,0025	0,00250	0,01750
-4,01	0,04	0,0226	0,00257	0,01743
-3,17	0,06	0,0424	0,00241	0,01759
-2,6	0,08	0,0626	0,00260	0,01740
-2,17	0,1	0,0823	0,00226	0,01774
-1,8	0,12	0,1026	0,00264	0,01736
-1,49	0,14	0,1223	0,00235	0,01765
-1,21	0,16	0,1423	0,00231	0,01769
-0,95	0,18	0,1627	0,00272	0,01728
-0,72	0,2	0,1823	0,00229	0,01771
-0,5	0,22	0,2023	0,00233	0,01767
-0,29	0,24	0,2226	0,00263	0,01737
-0,09	0,26	0,2430	0,00301	0,01699
0,09	0,28	0,2622	0,00217	0,01783

x	$F_N(x_i)$	$F(x_i)$	$ F(x_i) - F_N(x_{i-1}) $	$ F(x_i) - F_N(x_i) $
0,27	0,3	0,2821	0,00208	0,01792
0,45	0,32	0,3027	0,00269	0,01731
0,62	0,34	0,3228	0,00276	0,01724
0,78	0,36	0,3421	0,00213	0,01787
0,94	0,38	0,3619	0,00192	0,01808
1,1	0,4	0,3821	0,00209	0,01791
1,26	0,42	0,4026	0,00258	0,01742
1,41	0,44	0,4220	0,00204	0,01796
1,57	0,46	0,4430	0,00301	0,01699
1,72	0,48	0,4628	0,00282	0,01718
1,87	0,5	0,4827	0,00272	0,01728
2,02	0,52	0,5027	0,00266	0,01734
2,17	0,54	0,5226	0,00259	0,01741
2,32	0,56	0,5425	0,00247	0,01753
2,5	0,58	0,5662	0,00618	0,01382
2,62	0,6	0,5819	0,00186	0,01814
2,78	0,62	0,6026	0,00257	0,01743
2,94	0,64	0,6230	0,00299	0,01701
3,1	0,66	0,6431	0,00307	0,01693
3,26	0,68	0,6628	0,00276	0,01724
3,42	0,7	0,6820	0,00201	0,01799
3,59	0,72	0,7019	0,00194	0,01806
3,77	0,74	0,7224	0,00240	0,01760
3,95	0,76	0,7422	0,00215	0,01785
4,14	0,78	0,7622	0,00218	0,01782
4,34	0,8	0,7823	0,00230	0,01770
4,55	0,82	0,8023	0,00234	0,01766
4,77	0,84	0,8221	0,00208	0,01792
5,01	0,86	0,8421	0,00215	0,01785
5,27	0,88	0,8621	0,00214	0,01786
5,56	0,9	0,8823	0,00232	0,01768

x	$F_N(x_i)$	$F(x_i)$	$ F(x_i) - F_N(x_{i-1}) $	$ F(x_i) - F_N(x_i) $
5,89	0,92	0,9026	0,00263	0,01737
6,27	0,94	0,9227	0,00268	0,01732
6,73	0,96	0,9426	0,00256	0,01744
7,34	0,98	0,9625	0,00246	0,01754
8,33	1	0,9826	0,00257	0,01743

Найбільшим модулем різниць для $|F(x_i) - F_N(x_{i-1})|$ та $|F(x_i) - F_N(x_i)|$ є 0,0181 та 0,0062 відповідно, тому $G_N = \sqrt{N} \max |F(x) - F_N(x)| = \sqrt{50} \cdot 0,01814 \approx 0,128269$.

За таблицею значень функції Колмогорова (табл. 4.7) для рівня значущості $\alpha = 1 - 0,9 = 0,1$ нижня критична величина приблизно дорівнює 0,575. Дійсно, усередині таблиці шукаємо значення, яке дорівнює $0,1 \times 10^5 = 10000$ (усередині таблиці всі значення ймовірностей представлені як число, помножене на 10^5). Найближчими значеннями будуть 9865,64 та 11039,44, що відповідають $\lambda = 0,57$ та 0,58 відповідно. Отже, беремо середнє значення $\lambda = 0,575$ і приймаємо нульову гіпотезу, оскільки значення $G_N \approx 0,128$ не перевищує 0,575 і потрапляє в довірчий інтервал $(0; 0,575)$. Таким чином гіпотезу про те, що вибірка розподілена за нормальним $N(2;9)$ розподілом, приймаємо з довірчою ймовірністю 0,9.

3.5.11. Перевірка гіпотез однорідності двох вибірок

Однорідність розподілів випадкових величин (однорідність вибірок) за їх спостереженням – це однаковість інтегральних характеристик цих вибірок та цих розподілів, тобто однорідність за MX , DX , функціями розподілу та за щільністю розподілу.

Однорідність

MX – рівність MX ;

DX – рівність DX ;

$F(x)$ – рівність $F(x)$;

$f(x)$ – рівність $f(x)$.

Однорідність двох нормальних сукупностей за MX при відомих дисперсіях

Отже, маємо дві нормально розподілені незалежні вибірки з об'ємами N_1 та N_2 відповідно, отримані з двох незалежних, нормально розподілених генеральних сукупностей. Висуваємо гіпотезу, що **середні у цих генеральних сукупностей збігаються.**

Дисперсії збігаються. Якщо дисперсії обох сукупностей збігаються, то маємо дві вибірки з двох нормально розподілених генеральних сукупностей X та Y . Дисперсії у генеральних сукупностей відомі, збігаються та дорівнюють σ^2 . Невідомі математичні сподівання генеральних сукупностей позначимо через m_1 та m_2 відповідно.

Отримаємо

$$X: N(m_1; \sigma^2), \sigma^2 - \text{відоме} \rightarrow x_1, \dots, x_{N_1};$$

$$Y: N(m_2; \sigma^2), \sigma^2 - \text{відоме} \rightarrow y_1, \dots, y_{N_2}.$$

Нульовою гіпотезою виступає рівність математичних сподівань генеральних сукупностей

$$H_0: m_1 = m_2.$$

Статистикою критерію G та її наближенням розподілом ϵ

$$G = \frac{\bar{x} - \bar{y}}{\sqrt{\frac{\sigma^2}{N_1} + \frac{\sigma^2}{N_2}}} = \frac{\bar{x} - \bar{y}}{\sigma \sqrt{\frac{1}{N_1} + \frac{1}{N_2}}} \cong Z. \quad (3.38)$$

Відповідна довірча область матиме вигляд

$$D_{1-\alpha} = \left(x_{\frac{\alpha}{2}}^L; x_{\frac{\alpha}{2}}^U \right) = \left(-Z_{\frac{\alpha}{2}}; Z_{\frac{\alpha}{2}} \right). \quad (3.39)$$

Запишемо довірчий інтервал для параметра $m_1 - m_2$:

$$I_{1-\alpha}^{m_1 - m_2} = \left(-Z_{\frac{\alpha}{2}} \sqrt{\frac{\sigma^2}{N_1} + \frac{\sigma^2}{N_2}}; Z_{\frac{\alpha}{2}} \sqrt{\frac{\sigma^2}{N_1} + \frac{\sigma^2}{N_2}} \right). \quad (3.40)$$

Якщо нульовою гіпотезою є

$$H_0: m_1 - m_2 = \delta,$$

тобто математичні сподівання генеральних сукупностей відрізняються на число δ , то слід скористатись статистикою

$$G = \frac{(\bar{x} - \bar{y}) - \delta}{\sigma \sqrt{\frac{1}{N_1} + \frac{1}{N_2}}} \cong Z, \quad (3.41)$$

яка є більш загальним випадком (3.38).

Довірчий інтервал для параметра $m_1 - m_2$ має вигляд

$$I_{1-\alpha}^{m_1 - m_2} = \left(\delta - Z_{\frac{\alpha}{2}} \sqrt{\frac{\sigma^2}{N_1} + \frac{\sigma^2}{N_2}}; \delta + Z_{\frac{\alpha}{2}} \sqrt{\frac{\sigma^2}{N_1} + \frac{\sigma^2}{N_2}} \right). \quad (3.42)$$

Приклад 3.25

У двох містах, де тестували знання учнів, отримано вибірки із середніми $\bar{x}=140$ та $\bar{y}=155$ балів відповідно, з об'ємами $N_1=25$ та $N_2=27$ учнів та дисперсією $\sigma^2 = 60$. Вважаючи, що бали розподілені нормально, для рівня значущості $\alpha = 0,01$ перевірити гіпотезу про те, що в середньому рівень знань учнів першого міста нижчий за середній бал учнів другого міста на $\delta = 10$ балів.

Нульовою гіпотезою є $H_0: m_2 - m_1 = \delta$, де m_1 та m_2 – математичні сподівання генеральних сукупностей.

Отже, із (3.41) статистикою критерію є $G = \frac{(\bar{y} - \bar{x}) - \delta}{\sigma \sqrt{\frac{1}{N_2} + \frac{1}{N_1}}} =$

$$= \frac{(155 - 140) - 10}{\sqrt{60} \sqrt{\frac{1}{27} + \frac{1}{25}}} \approx 2,326. \text{ Далі з (3.39) знаходимо довірчий інтер-}$$

вал $D_{1-\alpha} = \left(-Z_{\frac{\alpha}{2}}; Z_{\frac{\alpha}{2}} \right) = (-2,58; 2,58)$. Оскільки значення статисти-

ки потрапляє до довірчого інтервалу, то з імовірністю 0,99 можемо стверджувати, що відмінність між середніми рівнями успішності учнів двох міст за даним тестуванням становить 10 балів, тобто, гіпотеза підтверджується.

Дисперсії не збігаються. Якщо дисперсії обох сукупностей не збігаються, то маємо дві генеральні сукупності з відомими

дисперсіями σ_1^2 та σ_2^2 та невідомими математичними сподіваннями. Із цих генеральних сукупностей доступними є дві вибірки з об'ємами N_1 та N_2 . Потрібно перевірити гіпотезу про значення математичних сподівань генеральних сукупностей. Отже,

$$X: N(m_1; \sigma_1^2), \quad \sigma_1^2 - \text{відоме} \rightarrow x_1, \dots, x_{N_1};$$

$$Y: N(m_2; \sigma_2^2), \quad \sigma_2^2 - \text{відоме} \rightarrow y_1, \dots, y_{N_2},$$

де $m_1 = m_2$ – гіпотетичні математичні сподівання генеральних сукупностей. Гіпотезою є

$$H_0: m_1 = m_2.$$

Знайдемо статистику критерію G та її наближений розподіл

$$G = \frac{\bar{x} - \bar{y}}{\sqrt{\frac{\sigma_1^2}{N_1} + \frac{\sigma_2^2}{N_2}}} \cong Z. \quad (3.43)$$

Довірча область має вигляд

$$D_{1-\alpha} = \left(x_{\frac{\alpha}{2}}^L; x_{\frac{\alpha}{2}}^U \right) = \left(-Z_{\frac{\alpha}{2}}; Z_{\frac{\alpha}{2}} \right). \quad (3.44)$$

Якщо нульовою гіпотезою є $H_0: m_1 - m_2 = \delta$, то слід використовувати більш загальний за (3.43) вигляд статистики

$$G = \frac{(\bar{x} - \bar{y}) - \delta}{\sqrt{\frac{\sigma_1^2}{N_1} + \frac{\sigma_2^2}{N_2}}} \cong Z. \quad (3.45)$$

Довірчий інтервал для параметра $m_1 - m_2$ має вигляд

$$I_{1-\alpha}^{m_1 - m_2} = \left(\delta - Z_{\frac{\alpha}{2}} \sqrt{\frac{\sigma_1^2}{N_1} + \frac{\sigma_2^2}{N_2}}; \delta + Z_{\frac{\alpha}{2}} \sqrt{\frac{\sigma_1^2}{N_1} + \frac{\sigma_2^2}{N_2}} \right). \quad (3.46)$$

Приклад 3.26

У двох містах X та Y проведено дослідження рівня заробітної плати (в умовних одиницях). Відібрано відповідно $N_1 = 40$ та $N_2 = 50$ респондентів із нормальних генеральних сукупностей (вважаємо, що розмір зарплати має нормальний розподіл) та знайдено вибіркові середні $\bar{x} = 130$ та $\bar{y} = 140$. Генеральні диспе-

рсії відомі: $\sigma_1^2 = 80$ та $\sigma_2^2 = 100$ відповідно. При рівні значущості $\alpha = 0,01$ перевірити нульову гіпотезу про рівність математичних сподівань генеральних сукупностей $H_0: m_1 = m_2$ при альтернативній гіпотезі $H_1: m_1 \neq m_2$.

Знайдемо спочатку з (3.43) значення функції G :

$$G = \frac{\bar{x} - \bar{y}}{\sqrt{\frac{\sigma_1^2}{N_1} + \frac{\sigma_2^2}{N_2}}} = \frac{130 - 140}{\sqrt{\frac{80}{40} + \frac{100}{50}}} = -5.$$

Оскільки нульова гіпотеза має вигляд $H_0: m_1 = m_2$, а альтернативна $H_1: m_1 \neq m_2$, то довірчий інтервал є двостороннім і з (3.44)

має вигляд $D_{1-\alpha} = \left(-Z_{\frac{\alpha}{2}}; Z_{\frac{\alpha}{2}} \right)$, тому в нашому випадку при

$\alpha = 0,01$, отримаємо з таблиці значень функції Лапласа (табл. 4.2), що $Z_{\frac{\alpha}{2}} = 2,58$. Дійсно, потрібно знайти точку, яка відтинає

праворуч графіка функції Лапласа площу $\frac{\alpha}{2} = 0,005$. Однак у

таблиці значень функції Лапласа наведено не верхні критичні величини, а ймовірність $\Phi(x) = P(Z < x)$, тобто, площу під графіком функції Лапласа, відрізану праворуч точкою x . Отже, у самій таблиці нам потрібно знайти значення $1 - 0,005 = 0,995$, за яким відшукати точку x , що відповідатиме цьому значенню. Такою і буде точка $x = Z_{\frac{\alpha}{2}} = 2,58$. Таким чином, довірчий інтервал матиме вигляд $D_{1-\alpha} = (-2,58; 2,58)$.

Бачимо, що значення функції G не потрапляє в довірчий інтервал, тому відхиляємо нульову гіпотезу і можемо стверджувати, що відмінності між середніми генеральних сукупностей є значущими з рівнем $\alpha = 0,01$.

Однорідність двох нормальних сукупностей за МХ при невідомих дисперсіях. Критерій Стьюдента

Дисперсії збігаються. Обидві вибірки об'ємами N_1 та N_2 отримані з сукупностей, що мають нормальні розподіли з **неві-**

домими, але однаковими дисперсіями (тобто дисперсії рівні між собою). Позначимо математичні сподівання генеральних сукупностей як m_1 та m_2 відповідно. Маємо

$$X: N(m_1; \sigma^2), \sigma^2 - \text{невідоме} \rightarrow S_X^2 \rightarrow x_1, \dots, x_{N_1};$$

$$Y: N(m_2; \sigma^2), \sigma^2 - \text{невідоме} \rightarrow S_Y^2 \rightarrow y_1, \dots, y_{N_2};$$

$$H_0: m_1 = m_2.$$

Статистикою критерію G та її наближеним розподілом є вираз

$$G = \frac{\bar{x} - \bar{y}}{\sqrt{(N_1 - 1)S_X^2 + (N_2 - 1)S_Y^2}} \sqrt{\frac{N_1 N_2 (N_1 + N_2 - 2)}{N_1 + N_2}} \cong t_{N_1 + N_2 - 2}. \quad (3.47)$$

Довірча область має вигляд

$$D_{1-\alpha} = \left(-t_{N_1 + N_2 - 2; \alpha/2}; t_{N_1 + N_2 - 2; \alpha/2} \right). \quad (3.48)$$

У загальному випадку, коли гіпотезу формують як $H_0: m_1 - m_2 = \delta$, функція G має вигляд

$$G = \frac{(\bar{x} - \bar{y}) - \delta}{\sqrt{(N_1 - 1)S_X^2 + (N_2 - 1)S_Y^2}} \sqrt{\frac{N_1 N_2 (N_1 + N_2 - 2)}{N_1 + N_2}} \cong t_{N_1 + N_2 - 2}. \quad (3.49)$$

Довірчий інтервал для параметра $m_1 - m_2$ має вигляд

$$I_{1-\alpha}^{m_1 - m_2} = \left(\delta - t_{N_1 + N_2 - 2; \frac{\alpha}{2}} \frac{\sqrt{(N_1 - 1)S_X^2 + (N_2 - 1)S_Y^2}}{\sqrt{\frac{N_1 N_2 (N_1 + N_2 - 2)}{N_1 + N_2}}}, \delta + t_{N_1 + N_2 - 2; \frac{\alpha}{2}} \frac{\sqrt{(N_1 - 1)S_X^2 + (N_2 - 1)S_Y^2}}{\sqrt{\frac{N_1 N_2 (N_1 + N_2 - 2)}{N_1 + N_2}}} \right). \quad (3.50)$$

Цей критерій розроблено Вільямом Госетом (1876–1937) для оцінювання якості пива в компанії "Arthur Guinness & Son", в якій він працював з 1899 р. Через зобов'язання перед компанією про нерозповсюдження комерційної таємниці, якою керівництво пивної компанії вважало використання статистичних методів, статтю Госета опубліковано в журналі "Биометрика" в 1908 р. під псевдонімом "Student". Тому цей критерій називають критерієм Стьюдента, а сам розподіл – розподілом Стьюдента.

Приклад 3.27

Із двох генеральних сукупностей заміжніх та незаміжніх жінок було відібрано відповідно 10 та 12 осіб. Отримано такі результати:

Заміжні жінки	Вік x_i	26	33	35	40	41
	Кількість n_i	1	2	3	2	2
Незаміжні жінки	Вік y_i	21	22	26	34	
	Кількість k_i	3	4	3	2	

Для рівня значущості $\alpha = 0,02$ перевірити гіпотезу про рівність математичних сподівань генеральних сукупностей $H_0: m_1 = m_2$, при альтернативній гіпотезі $H_1: m_1 \neq m_2$. Вважаємо, що дисперсії генеральних сукупностей збігаються, а самі генеральні сукупності розподілено за нормальними законами.

Отже, $1 - \alpha = 0,98$, $N_1 = 10$ та $N_2 = 12$. Для обчислення статистики критерію G спочатку потрібно знайти \bar{x} та \bar{y} . Skorистаємось формулою для знаходження вибіркового середнього за рядом розподілу:

$$\bar{x} = \frac{\sum x_i n_i}{N_1} \quad \text{та} \quad \bar{y} = \frac{\sum y_i k_i}{N_2}. \quad \text{Отримаємо} \quad \bar{x} = 35,9 \quad \text{та} \quad \bar{y} = 24,75.$$

Оскільки математичні сподівання генеральних сукупностей невідомі, то оцінками дисперсії будуть S_X^2 та S_Y^2 , які обчислюють

за формулами вибірових дисперсії $S_X^2 = \frac{1}{N_1 - 1} \cdot \sum_{i=1}^5 n_i (x_i - \bar{x})^2$ та

$$S_Y^2 = \frac{1}{N_2 - 1} \cdot \sum_{i=1}^4 k_i (y_i - \bar{y})^2, \quad \text{звідки маємо} \quad S_X^2 \approx 22,54 \quad \text{та} \quad S_Y^2 \approx 22,56.$$

Вибіркові дисперсії тут не збігаються (а ми перевіряємо гіпотезу про збігання середніх при однакових дисперсіях), тому їх розбіжність потрібно дослідити (маємо на увазі, що дослідити треба таке: чи значущою є розбіжність вибірових дисперсій) за критерієм Фішера – Снедекора (див. нижче критерії для перевірки рівності дисперсій двох генеральних сукупностей при невідомих середніх). Для цього з (3.58) обчислюємо частку

$$\frac{S_X^2}{S_Y^2} \cong F_{N_1-1; N_2-1}$$

і за формулою (3.59) перевіряємо, чи потрапляє вона до області $(F_{N_1-1; N_2-1; 1-\alpha/2}; F_{N_1-1; N_2-1; \alpha/2})$.

Отже, $\frac{S_X^2}{S_Y^2} = \frac{22,54}{22,56} \approx 0,999$. Як бачимо ця частка менше за 1,

тому розглядаємо частку $\frac{S_Y^2}{S_X^2} = \frac{22,56}{22,54} \approx 1,001$ і згідно з виразом

(3.60) будуємо довірчий інтервал

$$\left(F_{N_2-1; N_1-1; 1-\alpha/2}; F_{N_2-1; N_1-1; \alpha/2} \right) =$$

(за властивістю граничних величин для розподілу Фішера – Снедекора)

$$= \left(\frac{1}{F_{N_1-1; N_2-1; \alpha/2}}; F_{N_2-1; N_1-1; \alpha/2} \right) = (0,1036; 5,18).$$

Легко бачити, що частка потрапляє до довірчої області, тому вважаємо, що дисперсії генеральних сукупностей відрізняються не значуще, а отже, ці дисперсії дорівнюють одна одній і ми можемо перейти до обчислення статистики G критерію Стьюдента:

$$\begin{aligned} G &= \frac{\bar{x} - \bar{y}}{\sqrt{(N_1 - 1)S_X^2 + (N_2 - 1)S_Y^2}} \sqrt{\frac{N_1 N_2 (N_1 + N_2 - 2)}{N_1 + N_2}} = \\ &= \frac{35,9 - 24,75}{\sqrt{(10 - 1) \cdot 22,54 + (12 - 1) \cdot 22,56}} \sqrt{\frac{10 \cdot 12 (10 + 12 - 2)}{10 + 12}} \approx 5,484. \end{aligned}$$

Знайдемо тепер довірчу область для $1 - \alpha = 0,98$. Із (3.48)

$$\begin{aligned} D_{1-\alpha} &= \left(-t_{N_1+N_2-2; \alpha/2}; t_{N_1+N_2-2; \alpha/2} \right) = \left(-t_{10+12-2; 0,02/2}; t_{10+12-2; 0,02/2} \right) = \\ &= \left(-t_{20; 0,01}; t_{20; 0,01} \right) = (-2,5280; 2,5280). \end{aligned}$$

Значення критеріальної функції G не потрапляє до довірчої області, тому гіпотезу про рівність середніх відхиляють і можемо стверджувати, що з імовірністю 0,98 математичні сподівання генеральних сукупностей відрізняються значуще.

Дисперсії не збігаються. Обидві вибірки отримано з генеральних сукупностей, які мають нормальні розподіли з невідомими дисперсіями, що **не рівні** між собою.

$$X: N(m_1, \sigma_1^2) \quad , \quad \sigma_1^2 - \text{невідоме} \rightarrow S_X^2 \rightarrow x_1, \dots, x_{N_1};$$

$Y: N(m_2, \sigma_2^2)$, σ_2^2 – невідоме $\rightarrow S_Y^2 \rightarrow y_1, \dots, y_{N_2}$;

$$H_0: m_1 = m_2.$$

Статистикою критерію G та її наближеним розподілом ϵ

$$G = \frac{\bar{x} - \bar{y}}{\sqrt{\frac{S_X^2}{N_1} + \frac{S_Y^2}{N_2}}} \cong t_v. \quad (3.51)$$

де

$$v = \frac{\left(\frac{S_X^2}{N_1} + \frac{S_Y^2}{N_2}\right)^2}{\left(\frac{S_X^2}{N_1}\right)^2 \left(\frac{1}{N_1 - 1}\right) + \left(\frac{S_Y^2}{N_2}\right)^2 \left(\frac{1}{N_2 - 1}\right)}. \quad (3.52)$$

Довірча область буде така:

$$D_{1-\alpha} = (-t_{v, \alpha/2}, t_{v, \alpha/2}). \quad (3.53)$$

Зауваження

Цей критерій перевірки гіпотези про однорідність математичних сподівань двох генеральних сукупностей за невідомих дисперсій, що можуть не збігатися, називають критерієм Уелча. Він має емпіричний характер, а отже, є досить наближеним. Якщо немає вагомих підстав припускати, що дисперсії не дорівнюють одна одній, то слід застосовувати більш точний критерій, описаний вище для випадку дисперсій, що збігаються.

У багатьох випадках і без обчислення значення v зрозуміло, чи буде значущою величина t_v . Якщо ж сума $N_1 + N_2$ більша за 30, то замість наближення t_v використовують стандартне нормальне наближення, тоді довірчим інтервалом є $D_{1-\alpha} = (-Z_{\alpha/2}; Z_{\alpha/2})$.

Якщо перевіряють гіпотезу вигляду $m_1 - m_2 = \delta$, то використовують критеріальну статистику та її наближений розподіл

$$G = \frac{(\bar{x} - \bar{y}) - \delta}{\sqrt{\frac{S_X^2}{N_1} + \frac{S_Y^2}{N_2}}} \cong t_v. \quad (3.54)$$

ЗАДАЧІ

№ 1

За вибіркою об'єму $N_1 = 30$ мешканців міста A знайдено середній розмір зарплати $\bar{x} = 1300$ грн, а за вибіркою об'єму $N_2 = 40$ мешканців міста B знайдено середній розмір зарплати $\bar{y} = 1250$ грн. Генеральні дисперсії відомі – $\sigma_1^2 = 600$ та $\sigma_2^2 = 800$ відповідно. Потрібно при рівні значущості $\alpha = 0,05$ перевірити нульову гіпотезу $H_0: m_1 = m_2$, де m_1 та m_2 – математичні сподівання генеральних сукупностей, при альтернативній гіпотезі $H_1: m_1 \neq m_2$. Вважають, що генеральні сукупності незалежні й нормально розподілені.

№ 2

Розв'язати попередню задачу для $\bar{x} = 1300$ та $\bar{y} = 1290$.

№ 3

Розв'язати задачу № 1 для $\bar{x} = 1300$ та $\bar{y} = 1290$ і $\sigma_1^2 = 600$ та $\sigma_2^2 = 700$.

№ 4

За вибіркою об'єму $N_1 = 50$ робітників заводу A знайдено середній коефіцієнт задоволеності матеріальним становищем $\bar{x} = 20,1$, а за вибіркою об'єму $N_2 = 50$ робітників заводу B знайдено середній коефіцієнт задоволеності матеріальним становищем $\bar{y} = 19,8$. Генеральні дисперсії відомі: $\sigma_1^2 = 1,75$ та $\sigma_2^2 = 1,375$ відповідно. Потрібно при рівні значущості $\alpha = 0,05$ перевірити нульову гіпотезу $H_0: m_1 = m_2$, де m_1 та m_2 – математичні сподівання генеральних сукупностей, при альтернативній гіпотезі $H_1: m_1 \neq m_2$. Вважають, що генеральні сукупності незалежні й нормально розподілені.

№ 5

Розв'язати задачу № 4 для $\bar{x} = 22$ та $\bar{y} = 19$.

№ 6

Розв'язати задачу № 4 при $\sigma_1^2 = 1$ та $\sigma_2^2 = 3$.

№ 7

Розв'язати задачу № 1 при $\sigma_1^2 = \sigma_2^2 = 700$.

№ 8

Розв'язати задачу № 4 при $\sigma_1^2 = \sigma_2^2 = 1,5$.

№ 9

Розв'язати задачу №1 при невідомих σ_1^2 та σ_2^2 , $\sigma_1^2 = \sigma_2^2$.

№ 10

Розв'язати задачу № 4 при невідомих σ_1^2 та σ_2^2 , $\sigma_1^2 = \sigma_2^2$.

Перевірка гіпотез про однорідність D_X

Математичні сподівання відомі. Маємо дві нормально розподілені незалежні вибірки об'ємами N_1 та N_2 відповідно, отримані з двох незалежних, нормально розподілених генеральних сукупностей із математичними сподіваннями m_1 та m_2 відповідно. Висувають гіпотезу, що дисперсії цих генеральних сукупностей збігаються. Маємо

$$X: N(m_1; \sigma_1^2), m_1 - \text{відоме} \rightarrow \hat{\sigma}_X^2 \rightarrow x_1, \dots, x_{N_1};$$

$$Y: N(m_2; \sigma_2^2), m_2 - \text{відоме} \rightarrow \hat{\sigma}_Y^2 \rightarrow y_1, \dots, y_{N_2};$$

$$H_0: \sigma_1^2 = \sigma_2^2.$$

Статистикою критерію G та її наближеним розподілом ϵ

$$G = \frac{\hat{\sigma}_X^2}{\hat{\sigma}_Y^2} \cong F_{N_1-1; N_2-1}. \tag{3.55}$$

Якщо значення статистики G більше за 1, то довірча область

$$D_{1-\alpha} = (F_{N_1-1; N_2-1; 1-\alpha/2}; F_{N_1-1; N_2-1; \alpha/2}). \tag{3.56}$$

Якщо значення статистики G менше за 1, то для перевірки гіпотези використовують довірчу область

$$D_{1-\alpha} = (F_{N_2-1; N_1-1; 1-\alpha/2}; F_{N_2-1; N_1-1; \alpha/2}). \tag{3.57}$$

Загалом при побудові довірчих областей для розподілу Фішера слід звертати увагу на співвідношення (3.4).

Математичні сподівання не відомі. Маємо дві нормально розподілені незалежні вибірки об'ємами N_1 та N_2 відповідно, отримані із двох незалежних, нормально розподілених генеральних сукупностей. Висувають гіпотезу, що дисперсії в цих

генеральних сукупностях збігаються. Цей критерій ще називають критерієм Фішера – Снедекора. Маємо

$$X: N(m_1; \sigma_1^2), m_1 - \text{невідоме} \rightarrow S_X^2 \rightarrow x_1, \dots, x_{N_1};$$

$$Y: N(m_2; \sigma_2^2), m_2 - \text{невідоме} \rightarrow S_Y^2 \rightarrow y_1, \dots, y_{N_2};$$

$$H_0: \sigma_1^2 = \sigma_2^2.$$

Статистикою критерію G та її наближенням розподілом ϵ

$$G = \frac{S_X^2}{S_Y^2} \cong F_{N_1-1; N_2-1}. \quad (3.58)$$

Якщо значення статистики G більше за 1, то довірна область

$$D_{1-\alpha} = (F_{N_1-1; N_2-1; 1-\alpha/2}; F_{N_1-1; N_2-1; \alpha/2}). \quad (3.59)$$

У разі, коли значення статистики G менше за 1, то розглядають величину $G = \frac{S_Y^2}{S_X^2} \cong F_{N_2-1; N_1-1}$, і для перевірки гіпотези слід використовувати довірчу область

$$D_{1-\alpha} = (F_{N_2-1; N_1-1; 1-\alpha/2}; F_{N_2-1; N_1-1; \alpha/2}). \quad (3.60)$$

ЗАДАЧІ

№ 1

За вибіркою об'єму $N_1=30$ мешканців міста X знайдено середній розмір зарплати $\bar{x}=1300$ грн, а за вибіркою об'єму $N_2=40$ мешканців міста Y знайдено середній розмір зарплати $\bar{y}=1250$ грн. Вибіркові дисперсії відомі – $S_X^2=600$ та $S_Y^2=800$ відповідно. Потрібно при рівні значущості $\alpha=0,05$ перевірити нульову гіпотезу $H_0: \sigma_1^2 = \sigma_2^2$, де σ_1^2 та σ_2^2 – дисперсії генеральних сукупностей, при альтернативній гіпотезі $H_1: \sigma_1^2 \neq \sigma_2^2$. Вважають, що генеральні сукупності є незалежними і нормально розподіленими.

№ 2

За двома незалежними вибірками X та Y об'ємами $N_1=9$ та $N_2=16$, отриманими з нормальних генеральних сукупностей, знайдено виправлені вибіркові дисперсії $S_X^2=34,02$ та $S_Y^2=12,15$. Для рівня значущості $\alpha=0,01$ перевірити нульову гіпотезу H_0 :

$\sigma_1^2 = \sigma_2^2$, де σ_1^2 та σ_2^2 – дисперсії генеральних сукупностей, при альтернативній гіпотезі $H_1: \sigma_1^2 \neq \sigma_2^2$.

№ 3

Розв'язати задачу № 2 при $N_1=11$, $N_2=14$, $\alpha=0,05$, $S_X^2=0,76$ та $S_Y^2=0,38$.

№ 4

Розв'язати задачу № 2 при $N_1=14$, $N_2=10$, $\alpha=0,1$, $S_X^2=0,84$ та $S_Y^2=0,52$.

**Перевірка гіпотез
однорідності двох вибірок
(таблиця критеріїв)**

Зведемо результати цього підрозділу в табл. 3.3–3.5.

Однорідність MX

Таблиця 3.3

**Перевірка гіпотези однорідності
математичного сподівання при відомих дисперсіях**

Дисперсії збігаються	Дисперсії не збігаються
$X: N(m_1; \sigma^2)$, σ^2 – відоме $\rightarrow x_1, \dots, x_{N_1}$ $Y: N(m_2; \sigma^2)$, σ^2 – відоме $\rightarrow y_1, \dots, y_{N_2}$ $H_0: m_1 = m_2$ або $H_0: m_1 - m_2 = 0$ $G = \frac{\bar{x} - \bar{y}}{\sqrt{\frac{\sigma^2}{N_1} + \frac{\sigma^2}{N_2}}} = \frac{\bar{x} - \bar{y}}{\sigma \sqrt{\frac{1}{N_1} + \frac{1}{N_2}}} \cong Z$ $D_{1-\alpha} = \left(x_{\frac{\alpha}{2}}^L; x_{\frac{\alpha}{2}}^U \right) = \left(-Z_{\frac{\alpha}{2}}; Z_{\frac{\alpha}{2}} \right)$	$X: N(m_1, \sigma_1^2)$, σ_1^2 – відоме $\rightarrow x_1, \dots, x_{N_1}$ $Y: N(m_2, \sigma_2^2)$, σ_2^2 – відоме $\rightarrow y_1, \dots, y_{N_2}$ $H_0: m_1 = m_2$ або $H_0: m_1 - m_2 = 0$ $G = \frac{\bar{x} - \bar{y}}{\sqrt{\frac{\sigma_1^2}{N_1} + \frac{\sigma_2^2}{N_2}}} \cong Z$ $D_{1-\alpha} = \left(x_{\frac{\alpha}{2}}^L; x_{\frac{\alpha}{2}}^U \right) = \left(-Z_{\frac{\alpha}{2}}; Z_{\frac{\alpha}{2}} \right)$

Закінчення табл. 3.3

Дисперсії збігаються	Дисперсії не збігаються
$I_{1-\alpha}^{m_1-m_2} = \left(\begin{array}{l} -Z_{\frac{\alpha}{2}} \sqrt{\frac{\sigma^2}{N_1} + \frac{\sigma^2}{N_2}}; \\ Z_{\frac{\alpha}{2}} \sqrt{\frac{\sigma^2}{N_1} + \frac{\sigma^2}{N_2}} \end{array} \right)$ <p> $H_0: m_1 - m_2 = \delta$ $G = \frac{(\bar{x} - \bar{y}) - \delta}{\sigma \sqrt{\frac{1}{N_1} + \frac{1}{N_2}}} \cong Z$ </p> $I_{1-\alpha}^{m_1-m_2} = \left(\begin{array}{l} \delta - Z_{\frac{\alpha}{2}} \sqrt{\frac{\sigma^2}{N_1} + \frac{\sigma^2}{N_2}}; \\ \delta + Z_{\frac{\alpha}{2}} \sqrt{\frac{\sigma^2}{N_1} + \frac{\sigma^2}{N_2}} \end{array} \right)$	$I_{1-\alpha}^{m_1-m_2} = \left(\begin{array}{l} -Z_{\frac{\alpha}{2}} \sqrt{\frac{\sigma_1^2}{N_1} + \frac{\sigma_2^2}{N_2}}; \\ Z_{\frac{\alpha}{2}} \sqrt{\frac{\sigma_1^2}{N_1} + \frac{\sigma_2^2}{N_2}} \end{array} \right)$ <p> $H_0: m_1 - m_2 = \delta$ $G = \frac{(\bar{x} - \bar{y}) - \delta}{\sqrt{\frac{\sigma_1^2}{N_1} + \frac{\sigma_2^2}{N_2}}} \cong Z$ </p> $I_{1-\alpha}^{m_1-m_2} = \left(\begin{array}{l} \delta - Z_{\frac{\alpha}{2}} \sqrt{\frac{\sigma_1^2}{N_1} + \frac{\sigma_2^2}{N_2}}; \\ \delta + Z_{\frac{\alpha}{2}} \sqrt{\frac{\sigma_1^2}{N_1} + \frac{\sigma_2^2}{N_2}} \end{array} \right)$

Таблиця 3.4

**Перевірка гіпотези однорідності
математичного сподівання при невідомих дисперсіях**

Дисперсії збігаються	Дисперсії не збігаються
<p> $X: N(m_1; \sigma^2), \sigma^2$ – невідоме \rightarrow $S_X^2 \rightarrow x_1, \dots, x_{N_1}$ $Y: N(m_2; \sigma^2), \sigma^2$ – невідоме \rightarrow $S_Y^2 \rightarrow y_1, \dots, y_{N_2}$ $H_0: m_1 = m_2$ або $H_0: m_1 - m_2 = 0$ $G = \frac{\bar{x} - \bar{y}}{\sqrt{(N_1 - 1)S_X^2 + (N_2 - 1)S_Y^2}} \times$ $\times \sqrt{\frac{N_1 N_2 (N_1 + N_2 - 2)}{N_1 + N_2}} \cong t_{N_1 + N_2 - 2}$ </p>	<p> $X: N(m_1; \sigma_1^2), \sigma_1^2$ – невідоме \rightarrow $S_X^2 \rightarrow x_1, \dots, x_{N_1}$ $Y: N(m_2; \sigma_2^2), \sigma_2^2$ – невідоме \rightarrow $S_Y^2 \rightarrow y_1, \dots, y_{N_2}$ $H_0: m_1 = m_2$ або $H_0: m_1 - m_2 = 0$ $G = \frac{\bar{x} - \bar{y}}{\sqrt{\frac{S_X^2}{N_1} + \frac{S_Y^2}{N_2}}} \cong t_v$ </p>

Дисперсії збігаються	Дисперсії не збігаються
$D_{1-\alpha} = (-t_{N_1+N_2-2; \alpha/2}; t_{N_1+N_2-2; \alpha/2})$ $I_{1-\alpha}^{m_1-m_2} = \left(\begin{array}{l} -t_{N_1+N_2-2; \frac{\alpha}{2}} \times \\ \times \frac{\sqrt{(N_1-1)S_X^2 + (N_2-1)S_Y^2}}{\sqrt{\frac{N_1 N_2 (N_1 + N_2 - 2)}{N_1 + N_2}}}; \\ \\ t_{N_1+N_2-2; \frac{\alpha}{2}} \times \\ \times \frac{\sqrt{(N_1-1)S_X^2 + (N_2-1)S_Y^2}}{\sqrt{\frac{N_1 N_2 (N_1 + N_2 - 2)}{N_1 + N_2}}} \end{array} \right)$ <p>$H_0: m_1 - m_2 = \delta$</p> $G = \frac{(\bar{x} - \bar{y}) - \delta}{\sqrt{(N_1-1)S_X^2 + (N_2-1)S_Y^2}} \times$ $\times \sqrt{\frac{N_1 N_2 (N_1 + N_2 - 2)}{N_1 + N_2}} \cong t_{N_1+N_2-2}$ $I_{1-\alpha}^{m_1-m_2} = \left(\begin{array}{l} \delta - t_{N_1+N_2-2; \frac{\alpha}{2}} \times \\ \times \frac{\sqrt{(N_1-1)S_X^2 + (N_2-1)S_Y^2}}{\sqrt{\frac{N_1 N_2 (N_1 + N_2 - 2)}{N_1 + N_2}}}; \\ \\ \delta + t_{N_1+N_2-2; \frac{\alpha}{2}} \times \\ \times \frac{\sqrt{(N_1-1)S_X^2 + (N_2-1)S_Y^2}}{\sqrt{\frac{N_1 N_2 (N_1 + N_2 - 2)}{N_1 + N_2}}} \end{array} \right)$	<p>де</p> $v = \frac{\left(\frac{S_X^2}{N_1} + \frac{S_Y^2}{N_2} \right)^2}{\left(\frac{S_X^2}{N_1} \right)^2 \left(\frac{1}{N_1-1} \right) + \left(\frac{S_Y^2}{N_2} \right)^2 \left(\frac{1}{N_2-1} \right)}$ $D_{1-\alpha} = (-t_{v, \alpha/2}; t_{v, \alpha/2})$ $I_{1-\alpha}^{m_1-m_2} = \left(\begin{array}{l} -t_{v; \frac{\alpha}{2}} \sqrt{\frac{S_X^2}{N_1} + \frac{S_Y^2}{N_2}}; \\ \\ t_{v; \frac{\alpha}{2}} \sqrt{\frac{S_X^2}{N_1} + \frac{S_Y^2}{N_2}} \end{array} \right)$

Однорідність DX

Таблиця 3.5

Перевірка гіпотези однорідності дисперсій

Математичні сподівання відомі	Математичні сподівання невідомі
$X: N(m_1; \sigma_1^2),$ m_1 – відоме $\rightarrow \hat{\sigma}_X^2 \rightarrow x_1, \dots, x_{N_1}.$ $Y: N(m_2; \sigma_2^2),$ m_2 – відоме $\rightarrow \hat{\sigma}_Y^2 \rightarrow y_1, \dots, y_{N_2}.$ $H_0: \sigma_1^2 = \sigma_2^2$ $G = \frac{\hat{\sigma}_X^2}{\hat{\sigma}_Y^2} \cong F_{N_1-1; N_2-1}$ Якщо $G > 1$, то $D_{1-\alpha} = (F_{N_1-1; N_2-1; 1-\alpha/2}; F_{N_1-1; N_2-1; \alpha/2})$ Якщо $G < 1$, то $G = \frac{\hat{\sigma}_Y^2}{\hat{\sigma}_X^2} \cong F_{N_2-1; N_1-1}$ $D_{1-\alpha} = (F_{N_2-1; N_1-1; 1-\alpha/2}; F_{N_2-1; N_1-1; \alpha/2})$	$X: N(m_1; \sigma_1^2),$ m_1 – невідоме $\rightarrow S_X^2 \rightarrow x_1, \dots, x_{N_1}$ $Y: N(m_2; \sigma_2^2),$ m_2 – невідоме $\rightarrow S_Y^2 \rightarrow y_1, \dots, y_{N_2}$ $H_0: \sigma_1^2 = \sigma_2^2$ $G = \frac{S_X^2}{S_Y^2} \cong F_{N_1-1; N_2-1}$ Якщо $G > 1$, то $D_{1-\alpha} = (F_{N_1-1; N_2-1; 1-\alpha/2}; F_{N_1-1; N_2-1; \alpha/2})$ Якщо $G < 1$, то $G = \frac{S_Y^2}{S_X^2} \cong F_{N_2-1; N_1-1}$ $D_{1-\alpha} = (F_{N_2-1; N_1-1; 1-\alpha/2}; F_{N_2-1; N_1-1; \alpha/2})$

Перетворення Фішера для перевірки гіпотез про залежність

Маємо випадкову двовимірну вибірку об'єму N , отриману з двовимірної нормально розподіленої генеральної сукупності $(X; Y)$ із коефіцієнтом кореляції ρ_{XY} , тобто вибірку, кожен елемент якої є парою чисел, що відповідають ознакам X та Y (наприклад відповіді респондентів на питання X та Y).

Висувається гіпотеза про те, що коефіцієнт кореляції між цими ознаками X та Y генеральної сукупності дорівнює ρ_0 . Тобто маємо

$$(X; Y): N(m_1; \sigma_1^2) \times N(m_2; \sigma_2^2), \quad S_X^2, S_Y^2 \rightarrow r;$$

$$H_0: \rho_{XY} = \rho_0.$$

Запишемо статистику критерію G та її наближений розподіл

$$G = \sqrt{N-3} \left(\frac{1}{2} \ln \left(\frac{1+r}{1-r} \right) - \frac{1}{2} \ln \left(\frac{1+\rho_0}{1-\rho_0} \right) \right) \cong Z, \quad (3.61)$$

де r – вибірковий коефіцієнт кореляції, який обчислюють за формулою

$$r = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum (x_i - \bar{x})^2 \sum (y_i - \bar{y})^2}}. \quad (3.62)$$

Довірча область

$$D_{1-\alpha} = \left(-Z_{\frac{\alpha}{2}}, Z_{\frac{\alpha}{2}} \right). \quad (3.63)$$

Іноді для зручності функцію G у цьому випадку записують як

$$G = \sqrt{N-3} (z - \xi_0) \cong Z, \quad \text{де } z = \frac{1}{2} \ln \left(\frac{1+r}{1-r} \right), \quad \xi_0 = \frac{1}{2} \ln \left(\frac{1+\rho_0}{1-\rho_0} \right). \quad (3.64)$$

Значення z (його ще називають z -перетворенням Фішера) можна обчислити за табл. 4.6.

Зауваження

Критерій z -перетворення Фішера є наближеним, а точний розподіл статистики, який залежить від N та ρ_{XY} , дуже складний.

Перевірка гіпотези про нульовий коефіцієнт кореляції

Існує точний критерій для перевірки нульової гіпотези $H_0: \rho_{XY} = 0$ та її альтернативи $H_1: \rho_{XY} \neq 0$. Статистикою G та її наближеним розподілом у цьому випадку виступатиме

$$G = \sqrt{N-2} \frac{r}{\sqrt{1-r^2}} \cong t_{N-2} \quad (3.65)$$

із відповідною довірчою областю

$$D_{1-\alpha} = \left(-t_{N-2; \frac{\alpha}{2}}, t_{N-2; \frac{\alpha}{2}} \right). \quad (3.66)$$

Приклад 3.28

У таблиці наведено оцінки, отримані 15 студентами за відповіді на два питання на іспиті:

Студент	Оцінка перша	Оцінка друга	Студент	Оцінка перша	Оцінка друга	Студент	Оцінка перша	Оцінка друга
1	27	15	6	18	10	11	13	8
2	18	5	7	6	8	12	17	14
3	15	10	8	15	9	13	19	13
4	10	9	9	15	10	14	17	11
5	3	2	10	12	7	15	9	6

Перевіримо гіпотезу про те, що з довірчою ймовірністю 0,95 кореляція між оцінками за відповіді на питання 1 та питання 2 дорівнює 0,5.

Отже, дістаємо таке: $N=15$, $1-\alpha = 0,95$, $H_0: \rho_0= 0,5$.

Із формули (3.63) маємо довірчу область

$$D_{1-\alpha} = \left(-Z_{\frac{\alpha}{2}}; Z_{\frac{\alpha}{2}} \right) = \left(-Z_{\frac{1-0,95}{2}}; Z_{\frac{1-0,95}{2}} \right) = \left(-Z_{0,025}; Z_{0,025} \right) \approx (-1,96; 1,96).$$

Далі за (3.64) обчислюємо функцію $G = \sqrt{N-3}(z - \xi_0)$, для чого нам потрібно знайти r , $z = \frac{1}{2} \ln \left(\frac{1+r}{1-r} \right)$ та $\xi_0 = \frac{1}{2} \ln \left(\frac{1+\rho_0}{1-\rho_0} \right)$.

Отримаємо

$$\bar{x} \approx 14,2667, \quad \bar{y} \approx 9,1333,$$

$$r = \frac{\sum(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum(x_i - \bar{x})^2 \sum(y_i - \bar{y})^2}} \approx 0,763893 \approx 0,764.$$

За табл. 4.6 значень z -перетворень Фішера маємо

$$z = \frac{1}{2} \ln \left(\frac{1+r}{1-r} \right) = \frac{1}{2} \ln \left(\frac{1+0,764}{1-0,764} \right) \approx 1,0058 \text{ та}$$

$$\xi_0 = \frac{1}{2} \ln \left(\frac{1+\rho_0}{1-\rho_0} \right) = \frac{1}{2} \ln \left(\frac{1+0,5}{1-0,5} \right) \approx 0,5493.$$

Отже, значення критеріальної статистики

$$G = \sqrt{N-3}(z - \xi_0) \approx \sqrt{15-3}(1,0058 - 0,5493) = \sqrt{12} \cdot 0,4665 \approx 1,5814.$$

Легко бачити, що значення функції G потрапляє до довірчої області, тому можемо стверджувати, що з довірчою ймовірністю 0,95 ми приймаємо нульову гіпотезу про те, що кореляція між оцінками за відповіді на питання 1 та питання 2 дорівнює 0,5.

**Перевірка гіпотези
про рівність двох коефіцієнтів кореляції**

Маємо дві випадкові двовимірні вибірки об'ємами N_1 та N_2 , отримані з двох двовимірних нормально розподілених генеральних сукупностей (X_1, Y_1) , та (X_2, Y_2) із коефіцієнтами кореляції ρ_1 та ρ_2 відповідно.

Висуваємо гіпотезу про те, що **коефіцієнти кореляції генеральних сукупностей рівні між собою:**

$$(X_1; Y_1): N(m_{11}; \sigma_{11}^2) \times N(m_{12}; \sigma_{12}^2) \rightarrow S_{X_1}^2,$$

$$S_{Y_1}^2 \rightarrow r_1;$$

$$(X_2; Y_2): N(m_{21}; \sigma_{21}^2) \times N(m_{22}; \sigma_{22}^2) \rightarrow S_{X_2}^2,$$

$$S_{Y_2}^2 \rightarrow r_2;$$

$$H_0: \rho_1 = \rho_2.$$

Запишемо статистику критерію G та її наближений розподіл

$$G = \frac{z_1 - z_2}{S} \cong Z, \tag{3.67}$$

де

$$z_1 = \frac{1}{2} \ln \left(\frac{1+r_1}{1-r_1} \right), \quad z_2 = \frac{1}{2} \ln \left(\frac{1+r_2}{1-r_2} \right), \quad S^2 = \frac{1}{N_1-3} + \frac{1}{N_2-3}.$$

Довірча область

$$D_{1-\alpha} = \left(-Z_{\frac{\alpha}{2}}; Z_{\frac{\alpha}{2}} \right). \tag{3.68}$$

Приклад 3.29

У таблиці записано дані про вагу та зріст десяти чоловіків та восьми жінок. Перевірити гіпотезу, що з довірчою ймовірністю 0,95 коефіцієнт кореляції між ростом та вагою для чоловіків та для жінок збігається.

Чоловіки		Жінки	
Зріст, см	Вага, кг	Зріст, см	Вага, кг
166	63	161	60
187	98	158	61
170	82	165	68
171	80	154	52
182	86	166	67
188	85	170	68
175	84	167	69
182	105	177	80
179	75		
178	76		

Отже, маємо $N_1=10$, $N_2=8$, $H_0: \rho_1 = \rho_2$.

Із (3.68) критичною областю є

$$D_{1-\alpha} = \left(-Z_{\frac{\alpha}{2}}; Z_{\frac{\alpha}{2}} \right) = \left(-Z_{\frac{1-0,95}{2}}; Z_{\frac{1-0,95}{2}} \right) = \left(-Z_{0,025}; Z_{0,025} \right) \approx (-1,96; 1,96).$$

Далі з (3.67) обчислюємо функцію $G = \frac{z_1 - z_2}{S}$, для чого підраховуємо r_1 , r_2 , $z_1 = \frac{1}{2} \ln \left(\frac{1+r_1}{1-r_1} \right)$, $z_2 = \frac{1}{2} \ln \left(\frac{1+r_2}{1-r_2} \right)$, $S^2 = \frac{1}{N_1-3} + \frac{1}{N_2-3}$.

Маємо для чоловіків

$$\bar{x}_1 \approx 177,8, \quad \bar{y}_1 \approx 83,4,$$

$$r_1 = \frac{\sum (x_{1i} - \bar{x}_1)(y_{1i} - \bar{y}_1)}{\sqrt{\sum (x_{1i} - \bar{x}_1)^2 \sum (y_{1i} - \bar{y}_1)^2}} \approx 0,665081 \approx 0,665.$$

За табл. 4.6 значень z -перетворень Фішера (див. розд. 4) для чоловіків

$$z_1 = \frac{1}{2} \ln \left(\frac{1+r_1}{1-r_1} \right) = \frac{1}{2} \ln \left(\frac{1+0,665}{1-0,665} \right) \approx 0,8017.$$

Для жінок отримаємо

$$\bar{x}_2 \approx 164,75, \quad \bar{y}_2 \approx 65,625,$$

$$r_2 = \frac{\sum(x_{2i} - \bar{x}_2)(y_{2i} - \bar{y}_2)}{\sqrt{\sum(x_{2i} - \bar{x}_2)^2 \sum(y_{2i} - \bar{y}_2)^2}} \approx 0,966548 \approx 0,967.$$

За табл. 4.6 значень z -перетворень Фішера (див. розд. 4) для жінок

$$z_2 = \frac{1}{2} \ln \left(\frac{1+r_2}{1-r_2} \right) = \frac{1}{2} \ln \left(\frac{1+0,967}{1-0,967} \right) \approx 2,0439,$$

$$S^2 = \frac{1}{N_1-3} + \frac{1}{N_2-3} = \frac{1}{10-3} + \frac{1}{8-3} = \frac{1}{7} + \frac{1}{5} = \frac{12}{35} \approx 0,342857 \approx 0,3429.$$

Отже,

$$G = \frac{z_1 - z_2}{\sqrt{S^2}} = \frac{0,8017 - 2,0439}{\sqrt{0,3429}} \approx -2,12133.$$

Таким чином, значення функції G не потрапляє в довірчу область $(-1,96; 1,96)$, тому з імовірністю 0,95 відхиляють нульову гіпотезу про рівність коефіцієнтів кореляції і стверджують, що коефіцієнти кореляції відрізняються значуще.

3.6. Лінійна регресія

У задачах прикладного характеру часто виникає проблема, коли досліджують випадкову величину, яка певним чином залежить від деякої кількості невідповідних змінних, а отже, із певною точністю (певною довірчою ймовірністю) можна одну змінну записати через математичний вираз як функцію від інших змінних. Тобто метою регресійного аналізу є, за наявними емпіричними даними (за вибіркою), вираження однієї змінної (ознаки) через одну (або кілька) інших змінних (ознак) у вигляді рівняння та оцінювання якості такого рівняння для генеральної сукупності. У цьому рівнянні змінну, яку намагаються виразити через інші, називають залежною, а інші, які беруть участь у самому виразі, – незалежними.

Регресійний аналіз – це сукупність статистичних методів обробки емпіричних даних, які дозволяють в умовах стохастич-

ної залежності досліджуваної ознаки від не випадкових або випадкових змінних визначити цю залежність. Результатом регресійного аналізу є регресійна модель, що записується стохастичним рівнянням, яке показує таку залежність. Ці моделі використовують як для прогнозу значень ознак, так і для аналізу самого зв'язку між ознаками.

До регресійних моделей у статистиці належать лінійні та нелінійні. Це пов'язано з лінійністю чи нелінійністю параметрів у самому стохастичному рівнянні. Ми розглядатимемо лише лінійну регресію.

Слід розрізнити стохастичну та функціональну залежність між ознаками. Пояснимо відмінність між цими залежностями на прикладі.

Приклад 3.30

Експериментальним шляхом визначити залежність між температурою тіла пацієнтів у градусах за шкалою Цельсія і за шкалою Фаренгейта та порівняти її з функціональною залежністю між цими шкалами.

Для цього потрібно кілька разів заміряти температури одних і тих самих осіб підряд двома типами термометрів – один вимірює у градусах Цельсія, другий – у градусах Фаренгейта. Після цього записати дані у таблицю і спробувати знайти цю залежність. Результати такого експерименту графічно зображають у вигляді діаграми розсіювання (рис. 3.13, *а*). На діаграмі видно, що точний зв'язок установити неможливо вже навіть через те, що, з одного боку виміри є неточними (кожен пристрій має похибку вимірювання), з іншого, самі умови вимірювання можуть відрізнитись. Зовсім інша картина спостерігається у разі функціонального зв'язку. На діаграмі розсіювання (рис. 3.13, *б*) вибіркою температури, виміряної за Цельсієм, виступає та сама вибірка, що і на рис. 3.13, *а*, а значення температури за Фаренгейтом перераховано за відомою формулою $F = \frac{9}{5}C + 32$, тобто побудовано таким чином, що між змінними C (градуси за Цельсієм) та F (градуси за Фаренгейтом) існує функціональний зв'язок. Відповідно всі пари таких значень лежать на одній прямій (рис. 3.13, *б*).

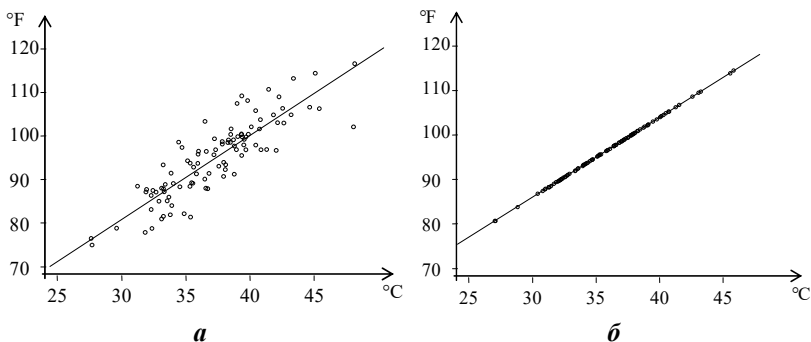


Рис. 3.13. Залежність температури у градусах Цельсія та Фаренгейта:
а – стохастична, б – функціональна

3.6.1. Проста (парна) лінійна регресія

Найпростішим випадком регресійної моделі є так звана проста (або парна) лінійна регресія. Це випадок, коли перед дослідником стоїть задача, на основі спостережень за двома змінними, визначити стохастичну залежність між ними, тобто за спостереженнями над змінними y та x записати стохастичне рівняння так, щоб, наприклад, змінну y можна було б формально записати через змінну x . Для зручності таку двовимірну вибірку записують як пари a_1, \dots, a_N , які представляють значення незалежної змінної x та залежної y : $a_i = (x_i, y_i)$, $i = \overline{1, N}$. Значення незалежної змінної x вважають не випадковими, а значення залежної змінної y – випадковими, збуреними помилками спостережень ε . Таким чином, загальний зв'язок між змінними представляють через стохастичне співвідношення, яке називають **моделлю спостережень**:

$$y = f(x; \beta) + \varepsilon. \quad (3.69)$$

У цьому виразі $f(x; \beta)$ є загальним записом, який у випадку лінійної регресії означає лінійний стосовно коефіцієнтів β (їх ще називають параметрами) вираз зі змінними x . Таким стохастичним рівнянням може бути, наприклад для парної лінійної регресії, $y = \frac{9}{5}x + 32 + \varepsilon$.

У загальному випадку функцію $f(x;\beta)$ у співвідношенні (3.67) повністю визначають набором коефіцієнтів, який охоплюють одним позначенням β і називають параметром. Для парної лінійної регресії

$$f(x;\beta) = \beta_0 + \beta_1 x, \quad \beta = \begin{pmatrix} \beta_0 \\ \beta_1 \end{pmatrix}, \quad (3.70)$$

тобто можна (3.69) переписати у вигляді $y = \beta_0 + \beta_1 x + \varepsilon$. Якщо для моделі $y = \frac{9}{5}x + 32 + \varepsilon$, наприклад, значеннями параметра

$$\text{парної лінійної регресії } \beta = \begin{pmatrix} 32 \\ \frac{9}{5} \end{pmatrix}.$$

Функцію $y = \beta_0 + \beta_1 x + \varepsilon$, яку повністю визначають параметром β і спостерігають зі значеннями, що знаходяться згідно зі співвідношенням (3.69), називають **простою лінійною регресією**.

Функцію $f(x;\beta)$ у виразі (3.69) називають **істинною регресією**, а ε – **помилкою спостереження**.

Отже, для найпростішої регресії істинною регресією є функція, яка визначається співвідношенням (3.70).

Кожну з пар $a_i = (x_i, y_i)$, $i = \overline{1, N}$, вибірки, називатимемо **спостереженням регресії**. Відповідно до (3.69) елементи спостереження регресії $a_i = (x_i, y_i)$ зв'язані співвідношенням

$$y_i = f(x_i; \beta) + \varepsilon_i, \quad i = \overline{1, N}, \quad (3.71)$$

яке називають **моделлю спостережень** для регресії. Тобто моделлю спостереження для простої лінійної регресії є система з N (кількість елементів вибірки) рівнянь вигляду

$$y_i = \beta_0 + \beta_1 x_i + \varepsilon_i, \quad i = \overline{1, N}. \quad (3.72)$$

Регресію, лінійну за елементами вектора параметрів β , називають **лінійною**.

3.6.2. Основна задача регресії

Лінійну регресію використовують для опису залежностей, представлених емпіричними даними, наприклад вибіркою $a_i = (x_i, y_i)$, $i = \overline{1, N}$, а також для прогнозу значень таких залежностей на аргументах x^* , що не представлені у вибірці: $x^* \neq x_i$ ($i = \overline{1, N}$). Тобто у задачі прогнозу за отриманими новими значеннями x_i завдяки регресійній моделі здійснюють оцінювання можливого значення y_i . Для розв'язування цієї задачі за спостережуваними даними оцінюють істинну регресію. Цю оцінку позначають через $\hat{f}(x, \beta)$. Така задача зводиться до оцінювання параметра β $\hat{f}(x, \beta) = f(x, \hat{\beta})$. Функцію $\hat{y} = f(x, \hat{\beta})$ називають **емпіричною регресією**.

Отже, результатом лінійного регресійного аналізу має бути пряма істинної регресії. Але, як видно з рис. 3.13, *а*, емпіричні дані, навіть маючи лінійну стохастичну залежність, не є функціонально лінійно залежними, як це зображено на рис. 3.13, *б*, тобто не лежать на одній прямій, а розташовані у деякій області навколо прямої. Це пов'язано головним чином із двома причинами: похибка вимірювання значень залежної та незалежної змінних (сюди також відносять "неточність" відповідей респондентів при опитуваннях); сам зв'язок не обов'язково має бути насправді лінійним – ми просто намагаємось побудувати лінійну модель, яка є наближеною.

Якщо таке лінійне регресійне рівняння побудоване, то ми отримаємо пряму, що проходить через "хмару" емпіричних даних, причому має проходити через точку з координатами (\bar{x}, \bar{y}) – чорна точка на перетині вертикальної та горизонтальної пунктирних прямих, що проходять через вибіркові середні \bar{x} та \bar{y} відповідно (рис. 3.14).

Кожен елемент "хмари" емпіричних даних є точкою з координатами, що відповідають значенням елементів двовимірної вибірки. Як видно з рис. 3.14, більшість елементів вибірки не лежить на регресійній прямій. Це цілком природне явище оскільки вимірювання є неточними і не гарантована функціональна лінійна

залежність між ознаками. Відхилення елементів вибірки від регресійної прямої називають **нев'язками** (рис. 3.14) $y_i - f(x_i, \beta)$. Ідея регресійного рівняння полягає у побудові рівняння таким чином, аби мінімізувати невідповідності, тобто провести лінію так, щоб невідповідності

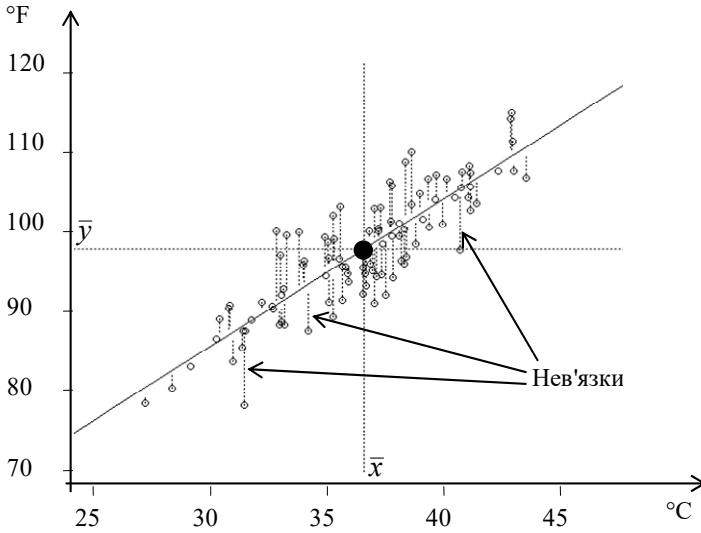


Рис. 3.14. Регресійна пряма та невідповідності

сукупно були мінімальними. Для кращого оцінювання такого рівняння мінімізують не суму N невідповідей, а суму їх квадратів і цей метод називають методом найменших квадратів.

**Метод найменших квадратів
у оцінюванні параметрів регресії,
рівняння Гаусса – Маркова**

Оцінювання параметра $\beta = \begin{pmatrix} \beta_0 \\ \beta_1 \end{pmatrix}$ регресії здійснюють методом найменших квадратів (МНК). Суть цього методу полягає в тому, що оцінкою вважають таке значення β , яке мінімізує суму квадратів невідповідей $y_i - f(x_i, \beta)$, $i = \overline{1, N}$, між спостережуваними значеннями регресії y_i та можливими значеннями $f(x_i, \beta)$, $i = \overline{1, N}$:

$$\hat{\beta} = \arg \min_{\beta} \sum_{i=1}^N (y_i - f(x_i, \beta))^2, \quad (3.73)$$

тобто таке β , при якому значення $\sum_{i=1}^N (y_i - f(x_i, \beta))^2$ (сума квадратів нев'язок) є найменшим з усіх можливих.

Розв'язують задачу пошуку мінімуму функціонала $A(\beta) = \sum_{i=1}^N (y_i - f(x_i, \beta))^2$ перевіркою необхідних умов екстремуму, тобто, прирівнюванням похідних за кожною компонентою вектора параметрів до нуля. У векторному вигляді відповідна система рівнянь скорочено запишеться у вигляді

$$\frac{\partial A(\beta)}{\partial \beta} = 0. \quad (3.74)$$

Систему рівнянь або векторне рівняння (3.74) називають **рівнянням Гаусса – Маркова для оцінювання параметрів регресії**.

Для розв'язків рівняння Гаусса – Маркова (3.74) ще треба доводити, що вони дійсно є точками мінімуму функціонала $A(\beta)$, адже перша похідна дорівнює 0 як у точках локального мінімуму, так і в точках локального максимуму.

Зазначимо, що для лінійної регресії функціонал $A(\beta)$ методу найменших квадратів представляють так:

$$A(\beta) = \|Y - X\beta\|^2.$$

Рівняння Гаусса – Маркова для найпростішої регресії

Для найпростішої регресії функціонал $A(\beta)$ має вигляд

$$A(\beta) = \sum_{i=1}^N (y_i - \beta_0 - \beta_1 x_i)^2.$$

Нескладно переконатися, що похідні за β_0 та β_1 записують відповідно:

$$\frac{\partial A(\beta)}{\partial \beta_0} = -2 \sum_{i=1}^N (y_i - \beta_0 - \beta_1 x_i),$$

$$\frac{\partial A(\beta)}{\partial \beta_1} = -2 \sum_{i=1}^N x_i (y_i - \beta_0 - \beta_1 x_i),$$

отже, система рівнянь набуває вигляду

$$\begin{cases} -2\sum_{i=1}^N (y_i - \beta_0 - \beta_1 x_i) = 0, \\ -2\sum_{i=1}^N x_i (y_i - \beta_0 - \beta_1 x_i) = 0. \end{cases} \quad (3.75)$$

Скоротимо на (-2) , розкриємо дужки та згрупуємо подібні члени у (3.75). Отримаємо зведену форму рівняння Гаусса – Маркова для найпростішої регресії

$$\begin{cases} N\beta_0 + \beta_1 \sum_{i=1}^N x_i = \sum_{i=1}^N y_i, \\ \beta_0 \sum_{i=1}^N x_i + \beta_1 \sum_{i=1}^N x_i^2 = \sum_{i=1}^N x_i y_i. \end{cases} \quad (3.76)$$

Система (3.76), в якій невідомі містяться в одній частині рівнянь, а константи – в іншій, називатимемо **зведеною формою рівнянь Гаусса – Маркова** чи просто зведеним рівнянням Гаусса – Маркова.

Розв'язання рівняння Гаусса – Маркова для найпростішої регресії

Значимо, що система (3.76) є системою лінійних рівнянь щодо β_0 та β_1 . Виключаючи β_0 , отримаємо розв'язок для β_1 , який позначають через $\hat{\beta}_1$:

$$\hat{\beta}_1 = \frac{N \sum_{i=1}^N x_i y_i - \sum_{i=1}^N x_i \cdot \sum_{i=1}^N y_i}{N \sum_{i=1}^N x_i^2 - \left(\sum_{i=1}^N x_i \right)^2}. \quad (3.77)$$

Підстановка значення $\hat{\beta}_1$ у перше рівняння системи (3.76) і розв'язання отриманого рівняння щодо β_0 подає розв'язок β_0 через $\hat{\beta}_1$:

$$\hat{\beta}_0 = \frac{1}{N} \sum_{i=1}^N y_i - \hat{\beta}_1 \left(\frac{1}{N} \sum_{i=1}^N x_i \right).$$

Позначаючи середні значення $\frac{1}{N} \sum_{i=1}^N y_i$ та $\frac{1}{N} \sum_{i=1}^N x_i$ через \bar{y} та \bar{x} відповідно, отримаємо остаточний вираз для $\hat{\beta}_0$:

$$\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x}. \quad (3.78)$$

Таким чином, розв'язок системи рівнянь Гаусса – Маркова для найпростішої регресії обчислюють за формулами (3.77) та (3.78).

Розв'язок (3.77) та (3.78) системи рівнянь Гаусса – Маркова є точкою мінімуму, про що свідчить таке: значення суми квадратичних відхилень $A(\beta)$ обмежено знизу нулем і неперервно залежить від параметрів. Це зауваження робить розв'язок системи Гаусса – Маркова оцінкою методу найменших квадратів для найпростішої регресії.

Обчислювальна форма оцінок коефіцієнтів найпростішої регресії

Нескладними перетвореннями вираз (3.78), що визначає β_1 , можна представити як

$$\begin{aligned} \hat{\beta}_1 &= \frac{\sum_{i=1}^N x_i y_i - \frac{1}{N} \sum_{i=1}^N x_i \cdot \sum_{i=1}^N y_i}{\sum_{i=1}^N x_i^2 - \frac{1}{N} \left(\sum_{i=1}^N x_i \right)^2} = \\ &= \frac{\sum_{i=1}^N (x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^N (x_i - \bar{x})^2} = \frac{\frac{1}{N-1} \sum_{i=1}^N (x_i - \bar{x})(y_i - \bar{y})}{\frac{1}{N-1} \sum_{i=1}^N (x_i - \bar{x})^2} = \frac{C_{xy}}{S_x^2}, \quad (3.79) \end{aligned}$$

де $S_x^2 = \frac{1}{N-1} \sum_{i=1}^N (x_i - \bar{x})^2$ – вибіркова дисперсія незалежної змінної,

$C_{xy} = \frac{1}{N-1} \sum_{i=1}^N (x_i - \bar{x})(y_i - \bar{y})$ – коваріація між залежною та незалежною змінними.

Вираз (3.79) використовують як формулу для обчислення коефіцієнта $\hat{\beta}_1$. Усі обчислення записують у таблицю, перші два стовпчики якої відповідають спостережуваним значенням x та y , а останні два – значенням $x \cdot y$ та x^2 .

Приклад 3.31

Нехай вихідні дані спостережень регресії і необхідні додаткові характеристики представлено в таблиці:

Сума
Середнє

x	y	$x \cdot y$	x^2
1	3	3	1
2	4	8	4
3	6	18	9
4	8	32	16
10	21	61	30
2,5	5,25	15,25	7,5

Оцінити параметр $\beta = \begin{pmatrix} \beta_0 \\ \beta_1 \end{pmatrix}$.

Модель спостережень із (3.72) має вигляд

$$\begin{pmatrix} 3 \\ 4 \\ 6 \\ 8 \end{pmatrix} = \begin{pmatrix} 1 & 1 \\ 1 & 2 \\ 1 & 3 \\ 1 & 4 \end{pmatrix} \begin{pmatrix} \beta_0 \\ \beta_1 \end{pmatrix} + \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \varepsilon_3 \\ \varepsilon_4 \end{pmatrix}.$$

Отже, відповідно до (3.79)

$$\hat{\beta}_1 = \frac{\sum_{i=1}^N x_i y_i - \frac{1}{N} \sum_{i=1}^N x_i \cdot \sum_{i=1}^N y_i}{\sum_{i=1}^N x_i^2 - \frac{1}{N} \left(\sum_{i=1}^N x_i \right)^2} = \frac{61 - \frac{1}{4} \cdot 10 \cdot 21}{30 - \frac{1}{4} \cdot 10^2} = 1,7,$$

тоді згідно з (3.78) $\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x} = 5,25 - 1,7 \cdot 2,5 = 1$.

3.6.3. Емпірична регресія для найпростішої регресії

Зважаючи на формули (3.77) та (3.78) оцінки параметрів, отримаємо формулу емпіричної регресії для найпростішої регресії у вигляді

$$\hat{y} = f(x, \hat{\beta}) = \hat{\beta}_0 + \hat{\beta}_1 x. \quad (3.80)$$

Використовуючи зв'язок між $\hat{\beta}_0$ та $\hat{\beta}_1$ (формула (3.78)), вираз емпіричної регресії (3.80) представляють як

$$\hat{y} - \bar{y} = \hat{\beta}_1 (x - \bar{x}). \quad (3.81)$$

Зауважимо, що вираз (3.80) для емпіричної регресії являє собою рівняння прямої, що проходить через точку $(\bar{x}; \bar{y})$ і має кутовий коефіцієнт $\hat{\beta}_1$.

Таким чином, для прикладу 3.31 емпірична регресія (оцінка істинної регресії) має вигляд $\hat{y} = f(x, \hat{\beta}) = 1 + 1,7x$. Графічно емпірична регресія є прямою, що групує спостережувані дані найкращим чином (рис. 3.15).

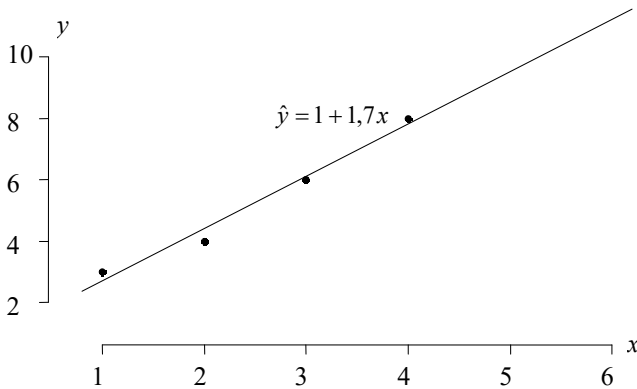


Рис. 3.15. Емпірична регресія

3.6.4. Прогноз за допомогою регресії

Оцінювання істинної регресії за допомогою емпіричної дає можливість не тільки стискати дані (замість таблиці спостережень отримувати значення коефіцієнтів $\hat{\beta}_0$ та $\hat{\beta}_1$), але і прогнозувати значення істинної регресії значеннями емпіричної регресії для довільного аргументу x^* :

$$\hat{f}(x^*, \hat{\beta}) = f(x^*, \hat{\beta}). \quad (3.82)$$

Якщо у прикладі 3.31 треба спрогнозувати значення істинної регресії, наприклад для $x^* = 5$, то $\hat{f}(5, \hat{\beta}) = f(5, \hat{\beta}) = 1 + 1,7 \cdot 5 = 9,5$.

Результати прогнозування зображено на рис. 3.16.

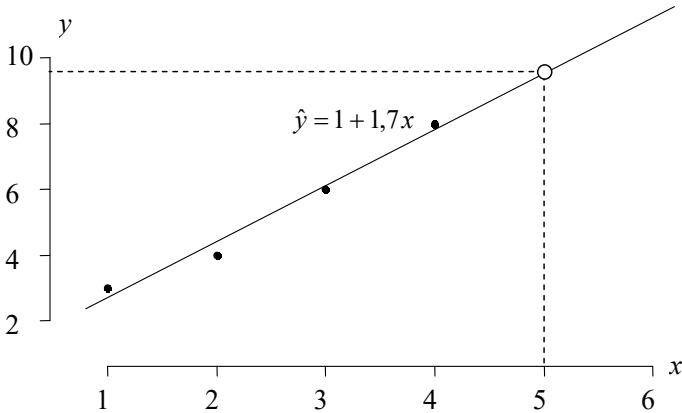


Рис. 3.16. Прогнозування за допомогою регресії

Графічно прогнозування за емпіричною регресією означає прогноз за тенденцією, яка міститься в емпіричних даних (у спостереженнях).

Припущення класичної нормальної регресії

Задачі оцінювання в лінійній регресії (у тому числі й у найпростішому варіанті) розглядають за певних припущень, які називають *припущеннями класичної нормальної регресії*. До них належать:

- припущення про незалежність спостережень – помилки спостережень ε_i , $i = \overline{1, N}$, є незалежними випадковими величинами;
- припущення про рівноточність (гомоскедастичність, збігання дисперсії) спостережень – помилки спостережень ε_i , $i = \overline{1, N}$, однаково розподілені з однією і тією самою дисперсією σ^2 ; математичне сподівання помилок спостережень вважають нульовим за відсутності систематичної складової в цих помилках;
- припущення про детермінованість (невипадковість) аргументів x_i , $i = \overline{1, N}$;
- припущення про те, що ранг матриці плану X збігається з кількістю параметрів, що підлягають оцінюванню, – $\text{rank} X = p$;
- припущення про нормальний розподіл помилок спостережень ε_i , $i = \overline{1, N}$.

Враховуючи наведені вище припущення, вважаємо, що ε_i , $i = \overline{1, N}$, мають $N(0; \sigma^2)$ -розподіл.

Як випливає з розглянутих припущень, розподілом компоненти у спостереженні $(x, y) \in N(f(x, \beta); \sigma^2)$. Зважаючи на це, задачу оцінювання істинної регресії $f(x, \beta)$ описують як задачу оцінювання математичного сподівання y .

3.6.5. Точність точкового оцінювання параметрів регресії та прогнозу за емпіричною регресією

У розв'язанні задачі точності точкового оцінювання (задачі ТТО) для регресії вирішальну роль відіграє оцінка дисперсії оцінки спостереження. Її визначають за допомогою характеристики, яку називають сумою квадратів нев'язок, позначають RSS (*Residual Sum of Squares*) і отримують через співвідношення

$$RSS = \sum_{i=1}^N (y_i - \hat{y}_i)^2 = \sum_{i=1}^N \left(y_i - f(x_i, \hat{\beta}) \right)^2. \quad (3.83)$$

Оцінка дисперсії визначається рівністю

$$\hat{\sigma}^2 = \frac{RSS}{N-p} = \frac{1}{N-p} \sum_{i=1}^N (y_i - \hat{y}_i)^2, \quad (3.84)$$

де p – кількість параметрів регресійного рівняння. Для простої лінійної регресії $p = 2$ (β_0 та β_1).

Для найпростішої регресії точність точкової оцінки $\varepsilon_{1-\alpha}^{f(x, \beta)}$ для прогнозованого значення визначають виразом

$$\varepsilon_{1-\alpha}^{f(x, \beta)} = t_{N-2; \frac{\alpha}{2}} \cdot \hat{\sigma} \cdot \sqrt{\frac{1}{N} + \frac{(x - \bar{x})^2}{\sum_{i=1}^N (x_i - \bar{x})^2}}. \quad (3.85)$$

У співвідношенні (3.85) бачимо таке:

- $t_{N-2; \frac{\alpha}{2}}$ – верхня критична величина розподілу Стьюдента з

$N-2$ ступенями вільності для рівня значущості $\frac{\alpha}{2}$;

- $\hat{\sigma}$ – корінь з оцінки дисперсії (стандартне відхилення);
- \bar{x} – середнє спостережуваних значень аргументу.

Розв'язання задачі ТТО за допомогою довірчого інтервалу має вигляд

$$I_{1-\alpha}^{f(x,\beta)} = \left(f(x, \hat{\beta}) - \varepsilon_{1-\alpha}^{f(x,\beta)}, f(x, \hat{\beta}) + \varepsilon_{1-\alpha}^{f(x,\beta)} \right). \quad (3.86)$$

Приклад 3.32

За умов прикладу 3.31 оцінка дисперсії $\hat{\sigma}^2$, яку розраховують за формулою (3.84), має вигляд

$$\hat{\sigma}^2 = \frac{1}{N-p} \sum_{i=1}^N (y_i - \hat{y}_i)^2 = \frac{0,3}{4-2} = 0,15.$$

3.6.6. Довірчі смуги

Оскільки регресійне рівняння є стохастичним, то ми не можемо під час прогнозування точно говорити про значення, якого може набути y при новому значенні x . Ми навіть не можемо точно визначити істинне значення регресії, а лише діапазон таких значень. Тобто отримуємо смугу вздовж регресійної прямої (регресійна пряма міститиметься всередині цієї смуги), у межах якої з певною довірчою ймовірністю буде набувати значення істинне значення регресії чи значення залежної змінної відповідно. Тут слід розрізняти довірчий інтервал для значення істинної регресії $y = f(x; \beta)$ та довірчий інтервал для окремих значень залежної змінної $y_i = f(x_i; \beta) + \varepsilon_i$.

Довірча смуга значень істинної регресії, довірча смуга Воркінга – Готелінга

Ліву (нижню) границю довірчого інтервалу для можливих значень істинної регресії описують функцією, яку позначають $f^L(x)$:

$$f^L(x) = f(x, \hat{\beta}) - \varepsilon_{1-\alpha}^{f(x,\beta)}, \quad (3.87)$$

а праву (верхню) – функцією $f^U(x)$:

$$f^U(x) = f(x, \hat{\beta}) + \varepsilon_{1-\alpha}^{f(x,\beta)}. \quad (3.88)$$

Смуга на числовій площині між функціями, що визначені співвідношеннями (3.87) та (3.88), носить назву **довірчої смуги Воркінга – Готелінга**. Отже, верхня та нижня границі цієї смуги записуються у вигляді відповідних виразів

$$\begin{aligned} f^L(x) &= f(x, \hat{\beta}) - \varepsilon_{1-\alpha}^{f(x, \beta)} = \\ &= \hat{\beta}_0 + \hat{\beta}_1 x - t_{N-2; \frac{\alpha}{2}} \hat{\sigma} \sqrt{\frac{1}{N} + \frac{(x - \bar{x})^2}{\sum_{i=1}^N (x_i - \bar{x})^2}}, \end{aligned} \quad (3.89)$$

та

$$\begin{aligned} f^U(x) &= f(x, \hat{\beta}) + \varepsilon_{1-\alpha}^{f(x, \beta)} = \\ &= \hat{\beta}_0 + \hat{\beta}_1 x + t_{N-2; \frac{\alpha}{2}} \hat{\sigma} \sqrt{\frac{1}{N} + \frac{(x - \bar{x})^2}{\sum_{i=1}^N (x_i - \bar{x})^2}}. \end{aligned} \quad (3.90)$$

За змістом перетин довільного перпендикуляра до осі OX у точці x із довірчою смугою Воркінга – Готелінга дає довірчий інтервал істинного значення регресії для цього значення аргументу x .

Зазначимо, що вигляд верхньої та нижньої границь для довірчої смуги, яка накриває істинне значення регресії в цілому для довірчої ймовірності β_c , визначають аналогами співвідношень (3.89) та (3.90), в яких замість верхньої критичної величини $t_{N-2; \frac{\alpha}{2}}$ розподілу Стьюдента стоїть вираз $\sqrt{2F_{2; N-2; \frac{\alpha}{2}}}$, де

$F_{2; N-2; \frac{\alpha}{2}}$ – верхня критична величина з рівнем значущості $\frac{\alpha}{2}$

для F -розподілу із 2 та $N-2$ ступенями вільності.

Приклад 3.33

Для умов прикладу 3.31 емпірична регресія має вигляд $\hat{y} = f(x, \hat{\beta}) = 1 + 1,7x$, $\bar{x} = 2,5$. Границі довірчої смуги Воркінга – Готелінга (рис. 3.17) для довірчої ймовірності 0,9 ($N=4$, $t_{N-2; 0,05} = t_{2; 0,05} = 9,9248 \approx 9,9$), зважаючи на обчислену раніше оцінку дисперсії $\hat{\sigma}^2 = 0,15$ ($\hat{\sigma} = 0,39$), визначають за формулами (3.89) та (3.90):

$$f^L(x) = f(x, \hat{\beta}) - \varepsilon_{1-\alpha}^{f(x, \hat{\beta})} = 1 + 1,7x - t_{N-2; \frac{1-0,9}{2}} \cdot 0,39 \sqrt{\frac{1}{4} + \frac{(x-2,5)^2}{\sum_{i=1}^4 (x_i - 2,5)^2}},$$

$$f^U(x) = f(x, \hat{\beta}) + \varepsilon_{1-\alpha}^{f(x, \hat{\beta})} = 1 + 1,7x + t_{N-2; \frac{1-0,9}{2}} \cdot 0,39 \sqrt{\frac{1}{4} + \frac{(x-2,5)^2}{\sum_{i=1}^4 (x_i - 2,5)^2}}.$$

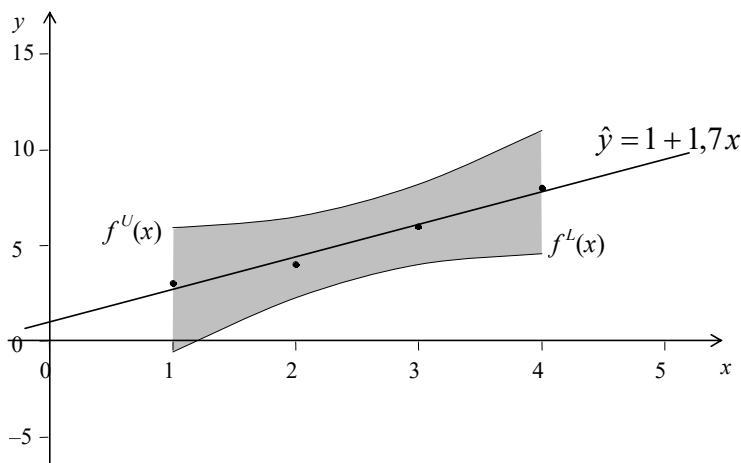


Рис. 3.17. Довірча смуга Воркінга – Готелінга

Далі підраховують суму

$$\sum_{i=1}^4 (x_i - 2,5)^2 = \sum_{i=1}^4 x_i^2 - \frac{\left(\sum_{i=1}^4 x_i\right)^2}{4} = 30 - \frac{10^2}{4} = 5.$$

Отже,

$$f^L(x) = 1 + 1,7x - 9,9 \cdot 0,39 \sqrt{\frac{1}{4} + \frac{(x-2,5)^2}{5}},$$

$$f^U(x) = 1 + 1,7x + 9,9 \cdot 0,39 \sqrt{\frac{1}{4} + \frac{(x-2,5)^2}{5}}.$$

Таким чином, остаточно маємо

$$f^L(x) = 1 + 1,7x - 3,861 \sqrt{\frac{1}{4} + \frac{(x-2,5)^2}{5}},$$

$$f^U(x) = 1 + 1,7x + 3,861 \sqrt{\frac{1}{4} + \frac{(x-2,5)^2}{5}}.$$

Довірча смуга значень залежної змінної

Оскільки при оцінці довірчого інтервалу для значення залежної змінної доводиться враховувати ще і стандартну похибку нев'язок, то дисперсія такої оцінки має вигляд

$$\hat{\sigma}^2 \cdot \left(\frac{1}{N} + \frac{(x_i - \bar{x})^2}{\sum_{i=1}^N (x_i - \bar{x})^2} \right) + \hat{\sigma}^2 = \hat{\sigma}^2 \cdot \left(1 + \frac{1}{N} + \frac{(x_i - \bar{x})^2}{\sum_{i=1}^N (x_i - \bar{x})^2} \right),$$

а точність точкової оцінки $\hat{\varepsilon}_{1-\alpha}^{y_i}$ для прогнозованого значення залежної змінної у точці x_i запишеться так:

$$\hat{\varepsilon}_{1-\alpha}^{y_i} = t_{N-2; \frac{\alpha}{2}} \cdot \hat{\sigma} \cdot \sqrt{1 + \frac{1}{N} + \frac{(x_i - \bar{x})^2}{\sum_{i=1}^N (x_i - \bar{x})^2}}. \quad (3.91)$$

Відповідно, довірча смуга буде дещо ширшою за смугу оцінки істинного значення і матиме верхню та нижню критичні функції

$$f_{\hat{y}}^L(x) = \hat{\beta}_0 + \hat{\beta}_1 x - t_{N-2; \frac{\alpha}{2}} \hat{\sigma} \sqrt{1 + \frac{1}{N} + \frac{(x - \bar{x})^2}{\sum_{i=1}^N (x_i - \bar{x})^2}} \quad (3.92)$$

та

$$f_{\hat{y}}^U(x) = \hat{\beta}_0 + \hat{\beta}_1 x + t_{N-2; \frac{\alpha}{2}} \hat{\sigma} \sqrt{1 + \frac{1}{N} + \frac{(x - \bar{x})^2}{\sum_{i=1}^N (x_i - \bar{x})^2}}. \quad (3.93)$$

Приклад 3.34

Для умов прикладу 3.33 границі довірчої смуги (рис. 3.18) для значень залежної змінної у визначають за формулами (3.92) та (3.93):

$$f_{\hat{y}}^L(x) = 1 + 1,7x - t_{N-2; \frac{1-0,9}{2}} \cdot 0,39 \sqrt{1 + \frac{1}{4} + \frac{(x-2,5)^2}{\sum_{i=1}^4 (x_i - 2,5)^2}},$$

$$f_{\hat{y}}^U(x) = 1 + 1,7x + t_{N-2; \frac{1-0,9}{2}} \cdot 0,39 \sqrt{1 + \frac{1}{4} + \frac{(x-2,5)^2}{\sum_{i=1}^4 (x_i - 2,5)^2}},$$

або остаточно

$$f_{\hat{y}}^L(x) = 1 + 1,7x - 3,861 \sqrt{1 + \frac{1}{4} + \frac{(x-2,5)^2}{5}},$$

$$f_{\hat{y}}^U(x) = 1 + 1,7x + 3,861 \sqrt{1 + \frac{1}{4} + \frac{(x-2,5)^2}{5}}.$$

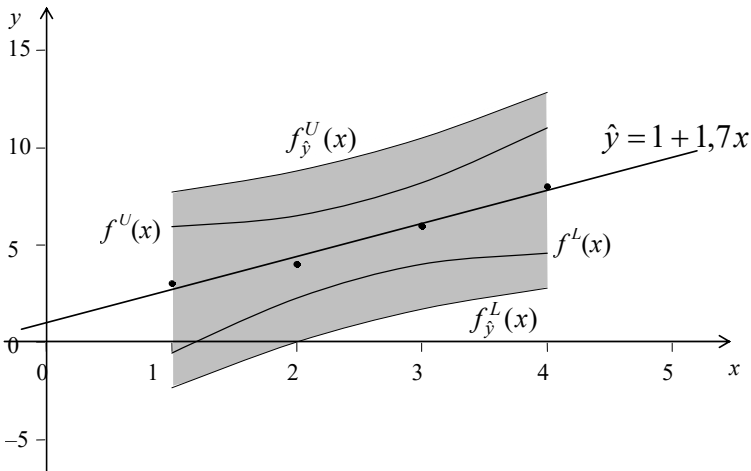


Рис. 3.18. Довірча смуга значень залежної змінної

Довірча смуга значень істинної регресії є довірчою смугою середніх значень залежної змінної.

3.6.7. Множинна та поліноміальна лінійні регресії

Якщо аргументів багато, то лінійний за аргументами тип функцій, зображений у (3.70), представляють виразом

$$f(x; \beta) = \beta_0 + \beta_1 x_1 + \dots + \beta_{p-1} x_{p-1}. \quad (3.94)$$

Аргумент цієї регресії є векторним $x = \begin{pmatrix} x_1 \\ \dots \\ x_{p-1} \end{pmatrix}$ так само, як і

параметр $\beta = \begin{pmatrix} \beta_0 \\ \dots \\ \beta_{p-1} \end{pmatrix}$.

Дуже поширене використання регресії, яку називають **поліноміальною**. У ній істинною регресією є

$$f(x, \beta) = \beta_0 + \beta_1 x^1 + \dots + \beta_{p-1} x^{p-1} \quad (3.95)$$

із числовим аргументом x і векторним параметром $\beta = \begin{pmatrix} \beta_0 \\ \vdots \\ \beta_{p-1} \end{pmatrix}$.

Отже, для множинної регресії істинною регресією є функція, яка визначається співвідношенням (3.95).

Як і у випадку простої лінійної регресії, кожному з пар $a_i = (x_i, y_i)$, $i = \overline{1, N}$, називатимемо **спостереженням регресії**. Відповідно до (3.69) елементи спостереження регресії $a_i = (x_i, y_i)$ зв'язані співвідношенням

$$y_i = f(x_i; \beta) + \varepsilon_i, \quad i = \overline{1, N}, \quad (3.96)$$

яке називають **моделлю спостережень** для регресії.

Регресію, лінійну за елементами вектора параметрів β , називають **лінійною**. Лінійність за параметром β означає, що істинна регресія є сумою відомих функцій із коефіцієнтами, що збігаються з компонентами вектора параметрів β :

$$f(x; \beta) = \beta_0 f_0(x) + \beta_1 f_1(x) + \dots + \beta_{p-1} f_{p-1}(x). \quad (3.97)$$

Лінійну регресію, в якій $f_0(x) \equiv 1$, тобто регресію вигляду

$$f(x, \beta) = \beta_0 + \beta_1 f_1(x) + \dots + \beta_{p-1} f_{p-1}(x) \quad (3.98)$$

називають **регресією з одиницею**.

Відповідно до цього визначення всі приклади регресій, наведених вище (найпростіша, множинна, поліноміальна), є лінійними з одиницею.

Модель спостережень (3.71) для лінійної регресії така:

$$y_i = \beta_0 f_0(x_i) + \beta_1 f_1(x_i) + \dots + \beta_{p-1} f_{p-1}(x_i) + \varepsilon_i, \quad i = \overline{1, N}, \quad (3.99)$$

і може представлятися у векторно-матричному вигляді

$$Y = X\beta + \varepsilon, \quad (3.100)$$

де Y – вектор-стовпець, складений зі значень y_i , $i = \overline{1, N}$; ε – вектор, що утворений із помилок спостережень ε_i , $i = \overline{1, N}$; X – матриця, складена зі значень усіх функцій $f_j(x)$, $j = \overline{0, p-1}$, на всіх аргументах x_i , $i = \overline{1, N}$,

$$X = \begin{pmatrix} f_0(x_1) & f_1(x_1) & \dots & f_{p-1}(x_1) \\ f_0(x_2) & f_1(x_2) & \dots & f_{p-1}(x_2) \\ \dots & \dots & \dots & \dots \\ f_0(x_N) & f_1(x_N) & \dots & f_{p-1}(x_N) \end{pmatrix}.$$

ЗАДАЧІ

У задачах 1–5 побудувати:

- прямі лінії регресії з незалежною змінною x та дати інтерпретацію знайдених параметрів;
- довірчу смугу значень істинної регресії;
- довірчу смугу значень залежної змінної;
- довірчий інтервал істинної регресії для значення $x = 140$;
- довірчий інтервал значення y при $x = 140$.

№ 1

Зріст (см)	172	175	177	180
Вага (кг)	70	75	77	85

№ 2 X – кількість сторінок у книзі, Y – кількість друкарських помилок:

X	100	200	300	350
Y	12	14	20	25

№ 3 X – об'єм поставок (тис. грошових одиниць), Y – об'єм продажу (тис. грошових одиниць)

X	152	173	203	249	250
Y	251	285	333	359	380

№ 4 X – об'єм продажу (тис. грошових одиниць), Y – прибуток (тис. грошових одиниць)

X	100	120	150	200
Y	30	40	50	75

№ 5

У вибірці записана інформація про відстань, яку може проїхати автомобіль на 1 л пального (kml), та потужність двигуна у кінських силах (hp):

kml	hp	kml	hp	kml	hp
5,56	110	3,78	245	4,58	180
5,56	110	6,46	62	4,02	180
6,03	93	6,03	95	2,75	205
5,66	110	5,08	123	2,75	215
4,95	175	4,71	123	3,89	230
4,79	105	4,34	180	8,57	66

a)

1) побудувати прямі лінії регресії з незалежною змінною hp та дати інтерпретацію знайдених параметрів;

2) побудувати довірчу смугу значень істинної регресії;

3) побудувати довірчу смугу значень залежної змінної;

4) побудувати довірчий інтервал істинної регресії для значення $h_p = 165$;

5) побудувати довірчий інтервал значення y при $h_p = 165$;

б)

1) побудувати прямі лінії регресії з незалежною змінною kml та дати інтерпретацію знайдених параметрів;

2) побудувати довірчу смугу значень істинної регресії;

3) побудувати довірчу смугу значень залежної змінної;

4) побудувати довірчий інтервал істинної регресії для значення $kml = 1,5$;

5) побудувати довірчий інтервал значення y при $kml = 1,5$.

4. СТАТИСТИЧНІ ТАБЛИЦІ

4.1. Таблиця значень гамма-функції

Для отримання значення гамма-функції потрібно вибрати значення у крайньому лівому стовпчику (десяті) і у верхньому рядку (соті). На перетині цих стовпчика та рядка наведено значення гамма-функції від суми цих двох значень.

Таблиця 4.1 (а)

	0	0,01	0,02	0,03	0,04	0,05
0		99,4326	49,4422	32,785	24,461	19,4701
0,1	9,5135	8,6127	7,8633	7,2302	6,6887	6,2203
0,2	4,5908	4,3599	4,1505	3,9598	3,7855	3,6256
0,3	2,9916	2,8903	2,7958	2,7072	2,6242	2,5461
0,4	2,2182	2,1628	2,1104	2,0605	2,0132	1,9681
0,5	1,7725	1,7384	1,7058	1,6747	1,6448	1,6161
0,6	1,4892	1,4667	1,445	1,4242	1,4041	1,3848
0,7	1,2981	1,2825	1,2675	1,253	1,239	1,2254
0,8	1,1642	1,1532	1,1425	1,1322	1,1222	1,1125
0,9	1,0686	1,0607	1,053	1,0456	1,0384	1,0315
1	1	0,9943	0,9888	0,9835	0,9784	0,9735
1,1	0,9514	0,9474	0,9436	0,9399	0,9364	0,933
1,2	0,9182	0,9156	0,9131	0,9108	0,9085	0,9064
1,3	0,8975	0,896	0,8946	0,8934	0,8922	0,8912
1,4	0,8873	0,8868	0,8864	0,886	0,8858	0,8857
1,5	0,8862	0,8866	0,887	0,8876	0,8882	0,8889
1,6	0,8935	0,8947	0,8959	0,8972	0,8986	0,9001
1,7	0,9086	0,9106	0,9126	0,9147	0,9168	0,9191
1,8	0,9314	0,9341	0,9368	0,9397	0,9426	0,9456
1,9	0,9618	0,9652	0,9688	0,9724	0,9761	0,9799
2	1	1,0043	1,0086	1,0131	1,0176	1,0222
2,1	1,0465	1,0516	1,0568	1,0621	1,0675	1,073

Продовження табл. 4.1 (а)

	0	0,01	0,02	0,03	0,04	0,05
2,2	1,1018	1,1078	1,114	1,1202	1,1266	1,133
2,3	1,1667	1,1738	1,1809	1,1882	1,1956	1,2031
2,4	1,2422	1,2503	1,2586	1,267	1,2756	1,2842
2,5	1,3293	1,3388	1,3483	1,358	1,3678	1,3777
2,6	1,4296	1,4404	1,4514	1,4625	1,4738	1,4852
2,7	1,5447	1,5571	1,5696	1,5824	1,5953	1,6084
2,8	1,6765	1,6907	1,7051	1,7196	1,7344	1,7494
2,9	1,8274	1,8436	1,86	1,8767	1,8936	1,9108
3	2	2,0186	2,0374	2,0565	2,0759	2,0955
3,1	2,1976	2,2189	2,2405	2,2623	2,2845	2,3069
3,2	2,424	2,4483	2,4731	2,4981	2,5235	2,5493
3,3	2,6834	2,7114	2,7397	2,7685	2,7976	2,8272
3,4	2,9812	3,0133	3,0459	3,0789	3,1124	3,1463
3,5	3,3234	3,3603	3,3977	3,4357	3,4742	3,5133
3,6	3,717	3,7595	3,8027	3,8464	3,8908	3,9358
3,7	4,1707	4,2197	4,2694	4,3199	4,3711	4,423
3,8	4,6942	4,7508	4,8083	4,8666	4,9257	4,9857
3,9	5,2993	5,3648	5,4313	5,4988	5,5673	5,6368
4	6	6,0759	6,153	6,2312	6,3106	6,3912
4,1	6,8126	6,9008	6,9902	7,0811	7,1733	7,2669
4,2	7,7567	7,8592	7,9632	8,0689	8,1762	8,2851
4,3	8,8553	8,9747	9,096	9,2191	9,3441	9,471
4,4	10,1361	10,2754	10,4169	10,5606	10,7065	10,8548
4,5	11,6317	11,7946	11,96	12,128	12,2987	12,472
4,6	13,3813	13,5719	13,7656	13,9624	14,1624	14,3655
4,7	15,4314	15,655	15,8822	16,1131	16,3478	16,5862
4,8	17,8379	18,1006	18,3676	18,6389	18,9147	19,1951
4,9	20,6674	20,9766	21,2908	21,6103	21,9351	22,2652
5	24	24,3645	24,735	25,1118	25,4948	25,8843
5,1	27,9318	28,3621	28,7998	29,2448	29,6973	30,1575

Продовження табл. 4.1 (а)

	0	0,01	0,02	0,03	0,04	0,05
5,2	32,5781	33,0872	33,6049	34,1314	34,6669	35,2116
5,3	38,078	38,6811	39,2945	39,9186	40,5533	41,1991
5,4	44,5988	45,3145	46,0426	46,7834	47,537	48,3038
5,5	52,3428	53,1934	54,059	54,9397	55,8359	56,7478
5,6	61,5539	62,5666	63,5972	64,646	65,7134	66,7997
5,7	72,5276	73,7351	74,9641	76,2151	77,4884	78,7845
5,8	85,6217	87,0637	88,5317	90,0261	91,5474	93,0961
5,9	101,2702	102,9949	104,7509	106,5388	108,3593	110,2128
6	120	122,066	124,1698	126,3122	128,4939	130,7156
6,1	142,4519	144,9305	147,4549	150,0258	152,6443	155,3112
6,2	169,4061	172,3841	175,4175	178,5073	181,6548	184,861
6,3	201,8133	205,3965	209,047	212,766	216,5548	220,4149
6,4	240,8338	245,1516	249,5511	254,0338	258,6014	263,2555
6,5	287,8853	293,0958	298,4056	303,8166	309,3308	314,9503
6,6	344,7019	350,9986	357,4162	363,957	370,6235	377,4183
6,7	413,4075	421,0275	428,7948	436,7125	444,7834	453,0108
6,8	496,6061	505,8403	515,2543	524,8519	534,6367	544,6124
6,9	597,4941	608,6999	620,1255	631,7753	643,654	655,7663
7	720	733,6167	747,5024	761,6626	776,1029	790,8291
7,1	868,9569	885,5255	902,4237	919,6582	937,236	955,1641
7,2	1050,3178	1070,5051	1091,0966	1112,1007	1133,5259	1155,381
7,3	1271,4236	1296,0521	1321,177	1346,8086	1372,9576	1399,6347
7,4	1541,3362	1571,4219	1602,1179	1633,4372	1665,3927	1697,9978
7,5	1871,2543	1908,0536	1945,6043	1983,9221	2023,0234	2062,9246
7,6	2275,0327	2320,1008	2366,0949	2413,0347	2460,9401	2509,8317
7,7	2769,8304	2825,0945	2881,5013	2939,0748	2997,8402	3057,8227
7,8	3376,9213	3444,7723	3514,0346	3584,7385	3656,9147	3730,5949
7,9	4122,7095	4206,1163	4291,2683	4378,2031	4466,9591	4557,5755
8	5040	5142,6533	5247,4672	5354,4884	5463,7647	5575,3451
8,1	6169,5937	6296,0863	6425,2567	6557,1631	6691,865	6829,4233

Закінчення табл. 4.1 (а)

	0	0,01	0,02	0,03	0,04	0,05
8,2	7562,2883	7718,3416	7877,7172	8040,4877	8206,7275	8376,5123
8,3	9281,3925	9474,1407	9671,0154	9872,1073	10077,509	10287,3154
8,4	11405,888	11644,2361	11887,7152	12136,4384	12390,5217	12650,0834
8,5	14034,407	14329,4825	14630,9441	14938,9338	15253,5966	15575,0807
8,6	17290,249	17655,9672	18029,6434	18411,4548	18801,5826	19200,2127
8,7	21327,694	21781,4789	22245,1897	22719,0486	23203,283	23698,1257
8,8	26339,986	26903,6719	27479,7509	28068,5022	28670,2112	29285,1699
8,9	32569,405	33270,3796	33986,8446	34719,1503	35467,655	36232,7254
9	40320	41192,6529	42084,6867	42996,5417	43928,6684	44881,5277
9,1	49973,709	51061,2603	52173,0847	53309,7361	54471,7812	55659,8001
9,2	62010,764	63367,5843	64754,8351	66173,2139	67623,4343	69106,2269
9,3	77035,558	78730,1093	80462,8481	82234,6534	84046,4249	85899,0836
9,4	95809,458	97928,0254	100094,562	102310,176	104576,003	106893,205
9,5	119292,46	121943,896	124655,643	127429,105	130265,7147	133166,9396
9,6	148696,14	152017,877	155415,53	158890,8547	162445,674	166081,8396
9,7	185550,94	189716,68	193978,05	198337,294	202796,6936	207358,5998
9,8	231791,88	237021,35	242371,403	247844,874	253444,667	259173,7533
9,9	289867,70	296439,082	303162,654	310042,012	317080,8358	324282,8926
10	362880	371145,802	379603,874	388258,771	397115,1623	406177,8259

Таблиця 4.1 (б)

	0,06	0,07	0,08	0,09
0	16,1457	13,7736	11,9966	10,6162
0,1	5,8113	5,4512	5,1318	4,8468
0,2	3,4785	3,3426	3,2169	3,1001
0,3	2,4727	2,4036	2,3383	2,2765
0,4	1,9252	1,8843	1,8453	1,8081
0,5	1,5886	1,5623	1,5369	1,5126
0,6	1,3662	1,3482	1,3309	1,3142
0,7	1,2123	1,1997	1,1875	1,1757

Продовження табл. 4.1 (б)

	0,06	0,07	0,08	0,09
0,8	1,1031	1,0941	1,0853	1,0768
0,9	1,0247	1,0182	1,0119	1,0059
1	0,9687	0,9642	0,9597	0,9555
1,1	0,9298	0,9267	0,9237	0,9209
1,2	0,9044	0,9025	0,9007	0,899
1,3	0,8902	0,8893	0,8885	0,8879
1,4	0,8856	0,8856	0,8857	0,8859
1,5	0,8896	0,8905	0,8914	0,8924
1,6	0,9017	0,9033	0,905	0,9068
1,7	0,9214	0,9238	0,9262	0,9288
1,8	0,9487	0,9518	0,9551	0,9584
1,9	0,9837	0,9877	0,9917	0,9958
2	1,0269	1,0316	1,0365	1,0415
2,1	1,0786	1,0842	1,09	1,0959
2,2	1,1395	1,1462	1,1529	1,1598
2,3	1,2107	1,2184	1,2262	1,2341
2,4	1,293	1,3019	1,3109	1,3201
2,5	1,3878	1,3981	1,4084	1,419
2,6	1,4968	1,5085	1,5204	1,5325
2,7	1,6216	1,6351	1,6487	1,6625
2,8	1,7646	1,7799	1,7955	1,8113
2,9	1,9281	1,9457	1,9636	1,9817
3	2,1153	2,1355	2,1559	2,1766
3,1	2,3297	2,3528	2,3762	2,3999
3,2	2,5754	2,6018	2,6287	2,6559
3,3	2,8571	2,8875	2,9183	2,9495
3,4	3,1807	3,2156	3,251	3,2869
3,5	3,5529	3,593	3,6338	3,6751
3,6	3,9814	4,0277	4,0747	4,1223
3,7	4,4757	4,5291	4,5833	4,6384

Продовження табл. 4.1 (б)

	0,06	0,07	0,08	0,09
3,8	5,0466	5,1084	5,1711	5,2348
3,9	5,7073	5,7788	5,8515	5,9252
4	6,473	6,556	6,6403	6,7258
4,1	7,3619	7,4584	7,5563	7,6557
4,2	8,3957	8,508	8,622	8,7378
4,3	9,6	9,7309	9,8639	9,9989
4,4	11,0053	11,1583	11,3136	11,4714
4,5	12,6482	12,8271	13,0089	13,1936
4,6	14,5719	14,7817	14,9948	15,2114
4,7	16,8285	17,0748	17,325	17,5794
4,8	19,48	19,7696	20,064	20,3632
4,9	22,6008	22,942	23,2889	23,6415
5	26,2802	26,6829	27,0922	27,5085
5,1	30,6255	31,1014	31,5854	32,0776
5,2	35,7656	36,329	36,9021	37,4851
5,3	41,8559	42,5241	43,2038	43,8954
5,4	49,0839	49,8776	50,6851	51,5068
5,5	57,6758	58,62	59,581	60,5588
5,6	67,9053	69,0305	70,1757	71,3413
5,7	80,1038	81,4467	82,8137	84,2052
5,8	94,6729	96,2781	97,9124	99,5762
5,9	112,1001	114,0218	115,9786	117,9711
6	132,978	135,2821	137,6286	140,0182
6,1	158,0276	160,7942	163,6122	166,4825
6,2	188,127	191,454	194,8433	198,2959
6,3	224,3476	228,3544	232,4367	236,596
6,4	267,9978	272,8302	277,7544	282,7721
6,5	320,6772	326,5136	332,4617	338,5237
6,6	384,3439	391,4029	398,598	405,932
6,7	461,3977	469,9473	478,6631	487,5482

Продовження табл. 4.1 (б)

	0,06	0,07	0,08	0,09
6,8	554,783	565,1524	575,7246	586,5038
6,9	668,1167	680,7103	693,552	706,6469
7	805,8469	821,1625	836,7818	852,7111
7,1	973,4498	992,1003	1011,1234	1030,5266
7,2	1177,6749	1200,4168	1223,6158	1247,2815
7,3	1426,851	1454,6177	1482,9462	1511,8484
7,4	1731,2661	1765,2115	1799,8483	1835,1909
7,5	2103,6425	2145,1945	2187,5981	2230,8714
7,6	2559,7304	2610,6575	2662,6349	2715,6849
7,7	3119,0483	3181,5435	3245,3356	3310,452
7,8	3805,8113	3882,5969	3960,9854	4041,0112
7,9	4650,0925	4744,5511	4840,9932	4939,4615
8	5689,2793	5805,6185	5924,4149	6045,7218
8,1	6969,9003	7113,3595	7259,866	7409,4862
8,2	8549,92	8727,0298	8907,9228	9092,6821
8,3	10501,6234	10720,5322	10944,143	11172,5596
8,4	12915,2448	13186,1298	13462,8649	13745,58
8,5	15903,5376	16239,1225	16581,9939	16932,314
8,6	19607,5347	20023,7429	20449,036	20883,617
8,7	24203,8146	24720,5932	25248,7106	25788,4214
8,8	29913,6768	30556,0376	31212,5649	31883,5782
8,9	37014,7367	37814,0725	38631,1256	39466,2978
9	45855,5913	46851,3417	47869,2726	48909,8894
9,1	56874,3862	58116,1469	59385,7036	60683,6923
9,2	70622,3392	72172,5363	73757,6011	75378,335
9,3	87793,5719	89730,8545	91711,9187	93737,775
9,4	109262,9711	111686,519	114165,0947	116699,9738
9,5	136134,2822	139169,2802	142273,508	145448,577
9,6	169801,2505	173605,8513	177497,6324	181478,6318
9,7	212025,416	216799,6026	221683,6789	226680,2243

Закінчення табл. 4.1 (б)

	0,06	0,07	0,08	0,09
9,8	265035,1767	271032,0538	277167,5759	283445,0105
9,9	331652,0405	339192,2304	346907,5081	354802,0169
10	415451,6572	424941,6691	434652,9955	444590,8946

4.2. Таблиця значень функції Лапласа

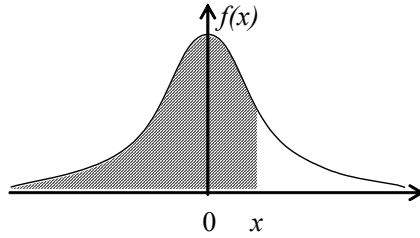
$$\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^x e^{-\frac{y^2}{2}} dy = P(Z < x).$$

Для отримання значення функції Лапласа потрібно вибрати значення у крайньому лівому стовпчику (десяті) та у верхньому рядку (соті).

На перетині цих стовпчика та рядка наведено значення функції $\Phi(x)$ – функції Лапласа від суми x цих двох значень.

Якщо треба знайти $\Phi(-x)$, де x – додатне число, то слід використати властивість функції Лапласа

$$\Phi(-x) = 1 - \Phi(x).$$



Таблиця 4.2

	0	0,01	0,02	0,03	0,04	0,05	0,06	0,07	0,08	0,09
0	0,5	0,504	0,508	0,512	0,516	0,5199	0,5239	0,5279	0,5319	0,5359
0,1	0,5398	0,5438	0,5478	0,5517	0,5557	0,5596	0,5636	0,5675	0,5714	0,5753
0,2	0,5793	0,5832	0,5871	0,591	0,5948	0,5987	0,6026	0,6064	0,6103	0,6141
0,3	0,6179	0,6217	0,6255	0,6293	0,6331	0,6368	0,6406	0,6443	0,648	0,6517
0,4	0,6554	0,6591	0,6628	0,6664	0,67	0,6736	0,6772	0,6808	0,6844	0,6879
0,5	0,6915	0,695	0,6985	0,7019	0,7054	0,7088	0,7123	0,7157	0,719	0,7224
0,6	0,7257	0,7291	0,7324	0,7357	0,7389	0,7422	0,7454	0,7486	0,7517	0,7549

Закінчення табл. 4.2

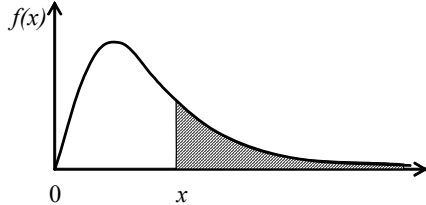
	0	0,01	0,02	0,03	0,04	0,05	0,06	0,07	0,08	0,09
0,7	0,758	0,7611	0,7642	0,7673	0,7704	0,7734	0,7764	0,7794	0,7823	0,7852
0,8	0,7881	0,791	0,7939	0,7967	0,7995	0,8023	0,8051	0,8078	0,8106	0,8133
0,9	0,8159	0,8186	0,8212	0,8238	0,8264	0,8289	0,8315	0,834	0,8365	0,8389
1	0,8413	0,8438	0,8461	0,8485	0,8508	0,8531	0,8554	0,8577	0,8599	0,8621
1,1	0,8643	0,8665	0,8686	0,8708	0,8729	0,8749	0,877	0,879	0,881	0,883
1,2	0,8849	0,8869	0,8888	0,8907	0,8925	0,8944	0,8962	0,898	0,8997	0,9015
1,3	0,9032	0,9049	0,9066	0,9082	0,9099	0,9115	0,9131	0,9147	0,9162	0,9177
1,4	0,9192	0,9207	0,9222	0,9236	0,9251	0,9265	0,9279	0,9292	0,9306	0,9319
1,5	0,9332	0,9345	0,9357	0,937	0,9382	0,9394	0,9406	0,9418	0,9429	0,9441
1,6	0,9452	0,9463	0,9474	0,9484	0,9495	0,9505	0,9515	0,9525	0,9535	0,9545
1,7	0,9554	0,9564	0,9573	0,9582	0,9591	0,9599	0,9608	0,9616	0,9625	0,9633
1,8	0,9641	0,9649	0,9656	0,9664	0,9671	0,9678	0,9686	0,9693	0,9699	0,9706
1,9	0,9713	0,9719	0,9726	0,9732	0,9738	0,9744	0,975	0,9756	0,9761	0,9767
2	0,9772	0,9778	0,9783	0,9788	0,9793	0,9798	0,9803	0,9808	0,9812	0,9817
2,1	0,9821	0,9826	0,983	0,9834	0,9838	0,9842	0,9846	0,985	0,9854	0,9857
2,2	0,9861	0,9864	0,9868	0,9871	0,9875	0,9878	0,9881	0,9884	0,9887	0,989
2,3	0,9893	0,9896	0,9898	0,9901	0,9904	0,9906	0,9909	0,9911	0,9913	0,9916
2,4	0,9918	0,992	0,9922	0,9925	0,9927	0,9929	0,9931	0,9932	0,9934	0,9936
2,5	0,9938	0,994	0,9941	0,9943	0,9945	0,9946	0,9948	0,9949	0,9951	0,9952
2,6	0,9953	0,9955	0,9956	0,9957	0,9959	0,996	0,9961	0,9962	0,9963	0,9964
2,7	0,9965	0,9966	0,9967	0,9968	0,9969	0,997	0,9971	0,9972	0,9973	0,9974
2,8	0,9974	0,9975	0,9976	0,9977	0,9977	0,9978	0,9979	0,9979	0,998	0,9981
2,9	0,9981	0,9982	0,9982	0,9983	0,9984	0,9984	0,9985	0,9985	0,9986	0,9986
3	0,9987	0,9987	0,9987	0,9988	0,9988	0,9989	0,9989	0,9989	0,999	0,999
3,1	0,999	0,9991	0,9991	0,9991	0,9992	0,9992	0,9992	0,9992	0,9993	0,9993
3,2	0,9993	0,9993	0,9994	0,9994	0,9994	0,9994	0,9994	0,9995	0,9995	0,9995
3,3	0,9995	0,9995	0,9995	0,9996	0,9996	0,9996	0,9996	0,9996	0,9996	0,9997
3,4	0,9997	0,9997	0,9997	0,9997	0,9997	0,9997	0,9997	0,9997	0,9997	0,9998
3,5	0,9998	0,9998	0,9998	0,9998	0,9998	0,9998	0,9998	0,9998	0,9998	0,9998
3,6	0,9998	0,9998	0,9999	0,9999	0,9999	0,9999	0,9999	0,9999	0,9999	0,9999

4.3. Таблиця значень верхніх критичних величин розподілу χ_n^2

$$P(\chi_n^2 > x) = \alpha,$$

де x – верхня критична величина рівня значущості α .

Для отримання значення верхньої критичної величини (точки, яка відтинає праворуч частку (імовірність) α більших значень)



потрібно вибрати кількість ступенів вільності у крайньому лівому стовпчику, а у верхньому – рівень значущості. На перетині цих стовпчика та рядка наведено значення верхньої критичної величини.

Таблиця 4.3(а)

$n \backslash \alpha$	0,995	0,99	0,98	0,975	0,95	0,9	0,8
1	0,000039	0,000157	0,000628	0,000982	0,003932	0,015791	0,064185
2	0,010025	0,0201	0,0404	0,0506	0,1026	0,2107	0,4463
3	0,07172	0,1148	0,1848	0,2158	0,3518	0,5844	1,0052
4	0,2070	0,2971	0,4294	0,4844	0,7107	1,0636	1,6488
5	0,4117	0,5543	0,7519	0,8312	1,1455	1,6103	2,3425
6	0,6757	0,8721	1,1344	1,2373	1,6354	2,2041	3,0701
7	0,9893	1,2390	1,5643	1,6899	2,1673	2,8331	3,8223
8	1,3444	1,6465	2,0325	2,1797	2,7326	3,4895	4,5936
9	1,7349	2,0879	2,5324	2,7004	3,3251	4,1682	5,3801
10	2,1559	2,5582	3,0591	3,2470	3,9403	4,8652	6,1791
11	2,6032	3,0535	3,6087	3,8157	4,5748	5,5778	6,9887
12	3,0738	3,5706	4,1783	4,4038	5,2260	6,3038	7,8073
13	3,5650	4,1069	4,7654	5,0088	5,8919	7,0415	8,6339
14	4,0747	4,6604	5,3682	5,6287	6,5706	7,7895	9,4673
15	4,6009	5,2293	5,9849	6,2621	7,2609	8,5468	10,3070
16	5,1422	5,8122	6,6142	6,9077	7,9616	9,3122	11,1521

Закінчення табл. 4.3 (а)

$n \backslash \alpha$	0,995	0,99	0,98	0,975	0,95	0,9	0,8
17	5,6972	6,4078	7,2550	7,5642	8,6718	10,0852	12,0023
18	6,2648	7,0149	7,9062	8,2307	9,3905	10,8649	12,8570
19	6,8440	7,6327	8,5670	8,9065	10,1170	11,6509	13,7158
20	7,4338	8,2604	9,2367	9,5908	10,8508	12,4426	14,5784
21	8,0337	8,8972	9,9146	10,2829	11,5913	13,2396	15,4446
22	8,6427	9,5425	10,6000	10,9823	12,3380	14,0415	16,3140
23	9,2604	10,1957	11,2926	11,6886	13,0905	14,8480	17,1865
24	9,8862	10,8564	11,9918	12,4012	13,8484	15,6587	18,0618
25	10,5197	11,5240	12,6973	13,1197	14,6114	16,4734	18,9398
26	11,1602	12,1981	13,4086	13,8439	15,3792	17,2919	19,8202
27	11,8076	12,8785	14,1254	14,5734	16,1514	18,1139	20,7030
28	12,4613	13,5647	14,8475	15,3079	16,9279	18,9392	21,5880
29	13,1211	14,2565	15,5745	16,0471	17,7084	19,7677	22,4751
30	13,7867	14,9535	16,3062	16,7908	18,4927	20,5992	23,3641
31	14,4578	15,6555	17,0423	17,5387	19,2806	21,4336	24,2551
32	15,1340	16,3622	17,7827	18,2908	20,0719	22,2706	25,1478
33	15,8153	17,0735	18,5271	19,0467	20,8665	23,1102	26,0422
34	16,5013	17,7891	19,2754	19,8063	21,6643	23,9523	26,9383
35	17,1918	18,5089	20,0274	20,5694	22,4650	24,7967	27,8359
36	17,8867	19,2327	20,7829	21,3359	23,2686	25,6433	28,7350
37	18,5858	19,9602	21,5419	22,1056	24,0749	26,4921	29,6355
38	19,2889	20,6914	22,3040	22,8785	24,8839	27,3430	30,5373
39	19,9959	21,4262	23,0693	23,6543	25,6954	28,1958	31,4405
40	20,7065	22,1643	23,8376	24,4330	26,5093	29,0505	32,3450
45	24,3110	25,9013	27,7203	28,3662	30,6123	33,3504	36,8844
50	27,9907	29,7067	31,6639	32,3574	34,7643	37,6886	41,4492
60	35,5345	37,4849	39,6994	40,4817	43,1880	46,4589	50,6406
80	51,1719	53,5401	56,2128	57,1532	60,3915	64,2778	69,2069
100	67,3276	70,0649	73,1422	74,2219	77,9295	82,3581	87,9453
120	83,8516	86,9233	90,3667	91,5726	95,7046	100,6236	106,8056
200	152,241	156,4320	161,1003	162,7280	168,2786	174,8353	183,0028

Таблица 4.3 (б)

$n \backslash \alpha$	0,75	0,5	0,25	0,2	0,125	0,1	0,075
1	0,10153	0,4549	1,3233	1,6424	2,3535	2,7055	3,1701
2	0,5754	1,3863	2,7726	3,2189	4,1589	4,6052	5,1805
3	1,2125	2,3660	4,1083	4,6416	5,7394	6,2514	6,9046
4	1,9226	3,3567	5,3853	5,9886	7,2140	7,7794	8,4963
5	2,6746	4,3515	6,6257	7,2893	8,6248	9,2364	10,0083
6	3,4546	5,3481	7,8408	8,5581	9,9917	10,6446	11,4659
7	4,2549	6,3458	9,0371	9,8032	11,3264	12,0170	12,8834
8	5,0706	7,3441	10,2189	11,0301	12,6361	13,3616	14,2697
9	5,8988	8,3428	11,3888	12,2421	13,9255	14,6837	15,6309
10	6,7372	9,3418	12,5489	13,4420	15,1982	15,9872	16,9714
11	7,5841	10,3410	13,7007	14,6314	16,4568	17,2750	18,2942
12	8,4384	11,3403	14,8454	15,8120	17,7033	18,5493	19,6020
13	9,2991	12,3398	15,9839	16,9848	18,9392	19,8119	20,8966
14	10,1653	13,3393	17,1169	18,1508	20,1658	21,0641	22,1795
15	11,0365	14,3389	18,2451	19,3107	21,3841	22,3071	23,4522
16	11,9122	15,3385	19,3689	20,4651	22,5949	23,5418	24,7155
17	12,7919	16,3382	20,4887	21,6146	23,7990	24,7690	25,9705
18	13,6753	17,3379	21,6049	22,7595	24,9970	25,9894	27,2178
19	14,5620	18,3377	22,7178	23,9004	26,1893	27,2036	28,4581
20	15,4518	19,3374	23,8277	25,0375	27,3765	28,4120	29,6920
21	16,3444	20,3372	24,9348	26,1711	28,5589	29,6151	30,9200
22	17,2396	21,3370	26,0393	27,3015	29,7369	30,8133	32,1424
23	18,1373	22,3369	27,1413	28,4288	30,9108	32,0069	33,3597
24	19,0373	23,3367	28,2412	29,5533	32,0809	33,1962	34,5723
25	19,9393	24,3366	29,3389	30,6752	33,2473	34,3816	35,7803
26	20,8434	25,3365	30,4346	31,7946	34,4104	35,5632	36,9841
27	21,7494	26,3363	31,5284	32,9117	35,5703	36,7412	38,1840
28	22,6572	27,3362	32,6205	34,0266	36,7272	37,9159	39,3801
29	23,5666	28,3361	33,7109	35,1394	37,8812	39,0875	40,5727
30	24,4776	29,3360	34,7997	36,2502	39,0326	40,2560	41,7619
31	25,3901	30,3359	35,8871	37,3591	40,1814	41,4217	42,9479
32	26,3041	31,3359	36,9730	38,4663	41,3278	42,5847	44,1309
33	27,2194	32,3358	38,0575	39,5718	42,4719	43,7452	45,3110

Закінчення табл. 4.3 (б)

$n \backslash \alpha$	0,75	0,5	0,25	0,2	0,125	0,1	0,075
34	28,1361	33,3357	39,1408	40,6756	43,6137	44,9032	46,4884
35	29,0540	34,3356	40,2228	41,7780	44,7535	46,0588	47,6631
36	29,9730	35,3356	41,3036	42,8788	45,8912	47,2122	48,8353
37	30,8933	36,3355	42,3833	43,9782	47,0270	48,3634	50,0051
38	31,8146	37,3355	43,4619	45,0763	48,1610	49,5126	51,1726
39	32,7369	38,3354	44,5395	46,1730	49,2931	50,6598	52,3378
40	33,6603	39,3353	45,6160	47,2685	50,4236	51,8051	53,5010
45	38,2910	44,3351	50,9849	52,7288	56,0523	57,5053	59,2872
50	42,9421	49,3349	56,3336	58,1638	61,6466	63,1671	65,0303
60	52,2938	59,3347	66,9815	68,9721	72,7508	74,3970	76,4113
80	71,1445	79,3343	88,1303	90,4053	94,7091	96,5782	98,8606
100	90,1332	99,3341	109,1412	111,6667	116,4327	118,4980	121,0166
120	109,2197	119,3340	130,0546	132,8063	137,9899	140,2326	142,9646
200	186,1717	199,3337	213,1022	216,6088	223,1863	226,0210	229,4658

Таблиця 4.3 (в)

$n \backslash \alpha$	0,05	0,025	0,0125	0,01	0,0075	0,005
1	3,8415	5,0239	6,2385	6,6349	7,1491	7,8794
2	5,9915	7,3778	8,7641	9,2103	9,7857	10,5966
3	7,8147	9,3484	10,8613	11,3449	11,9662	12,8382
4	9,4877	11,1433	12,7619	13,2767	13,9367	14,8603
5	11,0705	12,8325	14,5439	15,0863	15,7803	16,7496
6	12,5916	14,4494	16,2445	16,8119	17,5369	18,5476
7	14,0671	16,0128	17,8850	18,4753	19,2288	20,2777
8	15,5073	17,5345	19,4785	20,0902	20,8703	21,9550
9	16,9190	19,0228	21,0341	21,6660	22,4711	23,5894
10	18,3070	20,4832	22,5582	23,2093	24,0381	25,1882
11	19,6751	21,9200	24,0558	24,7250	25,5764	26,7568
12	21,0261	23,3367	25,5304	26,2170	27,0900	28,2995
13	22,3620	24,7356	26,9850	27,6882	28,5820	29,8195
14	23,6848	26,1189	28,4219	29,1412	30,0550	31,3193
15	24,9958	27,4884	29,8431	30,5779	31,5109	32,8013

Закінчення табл. 4.3 (в)

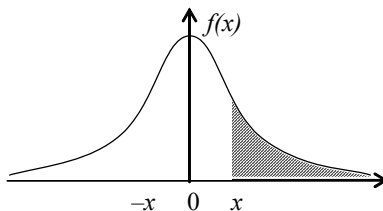
$n \backslash \alpha$	0,05	0,025	0,0125	0,01	0,0075	0,005
16	26,2962	28,8454	31,2501	31,9999	32,9516	34,2672
17	27,5871	30,1910	32,6443	33,4087	34,3785	35,7185
18	28,8693	31,5264	34,0268	34,8053	35,7927	37,1565
19	30,1435	32,8523	35,3986	36,1909	37,1954	38,5823
20	31,4104	34,1696	36,7605	37,5662	38,5875	39,9968
21	32,6706	35,4789	38,1134	38,9322	39,9698	41,4011
22	33,9244	36,7807	39,4578	40,2894	41,3429	42,7957
23	35,1725	38,0756	40,7943	41,6384	42,7075	44,1813
24	36,4150	39,3641	42,1235	42,9798	44,0642	45,5585
25	37,6525	40,6465	43,4458	44,3141	45,4134	46,9279
26	38,8851	41,9232	44,7616	45,6417	46,7556	48,2899
27	40,1133	43,1945	46,0713	46,9629	48,0912	49,6449
28	41,3371	44,4608	47,3752	48,2782	49,4207	50,9934
29	42,5570	45,7223	48,6738	49,5879	50,7442	52,3356
30	43,7730	46,9792	49,9671	50,8922	52,0621	53,6720
31	44,9853	48,2319	51,2556	52,1914	53,3747	55,0027
32	46,1943	49,4804	52,5393	53,4858	54,6823	56,3281
33	47,3999	50,7251	53,8187	54,7755	55,9851	57,6484
34	48,6024	51,9660	55,0938	56,0609	57,2833	58,9639
35	49,8018	53,2033	56,3648	57,3421	58,5771	60,2748
36	50,9985	54,4373	57,6319	58,6192	59,8667	61,5812
37	52,1923	55,6680	58,8954	59,8925	61,1522	62,8833
38	53,3835	56,8955	60,1552	61,1621	62,4340	64,1814
39	54,5722	58,1201	61,4117	62,4281	63,7120	65,4756
40	55,7585	59,3417	62,6648	63,6907	64,9864	66,7660
45	61,6562	65,4102	68,8851	69,9568	71,3095	73,1661
50	67,5048	71,4202	75,0389	76,1539	77,5605	79,4900
60	79,0819	83,2977	87,1839	88,3794	89,8865	91,9517
80	101,8795	106,6286	110,9900	112,3288	114,0145	116,3211
100	124,3421	129,5612	134,3417	135,8067	137,6500	140,1695
120	146,5674	152,2114	157,3708	158,9502	160,9360	163,6482
200	233,9943	241,0579	247,4838	249,4451	251,9075	255,2642

4.4. Таблиця значень верхніх критичних величин t_n -розподілу Стьюдента

$$P(t_n > x) = \alpha,$$

де x – верхня критична величина рівня значущості α .

Для отримання значення верхньої критичної величини (точки, яка відтинає праворуч частку (імовірність) α більших



за величиною значень) потрібно вибрати значення кількості степенів вільності у крайньому лівому стовпчику, а у верхньому – рівень значущості. На перетині цих стовпчика та рядка наведено значення верхньої критичної величини.

Таблиця 4.4(а)

$n \backslash \alpha$	0,001	0,00125	0,0025	0,00375	0,005	0,0075	0,01	0,0125	0,025
1	318,3088	254,6466	127,3213	84,8787	63,6567	42,4335	31,8205	25,4517	12,7062
2	22,3271	19,9625	14,089	11,482	9,9248	8,0728	6,9646	6,2053	4,3027
3	10,2145	9,4649	7,4533	6,4677	5,8409	5,0473	4,5407	4,1765	3,1824
4	7,1732	6,7583	5,5976	4,9982	4,6041	4,088	3,7469	3,4954	2,7764
5	5,8934	5,6042	4,7733	4,3299	4,0321	3,6338	3,3649	3,1634	2,5706
6	5,2076	4,9807	4,3168	3,9544	3,7074	3,3723	3,1427	2,9687	2,4469
7	4,7853	4,5946	4,0293	3,7155	3,4995	3,2032	2,998	2,8412	2,3646
8	4,5008	4,3335	3,8325	3,5509	3,3554	3,0851	2,8965	2,7515	2,306
9	4,2968	4,1458	3,6897	3,4307	3,2498	2,9982	2,8214	2,685	2,2622
10	4,1437	4,0045	3,5814	3,3392	3,1693	2,9316	2,7638	2,6338	2,2281
11	4,0247	3,8945	3,4966	3,2674	3,1058	2,8789	2,7181	2,5931	2,201
12	3,9296	3,8065	3,4284	3,2094	3,0545	2,8363	2,681	2,56	2,1788
13	3,852	3,7345	3,3725	3,1618	3,0123	2,801	2,6503	2,5326	2,1604
14	3,7874	3,6746	3,3257	3,1218	2,9768	2,7714	2,6245	2,5096	2,1448
15	3,7328	3,6239	3,286	3,0879	2,9467	2,7462	2,6025	2,4899	2,1314
16	3,6862	3,5805	3,252	3,0588	2,9208	2,7245	2,5835	2,4729	2,1199

Закінчення табл. 4.4 (а)

$n \backslash \alpha$	0,001	0,00125	0,0025	0,00375	0,005	0,0075	0,01	0,0125	0,025
17	3,6458	3,5429	3,2224	3,0334	2,8982	2,7056	2,5669	2,4581	2,1098
18	3,6105	3,5101	3,1966	3,0112	2,8784	2,6889	2,5524	2,445	2,1009
19	3,5794	3,4812	3,1737	2,9916	2,8609	2,6742	2,5395	2,4334	2,093
20	3,5518	3,4554	3,1534	2,9741	2,8453	2,6611	2,528	2,4231	2,086
21	3,5272	3,4325	3,1352	2,9585	2,8314	2,6493	2,5176	2,4138	2,0796
22	3,505	3,4118	3,1188	2,9443	2,8188	2,6387	2,5083	2,4055	2,0739
23	3,485	3,3931	3,104	2,9316	2,8073	2,629	2,4999	2,3979	2,0687
24	3,4668	3,3761	3,0905	2,9199	2,7969	2,6203	2,4922	2,3909	2,0639
25	3,4502	3,3606	3,0782	2,9093	2,7874	2,6122	2,4851	2,3846	2,0595
26	3,435	3,3464	3,0669	2,8996	2,7787	2,6049	2,4786	2,3788	2,0555
27	3,421	3,3334	3,0565	2,8906	2,7707	2,5981	2,4727	2,3734	2,0518
28	3,4082	3,3214	3,0469	2,8823	2,7633	2,5918	2,4671	2,3685	2,0484
29	3,3962	3,3102	3,038	2,8746	2,7564	2,586	2,462	2,3638	2,0452
30	3,3852	3,2999	3,0298	2,8675	2,75	2,5806	2,4573	2,3596	2,0423
31	3,3749	3,2903	3,0221	2,8608	2,744	2,5755	2,4528	2,3556	2,0395
32	3,3653	3,2813	3,0149	2,8546	2,7385	2,5708	2,4487	2,3518	2,0369
33	3,3563	3,2729	3,0082	2,8488	2,7333	2,5664	2,4448	2,3483	2,0345
34	3,3479	3,2651	3,002	2,8434	2,7284	2,5623	2,4411	2,3451	2,0322
35	3,34	3,2577	2,996	2,8383	2,7238	2,5584	2,4377	2,342	2,0301
36	3,3326	3,2507	2,9905	2,8334	2,7195	2,5547	2,4345	2,3391	2,0281
37	3,3256	3,2442	2,9852	2,8289	2,7154	2,5513	2,4314	2,3363	2,0262
38	3,319	3,238	2,9803	2,8246	2,7116	2,548	2,4286	2,3337	2,0244
39	3,3128	3,2322	2,9756	2,8206	2,7079	2,5449	2,4258	2,3313	2,0227
40	3,3069	3,2266	2,9712	2,8167	2,7045	2,542	2,4233	2,3289	2,0211
45	3,2815	3,2028	2,9521	2,8001	2,6896	2,5294	2,4121	2,3189	2,0141
50	3,2614	3,184	2,937	2,787	2,6778	2,5193	2,4033	2,3109	2,0086
60	3,2317	3,1562	2,9146	2,7675	2,6603	2,5044	2,3901	2,299	2,0003
80	3,1953	3,122	2,887	2,7435	2,6387	2,486	2,3739	2,2844	1,9901
100	3,1737	3,1018	2,8707	2,7293	2,6259	2,4751	2,3642	2,2757	1,984
120	3,1595	3,0885	2,8599	2,7199	2,6174	2,4679	2,3578	2,2699	1,9799
200	3,1315	3,0621	2,8385	2,7013	2,6006	2,4536	2,3451	2,2584	1,9719
∞	3,0902	3,0233	2,807	2,6738	2,5758	2,4324	2,3264	2,2414	1,96

Таблица 4.4 (б)

$n \backslash \alpha$	0,05	0,075	0,1	0,125	0,2	0,25	0,375	0,4
1	6,3138	4,1653	3,0777	2,4142	1,3764	1	0,4142	0,3249
2	2,92	2,2819	1,8856	1,6036	1,0607	0,8165	0,3651	0,2887
3	2,3534	1,9243	1,6377	1,4226	0,9785	0,7649	0,3492	0,2767
4	2,1318	1,7782	1,5332	1,3444	0,941	0,7407	0,3414	0,2707
5	2,015	1,6994	1,4759	1,3009	0,9195	0,7267	0,3367	0,2672
6	1,9432	1,6502	1,4398	1,2733	0,9057	0,7176	0,3336	0,2648
7	1,8946	1,6166	1,4149	1,2543	0,896	0,7111	0,3315	0,2632
8	1,8595	1,5922	1,3968	1,2403	0,8889	0,7064	0,3298	0,2619
9	1,8331	1,5737	1,383	1,2297	0,8834	0,7027	0,3286	0,261
10	1,8125	1,5592	1,3722	1,2213	0,8791	0,6998	0,3276	0,2602
11	1,7959	1,5476	1,3634	1,2145	0,8755	0,6974	0,3267	0,2596
12	1,7823	1,538	1,3562	1,2089	0,8726	0,6955	0,3261	0,259
13	1,7709	1,5299	1,3502	1,2041	0,8702	0,6938	0,3255	0,2586
14	1,7613	1,5231	1,345	1,2001	0,8681	0,6924	0,325	0,2582
15	1,7531	1,5172	1,3406	1,1967	0,8662	0,6912	0,3246	0,2579
16	1,7459	1,5121	1,3368	1,1937	0,8647	0,6901	0,3242	0,2576
17	1,7396	1,5077	1,3334	1,191	0,8633	0,6892	0,3239	0,2573
18	1,7341	1,5037	1,3304	1,1887	0,862	0,6884	0,3236	0,2571
19	1,7291	1,5002	1,3277	1,1866	0,861	0,6876	0,3233	0,2569
20	1,7247	1,497	1,3253	1,1848	0,86	0,687	0,3231	0,2567
21	1,7207	1,4942	1,3232	1,1831	0,8591	0,6864	0,3229	0,2566
22	1,7171	1,4916	1,3212	1,1815	0,8583	0,6858	0,3227	0,2564
23	1,7139	1,4893	1,3195	1,1802	0,8575	0,6853	0,3225	0,2563
24	1,7109	1,4871	1,3178	1,1789	0,8569	0,6848	0,3223	0,2562
25	1,7081	1,4852	1,3163	1,1777	0,8562	0,6844	0,3222	0,2561
26	1,7056	1,4834	1,315	1,1766	0,8557	0,684	0,322	0,256
27	1,7033	1,4817	1,3137	1,1756	0,8551	0,6837	0,3219	0,2559
28	1,7011	1,4801	1,3125	1,1747	0,8546	0,6834	0,3218	0,2558
29	1,6991	1,4787	1,3114	1,1739	0,8542	0,683	0,3217	0,2557
30	1,6973	1,4774	1,3104	1,1731	0,8538	0,6828	0,3216	0,2556
31	1,6955	1,4761	1,3095	1,1723	0,8534	0,6825	0,3215	0,2555
32	1,6939	1,4749	1,3086	1,1716	0,853	0,6822	0,3214	0,2555
33	1,6924	1,4738	1,3077	1,171	0,8526	0,682	0,3213	0,2554

Закінчення табл. 4.4 (б)

$\alpha \backslash n$	0,05	0,075	0,1	0,125	0,2	0,25	0,375	0,4
34	1,6909	1,4728	1,307	1,1703	0,8523	0,6818	0,3212	0,2553
35	1,6896	1,4718	1,3062	1,1698	0,852	0,6816	0,3212	0,2553
36	1,6883	1,4709	1,3055	1,1692	0,8517	0,6814	0,3211	0,2552
37	1,6871	1,4701	1,3049	1,1687	0,8514	0,6812	0,321	0,2552
38	1,686	1,4692	1,3042	1,1682	0,8512	0,681	0,321	0,2551
39	1,6849	1,4685	1,3036	1,1677	0,8509	0,6808	0,3209	0,2551
40	1,6839	1,4677	1,3031	1,1673	0,8507	0,6807	0,3208	0,255
45	1,6794	1,4645	1,3006	1,1654	0,8497	0,68	0,3206	0,2549
50	1,6759	1,462	1,2987	1,1639	0,8489	0,6794	0,3204	0,2547
60	1,6706	1,4582	1,2958	1,1616	0,8477	0,6786	0,3201	0,2545
80	1,6641	1,4535	1,2922	1,1588	0,8461	0,6776	0,3197	0,2542
100	1,6602	1,4507	1,2901	1,1571	0,8452	0,677	0,3195	0,254
120	1,6577	1,4488	1,2886	1,1559	0,8446	0,6765	0,3194	0,2539
200	1,6525	1,4451	1,2858	1,1537	0,8434	0,6757	0,3191	0,2537
∞	1,6448	1,4395	1,2815	1,1503	0,8416	0,6745	0,3186	0,2533

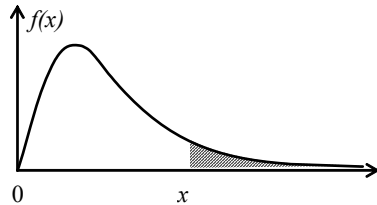
4.5. Таблиця значень верхніх критичних величин $F_{m;n}$ -розподілу Фішера

$$P(F_{m;n} > x) = \alpha,$$

де x – верхня критична величина рівня значущості α .

Для отримання значення верхньої критичної величини (точки, яка відтинає праворуч частку (імовірність) α більших

за величиною значень) потрібно вибрати спочатку таблицю, яка відповідає вказаному рівню значущості. Далі в цій таблиці обрати кількість ступенів вільності параметра n у крайньому



лівому стовпчику, а у верхньому – кількість ступенів вільності параметра m . На перетині стовпчика та рядка наведено значення верхньої критичної величини.

Якщо таблиці з шуканим рівнем значущості не виявлено, то

слід скористатись формулою $F_{m;n;\alpha} = \frac{1}{F_{n;m;1-\alpha}}$.

Таблиця 4.5.1(а)

$\alpha = 0,1$

$m \backslash n$	1	2	3	4	5	6	7	8	9	10
1	39,86	49,50	53,59	55,83	57,24	58,20	58,91	59,44	59,86	60,19
2	8,53	9,00	9,16	9,24	9,29	9,33	9,35	9,37	9,38	9,39
3	5,54	5,46	5,39	5,34	5,31	5,28	5,27	5,25	5,24	5,23
4	4,54	4,32	4,19	4,11	4,05	4,01	3,98	3,95	3,94	3,92
5	4,06	3,78	3,62	3,52	3,45	3,40	3,37	3,34	3,32	3,30
6	3,78	3,46	3,29	3,18	3,11	3,05	3,01	2,98	2,96	2,94
7	3,59	3,26	3,07	2,96	2,88	2,83	2,78	2,75	2,72	2,70
8	3,46	3,11	2,92	2,81	2,73	2,67	2,62	2,59	2,56	2,54
9	3,36	3,01	2,81	2,69	2,61	2,55	2,51	2,47	2,44	2,42
10	3,29	2,92	2,73	2,61	2,52	2,46	2,41	2,38	2,35	2,32
11	3,23	2,86	2,66	2,54	2,45	2,39	2,34	2,30	2,27	2,25
12	3,18	2,81	2,61	2,48	2,39	2,33	2,28	2,24	2,21	2,19
13	3,14	2,76	2,56	2,43	2,35	2,28	2,23	2,20	2,16	2,14
14	3,10	2,73	2,52	2,39	2,31	2,24	2,19	2,15	2,12	2,10
15	3,07	2,70	2,49	2,36	2,27	2,21	2,16	2,12	2,09	2,06
16	3,05	2,67	2,46	2,33	2,24	2,18	2,13	2,09	2,06	2,03
17	3,03	2,64	2,44	2,31	2,22	2,15	2,10	2,06	2,03	2,00
18	3,01	2,62	2,42	2,29	2,20	2,13	2,08	2,04	2,00	1,98
19	2,99	2,61	2,40	2,27	2,18	2,11	2,06	2,02	1,98	1,96
20	2,97	2,59	2,38	2,25	2,16	2,09	2,04	2,00	1,96	1,94
21	2,96	2,57	2,36	2,23	2,14	2,08	2,02	1,98	1,95	1,92
22	2,95	2,56	2,35	2,22	2,13	2,06	2,01	1,97	1,93	1,90

Закінчення табл. 45.2 (а)

$\begin{matrix} m \\ n \end{matrix}$	1	2	3	4	5	6	7	8	9	10
23	2,94	2,55	2,34	2,21	2,11	2,05	1,99	1,95	1,92	1,89
24	2,93	2,54	2,33	2,19	2,10	2,04	1,98	1,94	1,91	1,88
25	2,92	2,53	2,32	2,18	2,09	2,02	1,97	1,93	1,89	1,87
26	2,91	2,52	2,31	2,17	2,08	2,01	1,96	1,92	1,88	1,86
27	2,90	2,51	2,30	2,17	2,07	2,00	1,95	1,91	1,87	1,85
28	2,89	2,50	2,29	2,16	2,06	2,00	1,94	1,90	1,87	1,84
29	2,89	2,50	2,28	2,15	2,06	1,99	1,93	1,89	1,86	1,83
30	2,88	2,49	2,28	2,14	2,05	1,98	1,93	1,88	1,85	1,82
40	2,84	2,44	2,23	2,09	2,00	1,93	1,87	1,83	1,79	1,76
60	2,79	2,39	2,18	2,04	1,95	1,87	1,82	1,77	1,74	1,71
120	2,75	2,35	2,13	1,99	1,90	1,82	1,77	1,72	1,68	1,65
∞	2,71	2,30	2,08	1,94	1,85	1,77	1,72	1,67	1,63	1,60

Таблиця 4.5.1 (б)

$\begin{matrix} m \\ n \end{matrix}$	11	12	14	17	20	24	30	40	60	120	∞
1	60,47	60,71	61,07	61,46	61,74	62,00	62,26	62,53	62,79	63,06	63,33
2	9,40	9,41	9,42	9,43	9,44	9,45	9,46	9,47	9,47	9,48	9,49
3	5,22	5,22	5,20	5,19	5,18	5,18	5,17	5,16	5,15	5,14	5,13
4	3,91	3,90	3,88	3,86	3,84	3,83	3,82	3,80	3,79	3,78	3,76
5	3,28	3,27	3,25	3,22	3,21	3,19	3,17	3,16	3,14	3,12	3,11
6	2,92	2,90	2,88	2,85	2,84	2,82	2,80	2,78	2,76	2,74	2,72
7	2,68	2,67	2,64	2,61	2,59	2,58	2,56	2,54	2,51	2,49	2,47
8	2,52	2,50	2,48	2,45	2,42	2,40	2,38	2,36	2,34	2,32	2,29
9	2,40	2,38	2,35	2,32	2,30	2,28	2,25	2,23	2,21	2,18	2,16
10	2,30	2,28	2,26	2,22	2,20	2,18	2,16	2,13	2,11	2,08	2,06
11	2,23	2,21	2,18	2,15	2,12	2,10	2,08	2,05	2,03	2,00	1,97
12	2,17	2,15	2,12	2,08	2,06	2,04	2,01	1,99	1,96	1,93	1,90
13	2,12	2,10	2,07	2,03	2,01	1,98	1,96	1,93	1,90	1,88	1,85
14	2,07	2,05	2,02	1,99	1,96	1,94	1,91	1,89	1,86	1,83	1,80

Закінчення табл. 4.5.1 (б)

$\begin{matrix} m \\ n \end{matrix}$	11	12	14	17	20	24	30	40	60	120	∞
15	2,04	2,02	1,99	1,95	1,92	1,90	1,87	1,85	1,82	1,79	1,76
16	2,01	1,99	1,95	1,92	1,89	1,87	1,84	1,81	1,78	1,75	1,72
17	1,98	1,96	1,93	1,89	1,86	1,84	1,81	1,78	1,75	1,72	1,69
18	1,95	1,93	1,90	1,86	1,84	1,81	1,78	1,75	1,72	1,69	1,66
19	1,93	1,91	1,88	1,84	1,81	1,79	1,76	1,73	1,70	1,67	1,63
20	1,91	1,89	1,86	1,82	1,79	1,77	1,74	1,71	1,68	1,64	1,61
21	1,90	1,87	1,84	1,80	1,78	1,75	1,72	1,69	1,66	1,62	1,59
22	1,88	1,86	1,83	1,79	1,76	1,73	1,70	1,67	1,64	1,60	1,57
23	1,87	1,84	1,81	1,77	1,74	1,72	1,69	1,66	1,62	1,59	1,55
24	1,85	1,83	1,80	1,76	1,73	1,70	1,67	1,64	1,61	1,57	1,53
25	1,84	1,82	1,79	1,75	1,72	1,69	1,66	1,63	1,59	1,56	1,52
26	1,83	1,81	1,77	1,73	1,71	1,68	1,65	1,61	1,58	1,54	1,50
27	1,82	1,80	1,76	1,72	1,70	1,67	1,64	1,60	1,57	1,53	1,49
28	1,81	1,79	1,75	1,71	1,69	1,66	1,63	1,59	1,56	1,52	1,48
29	1,80	1,78	1,75	1,71	1,68	1,65	1,62	1,58	1,55	1,51	1,47
30	1,79	1,77	1,74	1,70	1,67	1,64	1,61	1,57	1,54	1,50	1,46
40	1,74	1,71	1,68	1,64	1,61	1,57	1,54	1,51	1,47	1,42	1,38
60	1,68	1,66	1,62	1,58	1,54	1,51	1,48	1,44	1,40	1,35	1,29
120	1,63	1,60	1,56	1,52	1,48	1,45	1,41	1,37	1,32	1,26	1,19
∞	1,57	1,55	1,50	1,46	1,42	1,38	1,34	1,30	1,24	1,17	1,01

Таблиця 4.5.2 (а)

$\alpha = 0,075$

$\begin{matrix} m \\ n \end{matrix}$	1	2	3	4	5	6	7	8	9	10
1	71,38	88,39	95,62	99,58	102,07	103,78	105,02	105,96	106,7	107,3
2	11,85	12,33	12,5	12,58	12,63	12,66	12,68	12,7	12,72	12,73
3	7,19	6,93	6,79	6,7	6,64	6,6	6,57	6,55	6,53	6,51
4	5,72	5,3	5,09	4,96	4,88	4,82	4,77	4,74	4,71	4,68
5	5,03	4,55	4,3	4,16	4,06	4	3,95	3,91	3,87	3,85

Закінчення табл. 4.5.3 (б)

$n \backslash m$	1	2	3	4	5	6	7	8	9	10
6	4,63	4,11	3,86	3,71	3,6	3,53	3,48	3,44	3,4	3,37
7	4,37	3,84	3,57	3,42	3,31	3,24	3,18	3,13	3,1	3,07
8	4,19	3,64	3,38	3,21	3,11	3,03	2,97	2,93	2,89	2,86
9	4,05	3,5	3,23	3,07	2,96	2,88	2,82	2,77	2,74	2,7
10	3,95	3,39	3,12	2,95	2,84	2,76	2,7	2,66	2,62	2,59
11	3,87	3,31	3,03	2,87	2,75	2,67	2,61	2,56	2,52	2,49
12	3,8	3,24	2,96	2,79	2,68	2,6	2,54	2,49	2,45	2,42
13	3,75	3,18	2,9	2,74	2,62	2,54	2,48	2,43	2,39	2,35
14	3,7	3,13	2,85	2,69	2,57	2,49	2,43	2,38	2,34	2,3
15	3,66	3,09	2,81	2,64	2,53	2,45	2,38	2,33	2,29	2,26
16	3,63	3,06	2,78	2,61	2,49	2,41	2,34	2,29	2,25	2,22
17	3,6	3,03	2,75	2,58	2,46	2,38	2,31	2,26	2,22	2,18
18	3,57	3	2,72	2,55	2,43	2,35	2,28	2,23	2,19	2,16
19	3,55	2,98	2,7	2,52	2,41	2,32	2,26	2,21	2,16	2,13
20	3,53	2,96	2,67	2,5	2,38	2,3	2,23	2,18	2,14	2,11
21	3,51	2,94	2,65	2,48	2,37	2,28	2,21	2,16	2,12	2,09
22	3,49	2,92	2,64	2,46	2,35	2,26	2,2	2,14	2,1	2,07
23	3,48	2,91	2,62	2,45	2,33	2,25	2,18	2,13	2,08	2,05
24	3,46	2,89	2,61	2,43	2,32	2,23	2,16	2,11	2,07	2,03
25	3,45	2,88	2,59	2,42	2,3	2,22	2,15	2,1	2,06	2,02
26	3,44	2,87	2,58	2,41	2,29	2,2	2,14	2,09	2,04	2,01
27	3,43	2,86	2,57	2,4	2,28	2,19	2,13	2,07	2,03	1,99
28	3,42	2,85	2,56	2,39	2,27	2,18	2,12	2,06	2,02	1,98
29	3,41	2,84	2,55	2,38	2,26	2,17	2,11	2,05	2,01	1,97
30	3,4	2,83	2,54	2,37	2,25	2,16	2,1	2,04	2	1,96
40	3,34	2,77	2,48	2,3	2,18	2,1	2,03	1,98	1,93	1,89
60	3,28	2,71	2,42	2,24	2,12	2,03	1,96	1,91	1,86	1,83
120	3,23	2,65	2,36	2,18	2,06	1,97	1,9	1,85	1,8	1,76
∞	3,17	2,59	2,3	2,12	2	1,91	1,84	1,78	1,74	1,7

Таблица 4.5.2 (б)

$m \backslash n$	11	12	14	17	20	24	30	40	60	120	∞
1	107,79	108,2	108,85	109,55	110,03	110,5	110,96	111,43	111,9	112,37	112,84
2	12,74	12,74	12,76	12,77	12,78	12,79	12,79	12,8	12,81	12,82	12,83
3	6,5	6,49	6,47	6,45	6,44	6,42	6,41	6,4	6,38	6,37	6,36
4	4,67	4,65	4,62	4,6	4,58	4,56	4,54	4,52	4,5	4,48	4,46
5	3,83	3,81	3,78	3,75	3,72	3,7	3,68	3,66	3,63	3,61	3,59
6	3,35	3,33	3,3	3,26	3,24	3,22	3,19	3,17	3,14	3,12	3,09
7	3,05	3,02	2,99	2,96	2,93	2,9	2,88	2,85	2,83	2,8	2,77
8	2,83	2,81	2,78	2,74	2,71	2,69	2,66	2,63	2,6	2,57	2,54
9	2,68	2,66	2,62	2,58	2,55	2,53	2,5	2,47	2,44	2,41	2,38
10	2,56	2,54	2,5	2,46	2,43	2,4	2,37	2,34	2,31	2,28	2,25
11	2,47	2,44	2,4	2,36	2,33	2,31	2,28	2,25	2,21	2,18	2,15
12	2,39	2,37	2,33	2,29	2,26	2,23	2,2	2,16	2,13	2,1	2,06
13	2,33	2,3	2,26	2,22	2,19	2,16	2,13	2,1	2,06	2,03	1,99
14	2,27	2,25	2,21	2,17	2,14	2,11	2,07	2,04	2,01	1,97	1,93
15	2,23	2,2	2,16	2,12	2,09	2,06	2,03	1,99	1,96	1,92	1,88
16	2,19	2,17	2,13	2,08	2,05	2,02	1,98	1,95	1,91	1,88	1,84
17	2,16	2,13	2,09	2,05	2,01	1,98	1,95	1,91	1,88	1,84	1,8
18	2,13	2,1	2,06	2,02	1,98	1,95	1,92	1,88	1,84	1,81	1,76
19	2,1	2,07	2,03	1,99	1,95	1,92	1,89	1,85	1,81	1,77	1,73
20	2,08	2,05	2,01	1,96	1,93	1,9	1,86	1,83	1,79	1,75	1,7
21	2,06	2,03	1,99	1,94	1,91	1,87	1,84	1,8	1,76	1,72	1,68
22	2,04	2,01	1,97	1,92	1,89	1,85	1,82	1,78	1,74	1,7	1,66
23	2,02	1,99	1,95	1,9	1,87	1,84	1,8	1,76	1,72	1,68	1,64
24	2	1,98	1,94	1,89	1,85	1,82	1,78	1,74	1,7	1,66	1,62
25	1,99	1,96	1,92	1,87	1,84	1,8	1,77	1,73	1,69	1,64	1,6
26	1,98	1,95	1,91	1,86	1,82	1,79	1,75	1,71	1,67	1,63	1,58
27	1,96	1,94	1,89	1,85	1,81	1,78	1,74	1,7	1,66	1,61	1,57
28	1,95	1,93	1,88	1,83	1,8	1,76	1,73	1,69	1,65	1,6	1,55
29	107,79	108,2	108,85	109,55	110,03	110,5	110,96	111,43	111,9	112,37	112,84
30	1,94	1,92	1,87	1,82	1,79	1,75	1,72	1,68	1,63	1,59	1,54

Закінчення табл. 4.5.2 (б)

$n \backslash m$	11	12	14	17	20	24	30	40	60	120	∞
40	1,93	1,91	1,86	1,81	1,78	1,74	1,7	1,66	1,62	1,58	1,53
60	1,86	1,84	1,79	1,74	1,7	1,67	1,63	1,58	1,54	1,49	1,43
120	1,79	1,77	1,72	1,67	1,63	1,59	1,55	1,5	1,45	1,4	1,33
∞	1,73	1,7	1,65	1,6	1,56	1,52	1,47	1,42	1,37	1,3	1,22

Таблиця 4.5.3 (а)

 $\alpha = 0,05$

$n \backslash m$	1	2	3	4	5	6	7	8	9	10
1	161,45	199,5	215,7	224,58	230,16	233,99	236,77	238,88	240,54	241,88
2	18,51	19,00	19,16	19,25	19,30	19,33	19,35	19,37	19,38	19,40
3	10,13	9,55	9,28	9,12	9,01	8,94	8,89	8,85	8,81	8,79
4	7,71	6,94	6,59	6,39	6,26	6,16	6,09	6,04	6,00	5,96
5	6,61	5,79	5,41	5,19	5,05	4,95	4,88	4,82	4,77	4,74
6	5,99	5,14	4,76	4,53	4,39	4,28	4,21	4,15	4,10	4,06
7	5,59	4,74	4,35	4,12	3,97	3,87	3,79	3,73	3,68	3,64
8	5,32	4,46	4,07	3,84	3,69	3,58	3,50	3,44	3,39	3,35
9	5,12	4,26	3,86	3,63	3,48	3,37	3,29	3,23	3,18	3,14
10	4,96	4,10	3,71	3,48	3,33	3,22	3,14	3,07	3,02	2,98
11	4,84	3,98	3,59	3,36	3,20	3,09	3,01	2,95	2,90	2,85
12	4,75	3,89	3,49	3,26	3,11	3,00	2,91	2,85	2,80	2,75
13	4,67	3,81	3,41	3,18	3,03	2,92	2,83	2,77	2,71	2,67
14	4,60	3,74	3,34	3,11	2,96	2,85	2,76	2,70	2,65	2,60
15	4,54	3,68	3,29	3,06	2,90	2,79	2,71	2,64	2,59	2,54
16	4,49	3,63	3,24	3,01	2,85	2,74	2,66	2,59	2,54	2,49
17	4,45	3,59	3,20	2,96	2,81	2,70	2,61	2,55	2,49	2,45
18	4,41	3,55	3,16	2,93	2,77	2,66	2,58	2,51	2,46	2,41
19	4,38	3,52	3,13	2,90	2,74	2,63	2,54	2,48	2,42	2,38
20	4,35	3,49	3,10	2,87	2,71	2,60	2,51	2,45	2,39	2,35
21	4,32	3,47	3,07	2,84	2,68	2,57	2,49	2,42	2,37	2,32
22	4,30	3,44	3,05	2,82	2,66	2,55	2,46	2,40	2,34	2,30

Закінчення табл. 4.5.3 (а)

$n \backslash m$	1	2	3	4	5	6	7	8	9	10
23	4,28	3,42	3,03	2,80	2,64	2,53	2,44	2,37	2,32	2,27
24	4,26	3,40	3,01	2,78	2,62	2,51	2,42	2,36	2,30	2,25
25	4,24	3,39	2,99	2,76	2,60	2,49	2,40	2,34	2,28	2,24
26	4,23	3,37	2,98	2,74	2,59	2,47	2,39	2,32	2,27	2,22
27	4,21	3,35	2,96	2,73	2,57	2,46	2,37	2,31	2,25	2,20
28	4,20	3,34	2,95	2,71	2,56	2,45	2,36	2,29	2,24	2,19
29	4,18	3,33	2,93	2,70	2,55	2,43	2,35	2,28	2,22	2,18
30	4,17	3,32	2,92	2,69	2,53	2,42	2,33	2,27	2,21	2,16
40	4,08	3,23	2,84	2,61	2,45	2,34	2,25	2,18	2,12	2,08
60	4,00	3,15	2,76	2,53	2,37	2,25	2,17	2,10	2,04	1,99
120	3,92	3,07	2,68	2,45	2,29	2,18	2,09	2,02	1,96	1,91
∞	3,84	3,00	2,61	2,37	2,21	2,10	2,01	1,94	1,88	1,83

Таблиця 4.5.3 (б)

$n \backslash m$	11	12	14	17	20	24	30	40	60	120	∞
1	243	243,9	245,4	246,9	248,02	249,1	250,1	251,1	252,2	253,3	254,3
2	19,4	19,41	19,42	19,44	19,45	19,45	19,46	19,47	19,48	19,49	19,5
3	8,76	8,74	8,71	8,68	8,66	8,64	8,62	8,59	8,57	8,55	8,53
4	5,94	5,91	5,87	5,83	5,80	5,77	5,75	5,72	5,69	5,66	5,63
5	4,70	4,68	4,64	4,59	4,56	4,53	4,50	4,46	4,43	4,40	4,37
6	4,03	4,00	3,96	3,91	3,87	3,84	3,81	3,77	3,74	3,70	3,67
7	3,60	3,57	3,53	3,48	3,44	3,41	3,38	3,34	3,30	3,27	3,23
8	3,31	3,28	3,24	3,19	3,15	3,12	3,08	3,04	3,01	2,97	2,93
9	3,10	3,07	3,03	2,97	2,94	2,90	2,86	2,83	2,79	2,75	2,71
10	2,94	2,91	2,86	2,81	2,77	2,74	2,70	2,66	2,62	2,58	2,54
11	2,82	2,79	2,74	2,69	2,65	2,61	2,57	2,53	2,49	2,45	2,40
12	2,72	2,69	2,64	2,58	2,54	2,51	2,47	2,43	2,38	2,34	2,30
13	2,63	2,60	2,55	2,50	2,46	2,42	2,38	2,34	2,30	2,25	2,21
14	2,57	2,53	2,48	2,43	2,39	2,35	2,31	2,27	2,22	2,18	2,13

Закінчення табл. 4.5.3 (б)

$n \backslash m$	11	12	14	17	20	24	30	40	60	120	∞
15	2,51	2,48	2,42	2,37	2,33	2,29	2,25	2,20	2,16	2,11	2,07
16	2,46	2,42	2,37	2,32	2,28	2,24	2,19	2,15	2,11	2,06	2,01
17	2,41	2,38	2,33	2,27	2,23	2,19	2,15	2,10	2,06	2,01	1,96
18	2,37	2,34	2,29	2,23	2,19	2,15	2,11	2,06	2,02	1,97	1,92
19	2,34	2,31	2,26	2,20	2,16	2,11	2,07	2,03	1,98	1,93	1,88
20	2,31	2,28	2,22	2,17	2,12	2,08	2,04	1,99	1,95	1,90	1,84
21	2,28	2,25	2,20	2,14	2,10	2,05	2,01	1,96	1,92	1,87	1,81
22	2,26	2,23	2,17	2,11	2,07	2,03	1,98	1,94	1,89	1,84	1,78
23	2,24	2,20	2,15	2,09	2,05	2,01	1,96	1,91	1,86	1,81	1,76
24	2,22	2,18	2,13	2,07	2,03	1,98	1,94	1,89	1,84	1,79	1,73
25	2,20	2,16	2,11	2,05	2,01	1,96	1,92	1,87	1,82	1,77	1,71
26	2,18	2,15	2,09	2,03	1,99	1,95	1,90	1,85	1,80	1,75	1,69
27	2,17	2,13	2,08	2,02	1,97	1,93	1,88	1,84	1,79	1,73	1,67
28	2,15	2,12	2,06	2,00	1,96	1,91	1,87	1,82	1,77	1,71	1,65
29	2,14	2,10	2,05	1,99	1,94	1,90	1,85	1,81	1,75	1,70	1,64
30	2,13	2,09	2,04	1,98	1,93	1,89	1,84	1,79	1,74	1,68	1,62
40	2,04	2,00	1,95	1,89	1,84	1,79	1,74	1,69	1,64	1,58	1,51
60	1,95	1,92	1,86	1,80	1,75	1,70	1,65	1,59	1,53	1,47	1,39
120	1,87	1,83	1,78	1,71	1,66	1,61	1,55	1,50	1,43	1,35	1,25
∞	1,79	1,75	1,69	1,62	1,57	1,52	1,46	1,39	1,32	1,22	1,01

Таблиця 4.5.4 (а)

 $\alpha = 0,025$

$n \backslash m$	1	2	3	4	5	6	7	8	9	10
1	647,8	799,5	864,15	899,6	921,83	937,1	948,2	956,64	963,28	968,63
2	38,51	39,00	39,17	39,25	39,30	39,33	39,36	39,37	39,39	39,40
3	17,44	16,04	15,44	15,10	14,88	14,73	14,62	14,54	14,47	14,42
4	12,22	10,65	9,98	9,60	9,36	9,20	9,07	8,98	8,90	8,84
5	10,01	8,43	7,76	7,39	7,15	6,98	6,85	6,76	6,68	6,62

Закінчення табл. 4.5.4 (а)

$n \backslash m$	1	2	3	4	5	6	7	8	9	10
6	8,81	7,26	6,60	6,23	5,99	5,82	5,70	5,60	5,52	5,46
7	8,07	6,54	5,89	5,52	5,29	5,12	4,99	4,90	4,82	4,76
8	7,57	6,06	5,42	5,05	4,82	4,65	4,53	4,43	4,36	4,30
9	7,21	5,71	5,08	4,72	4,48	4,32	4,20	4,10	4,03	3,96
10	6,94	5,46	4,83	4,47	4,24	4,07	3,95	3,85	3,78	3,72
11	6,72	5,26	4,63	4,28	4,04	3,88	3,76	3,66	3,59	3,53
12	6,55	5,10	4,47	4,12	3,89	3,73	3,61	3,51	3,44	3,37
13	6,41	4,97	4,35	4,00	3,77	3,60	3,48	3,39	3,31	3,25
14	6,30	4,86	4,24	3,89	3,66	3,50	3,38	3,29	3,21	3,15
15	6,20	4,77	4,15	3,80	3,58	3,41	3,29	3,20	3,12	3,06
16	6,12	4,69	4,08	3,73	3,50	3,34	3,22	3,12	3,05	2,99
17	6,04	4,62	4,01	3,66	3,44	3,28	3,16	3,06	2,98	2,92
18	5,98	4,56	3,95	3,61	3,38	3,22	3,10	3,01	2,93	2,87
19	5,92	4,51	3,90	3,56	3,33	3,17	3,05	2,96	2,88	2,82
20	5,87	4,46	3,86	3,51	3,29	3,13	3,01	2,91	2,84	2,77
21	5,83	4,42	3,82	3,48	3,25	3,09	2,97	2,87	2,80	2,73
22	5,79	4,38	3,78	3,44	3,22	3,05	2,93	2,84	2,76	2,70
23	5,75	4,35	3,75	3,41	3,18	3,02	2,90	2,81	2,73	2,67
24	5,72	4,32	3,72	3,38	3,15	2,99	2,87	2,78	2,70	2,64
25	5,69	4,29	3,69	3,35	3,13	2,97	2,85	2,75	2,68	2,61
26	5,66	4,27	3,67	3,33	3,10	2,94	2,82	2,73	2,65	2,59
27	5,63	4,24	3,65	3,31	3,08	2,92	2,80	2,71	2,63	2,57
28	5,61	4,22	3,63	3,29	3,06	2,90	2,78	2,69	2,61	2,55
29	5,59	4,20	3,61	3,27	3,04	2,88	2,76	2,67	2,59	2,53
30	5,57	4,18	3,59	3,25	3,03	2,87	2,75	2,65	2,57	2,51
40	5,42	4,05	3,46	3,13	2,90	2,74	2,62	2,53	2,45	2,39
60	5,29	3,93	3,34	3,01	2,79	2,63	2,51	2,41	2,33	2,27
120	5,15	3,80	3,23	2,89	2,67	2,52	2,39	2,30	2,22	2,16
∞	5,02	3,69	3,12	2,79	2,57	2,41	2,29	2,19	2,11	2,05

Таблица 4.5.4 (б)

$\begin{matrix} m \\ n \end{matrix}$	11	12	14	17	20	24	30	40	60	120	∞
1	973	976,7	982,55	988,7	993,1	997,27	1001,4	1005,6	1009,8	1014	1018
2	39,41	39,41	39,43	39,44	39,45	39,46	39,46	39,47	39,48	39,49	39,5
3	14,37	14,34	14,28	14,21	14,17	14,12	14,08	14,04	13,99	13,95	13,9
4	8,79	8,75	8,68	8,61	8,56	8,51	8,46	8,41	8,36	8,31	8,26
5	6,57	6,52	6,46	6,38	6,33	6,28	6,23	6,18	6,12	6,07	6,02
6	5,41	5,37	5,30	5,22	5,17	5,12	5,07	5,01	4,96	4,90	4,85
7	4,71	4,67	4,60	4,52	4,47	4,41	4,36	4,31	4,25	4,20	4,14
8	4,24	4,20	4,13	4,05	4,00	3,95	3,89	3,84	3,78	3,73	3,67
9	3,91	3,87	3,80	3,72	3,67	3,61	3,56	3,51	3,45	3,39	3,33
10	3,66	3,62	3,55	3,47	3,42	3,37	3,31	3,26	3,20	3,14	3,08
11	3,47	3,43	3,36	3,28	3,23	3,17	3,12	3,06	3,00	2,94	2,88
12	3,32	3,28	3,21	3,13	3,07	3,02	2,96	2,91	2,85	2,79	2,73
13	3,20	3,15	3,08	3,00	2,95	2,89	2,84	2,78	2,72	2,66	2,60
14	3,09	3,05	2,98	2,90	2,84	2,79	2,73	2,67	2,61	2,55	2,49
15	3,01	2,96	2,89	2,81	2,76	2,70	2,64	2,59	2,52	2,46	2,40
16	2,93	2,89	2,82	2,74	2,68	2,63	2,57	2,51	2,45	2,38	2,32
17	2,87	2,82	2,75	2,67	2,62	2,56	2,50	2,44	2,38	2,32	2,25
18	2,81	2,77	2,70	2,62	2,56	2,50	2,44	2,38	2,32	2,26	2,19
19	2,76	2,72	2,65	2,57	2,51	2,45	2,39	2,33	2,27	2,20	2,13
20	2,72	2,68	2,60	2,52	2,46	2,41	2,35	2,29	2,22	2,16	2,09
21	2,68	2,64	2,56	2,48	2,42	2,37	2,31	2,25	2,18	2,11	2,04
22	2,65	2,60	2,53	2,45	2,39	2,33	2,27	2,21	2,14	2,08	2,00
23	2,62	2,57	2,50	2,42	2,36	2,30	2,24	2,18	2,11	2,04	1,97
24	2,59	2,54	2,47	2,39	2,33	2,27	2,21	2,15	2,08	2,01	1,94
25	2,56	2,51	2,44	2,36	2,30	2,24	2,18	2,12	2,05	1,98	1,91
26	2,54	2,49	2,42	2,34	2,28	2,22	2,16	2,09	2,03	1,95	1,88
27	2,51	2,47	2,39	2,31	2,25	2,19	2,13	2,07	2,00	1,93	1,85
28	2,49	2,45	2,37	2,29	2,23	2,17	2,11	2,05	1,98	1,91	1,83
29	2,48	2,43	2,36	2,27	2,21	2,15	2,09	2,03	1,96	1,89	1,81

Закінчення табл. 4.5.4 (б)

$\begin{matrix} m \\ n \end{matrix}$	11	12	14	17	20	24	30	40	60	120	∞
30	2,46	2,41	2,34	2,26	2,20	2,14	2,07	2,01	1,94	1,87	1,79
40	2,33	2,29	2,21	2,13	2,07	2,01	1,94	1,88	1,80	1,72	1,64
60	2,22	2,17	2,09	2,01	1,94	1,88	1,82	1,74	1,67	1,58	1,48
120	2,10	2,05	1,98	1,89	1,82	1,76	1,69	1,61	1,53	1,43	1,31
∞	1,99	1,94	1,87	1,78	1,71	1,64	1,57	1,48	1,39	1,27	1,02

Таблиця 4.5.5(а)

 $\alpha = 0,0125$

$\begin{matrix} m \\ n \end{matrix}$	1	2	3	4	5	6	7	8	9	10
1	2593,2	3199,5	3458	3599,6	3688,6	3749,6	3794	3827,8	3854,3	3875,6
2	78,5	79	79,17	79,25	79,3	79,33	79,36	79,37	79,39	79,4
3	29,07	26,35	25,22	24,6	24,2	23,93	23,73	23,57	23,45	23,36
4	18,62	15,89	14,77	14,15	13,75	13,48	13,28	13,13	13,01	12,91
5	14,52	11,93	10,86	10,28	9,9	9,64	9,45	9,31	9,19	9,1
6	12,4	9,93	8,91	8,35	8	7,75	7,56	7,42	7,31	7,22
7	11,12	8,74	7,77	7,22	6,88	6,64	6,46	6,32	6,21	6,12
8	10,28	7,96	7,02	6,49	6,15	5,92	5,74	5,61	5,5	5,41
9	9,68	7,42	6,49	5,98	5,65	5,41	5,24	5,11	5	4,92
10	9,23	7,01	6,1	5,6	5,27	5,04	4,88	4,74	4,64	4,56
11	8,89	6,7	5,81	5,31	4,99	4,76	4,6	4,47	4,36	4,28
12	8,61	6,45	5,57	5,08	4,76	4,54	4,37	4,25	4,14	4,06
13	8,39	6,26	5,38	4,9	4,58	4,36	4,2	4,07	3,97	3,88
14	8,2	6,09	5,23	4,74	4,43	4,21	4,05	3,92	3,82	3,74
15	8,05	5,95	5,1	4,62	4,31	4,09	3,93	3,8	3,7	3,62
16	7,91	5,83	4,98	4,51	4,2	3,98	3,82	3,7	3,6	3,51
17	7,8	5,73	4,89	4,42	4,11	3,89	3,73	3,61	3,51	3,42
18	7,7	5,65	4,8	4,33	4,03	3,82	3,65	3,53	3,43	3,35
19	7,61	5,57	4,73	4,26	3,96	3,75	3,59	3,46	3,36	3,28
20	7,53	5,5	4,67	4,2	3,9	3,69	3,53	3,4	3,3	3,22
21	7,46	5,44	4,61	4,15	3,84	3,63	3,47	3,35	3,25	3,17

Закінчення табл. 4.5.5 (а)

$\begin{matrix} m \\ n \end{matrix}$	1	2	3	4	5	6	7	8	9	10
22	7,4	5,38	4,56	4,1	3,79	3,58	3,42	3,3	3,2	3,12
23	7,34	5,33	4,51	4,05	3,75	3,54	3,38	3,26	3,16	3,08
24	7,29	5,29	4,47	4,01	3,71	3,5	3,34	3,22	3,12	3,04
25	7,24	5,25	4,43	3,97	3,67	3,46	3,31	3,18	3,08	3
26	7,2	5,21	4,4	3,94	3,64	3,43	3,27	3,15	3,05	2,97
27	7,16	5,18	4,36	3,91	3,61	3,4	3,24	3,12	3,02	2,94
28	7,13	5,15	4,33	3,88	3,58	3,37	3,22	3,09	2,99	2,91
29	7,09	5,12	4,31	3,85	3,56	3,35	3,19	3,07	2,97	2,89
30	7,06	5,09	4,28	3,83	3,53	3,32	3,17	3,04	2,95	2,86
40	6,84	4,9	4,1	3,66	3,36	3,16	3	2,88	2,78	2,7
60	6,63	4,72	3,93	3,49	3,2	3	2,84	2,72	2,63	2,54
120	6,43	4,55	3,77	3,34	3,05	2,85	2,7	2,58	2,48	2,4
∞	6,24	4,38	3,62	3,19	2,91	2,71	2,55	2,43	2,34	2,26

Таблиця 4.5.5 (б)

$\begin{matrix} m \\ n \end{matrix}$	11	12	14	17	20	24	30	40	60	120	∞
1	3893,2	3907,9	3931,2	3956	3973,5	3990	4006,7	4023,4	4040,2	4057,1	4074
2	79,41	79,42	79,43	79,44	79,45	79,46	79,47	79,47	79,48	79,49	79,5
3	23,28	23,21	23,1	22,99	22,9	22,83	22,75	22,67	22,59	22,51	22,43
4	12,83	12,76	12,65	12,53	12,45	12,37	12,29	12,21	12,13	12,05	11,97
5	9,02	8,95	8,85	8,74	8,66	8,58	8,5	8,42	8,35	8,27	8,19
6	7,14	7,08	6,98	6,87	6,79	6,72	6,64	6,56	6,49	6,41	6,33
7	6,05	5,99	5,89	5,78	5,7	5,63	5,56	5,48	5,4	5,33	5,25
8	5,34	5,28	5,18	5,08	5	4,93	4,85	4,78	4,7	4,62	4,54
9	4,85	4,79	4,69	4,58	4,51	4,44	4,36	4,29	4,21	4,13	4,05
10	4,48	4,42	4,33	4,22	4,15	4,08	4	3,93	3,85	3,77	3,69
11	4,21	4,15	4,05	3,95	3,87	3,8	3,73	3,65	3,58	3,5	3,42
12	3,99	3,93	3,83	3,73	3,66	3,59	3,51	3,44	3,36	3,28	3,2
13	3,81	3,75	3,66	3,56	3,48	3,41	3,34	3,26	3,18	3,1	3,02

Закінчення табл. 4.5.5 (б)

$n \backslash m$	11	12	14	17	20	24	30	40	60	120	∞
14	3,67	3,61	3,51	3,41	3,34	3,27	3,19	3,12	3,04	2,96	2,87
15	3,55	3,49	3,39	3,29	3,22	3,14	3,07	2,99	2,91	2,83	2,75
16	3,44	3,39	3,29	3,19	3,11	3,04	2,97	2,89	2,81	2,73	2,64
17	3,36	3,3	3,2	3,1	3,02	2,95	2,88	2,8	2,72	2,64	2,55
18	3,28	3,22	3,13	3,02	2,95	2,87	2,8	2,72	2,64	2,56	2,47
19	3,21	3,15	3,06	2,95	2,88	2,81	2,73	2,65	2,57	2,49	2,4
20	3,15	3,09	3	2,89	2,82	2,75	2,67	2,59	2,51	2,43	2,34
21	3,1	3,04	2,94	2,84	2,77	2,69	2,62	2,54	2,46	2,37	2,28
22	3,05	2,99	2,9	2,79	2,72	2,65	2,57	2,49	2,41	2,32	2,23
23	3,01	2,95	2,85	2,75	2,68	2,6	2,53	2,45	2,36	2,28	2,18
24	2,97	2,91	2,82	2,71	2,64	2,56	2,49	2,41	2,32	2,24	2,14
25	2,93	2,87	2,78	2,68	2,6	2,53	2,45	2,37	2,29	2,2	2,1
26	2,9	2,84	2,75	2,64	2,57	2,49	2,42	2,34	2,25	2,16	2,07
27	2,87	2,81	2,72	2,61	2,54	2,46	2,39	2,31	2,22	2,13	2,04
28	2,84	2,79	2,69	2,59	2,51	2,44	2,36	2,28	2,19	2,1	2,01
29	2,82	2,76	2,67	2,56	2,49	2,41	2,33	2,25	2,17	2,08	1,98
30	2,8	2,74	2,64	2,54	2,46	2,39	2,31	2,23	2,14	2,05	1,95
40	2,63	2,57	2,48	2,37	2,3	2,22	2,14	2,06	1,97	1,87	1,76
60	2,48	2,42	2,32	2,21	2,14	2,06	1,98	1,89	1,8	1,69	1,57
120	2,33	2,27	2,17	2,06	1,98	1,9	1,82	1,73	1,63	1,51	1,36
∞	2,19	2,13	2,03	1,92	1,84	1,76	1,67	1,57	1,45	1,31	1

Таблиця 4.5.6(а)

 $\alpha = 0,01$

$n \backslash m$	1	2	3	4	5	6	7	8	9	10
1	4052	5000	5403	5625	5764	5859	5928	5981	6022	6056
2	98,50	99,00	99,17	99,25	99,30	99,33	99,36	99,37	99,39	99,40
3	34,12	30,82	29,46	28,71	28,24	27,91	27,67	27,49	27,35	27,23
4	21,20	18,00	16,69	15,98	15,52	15,21	14,98	14,80	14,66	14,55
5	16,26	13,27	12,06	11,39	10,97	10,67	10,46	10,29	10,16	10,05

Закінчення табл. 4.5.6 (а)

$n \backslash m$	1	2	3	4	5	6	7	8	9	10
6	13,75	10,92	9,78	9,15	8,75	8,47	8,26	8,10	7,98	7,87
7	12,25	9,55	8,45	7,85	7,46	7,19	6,99	6,84	6,72	6,62
8	11,26	8,65	7,59	7,01	6,63	6,37	6,18	6,03	5,91	5,81
9	10,56	8,02	6,99	6,42	6,06	5,80	5,61	5,47	5,35	5,26
10	10,04	7,56	6,55	5,99	5,64	5,39	5,20	5,06	4,94	4,85
11	9,65	7,21	6,22	5,67	5,32	5,07	4,89	4,74	4,63	4,54
12	9,33	6,93	5,95	5,41	5,06	4,82	4,64	4,50	4,39	4,30
13	9,07	6,70	5,74	5,21	4,86	4,62	4,44	4,30	4,19	4,10
14	8,86	6,51	5,56	5,04	4,69	4,46	4,28	4,14	4,03	3,94
15	8,68	6,36	5,42	4,89	4,56	4,32	4,14	4,00	3,89	3,80
16	8,53	6,23	5,29	4,77	4,44	4,20	4,03	3,89	3,78	3,69
17	8,40	6,11	5,18	4,67	4,34	4,10	3,93	3,79	3,68	3,59
18	8,29	6,01	5,09	4,58	4,25	4,01	3,84	3,71	3,60	3,51
19	8,18	5,93	5,01	4,50	4,17	3,94	3,77	3,63	3,52	3,43
20	8,10	5,85	4,94	4,43	4,10	3,87	3,70	3,56	3,46	3,37
21	8,02	5,78	4,87	4,37	4,04	3,81	3,64	3,51	3,40	3,31
22	7,95	5,72	4,82	4,31	3,99	3,76	3,59	3,45	3,35	3,26
23	7,88	5,66	4,76	4,26	3,94	3,71	3,54	3,41	3,30	3,21
24	7,82	5,61	4,72	4,22	3,90	3,67	3,50	3,36	3,26	3,17
25	7,77	5,57	4,68	4,18	3,85	3,63	3,46	3,32	3,22	3,13
26	7,72	5,53	4,64	4,14	3,82	3,59	3,42	3,29	3,18	3,09
27	7,68	5,49	4,60	4,11	3,78	3,56	3,39	3,26	3,15	3,06
28	7,64	5,45	4,57	4,07	3,75	3,53	3,36	3,23	3,12	3,03
29	7,60	5,42	4,54	4,04	3,73	3,50	3,33	3,20	3,09	3,00
30	7,56	5,39	4,51	4,02	3,70	3,47	3,30	3,17	3,07	2,98
40	7,31	5,18	4,31	3,83	3,51	3,29	3,12	2,99	2,89	2,80
60	7,08	4,98	4,13	3,65	3,34	3,12	2,95	2,82	2,72	2,63
120	6,85	4,79	3,95	3,48	3,17	2,96	2,79	2,66	2,56	2,47
∞	6,64	4,61	3,78	3,32	3,02	2,80	2,64	2,51	2,41	2,32

Таблица 4.5.6 (б)

$n \backslash m$	11	12	14	17	20	24	30	40	60	120	∞
1	6083	6106	6143	6181	6209	6235	6261	6287	6313	6339	6366
2	99,41	99,42	99,43	99,44	99,45	99,46	99,47	99,47	99,48	99,49	99,50
3	27,13	27,05	26,92	26,79	26,69	26,60	26,50	26,41	26,32	26,22	26,13
4	14,45	14,37	14,25	14,11	14,02	13,93	13,84	13,75	13,65	13,56	13,46
5	9,96	9,89	9,77	9,64	9,55	9,47	9,38	9,29	9,20	9,11	9,02
6	7,79	7,72	7,60	7,48	7,40	7,31	7,23	7,14	7,06	6,97	6,88
7	6,54	6,47	6,36	6,24	6,16	6,07	5,99	5,91	5,82	5,74	5,65
8	5,73	5,67	5,56	5,44	5,36	5,28	5,20	5,12	5,03	4,95	4,86
9	5,18	5,11	5,01	4,89	4,81	4,73	4,65	4,57	4,48	4,40	4,31
10	4,77	4,71	4,60	4,49	4,41	4,33	4,25	4,17	4,08	4,00	3,91
11	4,46	4,40	4,29	4,18	4,10	4,02	3,94	3,86	3,78	3,69	3,60
12	4,22	4,16	4,05	3,94	3,86	3,78	3,70	3,62	3,54	3,45	3,36
13	4,02	3,96	3,86	3,75	3,66	3,59	3,51	3,43	3,34	3,25	3,17
14	3,86	3,80	3,70	3,59	3,51	3,43	3,35	3,27	3,18	3,09	3,00
15	3,73	3,67	3,56	3,45	3,37	3,29	3,21	3,13	3,05	2,96	2,87
16	3,62	3,55	3,45	3,34	3,26	3,18	3,10	3,02	2,93	2,84	2,75
17	3,52	3,46	3,35	3,24	3,16	3,08	3,00	2,92	2,83	2,75	2,65
18	3,43	3,37	3,27	3,16	3,08	3,00	2,92	2,84	2,75	2,66	2,57
19	3,36	3,30	3,19	3,08	3,00	2,92	2,84	2,76	2,67	2,58	2,49
20	3,29	3,23	3,13	3,02	2,94	2,86	2,78	2,69	2,61	2,52	2,42
21	3,24	3,17	3,07	2,96	2,88	2,80	2,72	2,64	2,55	2,46	2,36
22	3,18	3,12	3,02	2,91	2,83	2,75	2,67	2,58	2,50	2,40	2,31
23	3,14	3,07	2,97	2,86	2,78	2,70	2,62	2,54	2,45	2,35	2,26
24	3,09	3,03	2,93	2,82	2,74	2,66	2,58	2,49	2,40	2,31	2,21
25	3,06	2,99	2,89	2,78	2,70	2,62	2,54	2,45	2,36	2,27	2,17
26	3,02	2,96	2,86	2,75	2,66	2,58	2,50	2,42	2,33	2,23	2,13
27	2,99	2,93	2,82	2,71	2,63	2,55	2,47	2,38	2,29	2,20	2,10
28	2,96	2,90	2,79	2,68	2,60	2,52	2,44	2,35	2,26	2,17	2,06
29	2,93	2,87	2,77	2,66	2,57	2,49	2,41	2,33	2,23	2,14	2,03

Закінчення табл. 4.5.6 (б)

$n \backslash m$	11	12	14	17	20	24	30	40	60	120	∞
30	2,91	2,84	2,74	2,63	2,55	2,47	2,39	2,30	2,21	2,11	2,01
40	2,73	2,66	2,56	2,45	2,37	2,29	2,20	2,11	2,02	1,92	1,80
60	2,56	2,50	2,39	2,28	2,20	2,12	2,03	1,94	1,84	1,73	1,60
120	2,40	2,34	2,23	2,12	2,03	1,95	1,86	1,76	1,66	1,53	1,38
∞	2,25	2,19	2,08	1,97	1,88	1,79	1,70	1,59	1,47	1,33	1,02

Таблиця 4.5.7 (а)

 $\alpha = 0,005$

$n \backslash m$	1	2	3	4	5	6	7	8	9	10
1	16211	19999	21615	22500	23056	23437	23715	23925	24091	24224
2	198,50	199,00	199,17	199,25	199,30	199,33	199,36	199,37	199,39	199,40
3	55,55	49,80	47,47	46,19	45,39	44,84	44,43	44,13	43,88	43,69
4	31,33	26,28	24,26	23,15	22,46	21,97	21,62	21,35	21,14	20,97
5	22,78	18,31	16,53	15,56	14,94	14,51	14,20	13,96	13,77	13,62
6	18,63	14,54	12,92	12,03	11,46	11,07	10,79	10,57	10,39	10,25
7	16,24	12,40	10,88	10,05	9,52	9,16	8,89	8,68	8,51	8,38
8	14,69	11,04	9,60	8,81	8,30	7,95	7,69	7,50	7,34	7,21
9	13,61	10,11	8,72	7,96	7,47	7,13	6,88	6,69	6,54	6,42
10	12,83	9,43	8,08	7,34	6,87	6,54	6,30	6,12	5,97	5,85
11	12,23	8,91	7,60	6,88	6,42	6,10	5,86	5,68	5,54	5,42
12	11,75	8,51	7,23	6,52	6,07	5,76	5,52	5,35	5,20	5,09
13	11,37	8,19	6,93	6,23	5,79	5,48	5,25	5,08	4,94	4,82
14	11,06	7,92	6,68	6,00	5,56	5,26	5,03	4,86	4,72	4,60
15	10,80	7,70	6,48	5,80	5,37	5,07	4,85	4,67	4,54	4,42
16	10,58	7,51	6,30	5,64	5,21	4,91	4,69	4,52	4,38	4,27
17	10,38	7,35	6,16	5,50	5,07	4,78	4,56	4,39	4,25	4,14
18	10,22	7,21	6,03	5,37	4,96	4,66	4,44	4,28	4,14	4,03
19	10,07	7,09	5,92	5,27	4,85	4,56	4,34	4,18	4,04	3,93
20	9,94	6,99	5,82	5,17	4,76	4,47	4,26	4,09	3,96	3,85

Закінчення табл. 4.5.7 (а)

$n \backslash m$	1	2	3	4	5	6	7	8	9	10
21	9,83	6,89	5,73	5,09	4,68	4,39	4,18	4,01	3,88	3,77
22	9,73	6,81	5,65	5,02	4,61	4,32	4,11	3,94	3,81	3,70
23	9,63	6,73	5,58	4,95	4,54	4,26	4,05	3,88	3,75	3,64
24	9,55	6,66	5,52	4,89	4,49	4,20	3,99	3,83	3,69	3,59
25	9,48	6,60	5,46	4,84	4,43	4,15	3,94	3,78	3,64	3,54
26	9,41	6,54	5,41	4,79	4,38	4,10	3,89	3,73	3,60	3,49
27	9,34	6,49	5,36	4,74	4,34	4,06	3,85	3,69	3,56	3,45
28	9,28	6,44	5,32	4,70	4,30	4,02	3,81	3,65	3,52	3,41
29	9,23	6,40	5,28	4,66	4,26	3,98	3,77	3,61	3,48	3,38
30	9,18	6,35	5,24	4,62	4,23	3,95	3,74	3,58	3,45	3,34
40	8,83	6,07	4,98	4,37	3,99	3,71	3,51	3,35	3,22	3,12
60	8,49	5,79	4,73	4,14	3,76	3,49	3,29	3,13	3,01	2,90
120	8,18	5,54	4,50	3,92	3,55	3,28	3,09	2,93	2,81	2,71
∞	7,88	5,30	4,28	3,72	3,35	3,09	2,90	2,74	2,62	2,52

Таблиця 4.5.7 (б)

$n \backslash m$	11	12	14	17	20	24	30	40	60	120	∞
1	24334	24426	24572	24727	24836	24940	25044	25148	25253	25359	25464
2	199,41	199,42	199,43	199,44	199,45	199,46	199,47	199,47	199,48	199,49	199,50
3	43,52	43,39	43,17	42,94	42,78	42,62	42,47	42,31	42,15	41,99	41,83
4	20,82	20,70	20,51	20,31	20,17	20,03	19,89	19,75	19,61	19,47	19,33
5	13,49	13,38	13,21	13,03	12,90	12,78	12,66	12,53	12,40	12,27	12,14
6	10,13	10,03	9,88	9,71	9,59	9,47	9,36	9,24	9,12	9,00	8,88
7	8,27	8,18	8,03	7,87	7,75	7,64	7,53	7,42	7,31	7,19	7,08
8	7,10	7,01	6,87	6,72	6,61	6,50	6,40	6,29	6,18	6,06	5,95
9	6,31	6,23	6,09	5,94	5,83	5,73	5,62	5,52	5,41	5,30	5,19
10	5,75	5,66	5,53	5,38	5,27	5,17	5,07	4,97	4,86	4,75	4,64
11	5,32	5,24	5,10	4,96	4,86	4,76	4,65	4,55	4,45	4,34	4,23
12	4,99	4,91	4,77	4,63	4,53	4,43	4,33	4,23	4,12	4,01	3,90

Закінчення табл. 4.5.7 (б)

$n \backslash m$	11	12	14	17	20	24	30	40	60	120	∞
13	4,72	4,64	4,51	4,37	4,27	4,17	4,07	3,97	3,87	3,76	3,65
14	4,51	4,43	4,30	4,16	4,06	3,96	3,86	3,76	3,66	3,55	3,44
15	4,33	4,25	4,12	3,98	3,88	3,79	3,69	3,58	3,48	3,37	3,26
16	4,18	4,10	3,97	3,83	3,73	3,64	3,54	3,44	3,33	3,22	3,11
17	4,05	3,97	3,84	3,71	3,61	3,51	3,41	3,31	3,21	3,10	2,98
18	3,94	3,86	3,73	3,60	3,50	3,40	3,30	3,20	3,10	2,99	2,87
19	3,84	3,76	3,64	3,50	3,40	3,31	3,21	3,11	3,00	2,89	2,78
20	3,76	3,68	3,55	3,42	3,32	3,22	3,12	3,02	2,92	2,81	2,69
21	3,68	3,60	3,48	3,34	3,24	3,15	3,05	2,95	2,84	2,73	2,61
22	3,61	3,54	3,41	3,27	3,18	3,08	2,98	2,88	2,77	2,66	2,55
23	3,55	3,47	3,35	3,21	3,12	3,02	2,92	2,82	2,71	2,60	2,48
24	3,50	3,42	3,30	3,16	3,06	2,97	2,87	2,77	2,66	2,55	2,43
25	3,45	3,37	3,25	3,11	3,01	2,92	2,82	2,72	2,61	2,50	2,38
26	3,40	3,33	3,20	3,07	2,97	2,87	2,77	2,67	2,56	2,45	2,33
27	3,36	3,28	3,16	3,03	2,93	2,83	2,73	2,63	2,52	2,41	2,29
28	3,32	3,25	3,12	2,99	2,89	2,79	2,69	2,59	2,48	2,37	2,25
29	3,29	3,21	3,09	2,95	2,86	2,76	2,66	2,56	2,45	2,33	2,21
30	3,25	3,18	3,06	2,92	2,82	2,73	2,63	2,52	2,42	2,30	2,18
40	3,03	2,95	2,83	2,70	2,60	2,50	2,40	2,30	2,18	2,06	1,93
60	2,82	2,74	2,62	2,49	2,39	2,29	2,19	2,08	1,96	1,83	1,69
120	2,62	2,54	2,42	2,29	2,19	2,09	1,98	1,87	1,75	1,61	1,43
∞	2,43	2,36	2,24	2,10	2,00	1,90	1,79	1,67	1,53	1,36	1,02

Таблиця 4.5.8 (а)

$\alpha = 0,0025$

$n \backslash m$	1	2	3	4	5	6	7	8	9	10
1	64845	79999	86460	90000	92224	93750	94859	95703	96365	96899
2	398,50	399,00	399,17	399,25	399,30	399,33	399,36	399,37	399,39	399,40
3	89,58	79,93	76,06	73,95	72,62	71,71	71,04	70,53	70,13	69,81
4	45,67	38,00	34,96	33,30	32,26	31,54	31,02	30,62	30,30	30,04

Закінчення табл. 4.5.8 (а)

$n \backslash m$	1	2	3	4	5	6	7	8	9	10
5	31,41	24,96	22,43	21,05	20,18	19,58	19,14	18,80	18,54	18,32
6	24,81	19,10	16,87	15,65	14,88	14,35	13,96	13,67	13,43	13,24
7	21,11	15,89	13,84	12,73	12,03	11,55	11,19	10,91	10,70	10,52
8	18,78	13,89	11,98	10,94	10,28	9,83	9,49	9,24	9,03	8,87
9	17,19	12,54	10,73	9,74	9,12	8,68	8,36	8,12	7,92	7,77
10	16,04	11,57	9,83	8,89	8,29	7,87	7,56	7,33	7,14	6,99
11	15,17	10,85	9,17	8,25	7,67	7,27	6,97	6,74	6,56	6,41
12	14,49	10,29	8,65	7,76	7,20	6,80	6,51	6,29	6,11	5,97
13	13,95	9,84	8,24	7,37	6,82	6,44	6,15	5,93	5,76	5,62
14	13,50	9,47	7,91	7,06	6,51	6,14	5,86	5,64	5,47	5,33
15	13,13	9,17	7,63	6,80	6,26	5,89	5,62	5,40	5,23	5,10
16	12,82	8,92	7,40	6,58	6,05	5,68	5,41	5,20	5,04	4,90
17	12,55	8,70	7,21	6,39	5,87	5,51	5,24	5,03	4,87	4,73
18	12,32	8,51	7,04	6,23	5,72	5,36	5,09	4,89	4,72	4,59
19	12,12	8,35	6,89	6,09	5,58	5,23	4,96	4,76	4,60	4,46
20	11,94	8,21	6,76	5,97	5,46	5,11	4,85	4,65	4,49	4,35
21	11,78	8,08	6,64	5,86	5,36	5,01	4,75	4,55	4,39	4,26
22	11,64	7,96	6,54	5,76	5,26	4,92	4,66	4,46	4,30	4,17
23	11,51	7,86	6,45	5,67	5,18	4,84	4,58	4,38	4,22	4,09
24	11,40	7,77	6,36	5,60	5,11	4,76	4,51	4,31	4,15	4,03
25	11,29	7,69	6,29	5,53	5,04	4,70	4,44	4,25	4,09	3,96
26	11,20	7,61	6,22	5,46	4,98	4,64	4,38	4,19	4,03	3,91
27	11,11	7,54	6,16	5,40	4,92	4,58	4,33	4,14	3,98	3,85
28	11,03	7,48	6,10	5,35	4,87	4,53	4,28	4,09	3,93	3,81
29	10,96	7,42	6,05	5,30	4,82	4,48	4,24	4,04	3,89	3,76
30	10,89	7,36	6,00	5,25	4,78	4,44	4,19	4,00	3,85	3,72
40	10,41	6,99	5,66	4,93	4,47	4,14	3,90	3,71	3,56	3,44
60	9,96	6,63	5,34	4,64	4,19	3,87	3,63	3,45	3,30	3,18
120	9,54	6,30	5,05	4,36	3,92	3,61	3,38	3,20	3,06	2,94
∞	9,14	5,99	4,77	4,11	3,68	3,38	3,15	2,97	2,83	2,71

Таблица 4.5.8 (б)

$n \backslash m$	11	12	14	17	20	24	30	40	60	120	∞
1	97339	97707	98288	98908	99345	99759	100176	100594	101014	101435	101858
2	399,41	399,42	399,43	399,44	399,45	399,46	399,47	399,47	399,48	399,49	399,50
3	69,54	69,32	68,96	68,58	68,31	68,06	67,80	67,54	67,28	67,02	66,75
4	29,83	29,66	29,37	29,07	28,86	28,66	28,45	28,24	28,03	27,82	27,61
5	18,14	17,99	17,76	17,50	17,32	17,15	16,98	16,80	16,62	16,44	16,26
6	13,08	12,95	12,74	12,51	12,35	12,19	12,04	11,88	11,72	11,56	11,39
7	10,38	10,25	10,06	9,85	9,70	9,56	9,41	9,26	9,12	8,96	8,81
8	8,73	8,61	8,43	8,23	8,09	7,95	7,82	7,68	7,54	7,39	7,25
9	7,63	7,52	7,35	7,16	7,02	6,89	6,76	6,62	6,49	6,35	6,21
10	6,86	6,75	6,58	6,40	6,27	6,14	6,01	5,88	5,75	5,61	5,47
11	6,29	6,18	6,02	5,84	5,71	5,58	5,46	5,33	5,20	5,07	4,93
12	5,85	5,74	5,58	5,40	5,28	5,16	5,03	4,91	4,78	4,65	4,51
13	5,50	5,40	5,24	5,06	4,94	4,82	4,70	4,57	4,44	4,31	4,18
14	5,21	5,12	4,96	4,79	4,66	4,55	4,42	4,30	4,17	4,04	3,91
15	4,98	4,88	4,73	4,56	4,44	4,32	4,20	4,08	3,95	3,82	3,69
16	4,79	4,69	4,54	4,37	4,25	4,13	4,01	3,89	3,76	3,63	3,50
17	4,62	4,52	4,37	4,21	4,09	3,97	3,85	3,73	3,60	3,47	3,34
18	4,48	4,38	4,23	4,07	3,95	3,83	3,71	3,59	3,47	3,34	3,20
19	4,35	4,26	4,11	3,94	3,83	3,71	3,59	3,47	3,35	3,22	3,08
20	4,24	4,15	4,00	3,84	3,72	3,61	3,49	3,37	3,24	3,11	2,98
21	4,15	4,06	3,91	3,74	3,63	3,51	3,40	3,27	3,15	3,02	2,88
22	4,06	3,97	3,82	3,66	3,54	3,43	3,31	3,19	3,07	2,93	2,80
23	3,99	3,89	3,75	3,58	3,47	3,35	3,24	3,12	2,99	2,86	2,72
24	3,92	3,83	3,68	3,52	3,40	3,29	3,17	3,05	2,92	2,79	2,65
25	3,85	3,76	3,62	3,46	3,34	3,23	3,11	2,99	2,86	2,73	2,59
26	3,80	3,71	3,56	3,40	3,28	3,17	3,06	2,93	2,81	2,68	2,54
27	3,75	3,66	3,51	3,35	3,23	3,12	3,00	2,88	2,76	2,62	2,48
28	3,70	3,61	3,46	3,30	3,19	3,07	2,96	2,84	2,71	2,58	2,44
29	3,66	3,56	3,42	3,26	3,14	3,03	2,92	2,79	2,67	2,53	2,39

Закінчення табл. 4.5.8 (б)

$n \backslash m$	11	12	14	17	20	24	30	40	60	120	∞
30	3,61	3,52	3,38	3,22	3,11	2,99	2,88	2,76	2,63	2,49	2,35
40	3,33	3,25	3,10	2,95	2,83	2,72	2,60	2,48	2,35	2,21	2,06
60	3,08	2,99	2,85	2,69	2,58	2,46	2,35	2,22	2,09	1,94	1,78
120	2,83	2,75	2,61	2,45	2,34	2,23	2,11	1,98	1,84	1,68	1,48
∞	2,61	2,53	2,39	2,23	2,12	2,00	1,88	1,74	1,59	1,40	1,03

4.6. Таблиця значень z -перетворень Фішера

$$z = z(r) = \frac{1}{2} \ln \left(\frac{1+r}{1-r} \right).$$

Для отримання значення z -перетворення потрібно вибрати значення r у крайньому лівому стовпчику (десяті та соті) і у верхньому рядку (тисячні). На перетині цих стовпчика та рядка наведено значення функції $z(r)$.

Таблиця 4.6

	0	0,001	0,002	0,003	0,004	0,005	0,006	0,007	0,008	0,009
0	0	0,001	0,002	0,003	0,004	0,005	0,006	0,007	0,008	0,009
0,01	0,01	0,011	0,012	0,013	0,014	0,015	0,016	0,017	0,018	0,019
0,02	0,02	0,021	0,022	0,023	0,024	0,025	0,026	0,027	0,028	0,029
0,03	0,03	0,031	0,032	0,033	0,034	0,035	0,036	0,037	0,038	0,039
0,04	0,04	0,041	0,042	0,043	0,044	0,045	0,046	0,047	0,048	0,049
0,05	0,05	0,051	0,052	0,053	0,0541	0,0551	0,0561	0,0571	0,0581	0,0591
0,06	0,0601	0,0611	0,0621	0,0631	0,0641	0,0651	0,0661	0,0671	0,0681	0,0691
0,07	0,0701	0,0711	0,0721	0,0731	0,0741	0,0751	0,0761	0,0772	0,0782	0,0792
0,08	0,0802	0,0812	0,0822	0,0832	0,0842	0,0852	0,0862	0,0872	0,0882	0,0892
0,09	0,0902	0,0913	0,0923	0,0933	0,0943	0,0953	0,0963	0,0973	0,0983	0,0993
0,1	0,1003	0,1013	0,1024	0,1034	0,1044	0,1054	0,1064	0,1074	0,1084	0,1094
0,11	0,1104	0,1115	0,1125	0,1135	0,1145	0,1155	0,1165	0,1175	0,1186	0,1196

Продовження табл. 4.6

	0	0,001	0,002	0,003	0,004	0,005	0,006	0,007	0,008	0,009
0,12	0,1206	0,1216	0,1226	0,1236	0,1246	0,1257	0,1267	0,1277	0,1287	0,1297
0,13	0,1307	0,1318	0,1328	0,1338	0,1348	0,1358	0,1368	0,1379	0,1389	0,1399
0,14	0,1409	0,1419	0,143	0,144	0,145	0,146	0,1471	0,1481	0,1491	0,1501
0,15	0,1511	0,1522	0,1532	0,1542	0,1552	0,1563	0,1573	0,1583	0,1593	0,1604
0,16	0,1614	0,1624	0,1634	0,1645	0,1655	0,1665	0,1676	0,1686	0,1696	0,1706
0,17	0,1717	0,1727	0,1737	0,1748	0,1758	0,1768	0,1779	0,1789	0,1799	0,1809
0,18	0,182	0,183	0,1841	0,1851	0,1861	0,1872	0,1882	0,1892	0,1903	0,1913
0,19	0,1923	0,1934	0,1944	0,1955	0,1965	0,1975	0,1986	0,1996	0,2007	0,2017
0,2	0,2027	0,2038	0,2048	0,2059	0,2069	0,2079	0,209	0,21	0,2111	0,2121
0,21	0,2132	0,2142	0,2153	0,2163	0,2174	0,2184	0,2195	0,2205	0,2216	0,2226
0,22	0,2237	0,2247	0,2258	0,2268	0,2279	0,2289	0,23	0,231	0,2321	0,2331
0,23	0,2342	0,2352	0,2363	0,2374	0,2384	0,2395	0,2405	0,2416	0,2427	0,2437
0,24	0,2448	0,2458	0,2469	0,248	0,249	0,2501	0,2512	0,2522	0,2533	0,2543
0,25	0,2554	0,2565	0,2575	0,2586	0,2597	0,2608	0,2618	0,2629	0,264	0,265
0,26	0,2661	0,2672	0,2683	0,2693	0,2704	0,2715	0,2726	0,2736	0,2747	0,2758
0,27	0,2769	0,2779	0,279	0,2801	0,2812	0,2823	0,2833	0,2844	0,2855	0,2866
0,28	0,2877	0,2888	0,2899	0,2909	0,292	0,2931	0,2942	0,2953	0,2964	0,2975
0,29	0,2986	0,2997	0,3008	0,3018	0,3029	0,304	0,3051	0,3062	0,3073	0,3084
0,3	0,3095	0,3106	0,3117	0,3128	0,3139	0,315	0,3161	0,3172	0,3183	0,3194
0,31	0,3205	0,3217	0,3228	0,3239	0,325	0,3261	0,3272	0,3283	0,3294	0,3305
0,32	0,3316	0,3328	0,3339	0,335	0,3361	0,3372	0,3383	0,3395	0,3406	0,3417
0,33	0,3428	0,344	0,3451	0,3462	0,3473	0,3484	0,3496	0,3507	0,3518	0,353
0,34	0,3541	0,3552	0,3564	0,3575	0,3586	0,3598	0,3609	0,362	0,3632	0,3643
0,35	0,3654	0,3666	0,3677	0,3689	0,37	0,3712	0,3723	0,3734	0,3746	0,3757
0,36	0,3769	0,378	0,3792	0,3803	0,3815	0,3826	0,3838	0,385	0,3861	0,3873
0,37	0,3884	0,3896	0,3907	0,3919	0,3931	0,3942	0,3954	0,3966	0,3977	0,3989
0,38	0,4001	0,4012	0,4024	0,4036	0,4047	0,4059	0,4071	0,4083	0,4094	0,4106
0,39	0,4118	0,413	0,4142	0,4153	0,4165	0,4177	0,4189	0,4201	0,4213	0,4225
0,4	0,4236	0,4248	0,426	0,4272	0,4284	0,4296	0,4308	0,432	0,4332	0,4344
0,41	0,4356	0,4368	0,438	0,4392	0,4404	0,4416	0,4428	0,4441	0,4453	0,4465

Продовження табл. 4.6

	0	0,001	0,002	0,003	0,004	0,005	0,006	0,007	0,008	0,009
0,42	0,4477	0,4489	0,4501	0,4513	0,4526	0,4538	0,455	0,4562	0,4574	0,4587
0,43	0,4599	0,4611	0,4624	0,4636	0,4648	0,466	0,4673	0,4685	0,4698	0,471
0,44	0,4722	0,4735	0,4747	0,476	0,4772	0,4784	0,4797	0,4809	0,4822	0,4834
0,45	0,4847	0,486	0,4872	0,4885	0,4897	0,491	0,4922	0,4935	0,4948	0,496
0,46	0,4973	0,4986	0,4999	0,5011	0,5024	0,5037	0,5049	0,5062	0,5075	0,5088
0,47	0,5101	0,5114	0,5126	0,5139	0,5152	0,5165	0,5178	0,5191	0,5204	0,5217
0,48	0,523	0,5243	0,5256	0,5269	0,5282	0,5295	0,5308	0,5321	0,5334	0,5347
0,49	0,5361	0,5374	0,5387	0,54	0,5413	0,5427	0,544	0,5453	0,5466	0,548
0,5	0,5493	0,5506	0,552	0,5533	0,5547	0,556	0,5573	0,5587	0,56	0,5614
0,51	0,5627	0,5641	0,5654	0,5668	0,5682	0,5695	0,5709	0,5722	0,5736	0,575
0,52	0,5763	0,5777	0,5791	0,5805	0,5818	0,5832	0,5846	0,586	0,5874	0,5888
0,53	0,5901	0,5915	0,5929	0,5943	0,5957	0,5971	0,5985	0,5999	0,6013	0,6027
0,54	0,6042	0,6056	0,607	0,6084	0,6098	0,6112	0,6127	0,6141	0,6155	0,6169
0,55	0,6184	0,6198	0,6213	0,6227	0,6241	0,6256	0,627	0,6285	0,6299	0,6314
0,56	0,6328	0,6343	0,6358	0,6372	0,6387	0,6401	0,6416	0,6431	0,6446	0,646
0,57	0,6475	0,649	0,6505	0,652	0,6535	0,655	0,6565	0,658	0,6595	0,661
0,58	0,6625	0,664	0,6655	0,667	0,6685	0,67	0,6716	0,6731	0,6746	0,6761
0,59	0,6777	0,6792	0,6807	0,6823	0,6838	0,6854	0,6869	0,6885	0,69	0,6916
0,6	0,6931	0,6947	0,6963	0,6978	0,6994	0,701	0,7026	0,7042	0,7057	0,7073
0,61	0,7089	0,7105	0,7121	0,7137	0,7153	0,7169	0,7185	0,7201	0,7218	0,7234
0,62	0,725	0,7266	0,7283	0,7299	0,7315	0,7332	0,7348	0,7365	0,7381	0,7398
0,63	0,7414	0,7431	0,7447	0,7464	0,7481	0,7498	0,7514	0,7531	0,7548	0,7565
0,64	0,7582	0,7599	0,7616	0,7633	0,765	0,7667	0,7684	0,7701	0,7718	0,7736
0,65	0,7753	0,777	0,7788	0,7805	0,7823	0,784	0,7858	0,7875	0,7893	0,791
0,66	0,7928	0,7946	0,7964	0,7981	0,7999	0,8017	0,8035	0,8053	0,8071	0,8089
0,67	0,8107	0,8126	0,8144	0,8162	0,818	0,8199	0,8217	0,8236	0,8254	0,8273
0,68	0,8291	0,831	0,8328	0,8347	0,8366	0,8385	0,8404	0,8423	0,8441	0,846
0,69	0,848	0,8499	0,8518	0,8537	0,8556	0,8576	0,8595	0,8614	0,8634	0,8653
0,7	0,8673	0,8693	0,8712	0,8732	0,8752	0,8772	0,8792	0,8812	0,8832	0,8852
0,71	0,8872	0,8892	0,8912	0,8933	0,8953	0,8973	0,8994	0,9014	0,9035	0,9056

Закінчення табл. 4.6

	0	0,001	0,002	0,003	0,004	0,005	0,006	0,007	0,008	0,009
0,72	0,9076	0,9097	0,9118	0,9139	0,916	0,9181	0,9202	0,9223	0,9245	0,9266
0,73	0,9287	0,9309	0,933	0,9352	0,9373	0,9395	0,9417	0,9439	0,9461	0,9483
0,74	0,9505	0,9527	0,9549	0,9571	0,9594	0,9616	0,9639	0,9661	0,9684	0,9707
0,75	0,973	0,9752	0,9775	0,9798	0,9822	0,9845	0,9868	0,9892	0,9915	0,9939
0,76	0,9962	0,9986	1,001	1,0034	1,0058	1,0082	1,0106	1,013	1,0154	1,0179
0,77	1,0203	1,0228	1,0253	1,0277	1,0302	1,0327	1,0352	1,0378	1,0403	1,0428
0,78	1,0454	1,0479	1,0505	1,0531	1,0557	1,0583	1,0609	1,0635	1,0661	1,0688
0,79	1,0714	1,0741	1,0768	1,0795	1,0822	1,0849	1,0876	1,0903	1,0931	1,0958
0,8	1,0986	1,1014	1,1042	1,107	1,1098	1,1127	1,1155	1,1184	1,1212	1,1241
0,81	1,127	1,1299	1,1329	1,1358	1,1388	1,1417	1,1447	1,1477	1,1507	1,1538
0,82	1,1568	1,1599	1,163	1,166	1,1692	1,1723	1,1754	1,1786	1,1817	1,1849
0,83	1,1881	1,1914	1,1946	1,1979	1,2011	1,2044	1,2077	1,2111	1,2144	1,2178
0,84	1,2212	1,2246	1,228	1,2315	1,2349	1,2384	1,2419	1,2454	1,249	1,2526
0,85	1,2562	1,2598	1,2634	1,2671	1,2707	1,2745	1,2782	1,2819	1,2857	1,2895
0,86	1,2933	1,2972	1,3011	1,305	1,3089	1,3129	1,3169	1,3209	1,3249	1,329
0,87	1,3331	1,3372	1,3414	1,3456	1,3498	1,354	1,3583	1,3626	1,367	1,3714
0,88	1,3758	1,3802	1,3847	1,3892	1,3938	1,3984	1,403	1,4077	1,4124	1,4171
0,89	1,4219	1,4268	1,4316	1,4365	1,4415	1,4465	1,4516	1,4566	1,4618	1,467
0,9	1,4722	1,4775	1,4828	1,4882	1,4937	1,4992	1,5047	1,5103	1,516	1,5217
0,91	1,5275	1,5334	1,5393	1,5453	1,5513	1,5574	1,5636	1,5698	1,5762	1,5826
0,92	1,589	1,5956	1,6022	1,6089	1,6157	1,6226	1,6296	1,6366	1,6438	1,651
0,93	1,6584	1,6658	1,6734	1,6811	1,6888	1,6967	1,7047	1,7129	1,7211	1,7295
0,94	1,738	1,7467	1,7555	1,7645	1,7736	1,7828	1,7923	1,8019	1,8117	1,8216
0,95	1,8318	1,8421	1,8527	1,8635	1,8745	1,8857	1,8972	1,909	1,921	1,9333
0,96	1,9459	1,9588	1,9721	1,9857	1,9996	2,0139	2,0287	2,0439	2,0595	2,0756
0,97	2,0923	2,1095	2,1273	2,1457	2,1649	2,1847	2,2054	2,2269	2,2494	2,2729
0,98	2,2976	2,3235	2,3507	2,3796	2,4101	2,4427	2,4774	2,5147	2,555	2,5987
0,99	2,6467	2,6996	2,7587	2,8257	2,9031	2,9945	3,1063	3,2504	3,4534	3,8002

4.7. Таблиця значень функції Колмогорова $K(\lambda)$

$$K(\lambda) = \sum_{k=-\infty}^{\infty} (-1)^k e^{-2k^2\lambda^2}, \text{ помножені на } 10^5.$$

Для отримання значення функції Колмогорова потрібно вибрати значення у крайньому лівому стовпчику (десяті) і у верхньому рядку (соті). На перетині цих стовпчика та рядка наведено значення функції Колмогорова від λ (суми цих двох значень), помножене на 10^5 .

Таблиця 4.7 (а)

λ	0,00	0,01	0,02	0,03	0,04
0,2	0	0	0	0	0
0,3	0,93	2,15	4,59	9,13	17,1
0,4	280,8	397,2	547,6	737,7	973
0,5	3605,48	4281,38	5030,57	5853,45	6749,70
0,6	13571,72	14922,90	16325,46	17775,18	19267,73
0,7	28876,48	30546,99	32226,50	33911,36	35598,05
0,8	45585,76	47203,86	48802,83	50380,90	51936,46
0,9	60726,93	62092,77	63428,50	64733,72	66008,11
1	73000,03	74056,58	75082,50	76078,06	77043,59
1,1	82228,18	82995,10	83735,60	84450,22	85139,50
1,2	88775,03	89302,99	89810,22	90297,31	90764,83
1,3	93190,78	93537,06	93868,20	94184,70	94487,07
1,4	96031,81	96248,67	96455,09	96651,50	96838,29
1,5	97778,20	97908,01	98031,02	98147,54	98257,86
1,6	98804,80	98879,12	98949,24	99015,36	99077,70
1,7	99382,26	99422,98	99461,24	99497,16	99530,87
1,8	99693,24	99714,61	99734,59	99753,28	99770,74
1,9	99853,64	99864,38	99874,38	99883,69	99892,35
2	99932,91	99938,08	99942,87	99947,32	99951,44
2,1	99970,45	99972,84	99975,04	99977,07	99978,95

Закінчення табл. 4.7 (а)

λ	0,00	0,01	0,02	0,03	0,04
2,2	99987,50	99988,55	99989,52	99990,41	99991,23
2,3	99994,92	99995,36	99995,77	99996,15	99996,49
2,4	99998,01	99998,20	99998,36	99998,51	99998,65
2,5	99999,25	99999,33	99999,39	99999,45	99999,50

Таблиця 4.7 (б)

λ	0,05	0,06	0,07	0,08	0,09
0,2	0,003	0,01	0,04	0,13	0,368
0,3	30,3	51,1	82,6	128,5	192,9
0,4	1258,9	1600,47	2002,2	2468,2	3001,7
0,5	7718,32	8757,70	9865,64	11039,44	12275,97
0,6	20798,70	22363,66	23958,22	25578,03	27218,83
0,7	37283,30	38963,97	40637,17	42300,19	43950,49
0,8	53468,08	54974,46	56454,46	57907,10	59331,50
0,9	67251,45	68463,60	69644,50	70794,15	71912,62
1	77979,44	78886,01	79763,71	80612,98	81434,31
1,1	85804,01	86444,32	87061,00	87654,62	88225,77
1,2	91213,36	91643,46	92055,70	92450,63	92828,81
1,3	94775,81	95051,41	95314,35	95565,09	95804,10
1,4	97015,85	97184,57	97344,81	97496,94	97641,29
1,5	98362,26	98461,02	98554,41	98642,67	98726,06
1,6	99136,43	99191,75	99243,83	99292,83	99338,92
1,7	99562,50	99592,16	99619,96	99646,01	99670,40
1,8	99787,05	99802,28	99816,49	99829,75	99842,11
1,9	99900,41	99907,90	99914,86	99921,33	99927,33
2	99955,25	99958,78	99962,05	99965,07	99967,87
2,1	99980,68	99982,28	99983,75	99985,10	99986,35
2,2	99991,99	99992,68	99993,31	99993,89	99994,43
2,3	99996,81	99997,09	99997,36	99997,60	99997,81
2,4	99998,78	99998,89	99999,00	99999,09	99999,18
2,5	99999,55	99999,59	99999,63	99999,67	99999,70

Список літератури

1. *Боровков А. А.* Теория вероятностей / А. А. Боровков. – М. : Наука, 1986.
2. *Вентцель Е. С.* Теория вероятностей. Задачи и упражнения / Е. С. Вентцель, Л. А. Овчаров. – М. : Наука, 1973.
3. *Гихман И. И.* Теория вероятностей и математическая статистика / И. И. Гихман, А. В. Скороход, М. И. Ядренко. – К. : Вища шк., 1988.
4. *Гмурман В. Е.* Руководство к решению задач по теории вероятностей и математической статистике / В. Е. Гмурман. – М. : Высш. шк., 2004.
5. *Гнеденко Б. В.* Курс теории вероятностей / Б. В. Гнеденко. – М. : Наука, 1988.
6. *Гурский Е. И.* Сборник задач по теории вероятностей и математической статистике / Е. И. Гурский. – Минск, Высш. шк., 1984.
7. *Емельянов Г. В.* Задачи по теории вероятностей и математической статистике / Г. В. Емельянов, В. П. Скитович. – Л. : изд-во Ленинградского ун-та, 1967.
8. *Ежов И. И.* Элементы комбинаторики / И. И. Ежов, А. В. Скороход, М. И. Ядренко. – М. : Наука, 1975.
9. *Ивченко Г. И.* Математическая статистика: учеб. пособие для ВТУЗов / Г. И. Ивченко, Ю. И. Медведев. – М. : Высш. шк., 1992.
10. *Паніотто В. І.* Статистичний аналіз соціологічних даних / В. І. Паніотто, В. С. Максименко, Н. М. Харченко. – К. : Вид. дім "Києво-Могилянська Академія", 2004.
11. *Поллард Дж.* Справочник по вычислительным методам статистики / Дж. Поллард. – М. : Финансы и статистика, 1982.
12. *Свешников А. А.* Сборник задач по теории вероятностей, математической статистике и теории случайных функций / А. А. Свешников. – М. : Наука, 1970.
13. *Севастьянов Б. А.* Курс теории вероятностей и математической статистики / Б. А. Севастьянов. – М. : Наука, 1982.
14. *Секей Г.* Парадоксы в теории вероятностей и математической статистике / Г. Секей. – М. : Мир, 1990.
15. *Скороход А. В.* Элементы теории вероятностей та теорії випадкових процесів / А. В. Скороход. – К. : Вища шк., 1975.
16. *Феллер В.* Введение в теорию вероятностей и ее приложения. В 2 т. / В. Феллер. – М. : Наука, 1984.
17. *Черняк О. І.* Теорія ймовірностей та математична статистика. Збірник задач / О. І. Черняк, О. М. Обушна, А. В. Ставицький. – К. : Знання, 2001.
18. *Шефтель З. Г.* Теорія ймовірностей / З. Г. Шефтель. – К. : Вища шк., 1994.
19. *Ширяев А. Н.* Вероятность / А. Н. Ширяев. – М. : Наука, 1989.
20. *Ядренко М. Й.* Дискретна математика: навч. посіб. / М. Й. Ядренко. – К. : ВПЦ "Експрес", 2003.

Навчальне видання

ДОНЧЕНКО Володимир Степанович
СИДОРОВ Микола Володимир-Станіславович

ТЕОРІЯ ЙМОВІРНОСТЕЙ ТА МАТЕМАТИЧНА СТАТИСТИКА ДЛЯ СОЦІАЛЬНИХ НАУК

Навчальний посібник

Редактор *Л. Магда*

Оригінал-макет виготовлено ВПЦ "Київський університет"
Виконавець *В. Гаркуша*



Формат 60x84^{1/16}. Ум. друк. арк. 23,25. Наклад 150. Зам. № 215-7518. Вид. № СЗ.
Гарнітура Times New Roman. Папір офсетний. Друк офсетний.

Підписано до друку 30.12.15

Видавець і виготовлювач
ВПЦ "Київський університет",
б-р Т. Шевченка, 14, м. Київ, 01601
☎ (38044) 239 32 22; (38044) 239 31 72; тел./факс (38044) 239 31 28
e-mail: vpc@univ.kiev.ua
http: vpc.univ.kiev.ua

Свідоцтво суб'єкта видавничої справи ДК № 1103 від 31.10.02