

# Теорія ймовірностей

## Конспект лекцій

Данило Тавров

# Зміст

<b>I. Основні поняття</b>	<b>1</b>
<b>1. Події та ймовірності</b>	<b>2</b>
1.1. Навіщо потрібна теорія ймовірностей . . . . .	2
1.2. Події . . . . .	3
1.3. Класичне визначення ймовірності . . . . .	8
1.4. Інтерпретації ймовірностей . . . . .	11
1.5. Вимірний простір . . . . .	13
1.5.1. Алгебри . . . . .	13
1.5.2. $\sigma$ -алгебри і вимірні простори . . . . .	16
1.5.3. Борелева $\sigma$ -алгебра на дійсній осі . . . . .	19
<b>2. Ймовірність як міра</b>	<b>23</b>
2.1. Аксіоматичне визначення міри . . . . .	23
2.2. Властивості міри . . . . .	24
2.2.1. Основні властивості міри . . . . .	24
2.2.2. Неперервність міри . . . . .	29
2.3. Ймовірнісна міра на зліченному просторі елементарних подій . . . . .	31
2.4. Міра Лебега на одиничному інтервалі . . . . .	32
2.5. Події з нульовою ймовірністю . . . . .	35
2.6. Доведення окремих тверджень . . . . .	37
<b>3. Умовна ймовірність</b>	<b>40</b>
3.1. Інтуїція та визначення . . . . .	40
3.2. Задача Гарднера про двох дітей . . . . .	43
3.3. Теорема Беєса та повна ймовірність . . . . .	46
3.4. Незалежні події . . . . .	51
<b>4. Дискретні випадкові величини. Схема Бернуллі</b>	<b>54</b>
4.1. Поняття про випадкову величину . . . . .	54
4.2. Дискретні випадкові величини . . . . .	57
4.3. Сподівання та дисперсія дискретної випадкової величини . . . . .	60
4.4. Дискретний рівномірний розподіл . . . . .	66
4.5. Схема Бернуллі . . . . .	67
4.5.1. Розподіл Бернуллі . . . . .	67
4.5.2. Біномний розподіл . . . . .	68
<b>5. Деякі поширені дискретні розподіли</b>	<b>74</b>
5.1. Розподіл Пуассона . . . . .	74
5.2. Геометричний розподіл . . . . .	80
5.3. Від'ємний біномний розподіл . . . . .	84

---

5.4. Гіпергеометричний розподіл . . . . .	88
5.5. Умовна функція ймовірності . . . . .	93
5.6. Доведення окремих тверджень . . . . .	94
<b>II. Випадкові величини</b>	<b>97</b>
<b>6. Неперервні випадкові величини</b>	<b>98</b>
6.1. Функції розподілу . . . . .	98
6.2. Зв'язок інтегралу та міри . . . . .	103
6.2.1. Крок 1. Індикаторні величини . . . . .	104
6.2.2. Крок 2. Прості випадкові величини . . . . .	104
6.2.3. Крок 3. Невід'ємні випадкові величини . . . . .	105
6.2.4. Крок 4. Довільні випадкові величини . . . . .	107
6.3. Абсолютно неперервні випадкові величини та щільність розподілу . . . . .	109
6.4. Сподівання та дисперсія як інтеграли . . . . .	116
6.5. Доведення окремих тверджень . . . . .	119
<b>7. Функції від випадкових величин. Деякі неперервні розподіли</b>	<b>122</b>
7.1. Стандартний рівномірний розподіл . . . . .	122
7.2. Стандартний нормальний розподіл . . . . .	123
7.3. Функції від випадкових величин . . . . .	127
7.4. Розподіл функції від випадкової величини . . . . .	132
7.4.1. Розподіл функції від дискретної випадкової величини . . . . .	133
7.4.2. Розподіл функції від абсолютно неперервної випадкової величини . . . . .	134
7.5. Сім'ї розподілів відносно зсуву-масштабування . . . . .	137
7.5.1. Рівномірний розподіл . . . . .	138
7.5.2. Нормальний розподіл . . . . .	139
7.6. Доведення окремих тверджень . . . . .	143
<b>8. Окремі застосування неперервних випадкових величин</b>	<b>146</b>
8.1. Інтегральне перетворення ймовірностей . . . . .	146
8.2. Бета-розподіл та елементи Беєсівського виведення . . . . .	150
8.3. Порядкові статистики . . . . .	155
8.4. Експоненційний розподіл . . . . .	157
8.5. Процес Пуассона . . . . .	162
<b>9. Випадкові вектори та незалежні випадкові величини</b>	<b>165</b>
9.1. Елементи теорії міри для добутків просторів . . . . .	165
9.1.1. Добуток просторів і добуток мір . . . . .	165
9.1.2. Теорема Фубіні . . . . .	167
9.1.3. Добутки вищих порядків . . . . .	169
9.2. Розподіли випадкових векторів . . . . .	171
9.2.1. Спільні функції та щільності розподілу . . . . .	171
9.2.2. Маржинальні функції та щільності розподілу . . . . .	173
9.3. Сподівання випадкового вектора та коваріація . . . . .	179
9.3.1. Сподівання випадкового вектора . . . . .	179
9.3.2. Коваріація та кореляція двох випадкових величин . . . . .	180

---

9.3.3. Матриця коваріацій випадкового вектора . . . . .	183
9.4. Незалежність випадкових величин . . . . .	185
9.4.1. Визначення і критерій незалежності . . . . .	185
9.4.2. Функції та сподівання від незалежних випадкових величин . . . . .	189
9.5. Суми незалежних випадкових величин . . . . .	192
9.5.1. Згортка щільностей розподілів . . . . .	192
9.5.2. Застосування згортки до відомих розподілів . . . . .	195
9.6. Доведення окремих тверджень . . . . .	197
9.6.1. Теорема про добуток мір . . . . .	197
9.6.2. Теорема Фубіні . . . . .	200
9.6.3. Властивості спільних функцій розподілу . . . . .	202
9.6.4. Критерій незалежності випадкових величин . . . . .	206
9.6.5. Згортки відомих розподілів . . . . .	209
<b>10. Функції від випадкових векторів та їхні розподіли</b>	<b>213</b>
10.1. Формула заміни змінних для випадкових векторів . . . . .	213
10.2. Гамма-розподіл . . . . .	219
10.3. Мультиномійний розподіл . . . . .	224
10.4. Багатовимірний нормальний розподіл . . . . .	229
10.5. Доведення окремих тверджень . . . . .	238
<b>11. Сподівання та інші характеристики випадкових величин</b>	<b>240</b>
11.1. Властивості сподівань . . . . .	240
11.1.1. Властивості, що випливають з поняття інтегралу . . . . .	240
11.1.2. Нерівність Єнсена . . . . .	241
11.1.3. Нерівність Гольдера . . . . .	245
11.1.4. Нерівності Маркова і Чебишова . . . . .	249
11.2. Міри центральної тенденції . . . . .	253
11.3. Сподівання і медіана як найліпші предиктори . . . . .	256
11.4. Моменти вищих порядків . . . . .	258
11.5. Квантилі . . . . .	264
<b>12. Умовні сподівання та умовні розподіли</b>	<b>266</b>
12.1. Умовні ймовірність і сподівання для дискретних випадкових величин . . . . .	266
12.2. Умовна ймовірність у загальному випадку . . . . .	268
12.2.1. Обумовлення $\sigma$ -алгеброю, породженою розбиттям . . . . .	268
12.2.2. Обумовлення довільною $\sigma$ -алгеброю . . . . .	270
12.2.3. Обумовлення випадковою величиною . . . . .	274
12.3. Умовні сподівання . . . . .	275
12.3.1. Сподівання за умови довільної $\sigma$ -алгебри . . . . .	275
12.3.2. Властивості умовних сподівань . . . . .	278
12.3.3. Умовні сподівання і медіана як найліпші предиктори . . . . .	281
12.3.4. Умовні дисперсії . . . . .	282
12.4. Умовні щільність і функція розподілу . . . . .	283

<b>III. Послідовності випадкових величин</b>	<b>289</b>
<b>13. Збіжність послідовностей випадкових величин</b>	<b>290</b>
13.1. Послідовності подій . . . . .	290
13.2. Послідовності випадкових величин . . . . .	295
13.2.1. Збіжність майже напевно . . . . .	296
13.2.2. Збіжність за ймовірністю . . . . .	298
13.2.3. Збіжність у середньому . . . . .	300
13.3. Зв'язок між видами збіжності випадкових величин . . . . .	300
13.3.1. Збіжність майже напевно і збіжність за ймовірністю . . . . .	300
13.3.2. Збіжність у середньому і збіжність за ймовірністю . . . . .	303
13.3.3. Збіжність у середньому і збіжність майже напевно . . . . .	304
13.4. Доведення окремих тверджень . . . . .	305
<b>14. Закон великих чисел</b>	<b>309</b>
14.1. Слабкий закон великих чисел . . . . .	309
14.2. Посиленій закон великих чисел . . . . .	310
14.3. Різниця між слабким та посиленім законами великих чисел . . . . .	314
14.4. Приклади застосування законів великих чисел . . . . .	317
14.4.1. Спроможні оцінки параметрів розподілів . . . . .	317
14.4.2. Емпіричні функції розподілів . . . . .	319
14.4.3. Гістограми розподілів . . . . .	322
14.4.4. Інтегрування методом Монте-Карло . . . . .	324
<b>15. Твірні функції моментів та характеристичні функції</b>	<b>327</b>
15.1. Визначення та основні властивості . . . . .	327
15.2. Зв'язок між моментами та твірною функцією моментів . . . . .	330
15.3. Оцінка Черноффа . . . . .	333
15.4. Єдиність твірної функції моментів . . . . .	336
15.5. Поняття про характеристичну функцію . . . . .	339
15.5.1. Комплексні числа та їхні властивості . . . . .	339
15.5.2. Властивості характеристичних функцій . . . . .	340
15.6. Доведення окремих тверджень . . . . .	342
<b>16. Центральна гранична теорема</b>	<b>344</b>
16.1. Збіжність за розподілом . . . . .	344
16.1.1. Визначення та інтерпретація . . . . .	344
16.1.2. Зв'язок з іншими видами збіжності . . . . .	347
16.1.3. Зв'язок із характеристичними функціями . . . . .	349
16.2. Центральна гранична теорема для незалежних однаково розподілених величин . . . . .	349
16.2.1. Формулювання та приклади . . . . .	349
16.2.2. Інтегральна теорема де Муавра-Лапласа . . . . .	357
16.2.3. Дельта-метод . . . . .	358
16.3. Інші варіанти центральної граничної теореми . . . . .	360
<b>Література</b>	<b>363</b>
<b>Предметний покажчик</b>	<b>364</b>



Частина I.

Основні поняття

# 1. Події та ймовірності

## 1.1. Навіщо потрібна теорія ймовірностей

Навколошній світ насичений випадковостями. Для опису випадкових ситуацій ми на інтуїтивному рівні послуговуємося поняттям *імовірності* (probability). Ми кажемо, дощ випаде з імовірністю 80% або обговорюємо шанси виграти в лотерею. Власне кажучи, свої початки вчення про ймовірності бере з практичних застосувань до азартних ігор.

Апарат імовірностей надає інструментарій для розуміння, опису та пояснення випадкових подій далеко не тільки в побуті. Без використання імовірнісних понять не обйтися у таких галузях науки та людської діяльності, де має місце «його величність випадок» [1]:

- *статистика*: маючи на руках деякі дані, ми намагаємося встановити характеристики процесу їх походження, з метою прогнозування чи класифікації нових даних;
- *фізика*: сучасні моделі з квантової фізики використовують імовірності для опису фундаментальних процесів;
- *біологія*: наука про спадковість — генетика — тісно пов’язана з імовірностями, адже як наслідування генів, так і їх мутації є випадковими;
- *економіка та економетрика*: на основі наявних даних за результатами певних спостережень дослідник намагається встановити причинно-наслідковий зв’язок між деякими чинниками (зниження податків, запровадження програм навчання для безробітних, наявність вищої освіти) та показниками, які становлять інтерес (економічне зростання, рівень заробітної плати);
- *медицина*: випробування нових ліків проходить в умовах випадкових клінічних випробувань, коли пацієнтів випадково розподіляють на групу, яка приймає ліки, та групу, яка приймає плацебо.

Теорія *ймовірностей* (probability theory) дає змогу формально і строго увести такі поняття, як імовірність, випадкова величина, розподіл тощо, без яких неможливо уявити собі опис будь-якого випадкового процесу чи явища. За допомогою апарату теорії ймовірностей можна давати відповідь, наприклад, на такі питання.

- Приклад 1.1.1.**
1. Нехай деякого пацієнта тестиють на наявність деякого захворювання, на яке страждає  $100p_1$  відсотків населення. Якщо результат тесту позитивний, то вважають, що пацієнт хворий. Нехай тест дає (помилково) негативний результат для хворої людини в  $100p_2$  відсотках випадків, і (помилково) позитивний результат для здорової людини в  $100p_3$  випадків. Чому дорівнює ймовірність, що пацієнт справді хворий, якщо тест дав позитивний результат?
  2. За страховим полісом, клієнт отримує 10 000 грн у випадку втрати багажу під час турystичної поїздки. Якщо ймовірність такої події — 1 на 200, то якою має бути вартість такого полісу?

3. Нехай для деякого конкретного типу раку після хіміотерапії пацієнт проживає щонайменше 5 років з імовірністю 75%. Нехай  $n$  пацієнтам було поставлено ранній діагноз і вони щойно почали курс хіміотерапії. Нехай  $X$  — число пацієнтів із-поміж загального числа  $n$ , які прожили 5 років і більше. Чому дорівнює ймовірність, що  $X = x$  для деякого конкретного  $x \leq n$ ? Чому дорівнює ймовірність, що принаймні 5 років проживають більше половини пацієнтів ( $X \geq n/2$ )? Яка середня кількість пацієнтів, що проживають принаймні 5 років?
4. Нехай тривалість роботи деякого пристроя дорівнює  $X$ . Якою ймовірнісною моделлю доречно описати тривалість його роботи? Чому дорівнює ймовірність, що пристрій пропрацює принаймні  $t_0$  часу ( $X \geq t_0$ )? Яка очікувана тривалість його роботи?

□

Нарешті, без теорії ймовірностей годі собі уявити галузь *машинного навчання* (machine learning) (та *науки про дані* (data science) в широкому сенсі), оскільки:

- задача класифікації полягає в присвоєнні об'єктам імовірності належності деякому класу;
- низку алгоритмів побудовано на принципах Беєсівської інтерпретації ймовірностей (розвізнавання образів, Беєсівські мережі, підбір гіперпараметрів на основі Беєсівських методів тощо);
- навчання відбувається на основі статистичних методів, наприклад методу максимальної правдоподібності, які у свою чергу базуються на теорії ймовірностей;
- для оцінки моделей використовують імовірнісні метрики (наприклад, ROC-крива або корінь середньоквадратичної похибки (root-mean-square error, RMSE)).

Придумати алгоритм машинного навчання — справа нехитра, але обґрунтувати його адекватність, показати, що для задач певного класу він є найліпший тощо неможливо без використання ймовірнісного і статистичного апарату.

Як це часто буває в математиці, поняття з теорії ймовірностей слугують для *моделювання* навколошнього світу, тобто є апроксимаціями реальних явищ. Наприклад, у фізиці використовують такі поняття, як ідеальний газ чи абсолютно чорне тіло, а в геометрії — точку і пряму. У рамках теорії ймовірностей ми також послуговуємося певними ідеальними об'єктами для опису подій реального світу. Зокрема, ми можемо говорити про ймовірність народження дитини з певною вагою, використовуючи для позначення ваги числа на дійсній осі, хоча вага кожної дитини фіксується з певною обмеженою точністю, а відтак множина всіх можливих ваг новонароджених є, за великим рахунком, хоч і дуже великою, але скінченою.

У цьому курсі лекцій ми будемо поєднувати викладення основних понять теорії ймовірностей як на інтуїтивному рівні, із використанням широкого спектру практичних прикладів, так і на рівні формальному, із застосуванням елементів теорії міри. Адже за влучним формулюванням із [1, с. 104], «у процесі вивчення теорії ймовірностей і статистики одне з найважливіших умінь, які варто опанувати — здатність переключатися з абстрактних понять на конкретні приклади».

Знайомство з поняттям імовірності почнемо з формалізації самого поняття події.

## 1.2. Подїї

Одним із найбазовіших понять у теорії ймовірностей є *випадковий експеримент* (random experiment), який можна розглядати як деяку процедуру, що відбувається за певних умов довільне число разів і результатом якого може бути один із деякого набору можливих варіантів.

Табл. 1.2.1.: Інтерпретація основних теоретико-множинних понять у контексті подій

Теоретико-множинне поняття	Інтерпретація
порожня множина (empty set) $\emptyset$	неможлива подія (impossible event)
універсальна множина (universal set) $\Omega$	певна подія (certain event)
належність $\omega \in A$	сталася подія $A$
включення $A \subseteq B$	якщо сталася подія $A$ , то сталася подія $B$
об'єднання (union) $A \cup B$	сталася або подія $A$ , або подія $B$ , або обидві
перетин (intersection) $A \cap B$	сталися обидві події $A$ і $B$
неперетинанні (disjoint) множини $A \cap B = \emptyset$	події $A$ і $B$ несумісні (mutually exclusive)
заперечення (complement) $A^c$	подія $A^c$ стається тоді і тільки тоді, коли подія $A$ не стається
різниця (difference) двох множин, $A \setminus B = A \cap B^c$	сталася тільки подія $A$ , точно не $B$
симетрична різниця (symmetric difference) двох множин, $A \Delta B = (A \cap B^c) \cup (A^c \cap B)$	сталася тільки одна з двох подій $A$ і $B$ , точно не обидві

**Визначення 1.2.1.** Простором елементарних подій (sample space)  $\Omega$  для деякого випадкового експерименту називатимемо множину всіх можливих його результатів. Простір  $\Omega$  може бути скінченим (finite), зліченним (countable) або незліченним (uncountable). Елементами  $\omega \in \Omega$  цього простору є елементарні події (samples). Тоді, подія (event)  $A$  — це деяка підмножина (subset) простору  $\Omega$ <sup>1</sup>.  $\square$

Ми кажемо, що *сталася подія  $A$* , якщо результат експерименту належить множині  $A$ . Елементарні події відіграють у теорії ймовірностей таку ж роль, як і абстрактні поняття точки та прямої — у геометрії. Оскільки події є підмножинами простору  $S$ , до них застосовні всі теоретико-множинні операції та їхні властивості, які в контексті подій мають особливу, інтуїтивно зрозумілу, інтерпретацію (Табл. 1.2.1).

Узагальнюючи ці поняття, можемо увести об'єднання і перетин довільного числа множин, у тому числі незліченного:

$$\bigcup_{i \in I} A_i, \quad \bigcap_{i \in I} A_i,$$

для деякої множини  $I$  індексів, яка може бути незліченною.

Кажемо, що події  $A_1, \dots, A_n$  утворюють *розвиття* (partition) простору  $\Omega$ , якщо вони є попарно несумісні —  $A_i \cap A_j = \emptyset$ ,  $i \neq j$  — а їх об'єднання дає весь простір:  $\bigcup_{i=1}^n A_i = \Omega$ .

Також послуговуватимемося поняттям потужності (cardinality) множини,  $|A|$ , яке узагальнює поняття «розміру множини». Потужність скінченної множини дорівнює кількості елементів у ній, потужність зліченої множини дорівнює  $\aleph_0$ , а потужність незліченної множини — потужності континууму  $\mathfrak{c}$ .

Операції над множинами часто ілюструють за допомогою *діаграм Венна* (Venn diagrams), приклади яких для поширеніших операцій над множинами наведено на Рис. 1.2.1.

<sup>1</sup>Україномовна термінологія є спадщиною радянських часів і не зовсім коректна, адже формально подію визначають як множину, тобто є сенс говорити про події  $\{\omega_1\}$ ,  $\{\omega_4\}$  тощо, проте елементарними подіями називають не відповідні множини, а елементи, які їх складають —  $\omega_1$ ,  $\omega_4$  тощо. У цьому сенсі англомовна термінологія послідовніша, адже елементи  $\Omega$ ,  $\omega$  називають samples, а відповідні множини  $\{\omega\}$  — простими подіями (simple events).

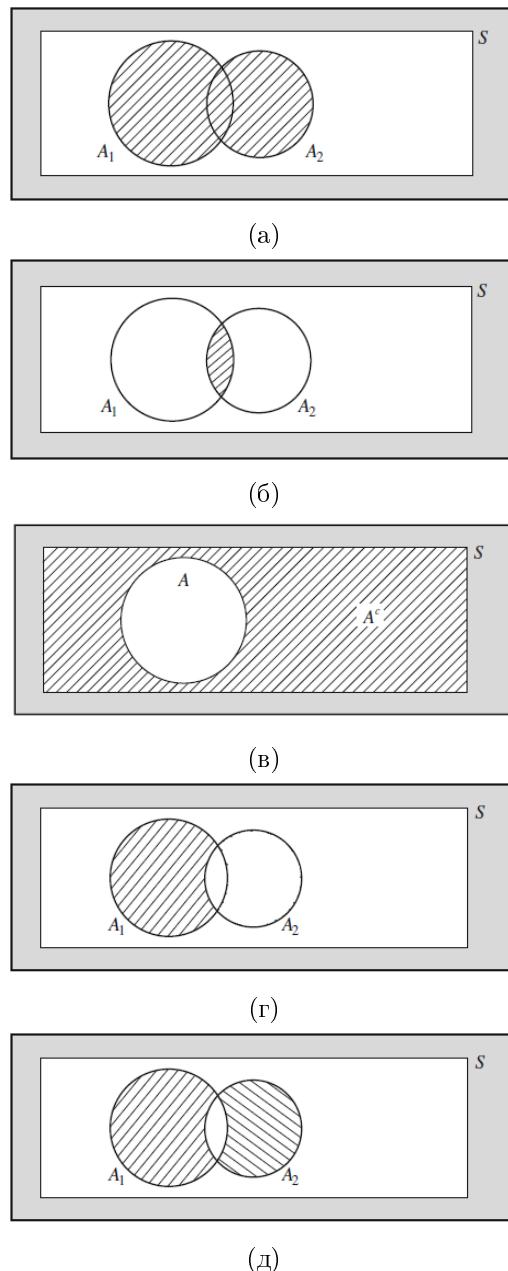


Рис. 1.2.1.: Операції над множинами. Заштриховані області являють собою результати операцій:  
 (а)  $A \cup B$ ; (б)  $A \cap B$ ; (в)  $A^c$ ; (г)  $A \setminus B$ ; (д)  $A \Delta B$  ([2], Рис. 2.2–2.6)

Також справедливі всі властивості та закони операцій над множинами:

- комутативність (commutativity) об'єднання і перетину:  $A \cup B = B \cup A$ ,  $A \cap B = B \cap A$ ;
- асоціативність (associativity) об'єднання і перетину:  $A \cup (B \cup C) = (A \cup B) \cup C$ ,  $A \cap (B \cap C) = (A \cap B) \cap C$ ;
- дистрибутивність (distributivity):  $A \cup (B \cap C) = (A \cup B) \cap (A \cup C)$ ,  $A \cap (B \cup C) = (A \cap B) \cup (A \cap C)$ ;
- закони де Моргана (de Morgan's laws):

$$\left( \bigcup_{i \in I} A_i \right)^c = \bigcap_{i \in I} A_i^c, \\ \left( \bigcap_{i \in I} A_i \right)^c = \bigcup_{i \in I} A_i^c.$$

Розгляньмо деякі приклади просторів елементарних подій, почавши зі скінчених.

**Приклад 1.2.2** (Підкидування монетки скінченну кількість разів). Класичним прикладом для ілюстрації уведених понять є підкидування правильної монетки. Вважатимемо, що підкид монетки може завершитися або випадом герба, або випадом числа. Монетка є правильною, якщо обидва ці варіанти є рівноЯмовірними. Позначмо випад герба через 1, а числа — через 0.

Нехай монетку підкидують 10 разів. Одним із можливих результатів такого експерименту є вектор  $(0, 1, 0, 0, 0, 1, 1, 0, 1, 0)^\top$ <sup>2</sup>, а весь простір елементарних подій у цьому випадку складається з усіх можливих послідовностей векторів:  $\Omega = \{(d_1, \dots, d_{10})^\top \in \{0, 1\}^{10}\}$ . Розгляньмо деякі приклади подій:

1. Подія  $A_1$  — перший випад завершився гербом:

$$A_1 = \{(1, d_2, \dots, d_{10})^\top \in \{0, 1\}^{10}\}.$$

За аналогією можемо ввести події  $A_j$ , що герб випав на  $j$ -ому підкиді,  $j = 2, \dots, 10$ .

2. Подія  $B$  — щонайменше один випад був гербом:

$$B = \bigcup_{j=1}^{10} A_j.$$

3. Подія  $C$  — усі випади були гербами:

$$C = \bigcap_{j=1}^{10} A_j.$$

<sup>2</sup>Як правило, «вектор» у лінійній алгебрі розуміють як вектор-стовпець, і в нашому курсі ми будемо послідовно цього дотримуватися.

4. Подія  $D$  — мали місце щонайменше два герби поспіль:

$$D = \bigcup_{j=1}^9 (A_j \cap A_{j+1}) .$$

□

**Приклад 1.2.3.** Для опису ймовірностей посадання першого, другого та третього місць серед  $n$  спортсменів у деяких змаганнях простором елементарних подій буде множина всіх можливих упорядкованих трійок натуральних чисел від 1 до  $n$  без повторень:

$$\Omega = \left\{ (i, j, k)^\top ; i \neq j, j \neq k, i \neq k, 1 \leq i, j, k \leq n \right\} .$$

□

У низці випадків скінченного числа елементарних подій стає недостатньо для описання явищ у реальному житті.

**Приклад 1.2.4** (Підкидання монетки до настання двох повторень). У розвиток Прикладу 1.2.2 розгляньмо випадок підкидання монетки доти, доки не випаде два герби або два числа поспіль. Позначмо такі елементарні події як  $H_j$  = «випало два герби поспіль, починаючи з  $j$ -го підкидання» та  $T_j$  = «випало два числа поспіль, починаючи з  $j$ -го підкидання». Тоді, наприклад,  $H_2 = (0, 1, 1)$ ,  $T_4 = (1, 0, 1, 0, 0)$ . Цілком очевидно, що простір елементарних подій у цьому випадку є зліченним.

Прикладом події в цьому контексті може слугувати  $A$  = «мало місце щонайменше чотири підкидання», яку можна подати як  $A = \bigcup_{j=3}^{\infty} (H_j \cup T_j)$ .

□

**Приклад 1.2.5.** Прикладом із реального життя, де постає зліченний простір елементарних подій — фіксування числа дорожньо-транспортних пригод, що стаються в певній локації протягом деякого періоду часу. Простором елементарних подій є множина невід'ємних цілих  $\Omega = \mathbb{Z}^+ = \{0, 1, 2, \dots\}$ . Зрозуміло, що в реальному житті кількість ДТП фізично обмежена, проте моделювання цієї ситуації за допомогою (нескінченно) зліченного простору спрощує практичні розрахунки.

□

Незліченні простори елементарних подій корисні як апроксимації в ситуаціях, де результати експерименту природно розглядати як дійсні числа.

**Приклад 1.2.6.** У задачах оцінювання ймовірності тієї чи тієї тривалості життя деякого пристроя простором елементарних подій може слугувати проміжок  $\Omega = (0; T]$  для деякого великого  $T$  або навіть множина додатних дійсних чисел  $\Omega = \mathbb{R}^+$ .

□

Простори такого роду виникають навіть у на перший погляд несподіваних випадках.

**Приклад 1.2.7** (Підкидування монетки нескінченну кількість разів). У розвиток попередніх прикладів розгляньмо підкидування монетки нескінченну кількість разів. Тоді простір подій можна записати як  $\Omega = \{\omega = (d_1, d_2, \dots), d_j \in \{0, 1\}, j \geq 1\}$ . Неважаючи на те, що процес побудови такого простору є «дискретним» (передбачає послідовні підкидання монетки), сам простір насправді є ізоморфним неперервному інтервалу, тобто є незліченним.

Це спостереження випливає з того факту, що будь-яке дійсне число  $0 < x \leq 1$  можна подати через двійкове розвинення<sup>3</sup>

$$x = \sum_{n=1}^{\infty} \frac{d_n}{2^n} = 0.d_1 d_2 \dots, \quad d_n \in \{0, 1\}, n \in \mathbb{N}.$$

Наприклад,  $\frac{1}{3} = 0.010101\dots = 0.(01)$ , де дужками позначаються періодично повторювані цифри.

Таким чином, будь-якому дійсному числу  $x \in (0, 1]$  відповідає елементарна подія  $\omega^4$ , а відтак увесь простір елементарних подій є незліченним.  $\square$

### 1.3. Класичне визначення ймовірності

Першим кроком до формалізації загальноприйнятих уявлень про ймовірність є класичне визначення ймовірності, яке історично бере початок із застосувань до азартних ігор.

**Визначення 1.3.1** (Класичне визначення ймовірності). Нехай простір елементарних подій є скінченим:  $\Omega = \{\omega_1, \dots, \omega_m\}$ . У цьому випадку за побудовою:

- $\{\omega_i\} \cap \{\omega_j\} = \emptyset, i \neq j$ , тобто елементарні події є попарно несумісні;
- $\bigcup_{i=1}^m \{\omega_i\} = \Omega$ , тобто експеримент завершується принаймні однією з елементарних подій;
- усі елементарні події однаково вірогідні.

Нехай  $A \subseteq \Omega$  — деяка подія. Тоді *класичною ймовірністю* (classical, naive probability) події вважатимемо відношення

$$\mathbb{P}(A) = \frac{|A|}{|\Omega|}. \quad (1.3.1)$$

$\square$

У цьому сенсі ймовірність можна вважати узагальненням пропорції, адже ймовірність події дорівнює відношенню кількості способів її настання (елементарних подій, що її складають) до загального числа можливих результатів експерименту. За влучним висловом відомого швейцарського математика й одного з засновників теорії ймовірностей Якова Бернуллі (Jacob Bernoulli, 1654–1705), «Імовірність — це ступінь визначеності, що так відноситься до повної визначеності, як частина відноситься до цілого».

<sup>3</sup>Детальне виведення цього факту наведено в [3], розділ А31 додатку.

<sup>4</sup>Деяким числам відповідають два розвинення, наприклад,  $\frac{1}{2} = 0.1(0) = 0.0(1)$ . Для визначеності беремо тільки розвинення, які не завершуються нулем.

**Приклад 1.3.2** (Кидання гральних кісточок). Розгляньмо випадок кидання двох шестигранних гральних кісточок. Вважатимемо, що кісточки «правильні», у тому сенсі, що вони можуть приземлитися будь-якою зі своїх граней догори з однаковою вірогідністю. У цьому випадку простір елементарних подій складається з упорядкованих пар  $\Omega = \{(i, j) : 1 \leq i \leq 6, 1 \leq j \leq 6\}$ . Загальне число таких пар, очевидно, дорівнює  $6 \cdot 6 = 36$ .

Підрахуймо ймовірності декількох подій:

1. Сума двох випалих чисел дорівнює 5. Ця подія є множиною  $A = \{(1, 4), (2, 3), (3, 2), (4, 1)\}$ , тому її ймовірність дорівнює  $\mathbb{P}(A) = 4/36 = 1/9$ .
2. На одній із кісточок випало 2, а на іншій — 4 (порядок кісточок несуттєвий). Цю подію можна записати як  $B = \{(2, 4), (4, 2)\}$ . Відповідно,  $\mathbb{P}(B) = 2/36 = 1/18$ .
3. Подія, що на другій кісточці випало більше число, ніж на першій. Це множина

$$C = \{(i, j) : i < j, 1 \leq i \leq 6, 1 \leq j \leq 6\}.$$

Потужність цієї множини дорівнює  $5 + 4 + 3 + 2 + 1 = 15$ , а відтак  $\mathbb{P}(C) = 15/36$ .

4. Подія, що на другій кісточці випало стільки ж, скільки на першій, або менше. Очевидно, що це подія  $D = \Omega \setminus C$ , а тому

$$\mathbb{P}(D) = \frac{|\Omega \setminus C|}{|\Omega|} = \frac{|\Omega| - |C|}{|\Omega|} = 1 - \mathbb{P}(C) = 21/36.$$

□

Нескладно бачити, що для класичних імовірностей справедливі такі властивості, для всіх подій  $A, B \subseteq \Omega$ :

- $\mathbb{P}(A) \in [0; 1]$ ;
- $\mathbb{P}(\Omega) = 1, \mathbb{P}(\emptyset) = 0$ ;
- якщо події взаємовиключні, то  $\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B)$ ;
- $\mathbb{P}(A^c) = 1 - \mathbb{P}(A)$ .

Ці властивості прямо випливають із визначення потужності скінченної множини та арифметичних правил.

Варто зазначити, що класичне визначення ймовірності є дуже обмеженим, адже воно вимагає, щоб простір елементарних подій  $\Omega$  був скінченим, а всі елементарні події  $\omega \in \Omega$  — однаково вірогідними. Некоректне використання цього визначення часто веде до помилкових тверджень, наприклад, що ймовірність життя на Marsі дорівнює 0.5, адже воно «або існує, або не існує». Аналогічно можна стверджувати, що ймовірність перемоги кандидата на виборах є 0.5, адже можливі тільки два варіанти розвитку подій — перемога або програш. Зрозуміло, що в цих ситуаціях використання (1.3.1) є недоречним, адже варіанти розвитку подій не є однаково вірогідними.

Разом із цим, класичне визначення ймовірності можна використовувати в низці практичних прикладів, наприклад, коли елементарні події є рівнозначні або в силу *симетрії* (скажімо, різні

варiанти роздачi карт iз добре тасованої колоди мають однакову вiрогiднiсть), або за побудовою (зокрема, коли проводять соцiологiчнi опитування на випадковiй вибiрцi респондентiв<sup>5</sup>).

Класичне визначення ймовiрностi можна узагальнити у злiченному випадку, якщо покласти за ймовiрнiсть границю вiдносних частот наставання певної подiї:

$$\mathbb{P}(A) = \lim_{N \rightarrow \infty} \frac{N(A)}{N}, \quad (1.3.2)$$

де  $N(A)$  — число наставань подiї  $A$  серед  $N$  повторень випадкового експерименту. Вищенаведенi властивостi ймовiрностей зберiгаються i в цьому випадку, що випливає з властивостей числових границь. Так, наприклад, якщо  $A \cap B = \emptyset$ :

$$\mathbb{P}(A \cup B) = \lim_{N \rightarrow \infty} \frac{N(A \cup B)}{N} = \lim_{N \rightarrow \infty} \left( \frac{N(A)}{N} + \frac{N(B)}{N} \right) = \mathbb{P}(A) + \mathbb{P}(B).$$

Скажiмо, у Прикладi 1.2.4 можемо покласти  $\mathbb{P}(\{H_j\}) = \mathbb{P}(\{T_j\}) = \frac{1}{2^{j+1}}$ , що вiдповiдало б iнтуїтивним очiкуванням. Так,  $\mathbb{P}(\{H_1\}) = \mathbb{P}((1, 1)) = \frac{1}{4}$ , адже цiлком «очiкувано», що будь-яка послiдовнiсть пiдкидань може початися або з двох нулiв, або з двох одиниць, або з одиницi пiслi нуля, або з нуля пiслi одиницi, тому ймовiрнiсть випаду двох одиниць на самому початку дорiвнює однiй чвертi. Щоправда, уже в цьому випадку постають питання обчислення ймовiрностей подiй, якi є злiченними комбiнацiями елементарних подiй, наприклад «мало мiсце щонайменше чотири пiдкидання», адже заздалегiдь немає гарантiї, що границя вiдповiдних сум iмовiрностей елементарних подiй iснуватиме.

На випадок незлiченних подiй можна *спробувати* узагальнити поняття класичної ймовiрностi як вiдношення деякої *miri* подiї  $A$  як множини (довжина вiдрiзку, площа пласкої фiгури, об'ем тiла тощо) до загальної мiri всього простору  $\Omega$  (довжини, площи, об'emu такого простору). Таке визначення ймовiрностi iнколи називають *геометричним*.

Проте така спроба формалiзувати iнтуїтивне уявлення про ймовiрнiсть на випадок незлiченного простору подiй натикається на певнi складностi. Зокрема, iмовiрнiсть деякої конкретної точки  $x \in [0; 1] \times [0; 1]$  дорiвнює 0, а проте ймовiрнiсть подiї  $A$ , що складається з таких точок, додатна. З iншого боку, якщо виколоти з подiї  $A$  одну точку  $x$ , то її ймовiрнiсть вiд цього не повинна змiнитися в будь-якому iнтуїтивному сенсi, а тим не менше  $A$  i  $A \setminus \{x\}$  — рiзni подiї. Це все пояснюється тим, що таких точок незлiченна кiлькiсть, а класичне визначення ймовiрностi передбачає можливiсть перерахувати всi «складники» деякої подiї.

**Приклад 1.3.3.** У Прикладi 1.2.7 можна визначити ймовiрнiсть подiї  $H_n$  = «першi  $n$  пiдкидань завершилися одиницями», помiтивши той факт, що  $\omega \in H_n$ , якщо

$$\sum_{i=1}^n \frac{1}{2^i} + \sum_{i=n+1}^{\infty} \frac{0}{2^i} < \omega \leq \sum_{i=1}^n \frac{1}{2^i} + \sum_{i=n+1}^{\infty} \frac{1}{2^i},$$

тобто що  $\omega$  належить промiжку всiх таких дiйсних чисел, першi  $n$  цифр дiйкового розвинення яких — одиницi, а решта — довiльнi. Оскiльки, за формулою суми нескiнченно спадної геометричної прогресiї,

$$\sum_{i=n+1}^{\infty} \frac{1}{2^i} = \frac{1/2^{n+1}}{1 - 1/2} = \frac{1}{2^n},$$

<sup>5</sup>На практицi побудова спрaвдi вiпадкової, або як її називають — *рeпрeзeнтативної*, вибiрки є задачeю непростою i для цього розробляють рiзноманiтнi методики.

довжина такого проміжку дорівнює  $\frac{1}{2^n}$ , що відповідає нашій інтуїції (пор. з імовірностями елементарних подій для Прикладу 1.2.4).

Але з відповідних маніпуляцій зовсім не випливає, як можна обчислити (і чи взагалі це можливо) ймовірності таких подій, як, скажімо,

$$\left\{ (d_1, d_2, \dots) : \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{i=1}^n d_i = \frac{1}{2} \right\}, \quad (1.3.3)$$

тобто що в нескінченій послідовності підкидів монетки міститься «однакова кількість» нулів і одиниць у тому розумінні, що границя емпіричних частот зустрічання нулів або одиниць дорівнює  $\frac{1}{2}$ . Це відповідає нашим уявленням про правильну монетку, і загальна ерудиція підказує, що ймовірність цієї події *повинна* бути «фактично 1», навіть незважаючи на те, що варіантів випадання решти послідовностей є незліченна кількість.  $\square$

Таким чином, існує потреба у формалізації поняття ймовірності в загальному випадку, яка б відповідала інтуїтивним уявленням про навколошній світ і до того ж була б математично строгою. Цьому присвятимо дальший виклад матеріалу.

## 1.4. Інтерпретації ймовірностей

Перед тим, як перейти від класичного визначення ймовірності до розгляду строго формальної побудови теорії ймовірностей, варто зазначити, що побутують два принципово відмінні підходи до *інтерпретації* самого поняття ймовірності.

Згідно з *частотним підходом* (frequentist approach), імовірність деякої події — це гранична частота настання події за умови нескінченного повторення одного й того ж експерименту, що відображене в (1.3.2). У такому контексті повторювані експерименти мають називу *випробувань* (trials).

**Приклад 1.4.1.** Продовжмо розгляд підкидання монеток з Прикладу 1.2.2. Імовірність випаду герба у розумінні (1.3.1) дорівнює 0.5. Згідно з частотною інтерпретацією ймовірностей, якщо ми будемо підкидати монетку велику кількість разів, *емпірична* частота випадів герба буде дорівнювати приблизно 0.5<sup>6</sup>.

Код на Лістингу 1.4.1 симулює підкидання монетки 1000 разів та будує графік емпіричних частот випадів герба як функцію від кількості випробувань. Як можна бачити з Рис. 1.4.1, емпіричні частоти зі збільшенням числа випробувань стають близькі до теоретичної ймовірності.

Лістинг 1.4.1: Код для Прикладу 1.4.1

```
import matplotlib.pyplot as plt
from matplotlib import rc
import numpy as np
from numpy.random import seed, random

5
rc('text', usetex=True)
rc('text.latex', unicode=True)
rc('text.latex', preamble=r'\usepackage[utf8]{inputenc}')
rc('text.latex', preamble=r'\usepackage[ukrainian]{babel}')
```

10

<sup>6</sup>Далі в цьому курсі ми розглянемо формалізацію цього спостереження у формі *закону великих чисел*.

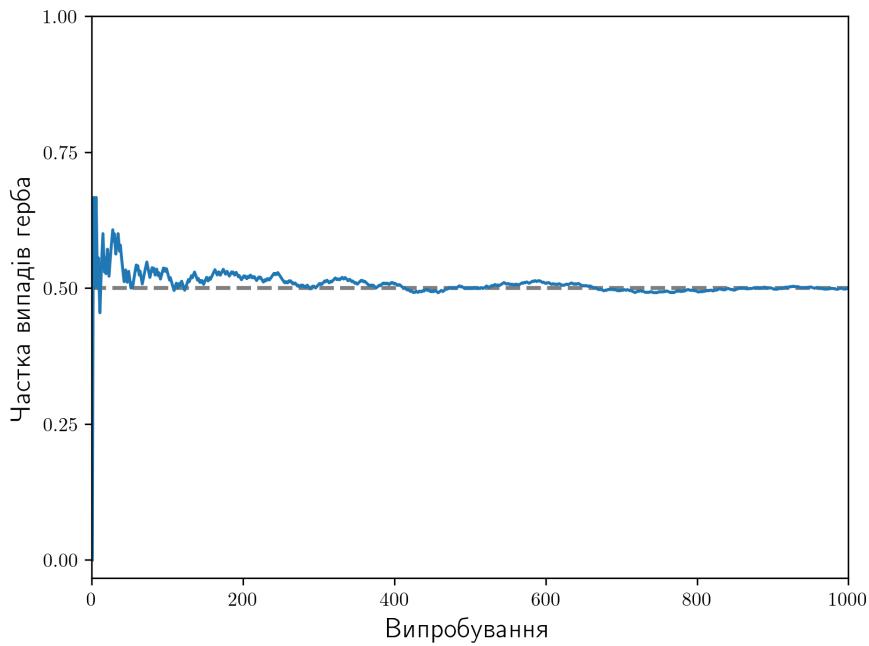


Рис. 1.4.1.: Графік емпіричних частот випадів герба як функції від кількості випробувань для Прикладу 1.4.1

```

seed(100)
prob = 0.5
n = 1000

15  tosses = [random() <= prob for i in range(n)]
frequencies = np.cumsum(tosses) / np.arange(1, n + 1)

plt.plot(np.arange(1, n + 1), frequencies, '-')
plt.hlines(prob, 0, n, linestyle='dashed', lw=2, alpha=0.5)
20  plt.xlim(0, n)
plt.xlabel("Випробування", fontsize=15)
plt.ylabel("Частка випадів герба", fontsize=15)
plt.xticks(np.arange(0, 1001, 200), fontsize=10)
plt.yticks([0.0, 0.25, 0.5, 0.75, 1.0], fontsize=10)
25  plt.tight_layout()
plt.savefig("../images/Lecture 1/frequencies.png", dpi=300)

```

□

Згідно з альтернативним підходом до інтерпретації ймовірностей — *Бессівським підходом* (Bayesian approach) — імовірність являє собою ступінь упевненості (belief) у настанні деякої події. У цьому сенсі можна говорити про ймовірності «смерти пацієнта після операції» або «визнання підсудного винним», хоча відповідні «експерименти» неможливо відтворити не те що нескінченну кількість разів, а навіть двічі. Навіть якщо відома історія виконання аналогічних операцій у миналому, і частка летальних випадків становить 10%, це мало стосується до ситуації конкретного пацієнта, якщо доктор має доступ до додаткової інформації про, скажімо, стан здоров'я пацієнта.

Обидва підходи до інтерпретації ймовірностей є взаємодоповнююальні і мають своє застосування в статистиці та машинному навчанні.

## 1.5. Вимірний простір

Розгляд формалізації поняття ймовірності почнемо зі з'ясування, якого роду події становлять для нас інтерес, тобто для яких підмножин простору елементарних подій ми прагнемо визначити поняття ймовірності.

Нехай маємо абстрактний простір елементарних подій  $\Omega$ . Позначмо через  $\mathcal{C}$  деякий клас (сім'ю) подій — тобто множину, яка містить інші множини<sup>7</sup>. Природним є бажання мати можливість визначити ймовірності для якомога більшого класу подій.

Використання теоретико-множинних операцій дає змогу утворювати з окремих подій інші події, отже наш клас повинен містити події, які можна утворити шляхом застосування елементарних операцій (заперечення, об'єднання та перетин) до інших подій. Оскільки закони де Моргана дають змогу подати перетин через комбінації заперечення та об'єднання, достатньо зосередити увагу на останніх двох операціях.

Також логічно, щоб клас  $\mathcal{C}$  був *замкненим* (closed) відносно цих операцій, тобто щоб застосування цих операцій не виводило нас за межі відповідного класу. У протилежному випадку з розгляду випадуть події, існування яких є цілком очікуваним. Такі класи множин мають назву алгебр.

### 1.5.1. Алгебри

**Визначення 1.5.1.** Клас  $\mathcal{F}$  підмножин деякого простору  $\Omega$  є *алгеброю* (algebra), якщо:

- (i)  $\Omega \in \mathcal{F}$ ;
- (ii) з  $A \in \mathcal{F}$  випливає  $A^c \in \mathcal{F}$ ;
- (iii) з  $A, B \in \mathcal{F}$  випливає  $A \cup B \in \mathcal{F}$ .

Також такий клас називають *полем* (field<sup>8</sup>).

□

**Твердження 1.5.2.** Із визначення алгебри безпосередньо випливають такі факти:

- (i)  $\emptyset \in \mathcal{F}$ ;
- (ii)  $\mathcal{F}$  замкнена відносно скінченногого числа об'єднань:

$$A_1, A_2, \dots, A_n \in \mathcal{F} \Rightarrow A_1 \cup A_2 \cup \dots \cup A_n \in \mathcal{F};$$

- (iii)  $\mathcal{F}$  замкнена відносно скінченногого числа перетинів:

$$A_1, A_2, \dots, A_n \in \mathcal{F} \Rightarrow A_1 \cap A_2 \cap \dots \cap A_n \in \mathcal{F}.$$

*Доведення.* (i) Згідно з властивостями (i) та (ii), враховуючи, що  $\Omega^c = \emptyset$ <sup>9</sup>;

<sup>7</sup>Події є підмножинами простору елементарних подій, проте є елементами  $\mathcal{C}$ :  $A \in \mathcal{C}$ , проте  $A \subseteq \Omega$ .

<sup>8</sup>Звідси відповідне позначення такого класу.

<sup>9</sup> $\mathcal{F}$  недарма називають алгеброю, адже  $\emptyset$  відіграє роль нуля,  $\Omega$  — одиниці, а операції  $\cap$  і  $\Delta$  — добутку і додавання, відповідно.

- (ii) це можна довести за методом математичної індукції. Справді, з визначення (iii) маємо виконання властивості для двох множин (база індукції). Нехай властивість виконується для  $n - 1$  множин. Тоді

$$A_1 \cup A_2 \cup \dots \cup A_n = (A_1 \cup \dots \cup A_{n-1}) \cup A_n.$$

Але оскільки  $A_1 \cup \dots \cup A_{n-1} \in \mathcal{F}$  за індукційним припущенням, то маємо об'єднання двох множин, а відтак можемо знову застосувати властивість (iii);

- (iii) за властивістю (ii),

$$(A_1 \cup A_2 \cup \dots \cup A_n)^c \in \mathcal{F},$$

тобто за законом де Моргана,

$$A_1^c \cap A_2^c \cap \dots \cap A_n^c \in \mathcal{F}.$$

Оскільки  $A_1, A_2, \dots, A_n \in \mathcal{F}$ , то за властивістю (ii) і  $A_1^c, A_2^c, \dots, A_n^c \in \mathcal{F}$ , а тому маємо:

$$A_1^c, A_2^c, \dots, A_n^c \in \mathcal{F} \Rightarrow A_1^c \cap A_2^c \cap \dots \cap A_n^c \in \mathcal{F},$$

отже маємо замкненість відносно скінченного числа перетинів.

□

**Приклад 1.5.3.** Деякі приклади алгебр:

- тривіальна алгебра:  $\mathcal{F} = \{\emptyset, \Omega\}$ ;
- повна алгебра: множина  $\mathcal{F} = 2^\Omega$  всіх підмножин  $\Omega$ ;
- алгебра на основі деякої множини  $A \subset \Omega$ :  $\mathcal{F} = \{\emptyset, A, A^c, \Omega\}$ ;
- алгебра на основі двох неперетинаних множин  $A, B \subset \Omega$ ,  $A \cap B = \emptyset$ :

$$\mathcal{F} = \{\emptyset, A, B, A^c, B^c, A \cup B, (A \cup B)^c, \Omega\} ;$$

- алгебра на основі скінченного розбиття: множина скінченних об'єднань множин зі скінченного розбиття деякого простору.

□

На дійсній осі нас цікавитиме конкретна алгебра  $\mathcal{B}_0$ , яку побудуємо так.

**Твердження 1.5.4.** Нехай  $\Omega = \mathbb{R}$ . Нехай  $\mathcal{I}$  — множина, яка містить порожню множину  $\emptyset$ , усі пів інтервали виду  $(a_1; a_2]$ ,  $a_1, a_2 \in \mathbb{R}$ ,  $a_1 < a_2$ , а також промені  $(-\infty; a]$ ,  $(b; \infty)$ ,  $a, b \in \mathbb{R}$ . Тоді клас  $\mathcal{B}_0$  скінченних об'єднань множин із  $\mathcal{I}$  є алгеброю.

*Доведення.* Перевірмо по черзі всі характеристики алгебр із Визначення 1.5.1:

- (i) нехай маємо деяке скінченне розбиття  $\mathbb{R}$ , яке складається з пів інтервалів та променів:

$$\Pi = \{(-\infty; a_1], (a_1; a_2], \dots, (a_n; \infty)\} .$$

Оскільки об'єднання елементів із  $\Pi$ , за визначенням розбиття, дає простір  $\Omega$ ,  $\Omega \in \mathcal{B}_0$ ;

(ii) розгляньмо декілька випадків. Якщо

$$A = (a_1; a'_1] \cup \dots \cup (a_n; a'_n] , \quad a_1 \leq \dots \leq a_n ,$$

де всі проміжки — неперетинанні (хоча деякі з них можуть бути порожніми, якщо  $a_i = a'_i$ ), то

$$A^c = (-\infty; a_1] \cup \dots \cup (a'_n; \infty) ,$$

тобто є скінченим об'єднанням пів інтервалів, а відтак належить  $\mathcal{B}_0$ .

Якщо  $A = (-\infty; a]$ , то  $A^c = (a; \infty)$ , і навпаки, тому обидва класи множин належать  $\mathcal{B}_0$ .

Таким чином, заперечення будь-якої множини  $A$ , яка є скінченим об'єднанням пів інтервалів усіх видів, належить  $\mathcal{B}_0$ ;

(iii) доведімо замкненість відносно перетину, що буде еквівалентно замкненості відносно об'єднання. Нехай

$$A = (-\infty; a_1] \cup (a_1; a'_1] \cup \dots \cup (a_m; a'_m] \cup (a'_m; \infty) , \quad a_1 \leq \dots \leq a_m ,$$

$$B = (-\infty; b_1] \cup (b_1; b'_1] \cup \dots \cup (b_n; b'_n] \cup (b'_n; \infty) , \quad b_1 \leq \dots \leq b_n ,$$

де всі проміжки неперетинанні. Тоді

$$\begin{aligned} A \cap B &= \left( (-\infty; a_1) \cup \left( \bigcup_{i=1}^m (a_i; a'_i] \right) \cup (a'_m; \infty) \right) \cap \left( (-\infty; b_1) \cup \left( \bigcup_{j=1}^n (b_j; b'_j] \right) \cup (b'_n; \infty) \right) \\ &= ((-\infty; a_1) \cap (-\infty; b_1)) \cup \left( \bigcup_{j=1}^n (-\infty; a_1) \cap (b_j; b'_j] \right) \cup ((-\infty; a_1) \cap (b'_n; \infty)) \\ &\quad \cup \left( \bigcup_{i=1}^m (a_i; a'_i] \cap (-\infty; b_1) \right) \cup \left( \bigcup_{i=1}^m \bigcup_{j=1}^n (a_i; a'_i] \cap (b_j; b'_j] \right) \cup \left( \bigcup_{i=1}^m (a_i; a'_i] \cap (b'_n; \infty) \right) \\ &\quad \cup ((a'_m; \infty) \cap (-\infty; b_1)) \cup \left( \bigcup_{j=1}^n (a'_m; \infty) \cap (b_j; b'_j] \right) \cup ((a'_m; \infty) \cap (b'_n; \infty)) . \end{aligned}$$

Таким чином, перетин  $A \cap B$  є об'єднанням перетинів або двох променів, або променя і пів інтервалу, або двох пів інтервалів. Кожний із цих перетинів сам по собі є або пів інтервалом, або променем, або порожньою множиною, відтак  $A \cap B$  належить  $\mathcal{B}_0$ .

□

**Зауваження 1.5.5.** Множини з окремих чисел  $\{x\}$ ,  $x \in \mathbb{R}$ , не належать  $\mathcal{B}_0$ , оскільки вони є результатом виконання (нескінченно) зліченних перетинів  $\bigcap_n (x - a_n; x]$ ,  $a_n \rightarrow 0$  (наприклад,  $a_n = 1/n$ ). □

**Зауваження 1.5.6.** Якщо розглядати тільки скінченні об'єднання й заперечення класу *обмежених* пів інтервалів  $\{(a; b]\}$ ,  $a, b \in \mathbb{R}$ , то вони не утворюватимуть алгебри, не тільки тому, що

весь простір  $\mathbb{R}$  неможливо утворити скінченним об'єднанням обмежених пів інтервалів, а й тому, що заперечення обмеженого пів інтервалу вже є об'єднанням двох необмежених:  $(a; b]^c = (-\infty; a] \cup (b, \infty)$ . Саме тому множина  $\mathcal{I}$ , окрім пів інтервалів, містить і промені.  $\square$

Із Прикладів 1.2.4–1.2.7 постає мотивація розширити клас цікавих для нас подій до результата (нескінченно) зліченного числа об'єднань (перетинів). Проте Визначення 1.5.1 не гарантує її замкненості відносно таких операцій.

**Приклад 1.5.7.** Розгляньмо алгебру  $\mathcal{B}_0$ . Збудуймо послідовність інтервалів

$$\left(0; \frac{1}{2}\right], \left(0; \frac{2}{3}\right], \dots, \left(0; 1 - \frac{1}{n}\right], \dots$$

Оскільки це є монотонно зростаюча послідовність вкладених інтервалів, скінченне об'єднання перших  $n$  із них дорівнює  $\left(0; 1 - \frac{1}{n}\right]$ , а відтак воно належить  $\mathcal{B}_0$ .

Будь-яка точка  $0 < x < 1$  належить усім інтервалам послідовності, починаючи з деякого  $n \geq \frac{1}{1-x}$ , а точка  $x = 1$  не належить жодному з них. Зліченним об'єднанням цих інтервалів є

$$\bigcup_{n=2}^{\infty} \left(0; 1 - \frac{1}{n}\right] = (0; 1),$$

але цей інтервал не належить  $\mathcal{B}_0$ , оскільки не містить своєї правої межі. Таким чином, алгебра  $\mathcal{B}_0$  не є замкненою відносно зліченного числа об'єднань.  $\square$

При цьому, очевидно, відкриті інтервали на кшталт  $(0; 1)$  є цілком бажаними подіями. Вимога замкненості відносно зліченного числа об'єднань веде до поняття  $\sigma$ -алгебри.

### 1.5.2. $\sigma$ -алгебри і вимірні простори

**Визначення 1.5.8.** Клас  $\mathcal{A}$  підмножин деякого простору  $\Omega$  є  $\sigma$ -алгеброю<sup>10</sup> ( $\sigma$ -algebra), якщо:

- (i)  $\Omega \in \mathcal{A}$ ;
- (ii) з  $A \in \mathcal{A}$  випливає  $A^c \in \mathcal{A}$ ;
- (iii) з  $A_1, A_2, \dots \in \mathcal{A}$  випливає  $\bigcup_{n \geq 1} A_n \in \mathcal{A}$ .

$\square$

Як у випадку алгебр, порожня множина належить  $\mathcal{A}$ , а сама  $\sigma$ -алгебра є замкненою також і відносно зліченного числа перетинів.

За визначенням, будь-яка  $\sigma$ -алгебра є одночасно алгеброю, але зворотне твердження не є справедливим, як показує Приклад 1.5.7.

**Визначення 1.5.9.** Пару  $(\Omega, \mathcal{A})$  називають *вимірним простором* (measurable space), а множини в  $\sigma$ -алгебрі  $\mathcal{A}$  — *вимірними множинами* (measurable sets).  $\square$

<sup>10</sup>Літера  $\sigma$  походить від першої літери німецького слова *Summe*.

Може виникнути питання: якщо основним об'єктом нашого інтересу є  $\sigma$ -алгебри як класи подій, для яких бажано визначити ймовірність, чи має самостійну роль поняття алгебри? Відповідь на це питання ствердна, але це стане зрозуміло тільки з наступної лекції.

Чому не можна покласти  $\mathcal{A} = 2^\Omega$ ? Як буде показано в наступній лекції, саме так і роблять у випадку скінченного (і навіть зліченного) простору елементарних подій. Проте в загальному випадку, як виявляється, неможливо визначити таку функцію, яка б зіставляла кожній множині з  $2^\Omega$  деяку ймовірність з інтуїтивно зрозумілими властивостями.

**Приклад 1.5.10.** Розгляньмо  $\Omega = (0; 1]$ . Покажімо, що на повній  $\sigma$ -алгебрі для цього проміжку неможливо визначити таку функцію, яка б кожній множині зіставляла ймовірність із такими інтуїтивно зрозумілими властивостями:

- $\mathbb{P}([a; b]) = b - a$ ,  $0 \leq a \leq b \leq 1$ ;
- якщо події  $A_1, A_2, \dots$  несумісні, то  $\mathbb{P}(\bigcup_n A_n) = \sum_n \mathbb{P}(A_n)$ , тобто ймовірності несумісних подій додаються для зліченного числа подій;
- $\mathbb{P}(A \otimes r) = \mathbb{P}(A)$ ,  $0 \leq r \leq 1$ , де

$$A \otimes r \equiv \{a + r ; a \in A, a + r \leq 1\} \cup \{a + r - 1 ; a \in A, a + r > 1\} ,$$

тобто ймовірності інваріантні відносно *циклічного зсуву на число  $r$  вправо* (адже довжина проміжку у цьому випадку не змінюється).

Побудуймо таку множину, для якої ймовірності з такими властивостями запропонувати неможливо<sup>11</sup>. Уведімо відношення еквівалентності (equivalence relation), таке, що  $x \sim y$  тоді й тільки тоді, коли  $x \otimes r = y$  для деякого ненульового раціонального числа з одиничного проміжку  $r$ :  $r \in \mathbb{Q} \cap (0; 1]$ . Це відношення утворює розбиття проміжку. Нехай  $H \subset (0; 1]$  містить в точності по одному представнику кожного класу<sup>12</sup>.

Множини  $H \otimes r_1$  і  $H \otimes r_2$  для  $r_1 \neq r_2$  неперетинанні. Дійсно: якщо цим множинам одночасно належить деяка точка  $h_1 \otimes r_1 = h_2 \otimes r_2$ , то виходить, що  $h_1 \sim h_2$ . Але це неможливо (якщо, звісно,  $h_1 \neq h_2$ ), тому що за побудовою всі точки  $H$  належать різним класам еквівалентності. Якщо ж  $h_1 = h_2$ , то тоді  $r_1 = r_2$ .

Більше того, кожна точка  $x \in (0; 1]$  належить якісь із цих множин, а відтак

$$(0; 1] = \bigcup_{r \in \mathbb{Q} \cap (0; 1]} (H \otimes r) .$$

Відповідно, імовірність обох множин повинна бути однакова:

$$\mathbb{P}((0; 1]) = \mathbb{P}\left(\bigcup_{r \in \mathbb{Q} \cap (0; 1]} (H \otimes r)\right) = \sum_r \mathbb{P}(H \otimes r) .$$

Оскільки ймовірність усього проміжку за визначенням дорівнює 1, а ймовірність  $\mathbb{P}(H \otimes r) = \mathbb{P}(H)$  завдяки інваріантності відносно зсуву,

$$1 = \sum_r \mathbb{P}(H) .$$

<sup>11</sup> Цей приклад запропонував італійський математик Джузеппе Віталі (Giuseppe Vitali, 1875–1932).

<sup>12</sup> Існування такої підмножини випливає з аксіоми вибору (Axiom of Choice).

Це неможливо, адже можуть існувати тільки два варіанти: або  $\mathbb{P}(H) = 0$ , і тоді  $1 = 0$ , або  $\mathbb{P}(H) = a > 0$ , і тоді  $1 = \infty$ .

Таким чином, існують такі підмножини одиничного проміжку, для яких неможливо визначити поняття ймовірності з бажаними властивостями.  $\square$

**Зауваження 1.5.11.** Як виявляється, існування таких проблемних множин є безпосереднім наслідком існування аксіоми вибору.  $\square$

**Зауваження 1.5.12.** Приклад 1.5.10 дає можливість зрозуміти, чому ми обмежилися замкненістю відносно саме зліченних об'єднань. Оскільки будь-яка множина є незліченним об'єднанням своїх елементів (наприклад,  $(0; 1] = \bigcup_{x \in (0; 1]} \{x\}$ ), дозвіл на використання незліченних об'єднань не дасть змоги виокремити множину подій, для яких можна визначити поняття ймовірності.  $\square$

Таким чином, ми доходимо висновку, що для коректного визначення ймовірності нам потрібно зосередитися на деякій  $\sigma$ -алгебрі подій, яка не дорівнює множині всіх підмножин простору елементарних подій. При цьому ми прагнемо, щоб у такій  $\sigma$ -алгебрі містилися всі події, які можуть становити практичний інтерес. Відтак ми розглянемо найменшу  $\sigma$ -алгебру, яка містить усі множини, що становлять інтерес.

**Визначення 1.5.13.** Для деякого довільного класу  $\mathcal{C}$  алгеброю, яку він породжує (algebra generated by  $\mathcal{C}$ ), називають найменшу алгебру, яка містить  $\mathcal{C}$ . Таку алгебру позначають через  $\mathcal{F}(\mathcal{C})$ .

Для деякого довільного класу  $\mathcal{C}$   $\sigma$ -алгеброю, яку він породжує ( $\sigma$ -algebra generated by  $\mathcal{C}$ ), називають найменшу  $\sigma$ -алгебру, яка містить  $\mathcal{C}$ . Таку  $\sigma$ -алгебру позначають через  $\mathcal{A}(\mathcal{C})$ .  $\square$

Виявляється,  $(\sigma)$ -алгебра, яку породжує деякий клас  $\mathcal{C}$ , не просто завжди існує, а й є єдиною.

**Лема 1.5.14.** Нехай  $\mathcal{C}$  — довільний клас підмножин простору  $\Omega$ . Тоді:

- (i) існує єдина найменша алгебра  $\mathcal{F}(\mathcal{C})$ , що містить  $\mathcal{C}$ ;
- (ii) існує єдина найменша  $\sigma$ -алгебра  $\mathcal{A}(\mathcal{C})$ , що містить  $\mathcal{C}$ .

**Доведення.** Доведемо тільки (i), адже (ii) можна довести аналогічно, замінивши всі скінченні об'єднання на зліченні.

Спочатку покажімо, що довільний перетин алгебр також є алгеброю. Дійсно, нехай  $\mathcal{F} = \bigcap_{i \in I} \mathcal{F}_i$ , де множина індексів  $I$  може бути незліченною, а кожна  $\mathcal{F}_i$  є алгеброю. Перевірмо усі властивості алгебр із Визначення 1.5.1:

- $\forall i \in I \ \Omega \in \mathcal{F}_i$ , а відтак  $\Omega \in \mathcal{F}$ ;
- якщо  $A \in \mathcal{F}$ , то, в силу властивостей перетину,  $\forall i \in I \ A \in \mathcal{F}_i$ . Тоді,  $\forall i \in I \ A^c \in \mathcal{F}_i$ , а відтак  $A^c \in \mathcal{F}$ ;
- якщо  $A_1, A_2, \dots, A_n \in \mathcal{F}$ , то  $\forall i \in I \ A_1, A_2, \dots, A_n \in \mathcal{F}_i$ , а відтак  $\forall i \in I \ \bigcup_n A_n \in \mathcal{F}_i$  за властивістю (iii) алгебр. За властивостями перетину, маємо  $\bigcup_n A_n \in \mathcal{F}$ .

Тоді можна покласти  $\mathcal{F}(\mathcal{C}) = \bigcap_{i \in I} \mathcal{F}_i$ , де  $\{\mathcal{F}_i, i \in I\}$  — непорожня множина всіх алгебр, що містять  $\mathcal{C}$ .  $\square$

Із Визначення 1.5.13 безпосередньо випливають деякі властивості.

**Лема 1.5.15.** Нехай  $\mathcal{C}, \mathcal{C}_1, \mathcal{C}_2$  — деякі класи множин. Тоді:

- (i)  $\mathcal{A}(\mathcal{A}(\mathcal{C})) = \mathcal{A}(\mathcal{C})$ ;
- (ii) якщо  $\mathcal{C}_1 \subseteq \mathcal{C}_2$ , то  $\mathcal{A}(\mathcal{C}_1) \subseteq \mathcal{A}(\mathcal{C}_2)$ ;
- (iii)  $\mathcal{A}(\mathcal{F}(\mathcal{C})) = \mathcal{A}(\mathcal{C})$ ;
- (iv) якщо  $\mathcal{C}_1 \subseteq \mathcal{C}_2 \subseteq \mathcal{A}(\mathcal{C}_1)$ , то  $\mathcal{A}(\mathcal{C}_2) = \mathcal{A}(\mathcal{C}_1)$ .

*Доведення.* (i) Очевидно;

- (ii) за визначенням,  $\mathcal{A}(\mathcal{C}_2)$  — найменша  $\sigma$ -алгебра, яка містить  $\mathcal{C}_2$ . Оскільки  $\mathcal{C}_2 \supseteq \mathcal{C}_1$ ,  $\mathcal{A}(\mathcal{C}_2)$  також є найменшою  $\sigma$ -алгеброю, яка містить  $\mathcal{C}_1$ , а відтак  $\mathcal{A}(\mathcal{C}_1) \subseteq \mathcal{A}(\mathcal{C}_2)$ ;
- (iii) оскільки  $\mathcal{C} \subseteq \mathcal{F}(\mathcal{C})$ , за властивістю (ii) маємо, що  $\mathcal{A}(\mathcal{C}) \subseteq \mathcal{A}(\mathcal{F}(\mathcal{C}))$ .

З іншого боку, за визначенням,  $\mathcal{A}(\mathcal{C}) = \bigcap_i \mathcal{A}_i$ , де кожна  $\sigma$ -алгебра  $\mathcal{A}_i \supseteq \mathcal{C}$ . Оскільки будь-яка  $\sigma$ -алгебра є алгеброю, а  $\mathcal{F}(\mathcal{C})$  — найменша з алгебр, що містить  $\mathcal{C}$ , то кожна  $\mathcal{A}_i \supseteq \mathcal{F}(\mathcal{C})$ . Відтак  $\mathcal{F}(\mathcal{C}) \subseteq \mathcal{A}(\mathcal{C})$ , оскільки належить кожній множині із відповідного перетину. Таким чином, за (i)–(ii),  $\mathcal{A}(\mathcal{F}(\mathcal{C})) \subseteq \mathcal{A}(\mathcal{A}(\mathcal{C})) = \mathcal{A}(\mathcal{C})$ .

$\mathcal{A}(\mathcal{C}) \subseteq \mathcal{A}(\mathcal{F}(\mathcal{C}))$  і  $\mathcal{A}(\mathcal{F}(\mathcal{C})) \subseteq \mathcal{A}(\mathcal{C})$  разом мають наслідком  $\mathcal{A}(\mathcal{C}) = \mathcal{A}(\mathcal{F}(\mathcal{C}))$ ;

- (iv) за (ii), з  $\mathcal{C}_1 \subseteq \mathcal{C}_2$  випливає  $\mathcal{A}(\mathcal{C}_1) \subseteq \mathcal{A}(\mathcal{C}_2)$ . Також за (i)–(ii) маємо, що з  $\mathcal{C}_2 \subseteq \mathcal{A}(\mathcal{C}_1)$  випливає  $\mathcal{A}(\mathcal{C}_2) \subseteq \mathcal{A}(\mathcal{C}_1)$ . Разом ці факти дають  $\mathcal{A}(\mathcal{C}_2) = \mathcal{A}(\mathcal{C}_1)$ .

$\square$

### 1.5.3. Борелева $\sigma$ -алгебра на дійсній осі

У Твердженні 1.5.4 ми увели поняття алгебри  $\mathcal{B}_0$ , яка містить скінченні об'єднання усіх можливих пів інтервалів і променів, які складають клас  $\mathcal{I}$ . Відповідно,  $\mathcal{B}_0 = \mathcal{F}(\mathcal{I})$ . Нас цікавить  $\sigma$ -алгебра, яку породжує  $\mathcal{B}_0$ . Виявляється, такою  $\sigma$ -алгеброю є так звана Борелева  $\sigma$ -алгебра.

**Визначення 1.5.16.**  $\sigma$ -алгебра  $\mathcal{B} = \mathcal{A}(\mathcal{O})$ , яку породжує клас усіх відкритих множин дійсної осі  $\mathcal{O}$ , має назву *Борелевої  $\sigma$ -алгебри* (Borel  $\sigma$ -algebra), а множини, що їй належать, називають *Борелевими множинами* (Borel sets).  $\square$

Які множини є Борелевими?

**Твердження 1.5.17.** (i) Усі замкнені множини є Борелевими;

- (ii) усі одноелементні множини  $\{x\}$ ,  $x \in \mathbb{R}$ , є Борелевими;

- (iii) усі зліченні множини є Борелевими;

- (iv) усі інтервали виду  $[a; b]$ ,  $(a; b]$ ,  $[a; b)$ ,  $(-\infty; b)$ ,  $[a; \infty)$  є Борелевими.

*Доведення.* (i) Будь-яка замкнена множина є запереченням відкритої множини;

(ii) кожну одноелементну множину  $\{x\}$  можна подати як злiченний перетин

$$\bigcap_{n=1}^{\infty} \left( x - \frac{1}{n}; x \right].$$

Альтернативно, кожна одноелементна множина є замкненою, тому належить  $\mathcal{B}_0$  за (i);

(iii) кожна злiченна множина є злiченним об'єднанням одноелементних множин;

(iv)  $[a; b] = (a; b) \cup \{a\} \cup \{b\}$ ,  $(a; b) = (a; b) \cup \{b\}$  тощо. □

Бiльше того, виявляється, що Борелевими є навiть множини, якi неможливо утворити злiченними теоретико-множинними операцiями над iнтервалами. Клас Борелевих множин є настiльки широким, що в рамках теорiї ймовiрностей бiльшiсть iз них niколи не виринає на практицi.

Доведiмо тепер, що Борелеву  $\sigma$ -алгебру може породити також алгебра  $\mathcal{B}_0$ .

**Лема 1.5.18.** Борелеву  $\sigma$ -алгебру породжує множина вiдкритих iнтервалiв  $\mathcal{I}_\mathcal{O} = \{(a; b) , a, b \in \mathbb{R}\}$ :  $\mathcal{B} = \mathcal{A}(\mathcal{I}_\mathcal{O})$ .

*Доведення.* Очевидно,  $\mathcal{I}_\mathcal{O} \subset \mathcal{O}$ , а тому  $\mathcal{A}(\mathcal{I}_\mathcal{O}) \subseteq \mathcal{A}(\mathcal{O}) = \mathcal{B}$  за Лемою 1.5.15.

З iншого боку,  $\mathcal{O} \subseteq \mathcal{A}(\mathcal{I}_\mathcal{O})$ , оскiльки будь-яку вiдкриту множину в  $\mathbb{R}$  можна утворити злiченними об'єднаннями вiдкритих iнтервалiв<sup>13</sup>. З  $\mathcal{O} \subseteq \mathcal{A}(\mathcal{I}_\mathcal{O})$  випливає  $\mathcal{B} = \mathcal{A}(\mathcal{O}) \subseteq \mathcal{A}(\mathcal{I}_\mathcal{O})$ .

Разом iз ранiше показаним вiключенням в iнший бiк констатуємо, що  $\mathcal{B} = \mathcal{A}(\mathcal{I}_\mathcal{O})$ . □

На основi цього твердження можна показати схожий результат.

**Лема 1.5.19.** Борелеву  $\sigma$ -алгебру породжує множина пiв iнтервалiв i променiв  $\mathcal{I}$ :  $\mathcal{B} = \mathcal{A}(\mathcal{I})$ .

*Доведення.* Доведення будуватимемо аналогiчно доведенню Леми 1.5.18. Спочатку потрiбно показати, що  $\mathcal{I}_\mathcal{O} \subseteq \mathcal{A}(\mathcal{I})$ , а вiдтак  $\mathcal{A}(\mathcal{I}_\mathcal{O}) = \mathcal{B} \subseteq \mathcal{A}(\mathcal{I})$ . Іншими словами, потрiбно показати, що вiдкритi iнтервали виду  $(a; b)$ ,  $a, b \in \mathbb{R}$ , можна утворити злiченними об'єднаннями i запереченнiми пiв iнтервалiв i променiв. Справдi, для всiх  $a, b \in \mathbb{R}$ ,  $(a; b) = \bigcup_{n=j}^{\infty} (a; b - 1/n]$  для деякого  $j$  такого, що  $b - 1/n > a$ .

Тепер покажiмо, що  $\mathcal{I} \subseteq \mathcal{A}(\mathcal{I}_\mathcal{O})$ , а вiдтак  $\mathcal{A}(\mathcal{I}) \subseteq \mathcal{A}(\mathcal{I}_\mathcal{O}) = \mathcal{B}$ . Справдi:

- $(a; b] = \bigcap_{n=1}^{\infty} (a; b + 1/n)$ ;
- $(-\infty; b] = \bigcap_{n=1}^{\infty} (-\infty; b + 1/n)$ ;
- $(a; \infty) = \bigcup_{n=1}^{\infty} (a; a + n)$ .

Із  $\mathcal{A}(\mathcal{I}_\mathcal{O}) = \mathcal{B} \subseteq \mathcal{A}(\mathcal{I})$  i  $\mathcal{A}(\mathcal{I}) \subseteq \mathcal{A}(\mathcal{I}_\mathcal{O}) = \mathcal{B}$  випливає  $\mathcal{A}(\mathcal{I}) = \mathcal{B}$ . □

**Твердження 1.5.20.** Борелеву  $\sigma$ -алгебру породжує алгебра  $\mathcal{B}_0$ :  $\mathcal{B} = \mathcal{A}(\mathcal{B}_0)$ .

*Доведення.* Це випливає з властивостi (iii) Леми 1.5.15, оскiльки  $\mathcal{B}_0 = \mathcal{F}(\mathcal{I})$ . □

<sup>13</sup>Цей результат вiдомий як лема Лiндельофа (Lindelöf Lemma), названа на честь фiнського математика Ернста Лiндельофа (Ernst Lindelöf, 1870–1946), яку ми доводити не будемо за браком часу, хоча доведення доволi простe.

**Зауваження 1.5.21.** Історично Борелеву  $\sigma$ -алгебру визначають як таку, яку породжує саме клас відкритих множин. Така побудова зручна тим, що відкриті множини мають визначення в будь-якому топологічному просторі. Проте множина скінчених операцій над відкритими множинами не є алгеброю (очевидно, ця множина не є замкненою відносно запереченння, адже неможливо утворити замкнену множину скінченним перетином відкритих множин), а тому, як буде видно з наступної лекції, не має практичної користі для побудови ймовірностей.

Тому в теорії ймовірностей загальноприйнято виводити  $\mathcal{B}$  з множини пів інтервалів (і променів). При цьому особливого значення не має, з якого боку інтервали відкриті чи замкнені. Проте сім'я множин  $\mathcal{I}$ , на відміну від тільки відкритих або тільки замкнених інтервалів, має корисну властивість замкненості відносно різниць множин. Той факт, що інтервал є відкритим саме зліва, є виключно даниною традиції.

Можна довести, що окрім пів інтервалів, незалежно від їх замкненості чи відкритості, породжують Борелеву  $\sigma$ -алгебру також *тільки* промені виду  $(-\infty; a]$ ,  $x \in \mathbb{R}$ , без пів інтервалів. Це випливає з того факту, що  $(a; b] = (-\infty; b] \setminus (-\infty; a])$ .  $\square$

Вимірний простір  $(\mathbb{R}, \mathcal{B})$  називають *Борелевою дійсною віссю* (Borel real line).

За аналогією з дійсною віссю можна увести поняття Борелевих множин у багатовимірному Евклідовому просторі.

**Визначення 1.5.22.**  $\sigma$ -алгебра  $\mathcal{B}^k = \mathcal{A}(\mathcal{O}^k)$ , яку породжує клас усіх відкритих множин  $\mathcal{O}^k \subseteq \mathbb{R}^k$ , має назву *Борелевої* (Borel), а множини, що їй належать, називають *Борелевими множинами*.  $\square$

Аналогічно одновимірному випадку, можна показати, що  $k$ -вимірну Борелеву  $\sigma$ -алгебру породжують також  $k$ -вимірні прямокутники  $I_1 \times I_2 \times \dots \times I_k \in \mathbb{R}^k$ , де кожний  $I_j \subset \mathcal{I}$ ,  $j = 1, \dots, k$ .

Також, услід за Зауваженням 1.5.21, можна стверджувати, що  $k$ -вимірну Борелеву  $\sigma$ -алгебру породжують  $k$ -вимірні аналоги променів  $(-\infty; x]$ , а саме множини виду

$$S_{\mathbf{x}} = \left\{ \mathbf{y} \in \mathbb{R}^k : y_i \leq x_i, i = 1, \dots, k \right\}, \quad \mathbf{x} = (x_1, \dots, x_k)^\top, \quad \mathbf{y} = (y_1, \dots, y_k)^\top.$$

Це можна бачити, зокрема, із того факту, що кожний прямокутник  $A = I_1 \times I_2 \times \dots \times I_k \in \mathbb{R}^k$  можна подати як

$$A = S_{(b_1, \dots, b_k)^\top} \setminus \left( S_{(a_1, b_2, \dots, b_k)^\top} \cup S_{(b_1, a_2, \dots, b_k)^\top} \cup \dots \cup S_{(b_1, b_2, \dots, a_k)^\top} \right).$$

На Рис. 1.5.1 наведено приклад у двовимірному просторі, де  $A = I_1 \times I_2 = (a_1; b_1] \times (a_2; b_2]$ .

**Визначення 1.5.23.** Вимірний простір  $(\mathbb{R}^k, \mathcal{B}^k)$  називають *Борелевим простором* (Borel space).  $\square$

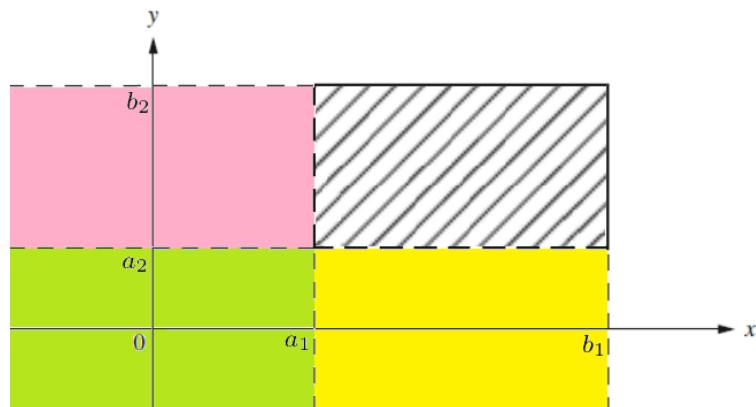


Рис. 1.5.1.: Подання обмеженого прямокутника  $A = (a_1; b_1] \times (a_2; b_2] \subset \mathbb{R}^2$  (штрихована область) через  $n$ -вимірні промені  $A = S_{(b_1, b_2)^\top} \setminus (S_{(a_1, b_2)^\top} \cup S_{(b_1, a_2)^\top})$ , де  $S_{(b_1, b_2)^\top}$  — об'єднання штрихованої та всіх зафарбованих областей,  $S_{(a_1, b_2)^\top}$  — об'єднання рожевої та зеленої областей, а  $S_{(b_1, a_2)^\top}$  — об'єднання зеленої та жовтої областей

## 2. Імовірність як міра

У цій лекції ми формально уведемо поняття ймовірності як міри відповідної множини і розглянемо її властивості. Почнемо ми з розглядом поняття міри в загальному випадку.

### 2.1. Аксіоматичне визначення міри

**Визначення 2.1.1.** *Функцією від множини* (set function) називають функцію  $f : \mathcal{C} \rightarrow \mathbb{R}$ , де  $\mathcal{C}$  — деякий клас підмножин простору  $\Omega$ :  $\mathcal{C} = \{A : A \subseteq \Omega\}$ .  $\square$

Цілком очевидно, що будь-яку функцію  $g : \mathbb{R} \rightarrow \mathbb{R}$  можна розглядати як функцію від множини таку, що  $g(\{x\}) \equiv g(x)$ . Відповідно в усьому нашому курсі ми використовуватимемо поняття «функція» для позначення будь-яких функцій, у тому числі функцій від множини.

Розгляньмо деякий вимірний простір  $(\Omega, \mathcal{A})$ . На такому просторі можна визначити поняття міри шляхом указання аксіом, які будь-яка корисна на практиці міра повинна задовільняти.

**Визначення 2.1.2.** Функція  $\mu : \mathcal{A} \rightarrow \mathbb{R}$  має назву *міри* (measure), якщо:

(i)  $\mu(A) \geq 0$  для всіх  $A \in \mathcal{A}$ ;

(ii)  $\mu$  є  $\sigma$ -адитивною ( $\sigma$ -additive): для деякої послідовності неперетинаних множин справедливо

$$\mu \left( \bigcup_{j=1}^{\infty} A_j \right) = \sum_{j=1}^{\infty} \mu(A_j), \quad A_j \in \mathcal{A}, \quad j = 1, 2, \dots, \quad A_i \cap A_j = \emptyset, \quad i \neq j. \quad (2.1.1)$$

$\square$

У загальному випадку мірою всього простору  $\mu(\Omega)$  може бути будь-яке дійсне число (і навіть нескінченно велике число). Міру називають *скінченою* (finite), якщо  $\mu(\Omega) < \infty$ . Також корисним у цьому курсі буде поняття  $\sigma$ -скінченої міри.

**Визначення 2.1.3.** Міра  $\mu$  є  $\sigma$ -скінченою ( $\sigma$ -finite), якщо  $\mu(\Omega) = \infty$ , але існує таке розбиття  $\{\mathcal{A}_i, i = 1, 2, \dots\}$  простору  $\Omega$ , що  $\mu(\mathcal{A}_i) < \infty$ ,  $i = 1, 2, \dots$ .  $\square$

**Приклад 2.1.4.** *Лічною мірою* (counting measure) на просторі  $(\Omega, \mathcal{A})$  називають таку міру  $\#$ , яка для будь-якої множини  $A \in \mathcal{A}$  зіставляє її потужність:

$$\#(A) = \begin{cases} n, & A = \{\omega_1, \dots, \omega_n\} \\ \infty, & \text{інакше} \end{cases}. \quad (2.1.2)$$

Якщо простір  $\Omega$  скінчений,  $|\Omega| < \infty$ , то лічна міра також є скінченою. Якщо ж  $\Omega$  містить (нескінченно) зліченну кількість елементів,  $|\Omega| = \aleph_0$ , то  $\#(\Omega) = \infty$ , а відтак лічна міра не є скінченою. При цьому вона є  $\sigma$ -скінченою, адже простір  $\Omega$  можна подати як зліченне об'єднання елементів, які він містить:  $\Omega = \{\omega_1\} \cup \{\omega_2\} \cup \dots$ . Кожна така множина має (скінченну) лічну міру 1.

При цьому варто зазначити, що якщо простір  $\Omega$  є незліченним,  $|\Omega| = \mathfrak{c}$ , то лічна міра вже не буде навіть  $\sigma$ -скінченою.  $\square$

**Визначення 2.1.5.** Міру  $\mathbb{P}$ , яка задовольняє всі аксіоматичні властивості з Визначення 2.1.2, і до того ж  $\mathbb{P}(\Omega) = 1$ , називають *імовірнісною мірою* (probability measure).  $\square$

**Визначення 2.1.6.** Трійку  $(\Omega, \mathcal{A}, \mu)$  називають *мірним простором* (measure space). Зокрема,  $(\Omega, \mathcal{A}, \mathbb{P})$  називають *імовірнісним простором* (probability space).  $\square$

Зазначені аксіоми відповідають інтуїтивним уявленням про ймовірність як математичну ідеалізацію поняття повторюваних незалежних випробувань — звідси випливало класичне визначення ймовірності (1.3.1), згідно з яким імовірність події, яка стається  $m$  разів із-поміж  $n$  незалежних випробувань, дорівнює  $\frac{m}{n}$ .

Те, що в силу властивостей правильного дробу  $0 \leq \frac{m}{n} \leq 1$ , веде до появи аксіоми (i) та відповідної нормалізації<sup>1</sup>. Також, якщо дві події стаються незалежно одна від одної, і їхні відносні частоти дорівнюють  $\frac{m_1}{n}$  і  $\frac{m_2}{n}$  відповідно, то імовірність настання принаймні однієї з них інтуїтивно повинна дорівнювати  $\frac{m_1+m_2}{n}$ . Цю інтуїцію формалізує аксіома (ii).

Незважаючи на те, що вищевказані аксіоми є дуже загальними і застосовні для будь-якого вимірного простору, уже їх самих достатньо, щоб довести низку важливих властивостей, якими володіє будь-яка міра, у тому числі ймовірнісна міра, яка нас цікавить у першу чергу.

## 2.2. Властивості міри

### 2.2.1. Основні властивості міри

Розгляньмо деякі властивості міри, які безпосередньо випливають із її аксіоматичного визначення.

**Теорема 2.2.1.** Будь-яка міра  $\mu$ , задана на мірному просторі  $(\Omega, \mathcal{A}, \mu)$ , має такі властивості:

- (i)  $\mu(\emptyset) = 0$ ;
- (ii) міра є скінченно-адитивною (адитивною, finitely additive):

$$\mu \left( \bigcup_{j=1}^n A_j \right) = \sum_{j=1}^n \mu(A_j), \quad A_j \in \mathcal{A}, \quad j = 1, 2, \dots, \quad A_i \cap A_j = \emptyset, \quad i \neq j; \quad (2.2.1)$$

<sup>1</sup>До речі, така нормалізація не єдина: у побуті ми часто послуговуємося відсотковими значеннями ймовірностей (напр., «імовірність дорівнює 50%»), тобто використовуємо нормалізацію  $\mathbb{P}(\Omega) = 100$ .

(iii)  $\mu$  є монотонно неспадною (monotonically nondecreasing):

$$\mu(A_1) \leq \mu(A_2) , \quad A_1, A_2 \in \mathcal{A} , \quad A_1 \subseteq A_2$$

Як наслідок, якщо  $A_1 \subseteq A_2$ , то  $\mu(A_2 \setminus A_1) = \mu(A_2 \cap A_1^c) = \mu(A_2) - \mu(A_1)$ ;

(iv)  $\mu$  є  $\sigma$ -субаддитивною ( $\sigma$ -subadditive)

$$\mu \left( \bigcup_{j=1}^{\infty} A_j \right) \leq \sum_{j=1}^{\infty} \mu(A_j) , \quad A_j \in \mathcal{A} , \quad j = 1, 2, \dots , \quad (2.2.2)$$

тобто відповідна властивість справедлива для будь-яких вимірних множин, *далеко не тільки неперетинаних*.

Ця нерівність також відома як *нерівність Була* (Boole's inequality)<sup>2</sup>;

(v)  $\mu(A \cup B) = \mu(A) + \mu(B) - \mu(A \cap B)$ .

*Доведення.* (i) Розгляньмо деяку непорожню множину  $A \in \mathcal{A}$  таку, що  $\mu(A) < \infty$ . Тоді можна покласти

$$A = A \cup \emptyset \cup \emptyset \dots ,$$

і за аксіомою (ii) матимемо, що

$$\mu(A) = \mu(A) + \mu(\emptyset) + \dots ,$$

звідки випливає, що  $\mu(\emptyset) = 0$ ;

(ii) формально можемо визначити таку послідовність неперетинаних множин  $A_i$ , що  $A_i = \emptyset$  для  $i > n$ . Тоді

$$\mu \left( \bigcup_{i=1}^n A_i \right) = \mu \left( \bigcup_{i=1}^n A_i \cup \bigcup_{n+1}^{\infty} \emptyset \right) = \sum_{i=1}^n \mu(A_i) + \sum_{n+1}^{\infty} 0 = \sum_{i=1}^n \mu(A_i) ;$$

(iii) якщо  $A_1 \subseteq A_2$ , то  $A_2 = A_1 \cup (A_2 \setminus A_1) = A_1 \cup (A_2 \cap A_1^c)$  — об'єднання двох неперетинаних множин. Відповідно,

$$\mu(A_2) = \mu(A_1) + \mu(A_2 \setminus A_1) \geq \mu(A_1) ,$$

оскільки міра невід'ємна за аксіомою (i). Звідси очевидно, що

$$\mu(A_2 \setminus A_1) = \mu(A_2) - \mu(A_1) ;$$

(iv) перепишімо зліченне об'єднання таким чином:

$$\bigcup_{i=1}^{\infty} A_i = A_1 + (A_2 \setminus A_1) + \dots + (A_{n+1} \setminus A_n \setminus \dots \setminus A_1) + \dots ,$$

де всі множини справа будуть неперетинаними (ілюстрацію для трьох множин наведено

<sup>2</sup>Названа так на честь видатного англійського математика Джорджа Була (George Boole, 1815–1864).

на Рис. 2.2.1). Тоді за аксіомою (ii) маємо:

$$\begin{aligned}\mu\left(\bigcup_{i=1}^{\infty} A_i\right) &= \mu(A_1) + \mu(A_2 \setminus A_1) + \dots + \mu(A_{n+1} \setminus A_n \setminus \dots \setminus A_1) + \dots \\ &\leq \mu(A_1) + \mu(A_2) + \dots + \mu(A_n) + \dots \\ &= \sum_{i=1}^{\infty} \mu(A_i),\end{aligned}$$

де ми використали властивість монотонності, щоб показати, що  $\mu(A \setminus B) \leq \mu(A)$ ;

- (v) нескладно бачити, що  $A \cup B = A \cup (B \cap A^c)$ , тобто маємо об'єднання двох неперетинаних множин. Тоді застосовна властивість адитивності:

$$\mu(A \cup B) = \mu(A \cup (B \cap A^c)) = \mu(A) + \mu(B \cap A^c).$$

Для обчислення  $\mu(B \cap A^c)$  потрібно помітити, що з властивостей теоретико-множинних операцій маємо подання множини  $B$  як об'єднання двох неперетинаних множин:

$$B = (B \cap A) \cup (B \cap A^c).$$

Тоді застосовна властивість адитивності:

$$\mu(B) = \mu((B \cap A) \cup (B \cap A^c)) = \mu(B \cap A) + \mu(B \cap A^c),$$

звідки випливає

$$\mu(B \cap A^c) = \mu(B) - \mu(A \cap B).$$

Таким чином, остаточно маємо:

$$\mu(A \cup B) = \mu(A) + \mu(B \cap A^c) = \mu(A) + \mu(B) - \mu(A \cap B).$$

□

**Зауваження 2.2.2.** Хід із поданням об'єднання в загальному випадку неперетинаних множин через об'єднання відповідних заперечень  $\mu(A_1) + \mu(A_2 \setminus A_1) + \dots$  дає змогу оперувати об'єднанням множин неперетинаних, що уможливлює застосування аксіоми (ii). Цей хід є стандартним у доведенні подібного роду властивостей. □

**Зауваження 2.2.3.** З урахуванням того, що для ймовірнісної міри  $\mathbb{P}$  справедлива нормалізація  $\mathbb{P}(\Omega) = 1$ , маємо дві властивості, специфічні саме для ймовірнісної міри:

- $\mathbb{P}(A^c) = 1 - \mathbb{P}(A)$ , оскільки  $\Omega = A \cup A^c$ , і тому  $1 = \mathbb{P}(\Omega) = \mathbb{P}(A) + \mathbb{P}(A^c)$ ;
- $\mathbb{P}(A) \leq 1$ , що випливає з невід'ємності ймовірності: оскільки  $\mathbb{P}(A) \geq 0$ ,  $\mathbb{P}(A) \leq \mathbb{P}(A) + \mathbb{P}(A^c) = 1$ .

□

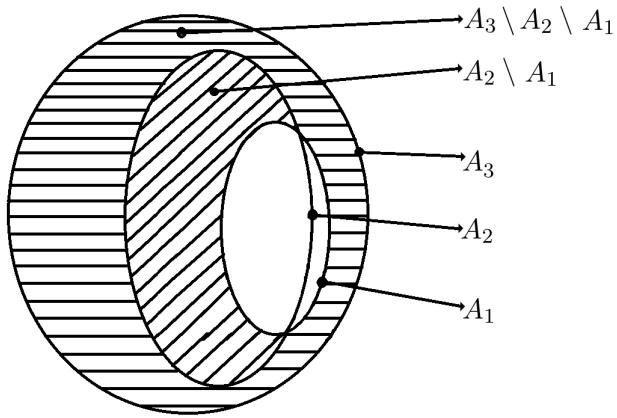


Рис. 2.2.1.: Ілюстрація до доведення Теореми 2.2.1: три множини  $A_1, A_2, A_3$  та їх відносні різниці — заштрихована косими лініями множина  $A_2 \setminus A_1$  та заштрихована горизонтальними лініями множина  $A_3 \setminus A_2 \setminus A_1$

**Приклад 2.2.4.** Нехай  $A$  і  $B$  — події, що два контракти, I і II, виконано в рамках своїх дедлайнів, відповідно. Нехай імовірність того, що принаймні один контракт виконано в рамках дедлайну, дорівнює 0.9, а імовірність того, що обидва контракти виконано в рамках дедлайнів, дорівнює 0.5. Чому дорівнює імовірність того, що тільки один із контрактів виконано в рамках дедлайну?

Надану інформацію можна формалізувати так:

$$\mathbb{P}(A \cup B) = 0.9, \quad \mathbb{P}(A \cap B) = 0.5.$$

Потрібно знайти імовірність події  $A \Delta B = (A \cap B^c) \cup (A^c \cap B)$ . Шляхом елементарних операцій можна показати, що ця подія також має форму  $(A \cup B) \setminus (A \cap B)$ , а відтак, за Теоремою 2.2.1(iii),

$$\mathbb{P}(A \Delta B) = \mathbb{P}((A \cap B^c) \cup (A^c \cap B)) = \mathbb{P}((A \cup B) \setminus (A \cap B)) = \mathbb{P}(A \cup B) - \mathbb{P}(A \cap B) = 0.9 - 0.5 = 0.4.$$

□

**Приклад 2.2.5.** У мішку міститься 6 талонів, пронумерованих від 1 до 6, із такими відносними частотами, які можна вважати імовірностями

номер $i$	1	2	3	4	5	6
$\mathbb{P}(i)$	$\frac{1}{4}$	$\frac{1}{8}$	$\frac{1}{8}$	$\frac{1}{8}$	$\frac{1}{8}$	$\frac{1}{4}$

Розгляньмо подію  $A$  = «випадково витягнутий талон має номер 3 або вищий». Події  $A_i$  = «витягнуто талон  $i$ » є несумісні (вони не можуть відбуватися одночасно), тому маємо:

$$\mathbb{P}(A) = \mathbb{P}(A_3 \cup A_4 \cup A_5 \cup A_6) = \mathbb{P}(A_3) + \mathbb{P}(A_4) + \mathbb{P}(A_5) + \mathbb{P}(A_6) = \frac{1}{8} + \frac{1}{8} + \frac{1}{8} + \frac{1}{4} = \frac{5}{8}.$$

Розгляньмо подію  $B$  = «витягнуто талон із парним номером». Імовірність цієї події така:

$$\mathbb{P}(B) = \mathbb{P}(A_2 + A_4 + A_6) = \frac{1}{8} + \frac{1}{8} + \frac{1}{4} = \frac{1}{2}.$$

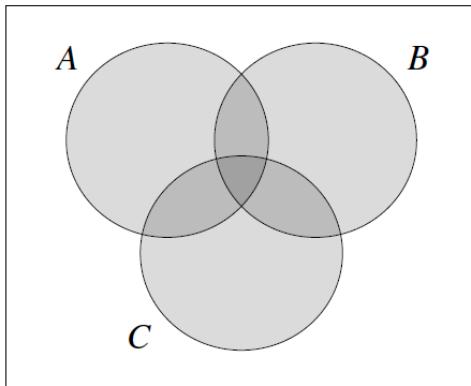


Рис. 2.2.2.: Діаграма Венна для трьох перетинаних множин ([1], стор. 24)

Тоді ймовірність, що витягнутий талон має парний номер чи його номер — 3 або вищий, обчислімо так:

$$\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B) - \mathbb{P}(A \cap B) .$$

Подія  $A \cap B$  має інтерпретацію витягу талону, який має парний номер, і одночасно цей номер щонайменше 3. Таким чином,

$$\mathbb{P}(A \cap B) = \mathbb{P}(A_4 \cup A_6) = \frac{1}{8} + \frac{1}{4} = \frac{3}{8} .$$

Остаточно маємо:

$$\mathbb{P}(A \cup B) = \frac{1}{2} + \frac{5}{8} - \frac{3}{8} = \frac{3}{4} .$$

□

Властивість із Теореми 2.2.1(v) можна узагальнити на довільну кількість множин. Сформулюймо відповідну теорему в застосуванні до ймовірнісних мір, хоча вона справедлива і для довільних мір також.

**Теорема 2.2.6** (Теорема включення-виключення (Inclusion-exclusion theorem)). Нехай  $(\Omega, \mathcal{A}, \mathbb{P})$  — деякий імовірнісний простір. Тоді

$$\mathbb{P}\left(\bigcup_{i=1}^n A_i\right) = \sum_i \mathbb{P}(A_i) - \sum_{i < j} \mathbb{P}(A_i \cap A_j) + \sum_{i < j < k} \mathbb{P}(A_i \cap A_j \cap A_k) - \dots + (-1)^{n+1} \mathbb{P}(A_1 \cap \dots \cap A_n) . \quad (2.2.3)$$

Теорему можна довести за індукцією, але ми цього тут робити не будемо.

Інтуїцію теореми легко показати у випадку трьох подій (Рис. 2.2.2). Для обчислення ймовірності об'єднання трьох перетинаних подій,  $\mathbb{P}(A \cup B \cup C)$ , ми спочатку додаємо ймовірності всіх подій окремо, але оскільки заштриховані області було враховано двічі, ми віднімаємо  $\mathbb{P}(A \cap B) + \mathbb{P}(A \cap C) + \mathbb{P}(B \cap C)$ , щоб уникнути подвійного підрахунку. Але в результаті маленький заштрихований перетин усіх трьох множин було спочатку тричі додано, а потім тричі віднято, і тому ми повинні додати його окремо. За цією ж логікою відбувається підрахунок імовірності об'єднання довільного числа множин.

### 2.2.2. Неперервність міри

Для доведення інших корисних властивостей дамо таке визначення.

**Визначення 2.2.7.** Міра  $\mu$ , задана на мірному просторі  $(\Omega, \mathcal{A}, \mu)$ , *неперервна знизу* (continuous from below), якщо для неспадної послідовності вимірних множин  $A_i \in \mathcal{A}$ ,  $i = 1, 2, \dots$  (тобто  $A_1 \subseteq A_2 \subseteq \dots$ ),

$$\lim_{i \rightarrow \infty} \mu(A_i) = \mu \left( \lim_{i \rightarrow \infty} A_i \right) = \mu \left( \bigcup_{i=1}^{\infty} A_i \right).$$

Міра  $\mu$  *неперервна зверху* (continuous from above), якщо для незростаючої послідовності вимірних множин  $A_i \in \mathcal{A}$ ,  $i = 1, 2, \dots$  (тобто  $A_1 \supseteq A_2 \supseteq \dots$ ),

$$\lim_{i \rightarrow \infty} \mu(A_i) = \mu \left( \lim_{i \rightarrow \infty} A_i \right) = \mu \left( \bigcap_{i=1}^{\infty} A_i \right).$$

Нарешті, міра  $\mu$  є просто *неперервною*, якщо вона одночасно неперервна як зверху, так і знизу.

Кажемо, що міра *неперервна в порожній множині* (continuous at the empty set), якщо вона неперервна зверху для незростаючої послідовності вимірних множин  $A_1 \supseteq A_2 \supseteq \dots$  такої, що  $A_i \rightarrow \emptyset$ , тобто якщо  $\lim_{i \rightarrow \infty} \mu(A_i) = 0$ .  $\square$

Наступна теорема з'єднує поняття адитивності та неперервності.

**Теорема 2.2.8.** (i) Будь-яка міра  $\mu$ , задана на мірному просторі  $(\Omega, \mathcal{A}, \mu)$ , неперервна;

(ii) якщо деяка скінченна функція  $\mu$  невід'ємна, адитивна і неперервна в порожній множині, а  $\mu(\emptyset) = 0$ , то ця функція є  $\sigma$ -адитивною (тобто є повноцінною мірою).

*Доведення.* (i) Спочатку покажімо неперервність знизу. Нехай маємо неспадну послідовність множин  $A_i \in \mathcal{A}$ ,  $i = 1, 2, \dots$ . Тоді, використовуючи хід із Заваження 2.2.2, маємо:

$$\begin{aligned} \lim_{i \rightarrow \infty} A_i &= \bigcup_{i=1}^{\infty} A_i = A_1 + (A_2 \setminus A_1) + \dots + (A_{n+1} \setminus A_n \setminus \dots \setminus A_1) + \dots \\ &= A_1 + (A_2 \setminus A_1) + \dots + (A_{n+1} \setminus A_n) + \dots, \end{aligned}$$

де ми врахували неспадну природу послідовності множин.

Позначивши  $B_1 \equiv A_1$ ,  $B_i \equiv A_i \setminus A_{i-1}$ ,  $i > 1$ , маємо:

$$\begin{aligned}
 \mu \left( \lim_{i \rightarrow \infty} A_i \right) &= \mu \left( \lim_{i \rightarrow \infty} B_i \right) \\
 &= \sum_{i=1}^{\infty} \mu(B_i) \\
 &= \lim_{n \rightarrow \infty} \sum_{i=1}^n \mu(B_i) \\
 &= \lim_{n \rightarrow \infty} (\mu(A_1) + \mu(A_2 \setminus A_1) + \dots + \mu(A_n \setminus A_{n-1})) \\
 &= \lim_{n \rightarrow \infty} (\mu(A_1) + \mu(A_2) - \mu(A_1) + \dots + \mu(A_n) - \mu(A_{n-1})) \\
 &= \lim_{i \rightarrow \infty} \mu(A_i) .
 \end{aligned}$$

Доведення неперервності зверху аналогічне, оскільки незростаючій послідовності  $A_i$  відповідає неспадна послідовність  $A_i^c$ , до того ж

$$\bigcup_{i=1}^{\infty} A_i^c = \left( \bigcap_{i=1}^{\infty} A_i \right)^c .$$

Із попереднього результату випливає, що

$$\begin{aligned}
 \lim_{i \rightarrow \infty} \mu(\Omega \setminus A_i) &= \lim_{i \rightarrow \infty} \mu(A_i^c) \\
 &= \mu \left( \lim_{i \rightarrow \infty} A_i^c \right) \\
 &= \mu \left( \bigcup_{i=1}^{\infty} A_i^c \right) \\
 &= \mu \left( \left( \bigcap_{i=1}^{\infty} A_i \right)^c \right) \\
 &= \mu \left( \Omega \setminus \bigcap_{i=1}^{\infty} A_i \right) \\
 &= \mu(\Omega) - \mu \left( \bigcap_{i=1}^{\infty} A_i \right) ,
 \end{aligned}$$

тобто  $\lim_{i \rightarrow \infty} \mu(A_i) = \mu \left( \bigcap_{i=1}^{\infty} A_i \right)$ ;

- (ii) нехай множини  $A_i \in \mathcal{A}$ ,  $i = 1, 2, \dots$ , неперетинанні і  $A = \bigcup_{i=1}^{\infty} A_i$ . Розгляньмо множини  $C_n = \bigcup_{i>n} A_i = A \setminus \left( \bigcup_{i \leq n} A_i \right)$ . Вони утворюють незростаючу послідовність, до того ж  $C_n \rightarrow \emptyset$ .

Ураховуючи неперервність у порожній множині, маємо:

$$\lim_{n \rightarrow \infty} \mu(C_n) = 0 \Rightarrow \lim_{n \rightarrow \infty} \mu \left( A \setminus \left( \bigcup_{i \leq n} A_i \right) \right) = \mu(A) - \lim_{n \rightarrow \infty} \sum_{i=1}^n \mu(A_i) = \mu(A) - \sum_{i=1}^{\infty} \mu(A_i) = 0 ,$$

а відтак

$$\mu(A) = \sum_{i=1}^{\infty} \mu(A_i) .$$

□

**Зауваження 2.2.9.** Визначення ймовірності, що безпосередньо випливає з Визначення 2.1.5 шляхом нормалізації, є сучасним (див., наприклад, [1], [3], [4] тощо).

Історично поняття ймовірності було введено аксіоматично радянським математиком Андрієм Колмогоровим (1903–1987) у 1933 р. Серед іншого, до аксіом імовірності належить не  $\sigma$ -адитивність, а просто адитивність:  $\mathbb{P}(\bigcup_{i=1}^n A_i) = \sum_{i=1}^n \mathbb{P}(A_i)$ , для деяких неперетинаних множин. Можна показати, що з адитивності та неперервності випливає  $\sigma$ -адитивність.

За великим рахунком, не принципово, що з чого випливає, головне, що всі потрібні властивості міри можна вивести з певного дуже обмеженого набору інтуїтивно зрозумілих аксіом. □

## 2.3. Імовірнісна міра на зліченному просторі елементарних подій

Для будь-якого зліченного простору елементарних подій  $\Omega$ , як було зазначено на попередній лекції, природною  $\sigma$ -алгеброю є повна  $\sigma$ -алгебра  $2^\Omega$  — множина всіх підмножин  $\Omega$ . Імовірнісну міру для вимірного простору  $(\Omega, 2^\Omega)$  можна задати так.

**Теорема 2.3.1.** Нехай  $(\Omega, 2^\Omega)$  — деякий злічений вимірний простір. Нехай функція  $p : \Omega \rightarrow [0; 1]$  така, що  $\sum_{\omega \in \Omega} p(\omega) = 1$ . Тоді функція  $\mathbb{P} : 2^\Omega \rightarrow [0; 1]$  така, що

$$\mathbb{P}(A) = \sum_{\omega \in A} p(\omega) , \quad (2.3.1)$$

є ймовірнісною мірою.

*Доведення.* По-перше, цілком очевидно, що відповідна функція є невід'ємною.

На всьому просторі функція набуває значення 1:

$$\mathbb{P}(\Omega) = \sum_{\omega \in \Omega} p(\omega) = 1 .$$

Для доведення властивості  $\sigma$ -адитивності розгляньмо  $A = \bigcup_{i=1}^{\infty} A_i$ , де  $A_i \in 2^\Omega$  і є неперетинаними. Кожну множину можна вписати явно:  $A_i = \{\omega_j, j \in I_i\}$ , де  $I_i$  — деяка множина індексів, до того ж  $I_i \cap I_j = \emptyset$ ,  $i \neq j$ . Використовуючи (2.3.1), маємо:

$$\mathbb{P}(A) = \sum_{\omega \in \bigcup_{i=1}^{\infty} A_i} p(\omega) = \sum_{i=1}^{\infty} \sum_{j \in I_i} p(\omega_j) = \sum_{i=1}^{\infty} \mathbb{P}(A_i) .$$

Таким чином, трійка  $(\Omega, 2^\Omega, \mathbb{P})$  утворює ймовірнісний простір.  $\square$

**Визначення 2.3.2.** Імовірнісну міру, утворену відповідно до Теореми 2.3.1, називають *дискретною* (discrete), як і сам імовірнісний простір.  $\square$

За допомогою Теореми 2.3.1 можна утворювати будь-які дискретні ймовірнісні міри.

**Приклад 2.3.3.** Найпростіший спосіб побудови дискретної ймовірнісної міри на *скінченному* просторі  $(\Omega, 2^\Omega)$  полягає у виборі  $p(\omega) = \frac{1}{|\Omega|}$ . Таке формулювання повністю відповідатиме поняттю класичної ймовірності.

Інші варіанти побудови дискретних імовірнісних мір розглядатимемо далі в цьому курсі.  $\square$

## 2.4. Міра Лебега на одиничному інтервалі

У випадку неперервних просторів елементарних подій визначення ймовірнісної міри можна коректно здійснити, використовуючи фундаментальний апарат теорії міри — *теорему Каратеодорі*<sup>3</sup>.

**Теорема 2.4.1** (Теорема Каратеодорі (Carathéodory theorem)<sup>4</sup>). Нехай на деякій алгебрі  $\mathcal{F}$  задана міра  $\mu$ . Тоді цю міру можна подовжити на  $\sigma$ -алгебру  $\mathcal{A}(\mathcal{F})$ , породжену цією алгеброю. Причому якщо міра скінчена (або  $\sigma$ -скінчена), то таке подовження єдине і скінченнє ( $\sigma$ -скінченнє).

Цей результат ми доводити не будемо, адже він є ключовим для теорії міри, а не теорії ймовірностей. Але без його застосування визначити корисні ймовірнісні міри неможливо.

Теорема 2.4.1 каже, що для міри  $\mu : \mathcal{F} \rightarrow \mathbb{R}^+$  існує *подовження* (extension)  $\nu : \mathcal{A}(\mathcal{F}) \rightarrow \mathbb{R}^+$  таке, що  $\mu(A) = \nu(A)$  для всіх  $A \in \mathcal{F}$ . Якщо ж міра  $\mu$  скінчена (або  $\sigma$ -скінчена), то подовження єдине в тому сенсі, що якщо існує інша міра  $\nu' : \mathcal{A}(\mathcal{F}) \rightarrow \mathbb{R}^+$  така, що  $\nu'(A) = \mu(A)$  для всіх  $A \in \mathcal{F}$ , то  $\nu(A) = \nu'(A)$  для всіх  $A \in \mathcal{A}(\mathcal{F})$ .

На практиці застосування Теореми 2.4.1 здійснюють так:

- спочатку визначають деяку функцію  $\mu$  на множинах деякого малого класу  $\mathcal{C}$ ;
- перевіряють, що функція  $\mu$  є мірою, тобто що вона є невід'ємною та  $\sigma$ -адитивною, на алгебрі  $\mathcal{F}(\mathcal{C})$ , породжений класом  $\mathcal{C}$ ;
- використовують Теорему 2.4.1 для подовження відповідної міри на  $\sigma$ -алгебру  $\mathcal{A}(\mathcal{C})$ , породжену класом  $\mathcal{C}$ , адже за Лемою 1.5.15 це буде також  $\sigma$ -алгебра, породжена алгеброю  $\mathcal{F}(\mathcal{C})$ .

Оскільки перевірку властивостей міри простіше зробити на множинах, утворених скінченими, а не (нескінченно) зліченими об'єднаннями деяких базових множин, поняття алгебри має самостійне значення для побудови ймовірнісних мір.

Для прикладу розгляньмо процес побудови ймовірнісної міри  $\mu$  на вимірному просторі  $((0; 1]; \mathcal{B}(0; 1])$ , тобто на одиничному пів інтервалі з підмножиною Борелевої  $\sigma$ -алгебри, визначеною на цьому пів інтервалі.

Для цього ми слідуватимемо такими кроками:

<sup>3</sup>У літературі під назвою теореми Каратеодорі вказують доволі різні результати, зокрема, варіант, запропонований у цьому курсі лекцій, часто називають *теоремою Хана-Колмогорова* (Hahn-Kolmogorov theorem). Цей варіант теореми розглядається, наприклад, у [3]—[6].

<sup>4</sup>Названа так на честь грецько-німецького математика Константина Каратеодорі (Constantin Carathéodory, 1873–1950).

- визначимо міру  $\mu$  на класі  $\mathcal{I}(0; 1]$ , який є підмножиною класу  $\mathcal{I}$ , розглянутого в Розд. 1.5 (клас пів інтервалів і променів), визначену на пів інтервалі  $(0; 1]$ . Зрозуміло, що на цьому пів інтервалі про промені не може бути мови, тому  $\mathcal{I}(0; 1]$  містить порожню множину  $\emptyset$  та пів інтервали виду  $(a_1; a_2]$ ,  $a_1, a_2 \in (0; 1]$ ;
- перевіримо, що функція  $\mu$  є мірою, тобто що вона є невід'ємною та  $\sigma$ -адитивною, на алгебрі  $\mathcal{B}_0(0; 1] = \mathcal{F}(\mathcal{I}(0; 1])$ , тобто на алгебрі скінченних заперечень та об'єднань пів інтервалів із  $(0; 1]$ ;
- використаємо Теорему 2.4.1 для подовження відповідної міри на  $\sigma$ -алгебру  $\mathcal{B}(0; 1]$ .

Міру  $\mu$  визначмо як  $\mu((a; b]) = b - a$ , тобто мірою деякого інтервалу буде його довжина. Оскільки міра (довжина) усього простору є  $\mu((0; 1]) = 1 - 0 = 1$ , то ми в результаті дістанемо не просто міру, а ймовірнісну міру, яка і буде строго формальним узагальненням поняття класичної ймовірності як відносної частоти настання події.

Тепер потрібно показати, що довжина пів інтервалу є мірою на алгебрі  $\mathcal{B}_0(0; 1]$  зі скінченних об'єднань і заперечень пів інтервалів із  $(0; 1]$ .

**Теорема 2.4.2.** Нехай маємо пару  $((0; 1], \mathcal{B}_0(0; 1])$ . Тоді міра на відповідній алгебрі, яку визна- чають як  $\mu((a; b]) = b - a$ <sup>5</sup>, є мірою в розумінні Визначення 2.1.2.

*Доведення.* Те, що довжина інтервалу є невід'ємною, є самоочевидним.

Для того, щоб показати  $\sigma$ -адитивність міри  $\mu$  на алгебрі  $\mathcal{B}_0(0; 1]$ , ми повинні використати властивість  $\sigma$ -адитивності довжин інтервалу, тобто якщо  $A = \bigcup_{i=1}^{\infty} I_i$ , де  $I_i, i = 1, \dots, n$  — неперетинані пів інтервали з  $(0; 1]$  (у тому числі порожні множини), то

$$\mu(A) = \sum_{i=1}^{\infty} \mu(I_i).$$

Формальне доведення цього факту наведено в Розд. 2.6.

Нехай  $A = \bigcup_{k=1}^{\infty} A_k$ , де  $A, A_1, A_2, \dots \in \mathcal{B}_0(0; 1]$ , і всі  $A_k$  неперетинані<sup>6</sup>. Оскільки ці множини належать алгебрі скінченних об'єднань інтервалів,  $A = \bigcup_{i=1}^n I_i$  і  $A_k = \bigcup_{j=1}^{m_k} J_{k,j}$ , де пів інтервали у відповідних сумах є неперетинані пів інтервали з  $\mathcal{I}(0; 1]$ . Використовуючи результат Леми 2.6.4, маємо:

$$\begin{aligned} \mu(A) &= \sum_{i=1}^n \mu(I_i) = \sum_{i=1}^n \sum_{k=1}^{\infty} \mu(I_i \cap A_k) = \sum_{i=1}^n \sum_{k=1}^{\infty} \mu\left(I_i \cap \left(\bigcup_{j=1}^{m_k} J_{k,j}\right)\right) \\ &= \sum_{i=1}^n \sum_{k=1}^{\infty} \sum_{j=1}^{m_k} \mu(I_i \cap J_{k,j}) = \sum_{k=1}^{\infty} \sum_{j=1}^{m_k} \mu(J_{k,j}) = \sum_{k=1}^{\infty} \mu(A_k). \end{aligned}$$

□

Використовуючи цей результат, можемо застосувати Теорему 2.4.1 і подовжити довжини інтервалів як міру на Борелеву  $\sigma$ -алгебру  $\mathcal{B}(0; 1]$ . Результат цього подовження має називати *міри Лебега* (Lebesgue measure)<sup>7</sup>, яку прийнято позначати як  $\lambda$ . Більше того, оскільки  $\lambda(\Omega) =$

<sup>5</sup>При цьому вважатимемо, що  $\mu(\emptyset) = 0$ .

<sup>6</sup>Хоча алгебра не обов'язково повинна бути замкнена відносно зліченного об'єднання множин, ми вимагаємо  $A \in \mathcal{B}_0(0; 1]$ , інакше доведення не має сенсу.

<sup>7</sup>На честь французького математика Анрі Лебега (Henri Lebesgue, 1875–1941), який її запропонував.

$\lambda((0; 1]) = 1$ , то міра Лебега на пів інтервалі  $(0; 1]$  є ймовірнісною мірою. Іншими словами, трійка  $((0; 1], \mathcal{B}(0; 1], \lambda)$  є ймовірнісним простором.

**Зауваження 2.4.3.** Цілком очевидно, що ніде у вищезгаданих доведеннях ми не використовували той факт, що  $\Omega = (0; 1]$ . Іншими словами, міру Лебега можна увести для деякого довільного інтервалу  $\Omega = (a; b]$ , щоправда, якщо  $b - a \neq 1$ , то така міра не буде ймовірнісною.

Міру Лебега також можна увести для всієї дійсної осі  $\mathbb{R}$ . У цьому випадку, очевидно, мірою всього простору буде нескінченно велике число, тобто міра вже не буде скінченою. Проте Теорема 2.4.1 залишається справедливою і для  $\sigma$ -скінчених мір. Дійсну вісь можна розбити на зліченне об'єднання інтервалів:

$$\mathbb{R} = \left( \bigcup_{n=0}^{\infty} (-n - 1; -n] \right) \cup \left( \bigcup_{n=0}^{\infty} (n; n + 1] \right),$$

міра Лебега (довжина) кожного з яких є скінченою (а саме 1).

Таким чином, міра Лебега  $\lambda$  — єдина  $\sigma$ -скінчена міра така, що  $\lambda((a; b]) = b - a$ .  $\square$

**Зауваження 2.4.4.** Аналогічним чином можна увести міру Лебега в  $k$ -вимірному просторі  $\mathbb{R}^k$ . Зокрема, для  $B = A_1 \times \dots \times A_k$ ,  $A_j \in \mathcal{B}_0(0; 1]$ ,  $j = 1, \dots, k$ , можна увести міру

$$\mu(B) = \prod_{j=1}^k \lambda(A_j).$$

Тоді якщо  $E = \bigcup_{j=1}^m B_j$ , де  $B_j$  неперетинані, то

$$\mu(E) = \sum_{j=1}^m \mu(B_j).$$

Подовження цієї міри на  $\sigma$ -алгебру  $\mathcal{B}^k$  називають  $k$ -вимірною мірою Лебега і позначають через  $\lambda_k$ .  $\square$

Констатуємо без доведення важливу властивість міри Лебега.

**Твердження 2.4.5.** Міра Лебега  $\lambda$  інваріантна відносно зсуву: якщо  $A \in \mathcal{B}^k$ , то

$$A + x \equiv \{a + x : a \in A\} \in \mathcal{B}^k,$$

а

$$\lambda_k(A) = \lambda_k(A + x), \quad \forall x \in \mathbb{R}^k.$$

Повертаючись до Прикладу 1.5.10, ми бачимо, що міра, властивості якої ми вимагали, фактично є мірою Лебега. Ми можемо тепер говорити про те, що множина Віталі є *невимірною* в розумінні міри Лебега.

## 2.5. Події з нульовою ймовірністю

**Визначення 2.5.1.** Множину  $A$ , міра Лебега якої  $\lambda(A) = 0$ , називають *множиною міри нуль за Лебегом* (Lebesgue measure zero set, null set).

За аналогією подію  $A$ , імовірнісна міра якої  $\mathbb{P}(A) = 0$ , називають *нульовою подією* (null event).  $\square$

Розгляньмо приклади множин міри нуль із  $\mathcal{B}(0; 1]$ .

**Твердження 2.5.2.** Множинами міри нуль за Лебегом є множини  $A$  такі, що

$$A \subset \bigcup_{k=1}^{\infty} I_k, \quad \forall \varepsilon > 0 \quad \sum_{k=1}^{\infty} \lambda(I_k) < \varepsilon,$$

де  $I_k$  — деякі пів інтервали з  $\mathcal{I}(0; 1]$ .

*Доведення.* Це безпосередньо випливає з властивостей монотонності і  $\sigma$ -субадитивності:

$$\lambda(A) < \lambda\left(\bigcup_{k=1}^{\infty} I_k\right) \leq \sum_{k=1}^{\infty} \lambda(I_k) < \varepsilon.$$

Оскільки  $\varepsilon$  є абсолютно довільний,  $\lambda(A) = 0$ .  $\square$

**Зауваження 2.5.3.** Із Твердження 2.5.2 випливає, що будь-яка одноелементна множина є множиною міри нуль. Дійсно, нехай  $x \in (0, 1]$ . Для будь-якого  $\varepsilon > 0$ ,  $x \in (x - \varepsilon; x]$ . За властивістю монотонності міри,  $0 \leq \lambda(\{x\}) \leq \lambda((x - \varepsilon; x]) = \varepsilon$ . Оскільки  $\varepsilon$  є довільним, випливає  $\lambda(\{x\}) = 0$ .

Будь-яка зліченна множина є зліченним об'єднанням одноелементних множин, і тому її міра також дорівнює нулю за властивістю  $\sigma$ -адитивності. Так, множинами міру нуль є множини натуральних, цілих, раціональних чисел тощо.

Відповідно, імовірність події, заданої інтервалом, не залежить від того, чи включено в інтервал його межі:  $\lambda((a; b)) = \lambda([a; b]) = \lambda((a; b]) = \lambda([a; b])$ .  $\square$

Класичним прикладом *незліченної* множини міри нуль є множина Кантора (Cantor set)<sup>8</sup>.

**Приклад 2.5.4** (Множина Кантора є множиною міри нуль). Розгляньмо замкнений інтервал  $[0; 1]$ . Виріжмо з нього середню третину — відкритий інтервал  $(1/3; 2/3)$ . У кожному закритому інтервалі —  $[0; 1/3]$  та  $[2/3; 1]$  — також виріжмо по середній третині (конкретно, виріжмо відкриті інтервали  $(1/9; 2/9)$  та  $(7/9; 8/9)$ ). Цей процес продовжуватимемо до нескінченності. Перші декілька кроків зображені на Рис. 2.5.1.

Після кожного кроку  $i$  буде вирізано  $1 + 2 + 4 + \dots + 2^{i-1} = 2^i - 1$  інтервалів, які позначмо через  $J_{i,j}$ ,  $j = 1, \dots, 2^i - 1$ . Множину всіх вирізаних інтервалів після кроку  $i$  позначмо через  $U_i = \bigcup_{j=1}^{2^i-1} J_{i,j}$ .

<sup>8</sup>Названа так на честь визначного німецького математика Георга Кантора (Georg Ferdinand Ludwig Philipp Cantor, 1845–1918).



Рис. 2.5.1.: Побудова множини Кантора: сірим позначені відкриті інтервали  $U_i$ , які вирізають на кожному кроці  $i$  [7, Рис. 11.2]

Ці множини утворюють зростаючу послідовність  $\{U_i\}$ , граничною множиною якої є

$$U = \bigcup_{i=1}^{\infty} U_i.$$

Множина  $U$  є відкритою, адже вона є зліченним об'єднанням (неперетинаних) відкритих інтервалів. Її мірою Лебега, з урахуванням властивості неперервності міри з Визначення 2.2.7, є

$$\begin{aligned} \lambda(U) &= \lambda\left(\bigcup_{i=1}^{\infty} U_i\right) = \lambda\left(\lim_{i \rightarrow \infty} U_i\right) = \lim_{i \rightarrow \infty} \lambda(U_i) = \\ &= \lim_{i \rightarrow \infty} \left( \frac{1}{3} + \frac{2}{9} + \frac{4}{27} + \dots + \frac{2^{i-1}}{3^i} \right) \\ &= \lim_{i \rightarrow \infty} \left( \frac{1}{2} \cdot \frac{2}{3} \cdot \frac{1 - (2/3)^i}{1 - 2/3} \right) = 1. \end{aligned}$$

Множиною Кантора називають множину  $C = [0; 1] \setminus U$ . У цьому випадку  $\lambda(C) = \lambda(U^c) = 1 - \lambda(U) = 0$ , тобто множина Кантора має міру нуль. Проте вона є незліченою замкненою множиною.

Конкретніше, множина  $C$  складається з усіх таких точок  $x \in [0; 1]$ , що їх можна подати у трійковому розвиненні  $x = \sum_{i=1}^{\infty} \frac{x_i}{3^i}$ , де  $x_i \in \{0, 2\}$ . Так, наприклад,  $\frac{1}{3} = 0.0(2)$ ,  $\frac{2}{3} = 0.2(0)$ ,  $\frac{1}{9} = 0.00(2)$  і т.д. Тоді можна відповідні дроби подати як двійкові, замінивши  $x_i = \frac{x_i}{2}$ ,  $x_i \in \{0, 1\}$ . Як було розглянуто раніше, множина відповідних двійкових розвинень має таку саму потужність, як і весь інтервал  $[0; 1]$ .  $\square$

**Визначення 2.5.5.** Твердження  $S$  про елементи простору  $\omega \in \Omega$  називають істинним *майже скрізь* (almost everywhere, a.e.), якщо

$$\mu(\{\omega \in \Omega \mid S(\omega) \text{ хибне}\}) = 0.$$

У контексті ймовірностей кажуть, що твердження істинне *майже напевно* (almost surely, a.s.) або *з імовірністю 1* (with probability 1, w.p.1).  $\square$

Це визначення є дуже важливим у теорії ймовірностей, адже дає змогу відкидати надзвичайно малоймовірні варіанти завершення деякого експерименту.

## 2.6. Доведення окремих тверджень

Для доведення властивості адитивності міри  $\mu((a; b]) = b - a$  на алгебрі  $\mathcal{B}_0(0; 1]$ . Властивість адитивності доводитимемо як одночасне виконання властивостей суперадитивності та субадитивності.

**Лема 2.6.1** (Суперадитивність довжин інтервалів). Нехай маємо інтервал  $I = (a; b]$ . Розгляньмо набір неперетинаних інтервалів  $I_1, I_2, \dots$ , де  $I_k = (a_k; b_k]$ ,  $k = 1, 2, \dots$ , до того ж  $a_1 \leq a_2 \dots \leq \dots$ , а  $I \supset \bigcup_{k=1}^{\infty} I_k$ . Тоді  $\sum_{k=1}^{\infty} \mu(I_k) \leq \mu(I)$ . Якщо кількість інтервалів скінчена ( $I_1, \dots, I_n$ ), то  $\sum_{k=1}^n \mu(I_k) \leq \mu(I)$ .

*Доведення.* Для скінченного числа інтервалів покажімо цей результат за індукцією. Справді, для  $n = 1$  це твердження тривіальне. Нехай воно справедливе для  $n - 1$ . Тоді,

$$\bigcup_{k=1}^{n-1} (a_k; b_k] \subset (a; a_n] ,$$

а відтак

$$\sum_{k=1}^{n-1} (b_k - a_k) \leq a_n - a$$

за індукційним припущенням. Тоді, очевидно,

$$\sum_{k=1}^n (b_k - a_k) \leq (a_n - a) + (b_n - a_n) \leq b - a .$$

Якщо кількість інтервалів зліченна,  $I_1, I_2, \dots$ , то потрібно показати, що

$$S \equiv \lim_{n \rightarrow \infty} \sum_{i=1}^n \mu(I_i) \equiv \lim_{n \rightarrow \infty} S_n \leq b - a .$$

Оскільки для кожної скінченної підпослідовності  $I_1, \dots, I_n$  субадитивність уже доведено, то  $S_n \leq b - a$  для всіх  $n \geq 1$ .

За визначенням границі послідовності,

$$\forall \varepsilon > 0 \ \exists N \in \mathbb{N} \ \forall n \geq N \ |S - S_n| < \varepsilon ,$$

тобто, серед іншого

$$\begin{aligned} S &< S_n + \varepsilon \leq b + \varepsilon , \\ a - \varepsilon &\leq S_n - \varepsilon < S . \end{aligned}$$

Спряженучи  $\varepsilon \rightarrow 0$ , маємо

$$a \leq S \leq b ,$$

що й потрібно було показати.  $\square$

**Лема 2.6.2** (Субадитивність довжин інтервалів). Нехай маємо інтервал  $I = (a; b]$ . Розгляньмо набір (необов'язково неперетинаних) інтервалів  $I_1, I_2, \dots$ , де  $I_k = (a_k; b_k]$ ,  $k = 1, 2, \dots$ , до того ж  $a_1 \leq a_2 \dots \leq \dots$ , а  $I \subset \bigcup_{k=1}^{\infty} I_k$ . Тоді  $\mu(I) \leq \sum_{k=1}^{\infty} \mu(I_k)$ . Якщо кількість інтервалів скінчена ( $I_1, \dots, I_n$ ), то  $\mu(I) \leq \sum_{k=1}^n \mu(I_k)$ .

*Доведення.* Для скінченного числа інтервалів покажімо цей результат за індукцією. Для  $n = 1$  твердження тривіально виконується. Нехай воно справедливе для  $n - 1$ , і нехай  $I \subset \bigcup_{k=1}^n I_k$ .

Якщо  $a \leq a_n < b_n \leq b$ , то це нічого не дає, адже ми вже показали субадитивність для  $n - 1$ , а в цьому випадку нічого, крім зайдового проміжку, не додається. Натомість ситуація, коли новий інтервал виходить за межі  $(a; b]$ , становить інтерес.

Достатньо розглянути вихід за межі з якогось одного боку. Тому нехай  $a_n < b \leq b_n$ . Тоді, якщо  $a_n \leq a$ , то результат тривіальний, адже фактично  $I \subset I_n$ , і тому зрозуміло, що  $\mu(I) \leq \mu(I_n) \leq \mu(\bigcup_{k=1}^n I_k)$ .

Тому нехай  $(a; a_n] \subset \bigcup_{k=1}^{n-1} (a_k; b_k]$ , а відтак

$$\sum_{k=1}^{n-1} (b_k - a_k) \geq a_n - a$$

за індукційним припущенням. Тоді, очевидно,

$$\sum_{k=1}^n (b_k - a_k) \geq (a_n - a) + (b_n - a_n) \geq b - a .$$

Якщо кількість інтервалів зліченна,  $I_1, I_2, \dots$ , і  $I \subset \bigcup_{k=1}^{\infty} I_k$ , то ми фактично маємо ситуацію *покриття* (coverage) інтервалу  $I$  зліченним об'єднанням інтервалів  $I_k$ . Щоб звести цей випадок до скінченного, який уже доведено, ми застосуємо *Теорему Гайне-Бореля* (Heine-Borel Theorem), згідно з якою кожна замкнена обмежена множина в  $\mathbb{R}$  є компактом (сомпакт), тобто з будь-якого зліченного покриття можна виділити скінченне підпокриття.

Відповідно, задача полягає в тому, щоб застосувати Теорему Гайне-Бореля до пів інтервалу  $I = (a; b]$ , який є обмеженим, але не є замкненим зліва. Нехай  $0 < \varepsilon < b - a$ . Тоді послідовність відкритих інтервалів  $(a_k; b_k + \varepsilon 2^{-k})$ ,  $k = 1, 2, \dots$ , покриває замкнений інтервал  $[a + \varepsilon; b]$ . За Теоремою Гайне-Бореля, із цього покриття можна виділити деяке скінченне підпокриття:

$$[a + \varepsilon; b] \subset \bigcup_{k=1}^n (a_k; b_k + \varepsilon 2^{-k}) .$$

Цілком очевидно, що це твердження залишиться справедливим, якщо виколоти ліву межу інтервалу:

$$(a + \varepsilon; b] \subset \bigcup_{k=1}^n (a_k; b_k + \varepsilon 2^{-k}) ,$$

а відтак, за вже доведеним скінченним випадком,

$$\mu((a + \varepsilon; b]) = b - a - \varepsilon \leq \sum_{k=1}^n \mu((a_k; b_k + \varepsilon 2^{-k})) = \sum_{k=1}^n (b_k + \varepsilon 2^{-k} - a_k) \leq \sum_{k=1}^{\infty} (b_k - a_k) + \varepsilon ,$$

оскільки  $\sum_{k=1}^{\infty} 2^{-k} = 1$  за властивостями нескінченно спадної геометричної прогресії.

Оскільки  $\varepsilon$  абсолютно довільне, випливає результат:

$$\mu((a + \varepsilon; b]) = \sum_{k=1}^{\infty} (b_k - a_k) + \varepsilon,$$

тобто

$$\mu((a + \varepsilon; b]) \leq \sum_{k=1}^{\infty} (b_k - a_k),$$

що й потрібно було довести.  $\square$

**Зауваження 2.6.3.** У доведенні Леми 2.6.1 перехід від скінченного до нескінченно зліченного випадку був простий. Для Леми 2.6.2 такий перехід вимагав застосування поняття компактності (для чого й була потрібна Теорема Гайнє-Бореля). Саме нескінченно зліченна частина Леми 2.6.2 свідчить, що інтервал додатної довжини не може мати нульову міру:  $\mu(I) = b - a$  є додатною нижньою межею суми довжин інтервалів у будь-якому покритті  $I$ .  $\square$

Результати Лем 2.6.1–2.6.2 справедливі для будь-яких пів інтервалів, і в тому числі пів інтервалу  $(0; 1]$ . Їх одночасне виконання є основою доведення такого твердження.

**Лема 2.6.4** (Адитивність довжин пів інтервалів). Нехай  $A = \bigcup_{i=1}^{\infty} I_i$ , де  $I_i$ ,  $i = 1, \dots, n$  — неперетинані пів інтервали з  $(0; 1]$  (у тому числі порожні множини). Тоді

$$\mu(A) = \sum_{i=1}^{\infty} \mu(I_i).$$

*Доведення.* Те, що  $\mu(A) = \sum_{i=1}^{\infty} \mu(I_i)$ , безпосередньо випливає з Лем 2.6.1–2.6.2.

Єдине, що потрібно додатково перевірити — коректність визначення міри множини  $A$ : оскільки множину  $A$  можна подати як скінченне об'єднання інших неперетинаних пів інтервалів  $A = \bigcup_{j=1}^{\infty} J_j$ <sup>9</sup>, потрібно показати, що  $\sum_{i=1}^{\infty} \mu(I_i) = \sum_{j=1}^{\infty} \mu(J_j)$ .

Справді, оскільки  $J_j$  утворюють розбиття множини  $A$ , можна подати кожний пів інтервал як перетин  $I_i = \bigcap_{j=1}^{\infty} (I_i \cap J_j)$ . Відповідно,

$$\sum_{i=1}^{\infty} \mu(I_i) = \sum_{i=1}^{\infty} \sum_{j=1}^{\infty} \mu(I_i \cap J_j).$$

Звернімо увагу, що деякі з перетинів можуть бути порожні, але це допустимо, адже  $\mu(\emptyset) = 0$ .

Точні такі самі міркування справедливі і в інший бік:

$$\sum_{j=1}^{\infty} \mu(J_j) = \sum_{j=1}^{\infty} \sum_{i=1}^{\infty} \mu(J_j \cap I_i),$$

що доводить коректність визначення міри  $\mu$ .  $\square$

<sup>9</sup>Наприклад,  $(0; 0.5] \cup (0.7; 1] = (0; 0.2] \cup (0.2; 0.5] \cup (0.7; 1] = (0; 0.5] \cup (0.7; 0.8] \cup (0.8; 1]$ .

## 3. Умовна ймовірність

### 3.1. Інтуїція та визначення

У попередніх лекціях ми формально увели поняття ймовірності (як міри) для опису ступенів упевненості та невизначеностей реальних подій і явищ. Проте на практиці ми часто стикаємося з ситуаціями, коли надходження нової інформації (нових даних, нових свідчень тощо) може змінити наші уявлення про ймовірності відповідних подій. Нове спостереження, сумісне з попередніми оцінками, може підсилити нашу впевненість, а спостереження, що вибивається з загальної картини, може похитнути наші попередні уявлення. Для формалізації цих ситуацій слугує поняття умовної ймовірності.

**Приклад 3.1.1.** Нехай одного ранку ми хочемо оцінити ймовірність події  $R$ , що піде дощ. Наша початкова оцінка такої ймовірності дорівнює  $\mathbb{P}(R)$ . Якщо ми виглянемо у вікно і побачимо, що небо затягнули хмари, ми вважатимемо, що ймовірність дощу вища. Цю оновлену ймовірність позначають через  $\mathbb{P}(R | C)$ , де  $C$  — подія, що небо затягнули хмари. Перехід від  $\mathbb{P}(R)$  до  $\mathbb{P}(R | C)$  називають *обумовленням* подією  $C$  (conditioning on  $C$ ).

Із плином дня можуть надходити нові й нові факти як — події  $B_1, \dots, B_n$  — і в результаті наша ймовірність буде дорівнювати  $\mathbb{P}(R | C, B_1, \dots, B_n) \equiv \mathbb{P}(R | C \cap B_1 \cap \dots \cap B_n)$ .  $\square$

Дамо формальне визначення умовної ймовірності.

**Визначення 3.1.2.** Нехай маємо вимірний простір  $(\Omega, \mathcal{A})$ . Якщо  $A$  і  $B$  — деякі події з  $\mathcal{A}$ , і  $\mathbb{P}(B) > 0$ , то *умовною ймовірністю*  $A$  за умови  $B$  (conditional probability of  $A$  given  $B$ ) є ймовірність

$$\mathbb{P}(A | B) = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)}. \quad (3.1.1)$$

Імовірність  $\mathbb{P}(A)$  називають *апріорною* (prior), а ймовірність  $\mathbb{P}(A | B)$  — *апостеріорною* (posterior).  $\square$

Нескладно бачити, що  $\mathbb{P}(A | A) = \mathbb{P}(A \cap A) / \mathbb{P}(A) = 1$ , тобто після спостереження, що подія  $A$  відбулася, її ймовірність дорівнює 1.

**Зауваження 3.1.3.** Варто зазначити, що запис  $\mathbb{P}(A | B)$  не означає, що існує подія  $A | B$ , імовірність якої ми міряємо. Такої події не існує. Насправді, ми визначили нову ймовірнісну міру,  $\mathbb{P}(\cdot | B)$ , яка відрізняється від  $\mathbb{P}(\cdot)$  тим, що враховує факт настання події  $B$ .

Формальніше, маючи ймовірнісний простір  $(\Omega, \mathcal{A}, \mathbb{P}(\cdot))$ ,  $B \in \mathcal{A}$ , ми визначаємо нову міру таку, що  $(\Omega, \mathcal{A}, \mathbb{P}(\cdot | B))$  також є ймовірнісним простором.  $\square$

**Твердження 3.1.4.** Нехай маємо вимірний простір  $(\Omega, \mathcal{A})$ ,  $B \in \mathcal{A}$ , і  $\mathbb{P}(B) > 0$ . Тоді умовна ймовірність  $\mathbb{P}(\cdot | B)$  є ймовірнісною мірою в розумінні Визначення 2.1.5.

*Доведення.* Як завжди, потрібно перевірити виконання всіх трьох аксіом імовірнісної міри.

Умовна ймовірність невід'ємна, адже вона є відношенням двох невід'ємних значень.

Імовірність усього простору дорівнює

$$\mathbb{P}(\Omega | B) = \frac{\mathbb{P}(\Omega \cap B)}{\mathbb{P}(B)} = \frac{\mathbb{P}(B)}{\mathbb{P}(B)} = 1.$$

Якщо  $A_1, A_2, \dots \in \mathcal{A}$  — послідовність несумісних подій, то

$$\begin{aligned} \mathbb{P}\left(\bigcup_{i=1}^{\infty} A_i | B\right) &= \frac{1}{\mathbb{P}(B)} \cdot \mathbb{P}\left(\left(\bigcup_{i=1}^{\infty} A_i\right) \cap B\right) = \frac{1}{\mathbb{P}(B)} \cdot \mathbb{P}\left(\bigcup_{i=1}^{\infty} (A_i \cap B)\right) \\ &= \frac{1}{\mathbb{P}(B)} \cdot \sum_{i=1}^{\infty} \mathbb{P}(A_i \cap B) = \sum_{i=1}^{\infty} \mathbb{P}(A_i | B), \end{aligned}$$

що доводить  $\sigma$ -адитивність умовної ймовірності.

Таким чином, умовна ймовірність задовольняє всі аксіоми ймовірнісної міри.  $\square$

**Приклад 3.1.5.** Нехай маємо добре перетасовану колоду з 52 гральних карт. Витягнімо випадковим чином дві карти без повторень. Нехай  $A$  = «перша карта — чири»,  $B$  = «друга карта червона». Знайдімо ймовірності  $\mathbb{P}(A | B)$  та  $\mathbb{P}(B | A)$ .

Імовірність перетину двох подій дорівнює

$$\mathbb{P}(A \cap B) = \frac{13}{52} \cdot \frac{25}{51} = \frac{25}{204},$$

оскільки ймовірність на момент першого витягу в колоді 13 чиризових карт, а після витягу однієї з них серед 51 карт, що лишилися, міститься 25 червоних.

Також можна порахувати ймовірності подій окремо:  $\mathbb{P}(A) = \frac{1}{4}$ , оскільки перша карта може бути будь-якої з чотирьох мастей, та  $\mathbb{P}(B) = \frac{1}{2}$ , оскільки незалежно від того, якою є перша карта, друга карта може бути будь-якого кольору<sup>1</sup>.

Тоді, за (3.1.1) маємо:

$$\begin{aligned} \mathbb{P}(A | B) &= \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)} = \frac{25}{102}, \\ \mathbb{P}(B | A) &= \frac{\mathbb{P}(B \cap A)}{\mathbb{P}(A)} = \frac{25}{51}. \end{aligned}$$

$\square$

**Зауваження 3.1.6.** Надзвичайно важливо пам'ятати, що обумовлення однієї події іншою не є комутативним:  $\mathbb{P}(A | B) \neq \mathbb{P}(B | A)$  у загальному випадку.

<sup>1</sup> Якщо це звучить непереконливо, що  $\mathbb{P}(B) = \frac{1}{2} \cdot \frac{26}{51} + \frac{1}{2} \cdot \frac{25}{51} = \frac{1}{2}$ , де окремо розглядаються випадки витягу першої червоної або чорної карти.

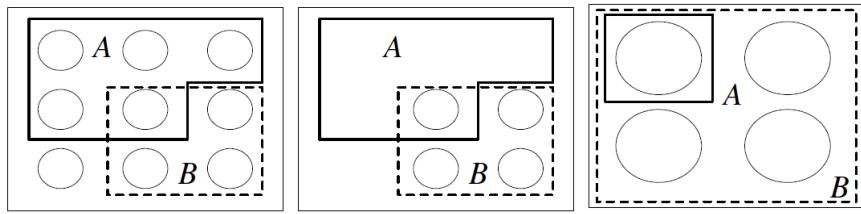


Рис. 3.1.1.: Інтуїтивна ілюстрація перенормалізації ймовірностей: спочатку події  $A$  і  $B$  є підмножинами простору елементарних подій (у цьому випадку скінченного). Оскільки подія  $B$  відбулася, ми ігноруємо елементарні події, які не належать  $B$ . Таким чином,  $\mathbb{P}(A) = 5/9$ , проте  $\mathbb{P}(A | B) = 1/4$  [1, Рис. 2.1]

Також потрібно розуміти, що обидві ймовірності,  $\mathbb{P}(A | B)$  і  $\mathbb{P}(B | A)$ , мають сенс, адже коли ми рахуємо умовні ймовірності, ми не розглядаємо причиново-наслідкові зв'язки між подіями, а тільки враховуємо інформацію, яка одна подія може пролити на іншу. Наприклад, якщо в Прикладі 3.1.5 уявити, що дві карти виймаються *одночасно*, але двома руками, то замість першої і другої карти можемо говорити про ліву та праву карти, але відповідні умовні ймовірності від цього не зміняться.  $\square$

Обумовлення події  $A$  подією  $B$  означає, що ми *перенормалізовуємо* ймовірнісну міру. Ілюстративно це зображенено на Рис. 3.1.1. Таке обумовлення події має своє пояснення для обох підходів до інтерпретації ймовірностей:

- частотна інтерпретація умовної ймовірності спонукає нас розглядати ймовірність  $\mathbb{P}(A | B)$  як відносну частоту настання події  $A$  *серед тих експериментів*, які закінчилися подією  $B$ ;
- Беєсівська інтерпретація є природною і відображає оновлення суб'єктивної ймовірності після надходження додаткової інформації.

Із визначення умовної ймовірності (3.1.1) випливає, що ймовірність перетину двох подій можна подати таким чином:

$$\mathbb{P}(A \cap B) = \mathbb{P}(B) \mathbb{P}(A | B) = \mathbb{P}(A) \mathbb{P}(B | A) , \quad \mathbb{P}(A) > 0 , \quad \mathbb{P}(B) > 0 . \quad (3.1.2)$$

Відповідну властивість можна узагальнити.

**Твердження 3.1.7.** Нехай  $A_1, \dots, A_n$  — деякі події такі, що  $\mathbb{P}(A_1 \cap \dots \cap A_n) > 0$ . Тоді

$$\begin{aligned} \mathbb{P}(A_1 \cap \dots \cap A_n) &= \mathbb{P}(A_1) \mathbb{P}(A_2 | A_1) \mathbb{P}(A_3 | A_1, A_2) \dots \mathbb{P}(A_n | A_1, \dots, A_{n-1}) \\ &= \mathbb{P}(A_2) \mathbb{P}(A_1 | A_2) \mathbb{P}(A_3 | A_1, A_2) \dots \mathbb{P}(A_n | A_1, \dots, A_{n-1}) \\ &= \dots \end{aligned} \quad (3.1.3)$$

і т.д. для всіх можливих комбінацій подій.

*Доведення.* Доведення здійснюватимемо за індукцією. Для  $n = 2$  теорема є просто повторенням (3.1.2).

Нехай теорема виконується для  $n - 1$ :

$$\mathbb{P}(A_1 \cap \dots \cap A_{n-1}) = \mathbb{P}(A_1) \mathbb{P}(A_2 | A_1) \dots \mathbb{P}(A_{n-1} | A_1, \dots, A_{n-2}) .$$

Тоді

$$\begin{aligned}\mathbb{P}(A_1 \cap \dots \cap A_n) &= \mathbb{P}((A_1 \cap \dots \cap A_{n-1}) \cap A_n) = \mathbb{P}(A_1 \cap \dots \cap A_{n-1}) \mathbb{P}(A_n | A_1, \dots, A_{n-1}) \\ &= \mathbb{P}(A_1) \mathbb{P}(A_2 | A_1) \dots \mathbb{P}(A_{n-1} | A_1, \dots, A_{n-2}) \mathbb{P}(A_n | A_1 \cap \dots \cap A_{n-1}) .\end{aligned}$$

□

Використання обумовлень дає змогу розв'язувати складні задачі, розбиваючи їх на простіші підзадачі.

**Приклад 3.1.8.** У вазі містяться 10 однакових кульок, 5 із яких — чорні, 3 — червоні і 2 — білі. По черзі і без повторень витягають 4 кульки. Підрахуймо ймовірність того, що перша кулька — чорна ( $B_1$ ), друга кулька — червона ( $R_2$ ), третя кулька — біла ( $W_3$ ), а четверта — знову чорна ( $B_4$ ):

$$\mathbb{P}(B_1 \cap R_2 \cap W_3 \cap B_4) = \mathbb{P}(B_4 | B_1, R_2, W_3) \mathbb{P}(W_3 | B_1, R_2) \mathbb{P}(R_2 | B_1) \mathbb{P}(B_1) .$$

Кожну окрему ймовірність легко обчислити:

$$\mathbb{P}(B_1) = \frac{5}{10}, \quad \mathbb{P}(R_2 | B_1) = \frac{3}{9}, \quad \mathbb{P}(W_3 | B_1, R_2) = \frac{2}{8}, \quad \mathbb{P}(B_4 | B_1, R_2, W_3) = \frac{4}{7},$$

а отже

$$\mathbb{P}(B_1 \cap R_2 \cap W_3 \cap B_4) = \frac{4}{7} \cdot \frac{2}{8} \cdot \frac{3}{9} \cdot \frac{5}{10} \approx 0.024 .$$

□

## 3.2. Задача Гарднера про двох дітей

Розгляньмо класичний приклад, запропонований американським популяризатором математики Мартіном Гарднером (Martin Gardner, 1914–2010) у журналі *Scientific American* у 1956 р.

Містер Джонс має двох дітей. Старша з них — дівчинка. Яка ймовірність того, що в нього дві дівчинки?

Містер Сміт має двох дітей. Принаймні один із них — хлопчик. Яка ймовірність того, що в нього два хлопчики?

Протягом десятиліть точилися суперечки, як правильно інтерпретувати відповідні твердження, і чи є обидві ймовірності однаковими. Сам Гарднер зазначав, що ймовірності дорівнюють  $1/2$  і  $1/3$ , відповідно. Спробуймо розібратися, чому так.

Позначмо хлопчиків літерою  $B$ , а дівчат — літерою  $G$ . Покладімо для простоти, що ймовірності народження хлопчика та дівчинки однакові, а народження дитини будь-якої статі не залежить від статі попередньої дитини. Також розглядатимемо задачу з містером Смітом у термінах дівчат замість хлопчиків (цей випадок повністю симетричний).

Простір елементарних подій для цієї задачі має вид  $\Omega = \{GG, GB, BG, BB\}$ , де  $XY = \text{«старша дитина — } X, \text{ молодша дитина — } Y\text{»}$ ,  $X, Y \in \{B, G\}$ . Використовуватимемо символ  $\cdot$  на позначення ситуації, коли непринципово, якої статі та чи та дитина. Наприклад, подія  $G \cdot = \{GG, GB\}$ .

Тоді подія  $G$  = «принаймні одна дитина — дівчинка» має вид  $G = (G \cdot) \cup (\cdot G)$ . Використовуючи (3.1.1), маємо:

$$\begin{aligned}\mathbb{P}(GG \mid G \cdot) &= \frac{\mathbb{P}(GG \cap (G \cdot))}{\mathbb{P}(G \cdot)} = \frac{\mathbb{P}(\{GG\})}{\mathbb{P}(\{GG, GB\})} = \frac{1/4}{1/2} = \frac{1}{2}, \\ \mathbb{P}(GG \mid G) &= \frac{\mathbb{P}(GG \cap G)}{\mathbb{P}(G)} = \frac{\mathbb{P}(\{GG\})}{\mathbb{P}(\{GG, GB, BG\})} = \frac{1/4}{3/4} = \frac{1}{3}.\end{aligned}$$

Звернімо увагу, що в силу симетрії  $\mathbb{P}(GG \mid G \cdot) = \mathbb{P}(GG \mid \cdot G) = 1/2$ , адже інформація про те, що саме *старша* дитина є дівчинкою, несе таке ж навантаження, що й те, що саме *молодша* дитина є дівчинкою. Натомість, імовірність  $\mathbb{P}(GG \mid G)$  відрізняється, оскільки ми обумовлюємо наявність двох дівчат подіюю, що *хоча б одна* з дітей є дівчинкою, *не уточнюючи*, яка саме. Іншими словами, обумовлюючи подіюю  $G \cdot$  або  $\cdot G$ , ми «викидаємо» з простору елементарних подій по два варіанти в кожному випадку ( $BB, BG$  та  $BB, GB$ , відповідно), у той час як обумовлення подіюю  $G$  відкидає тільки один варіант —  $BB$ .

Можна навести альтернативну інтерпретацію цієї задачі, яка призведе до зовсім іншого результату:

Містер Сміт гуляє в парку з дівчинкою. На питання, чи це його дочка, він відповідає ствердно. Яка ймовірність того, що в нього дві дівчинки, якщо на прогулянку він вибрав одного зі своїх дітей абсолютно випадково?

У цьому випадку простір елементарних подій інший:

$$\Omega = \{GG/g \cdot, GG/\cdot g, GB/g \cdot, GB/\cdot b, BG/b \cdot, BG/\cdot g, BB/b \cdot, BB/\cdot b\},$$

де маленькими літерами позначено, кого саме з дітей бачили на прогулянці.

У цьому випадку обумовлюємо подіюю  $G'$  = «на прогулянці містера Сміта побачили з дівчинкою»:  $G' = \{GG/g \cdot, GG/\cdot g, GB/g \cdot BG/\cdot g\}$ , а відтак

$$\mathbb{P}(GG \mid G') = \frac{\mathbb{P}(GG \cap G')}{\mathbb{P}(G')} = \frac{\mathbb{P}(\{GG/g \cdot, GG/\cdot g\})}{\mathbb{P}(\{GG/g \cdot, GG/\cdot g, GB/g \cdot BG/\cdot g\})} = \frac{2/8}{4/8} = \frac{1}{2}.$$

Розгляньмо варіант цієї задачі, де додатково враховуватимемо інформацію про пору року, у яку народилася дитина<sup>2</sup>. У цьому випадку простір елементарних подій містить усі можливі комбінації виду  $\Omega = \{XS, YT, X, Y \in \{G, B\}, S, T \in \{W, Sp, Su, F\}\}$ <sup>3</sup>, тобто додатково до статі дитини уточнююємо пору року її народження. Як і раніше, використовуватимемо символ  $\cdot$  для позначення того факту, що статі чи пори року відповідної дитини не має значення. Уесь можливий перелік із 64 варіантів зображенено на Рис. 3.2.1a.

Чому дорівнює ймовірність того, що обидві дитини — дівчатка, якщо принаймні одна з дітей — дівчинка, що народилася взимку?

Позначмо обумовлювальну подію через  $GW = (\cdot, GW) \cup (GW, \cdot)$ . За формулою (3.1.1) маємо:

$$\mathbb{P}(G \cdot, G \cdot \mid GW) = \frac{\mathbb{P}((G \cdot, G \cdot) \cap GW)}{\mathbb{P}(GW)}.$$

<sup>2</sup>У 2006 американські математики Майкл і Том Старберд (Michael and Tom Starbird) запропонували варіант із народженням дитини у вівторок. Тут розглядаємо аналогічний варіант, але з порами року, запропонований у [1].

<sup>3</sup>За першими літерами англійських відповідників слів зима, весна, літо, осінь — winter, spring, summer, fall.

BW,BW	BW,BSp	BW,BSu	BW,BF	BW,GW	BW,GSp	BW,GSu	BW,GF
BSp,BW	BSp,BSp	BSp,BSu	BSp,BF	BSp,GW	BSp,GSp	BSp,GSu	BSp,GF
BSu,BW	BSu,BSp	BSu,BSu	BSu,BF	BSu,GW	BSu,GSp	BSu,GSu	BSu,GF
BF,BW	BF,BSp	BF,BSu	BF,BF	BF,GW	BF,GSp	BF,GSu	BF,GF
GW,BW	GW,BSp	GW,BSu	GW,BF	GW,GW	GW,GSp	GW,GSu	GW,GF
GSp,BW	GSp,BSp	GSp,BSu	GSp,BF	GSp,GW	GSp,GSp	GSp,GSu	GSp,GF
GSu,BW	GSu,BSp	GSu,BSu	GSu,BF	GSu,GW	GSu,GSp	GSu,GSu	GSu,GF
GF,BW	GF,BSp	GF,BSu	GF,BF	GF,GW	GF,GSp	GF,GSu	GF,GF

(a)

BW,BW	BW,BSp	BW,BSu	BW,BF	BW,GW	BW,GSp	BW,GSu	BW,GF
BSp,BW	BSp,BSp	BSp,BSu	BSp,BF	BSp,GW	BSp,GSp	BSp,GSu	BSp,GF
BSu,BW	BSu,BSp	BSu,BSu	BSu,BF	BSu,GW	BSu,GSp	BSu,GSu	BSu,GF
BF,BW	BF,BSp	BF,BSu	BF,BF	BF,GW	BF,GSp	BF,GSu	BF,GF
GW,BW	GW,BSp	GW,BSu	GW,BF	GW,GW	GW,GSp	GW,GSu	GW,GF
GSp,BW	GSp,BSp	GSp,BSu	GSp,BF	GSp,GW	GSp,GSp	GSp,GSu	GSp,GF
GSu,BW	GSu,BSp	GSu,BSu	GSu,BF	GSu,GW	GSu,GSp	GSu,GSu	GSu,GF
GF,BW	GF,BSp	GF,BSu	GF,BF	GF,GW	GF,GSp	GF,GSu	GF,GF

(б)

BW,BW	BW,BSp	BW,BSu	BW,BF	BW,GW	BW,GSp	BW,GSu	BW,GF
BSp,BW	BSp,BSp	BSp,BSu	BSp,BF	BSp,GW	BSp,GSp	BSp,GSu	BSp,GF
BSu,BW	BSu,BSp	BSu,BSu	BSu,BF	BSu,GW	BSu,GSp	BSu,GSu	BSu,GF
BF,BW	BF,BSp	BF,BSu	BF,BF	BF,GW	BF,GSp	BF,GSu	BF,GF
GW,BW	GW,BSp	GW,BSu	GW,BF	GW,GW	GW,GSp	GW,GSu	GW,GF
GSp,BW	GSp,BSp	GSp,BSu	GSp,BF	GSp,GW	GSp,GSp	GSp,GSu	GSp,GF
GSu,BW	GSu,BSp	GSu,BSu	GSu,BF	GSu,GW	GSu,GSp	GSu,GSu	GSu,GF
GF,BW	GF,BSp	GF,BSu	GF,BF	GF,GW	GF,GSp	GF,GSu	GF,GF

(в)

Рис. 3.2.1.: Простір елементарних подій для задачі Гарднера: (а) повний перелік елементарних подій; (б) події «обидві дитини — дівчатка» (помаранчевий) та «принаймні одна дитина — дівчинка» (синій і помаранчевий разом); (в) події «обидві дитини — дівчатка, і одна з них народилася взимку» (помаранчевий) та «принаймні одна дитина — дівчинка, що народилася взимку» (синій і помаранчевий разом)

Покладімо для простоти, що дитина може народитися в будь-яку пору року з однаковою ймовірністю  $1/4$ . Тоді ймовірність, що дитина певної статі народиться у певну пору року, становить  $(1/2) \cdot (1/4) = (1/8)$ . Відповідно, імовірність, що це буде хтось завгодно, але не дівчинка взимку, дорівнює  $1 - (1/8) = 7/8$ . Тоді

$$\mathbb{P}(GW) = 1 - \mathbb{P}(GW^c) = 1 - \left(\frac{7}{8}\right)^2 = \frac{15}{64}.$$

Імовірність  $\mathbb{P}((G \cdot, G \cdot) \cap GW)$  можна обчислити з Рис. 3.2.1в, і вона очевидно дорівнює  $7/64$ . Таким чином,

$$\mathbb{P}(G \cdot, G \cdot | GW) = \frac{7/64}{15/64} = \frac{7}{15}.$$

На перший погляд цей результат здається несподіваним, адже інтуїція підказує, що додаткова інформація про пору року не несе ніякого навантаження на додачу до того факту, що одна з дітей — дівчинка. Проте насправді це має велике значення, адже дає змогу більше уточнити, про кого саме з двох дітей мова. Іншими словами, якби ми обумовлювали подією «принаймні одна з двох дітей — дівчинка, що народилася 31 березня о 8:30», то відповідна умовна ймовірність була б надзвичайно близька до  $1/2$ , як це було у випадку  $\mathbb{P}(GG | G \cdot)$ .

Ілюстративно це показано на Рис. 3.2.16–3.2.1в. Шукана ймовірність дорівнює відношенню кількості помаранчевих комірок до загальної кількості зафарбованих комірок. Зафарбовані комірки утворюють дві стрічки, горизонтальну й вертикальну, які разом відповідають події «принаймні одна дитина — дівчинка, що має певну властивість». Суто помаранчева область відповідає події «обидві діті — дівчатка, і одна з них має певну властивість». Стрічки перетинаються в події «обидві діті — дівчатка, і обидві з них мають певну властивість».

Якби стрічки взагалі не перетиналися, імовірність була б рівно  $1/2$ . У випадку властивості просто *бути дівчинкою*, стрічки дуже широкі, і перетин великий, тому ймовірність дорівнює  $1/3$ , як було показано вище. Що більше інформації додається, то більше елементів простору, то тонші стають стрічки, і то менше ймовірність, що обидві дочки матимуть одну й ту ж властивість. Іншими словами, дуже малоямовірно, щоб така деталізована інформація, як *народилася 31 березня о 8:30*, стосувалася одночасно двох дочок, і тому перетин двох стрічок дуже малий.

### 3.3. Теорема Беєса та повна ймовірність

Із визначення (3.1.2) безпосередньо випливають такі твердження.

**Теорема 3.3.1** (Теорема Беєса (Bayes' Theorem)<sup>4</sup>). Нехай маємо вимірний простір  $(\Omega, \mathcal{A})$ . Для деяких подій  $A, B \in \mathcal{A}$ ,  $\mathbb{P}(B) > 0$ , справедливо:

$$\mathbb{P}(A | B) = \frac{\mathbb{P}(B | A) \mathbb{P}(A)}{\mathbb{P}(B)}. \quad (3.3.1)$$

**Теорема 3.3.2** (Закон повної ймовірності (Law of total probability)). Нехай  $A_1, \dots, A_n \in \mathcal{A}$  утворюють розбиття вимірного простору  $(\Omega, \mathcal{A})$ , і  $\mathbb{P}(A_i) > 0$  для всіх  $i$ , тобто  $\Omega = \bigcup_{i=1}^n A_i$ ,  $A_i \cap A_j = \emptyset$ ,  $i \neq j$ . Тоді

$$\mathbb{P}(B) = \sum_{i=1}^n \mathbb{P}(B | A_i) \mathbb{P}(A_i). \quad (3.3.2)$$

<sup>4</sup>Названа так на честь англійського статистика, філософа і священика Томаса Беєса (Thomas Bayes, 1701–1761).

**Доведення.** Оскільки множини  $A_1, \dots, A_n$  утворюють розбиття  $\Omega$ , множину  $B$  можна подати як об'єднання неперетинаних множин:

$$B = (B \cap A_1) \cup \dots \cup (B \cap A_n).$$

Використовуючи адитивність імовірності:

$$\mathbb{P}(B) = \mathbb{P}(B \cap A_1) + \dots + \mathbb{P}(B \cap A_n).$$

Нарешті, використовуючи (3.1.2), маємо:

$$\mathbb{P}(B) = \mathbb{P}(B | A_1) \mathbb{P}(A_1) + \dots + \mathbb{P}(B | A_n) \mathbb{P}(A_n).$$

□

**Зауваження 3.3.3.** Фактично, Теорема 3.3.2 стверджує, що безумовну ймовірність  $\mathbb{P}(B)$  можна подати як *зважене середнє* (weighted average) імовірностей умовних, де вагами виступають імовірності обумовлюючих подій. □

Правило Беєса зручно застосовувати у випадках, коли обчислення умовної імовірності  $\mathbb{P}(A | B)$  простіше звести до обчислення умовної ймовірності  $\mathbb{P}(B | A)$ , і навпаки. При цьому дуже важливо розуміти, що обумовлення однієї події іншою зовсім не взаємозамінне.

**Приклад 3.3.4** (Помилка обвинувачення). В 1998 р. Селлі Кларк (Sally Clark) обвинуватили в умисному вбивстві двох синів невдовзі після їхнього народження. У судовому процесі свідок-експерт заявив, що ймовірність смерті немовляти від синдрому раптової дитячої смерті складає  $1/8500$ , а смерть обох дітей одночасно взагалі складає  $(1/8500)^2$ , або 1 на 73 мільйони випадків. Відтак, стверджував експерт, Селлі невинувата з імовірністю 1 на 73 мільйони.

Експерт припустився хибних суджень<sup>5</sup>, оскільки перепутав дві умовні ймовірності. У позначенннях  $I = \text{«обвинувачений винний»}$ ,  $E = \text{«дитина загинула внаслідок синдрому»}$ , маємо  $\mathbb{P}(I | E) \neq \mathbb{P}(E | I)$ . Експерт зазначив, що ймовірність  $\mathbb{P}(E | I)$  дуже мала. Але нас цікавить зовсім не ця імовірність, а ймовірність невинуватості *після того, як смерть уже сталася*:

$$\mathbb{P}(I | E) = \frac{\mathbb{P}(E | I) \mathbb{P}(I)}{\mathbb{P}(E)} = \frac{\mathbb{P}(E | I) \mathbb{P}(I)}{\mathbb{P}(E | I) \mathbb{P}(I) + \mathbb{P}(E | I^c) \mathbb{P}(I^c)},$$

де апріорна ймовірність  $\mathbb{P}(I^c)$  надзвичайно мала (осіб, які вчинили подвійне вбивство, не так ужай багато!), а відтак увесь дріб дуже близький до 1.

Селлі Кларк випустили через 3 роки перебування у в'язниці. □

**Приклад 3.3.5.** Деякого пацієнта тестують на наявність дуже рідкого захворювання, на яке страждає  $100p_1$  відсотків населення. Якщо результат тесту позитивний, то вважають, що пацієнт хворий. Нехай  $D = \text{«пацієнт хворий»}$ ,  $T = \text{«результат тесту позитивний»}$ .

<sup>5</sup>По-перше, експерт хибно припустив, що ймовірності двох дитячих смертей є незалежні, хоча на практиці вони могли бути обумовлені спільними генетичними особливостями тощо.

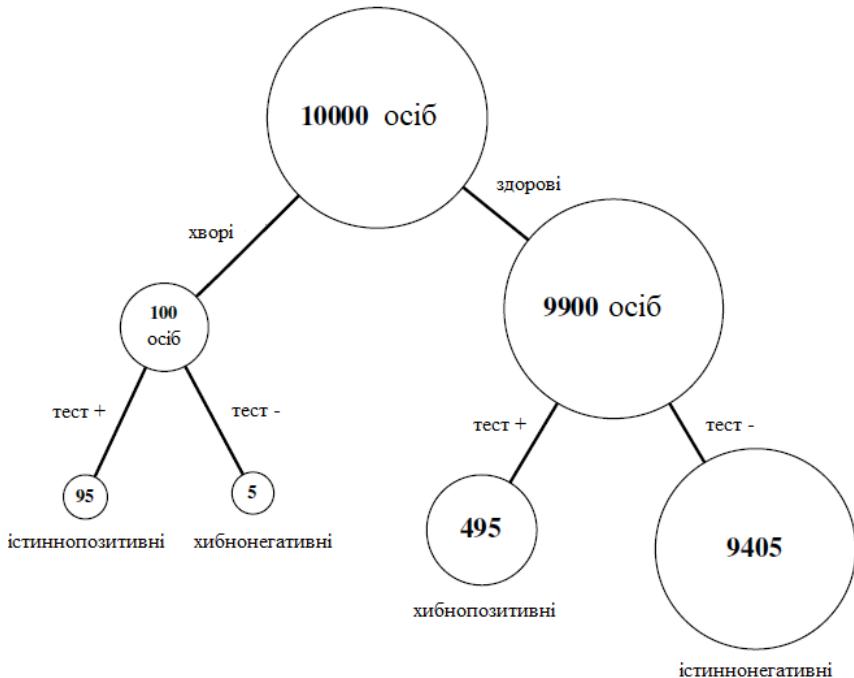


Рис. 3.3.1.: Тестування населення в 10 000 осіб [1, Рис. 2.4]

Нехай тест має чутливість  $\mathbb{P}(T | D) = 1 - p_2$  і специфічність  $\mathbb{P}(T^c | D^c) = 1 - p_3$ . Іншими словами, тест показує хибногативний результат з імовірністю  $p_2$  і хибнопозитивний результат з імовірністю  $p_3$ . Чому дорівнює ймовірність, що пацієнт із позитивним тестом *насправді* хворий?

Згідно з правилом Беєса та законом повної ймовірності,

$$\begin{aligned}
 \mathbb{P}(D | T) &= \frac{\mathbb{P}(T | D) \mathbb{P}(D)}{\mathbb{P}(T)} \\
 &= \frac{\mathbb{P}(T | D) \mathbb{P}(D)}{\mathbb{P}(T | D) \mathbb{P}(D) + \mathbb{P}(T | D^c) \mathbb{P}(D^c)} \\
 &= \frac{(1 - p_2) \cdot p_1}{(1 - p_2) \cdot p_1 + p_3 \cdot (1 - p_1)}.
 \end{aligned}$$

Наприклад, поклавши  $p_1 = 0.01$ ,  $p_2 = p_3 = 0.05$ , матимемо, що  $\mathbb{P}(D | T) \approx 0.16$ .

Таким чином, незважаючи на те, що тест має точність 95%, пацієнт насправді хворий усього лише з імовірністю 16%! Більшість людей бере подив. Проте насправді нічого дивного тут немає, адже апостеріорна ймовірність рахується на основі не тільки результатів тесту, а й *априорної* ймовірності поширення хвороби, яка в даному прикладі надзвичайно низька.

Добру ілюстрацію цих міркувань наведено на Рис. 3.3.1, на якій хворими є 100 осіб із 10 000 (що відповідає 1% захворюваності). Якби кожну людину протестували, приблизно 5 осіб зі 100 справді хворих було б (помилково) визнано здоровими. А з-поміж 9 900 здорових приблизно 495 ( $= 0.05 \cdot 9900$ ) осіб було б (помилково) визнано хворими. Розгляньмо тільки тих осіб, тест яких вийшов позитивним. Із них 95 істиннопозитивних випадків суттєво поступаються 495 хибнопозитивним випадкам. Відповідно, більшість осіб із позитивним результатом тесту насправді цілком здорові.

□

Правило Бееса залишається справедливими незалежно від того, чи обумовлення декількома подіями відбувається одночасно, чи крок за кроком.

**Приклад 3.3.6.** Продовжмо Приклад 3.3.5. Нехай після першого позитивного тесту пацієнт вирішив пройти тест повторно для одержання точнішого результату. Позначивши позитивні результати обох тестів через  $T_1$  і  $T_2$ , маємо:

$$\mathbb{P}(D | T_1 \cap T_2) = \frac{\mathbb{P}(T_1 \cap T_2 | D) \mathbb{P}(D)}{\mathbb{P}(T_1 \cap T_2 | D) \mathbb{P}(D) + \mathbb{P}(T_1 \cap T_2 | D^c) \mathbb{P}(D^c)} = \frac{0.95^2 \cdot 0.01}{0.95^2 \cdot 0.01 + 0.05^2 \cdot 0.99} \approx 0.78.$$

Якщо ж обчислювати відповідну інформацію послідовно, то спочатку одержимо  $\mathbb{P}(D | T_1) = 0.16$ , як і раніше, а тоді

$$\mathbb{P}(D | T_1, T_2) = \frac{\mathbb{P}(T_2 | D, T_1) \mathbb{P}(D | T_1)}{\mathbb{P}(T_2 | D, T_1) \mathbb{P}(D | T_1) + \mathbb{P}(T_2 | D^c, T_1) \mathbb{P}(D^c | T_1)} = \frac{0.95 \cdot 0.16}{0.95 \cdot 0.16 + 0.05 \cdot 0.84} \approx 0.78,$$

тобто результат той самий.

До речі, як можна бачити, повторне виконання тесту з позитивним результатом збільшує ймовірність того, що пацієнт справді хворий, із 0.16 до 0.78! □

Правило Бееса та закон повної ймовірності можна узагальнити на випадок обумовлення додатковими подіями.

**Твердження 3.3.7.** Нехай маємо вимірний простір  $(\Omega, \mathcal{A})$ . Для деяких подій  $A, B, C \in \mathcal{A}$ ,  $\mathbb{P}(B \cap C) > 0$ , справедливо:

$$\mathbb{P}(A | B, C) = \frac{\mathbb{P}(B | A, C) \mathbb{P}(A | C)}{\mathbb{P}(B | C)}. \quad (3.3.3)$$

**Твердження 3.3.8.** Нехай  $A_1, \dots, A_n$  утворюють розбиття вимірного простору  $(\Omega, \mathcal{A})$ , і до того ж  $\mathbb{P}(A_i \cap B) > 0$  для всіх  $i$ . Тоді

$$\mathbb{P}(B | C) = \sum_{i=1}^n \mathbb{P}(B | A_i, C) \mathbb{P}(A_i | C). \quad (3.3.4)$$

**Зауваження 3.3.9.** У формулі (3.3.3) непринципово, які саме події вважати обумовлюючими:

$$\mathbb{P}(A | B, C) = \frac{\mathbb{P}(C | A, B) \mathbb{P}(A | B)}{\mathbb{P}(C | B)} = \frac{\mathbb{P}(A \cap B \cap C)}{\mathbb{P}(B \cap C)}.$$

□

Табл. 3.3.1.: Результати лікування каменів у нирках для Прикладу 3.3.10

Операцій	Метод А			Метод В		
	Успішних	Відсоток успіхів	Операцій	Успішних	Відсоток успіхів	
Малі	87	81	93%	270	234	87%
Великі	263	192	73%	80	55	69%
Разом	350	273	78%	350	289	83%

**Приклад 3.3.10** (Парадокс Симпсона). Розгляньмо парадокс Симпсона (Simpson paradox)<sup>6</sup> на прикладі лікування каменів у нирках<sup>7</sup>. У Табл. 3.3.1 наведено результати лікування малих та великих каменів двома різними методами.

Для кожної категорії каменів метод А показав ліпші результати, проте загалом його результат нижчий від методу В (78% проти 83%). Така ситуація є парадоксальною. Спробуймо розібратися, чому так відбувається.

Нехай подія  $S$  = «лікування успішне»,  $A$  = «застосовано метод А»,  $B$  = «застосовано метод В» (очевидно, що  $A = B^c$ ),  $C$  = «лікували малі камені». Тоді, за (3.3.2),

$$\begin{aligned}\mathbb{P}(S | A) &= \mathbb{P}(S | C, A) \mathbb{P}(C | A) + \mathbb{P}(S | C^c, A) \mathbb{P}(C^c | A) , \\ \mathbb{P}(S | B) &= \mathbb{P}(S | C, B) \mathbb{P}(C | B) + \mathbb{P}(S | C^c, B) \mathbb{P}(C^c | B) .\end{aligned}$$

Відповідні ймовірності успішного лікування з застосуванням різних методів,  $\mathbb{P}(S | A)$  і  $\mathbb{P}(S | B)$ , є зваженими середніми ймовірностей успішного лікування каменів певного типу з застосуванням різних методів, а ваговими коефіцієнтами є  $\mathbb{P}(C | A)$ ,  $\mathbb{P}(C^c | A)$ ,  $\mathbb{P}(C | B)$  і  $\mathbb{P}(C^c | B)$ . Це нішо інше, як імовірності, що камені певного типу лікували певним методом, тобто простою мовою — частота лікування кожного типу хвороби кожним методом.

Якби ці ваги були рівними, то жодного парадоксу не виникло б, але оскільки

$$\mathbb{P}(C | B) > \mathbb{P}(C | A) , \quad \mathbb{P}(C^c | B) < \mathbb{P}(C^c | A) ,$$

то відбувається відповідний перекіс. На наших даних маємо:

$$\begin{aligned}\mathbb{P}(S | B) &= 0.87 \cdot \frac{270}{350} + 0.69 \cdot \frac{80}{350} \approx 0.83 , \\ \mathbb{P}(S | B^c) &= 0.93 \cdot \frac{87}{350} + 0.73 \cdot \frac{263}{350} \approx 0.78 ,\end{aligned}$$

як і вказано в таблиці. Ми бачимо, що в першому рівнянні значно вища вага першого доданку, а в другому рівнянні — другого доданку.

У загальному випадку, проблема виникає тоді, коли присутній *заважаючий чинник* (confounder), як тип каменю із нашого прикладу. Як правило, такі чинники визначають успішність того чи того підходу більше, ніж сам підхід.

Інший класичний приклад парадоксу Симпсона пов'язаний із аналізом результатів вступної кампанії 1973 р. в Університет Каліфорнії в Берклі. Загалом дані показали, що в аспірантуру та магістратуру набирали значно більше чоловіків, ніж жінок, що стало основною звинуваченю у

<sup>6</sup>Названий на честь британського статистика Едварда Симпсона (Edward H. Simpson) (1922–2019).

<sup>7</sup>C. R. Charig, D. R. Webb, S. R. Payne, J. E. Wickham. “Comparison of treatment of renal calculi by open surgery, percutaneous nephrolithotomy, and extracorporeal shockwave lithotripsy.” Br Med J (Clin Res Ed), 292 (6524), 879–882.

статевій дискримінації з боку університету. Проте, аналіз даних по окремих факультетах показав, що в більшості випадків жінок брали більше, ніж чоловіків. Як виявилося, загальна картина виявилася змазаною через те, що жінки переважно подавалися на ті факультети, де була вища конкуренція, а чоловіки надавали перевагу факультетам із низьким конкурсом.  $\square$

### 3.4. Незалежні події

За влучним висловом автора класичного підручника з теорії ймовірностей Р. Дарретта [5, стор. 43], теорія міри закінчується, а теорія ймовірності починається там, де з'являється визначення незалежності.

**Визначення 3.4.1.** Події  $A$  і  $B$  називають *незалежними* (independent) і позначають  $A \perp\!\!\!\perp B$ , якщо

$$\mathbb{P}(A \cap B) = \mathbb{P}(A) \mathbb{P}(B) . \quad (3.4.1)$$

Якщо  $\mathbb{P}(A) > 0$  і  $\mathbb{P}(B) > 0$ , то з (3.1.1) випливає, що події незалежні, якщо

$$\mathbb{P}(A | B) = \mathbb{P}(A) , \quad \mathbb{P}(B | A) = \mathbb{P}(B) .$$

$\square$

Іншими словами, події незалежні, коли обумовлення однієї події іншою жодним чином не впливає на її ймовірність.

Фактично, Визначення 3.4.1 формалізує інтуїтивне уявлення про незалежні події як результат незалежного вибору елементів. Наприклад, якщо  $A = \{a_1, \dots, a_m\}$ ,  $B = \{b_1, \dots, b_n\}$ , то існує  $m \cdot n$  можливих пар  $(a_i, b_j)$ , які можна утворити, «незалежно» вибираючи елементи з кожної множини.

**Зауваження 3.4.2.** Поняття незалежності не має жодного стосунку до поняття несумісності. Так, якщо  $\mathbb{P}(A \cap B) = 0$ , то події були би незалежні, якщо  $\mathbb{P}(A) = 0$  чи  $\mathbb{P}(B) = 0$ . Відповідно, якщо одночасно  $\mathbb{P}(A) > 0$ ,  $\mathbb{P}(B) > 0$ , то події не можуть бути незалежними, хоча вони несумісні. Тобто якщо ми знаємо, що сталася подія  $A$ , то через їх несумісність подія  $B$  ніяк не могла статися, і навпаки. Таким чином, ці події є залежними, адже інформація про одну з них впливає на ймовірність іншої.  $\square$

**Твердження 3.4.3.** Якщо події  $A \perp\!\!\!\perp B$ , то  $A \perp\!\!\!\perp B^c$ ,  $A^c \perp\!\!\!\perp B$ ,  $A^c \perp\!\!\!\perp B^c$ .

*Доведення.* Нехай події  $A \perp\!\!\!\perp B$ . Якщо  $\mathbb{P}(A) = 0$ , то  $A$  незалежна від будь-якої події, у тому числі  $B^c$ .

Якщо  $\mathbb{P}(A) > 0$ , то

$$\mathbb{P}(B^c | A) = 1 - \mathbb{P}(B | A) = 1 - \mathbb{P}(B) = \mathbb{P}(B^c) ,$$

а відтак події  $A$  і  $B^c$  незалежні.

Аналогічно доводять інші випадки.  $\square$

**Визначення 3.4.4.** Події  $A_1, \dots, A_n$  незалежні,  $A_1 \perp\!\!\!\perp \dots \perp\!\!\!\perp A_n$ , якщо  $\mathbb{P}(A_i \cap A_j) = \mathbb{P}(A_i)\mathbb{P}(A_j)$  для всіх  $i \neq j$ ,  $\mathbb{P}(A_i \cap A_j \cap A_k) = \mathbb{P}(A_i)\mathbb{P}(A_j)\mathbb{P}(A_k)$  для всіх  $i, j, k$  різних, і т.д.

Події зі зліченної послідовності  $A_1, \dots, A_n, \dots$  незалежні, якщо незалежними є будь-яке скінченне число подій із послідовності.  $\square$

**Приклад 3.4.5** (Попарна незалежність не означає незалежності). Нехай простір містить елементарні події

$$\Omega = \{abc, acb, bac, bca, cab, cba, aaa, bbb, ccc\} ,$$

імовірність кожної з яких  $1/9$ . Позначмо через  $A$  =«літера  $a$  опинилася на першій позиції»,  $B$  =«літера  $b$  опинилася на другій позиції» і  $C$  =«літера  $c$  опинилася на третьій позиції». Вочевидь,  $\mathbb{P}(A) = \mathbb{P}(B) = \mathbb{P}(C) = 1/3$ . Також маємо, що  $\mathbb{P}(A \cap B) = \mathbb{P}(A \cap C) = \mathbb{P}(B \cap C) = 1/9$ , тобто всі три події попарно незалежні.

Разом із тим,  $\mathbb{P}(A \cap B \cap C) = 1/9$ , оскільки тільки один елемент,  $abc$ , відповідає цій умові. Таким чином, оскільки  $1/9 \neq 1/27$ , події не є повністю незалежними.  $\square$

Між іншим, так само її інформація  $\mathbb{P}(A \cap B \cap C) = \mathbb{P}(A)\mathbb{P}(B)\mathbb{P}(C)$  не означає, що події незалежні, зокрема, що вони попарно незалежні. Адже якщо  $\mathbb{P}(A) = 0$ , то ця рівність є тотожністю, а відтак імовірності  $\mathbb{P}(B)$  і  $\mathbb{P}(C)$  можуть бути абсолютно довільні.

**Приклад 3.4.6.** Нехай у родині є  $n \geq 2$  дітей, а ймовірність народження дитини будь-якої статі дорівнює  $1/2$ . Нехай  $A_1$  =«у родині не більше від 1 дівчинки» і  $A_2$  =«у родині є і хлопчики, і дівчатка». Чи є ці події незалежними?

Позначивши  $(G = k)$  =«у родині рівно  $k$  дівчаток» і  $(B = k)$  =«у родині рівно  $k$  хлопчиків», маємо:

$$\mathbb{P}(A_1) = \mathbb{P}(G = 0) + \mathbb{P}(G = 1) = \frac{1}{2^n} + \frac{n}{2^n} ,$$

$$\mathbb{P}(A_2) = \mathbb{P}((G \geq 0) \cap (B \geq 0)) = 1 - \mathbb{P}((G = 0) \cup (B = 0)) = 1 - (\mathbb{P}(G = 0) + \mathbb{P}(B = 0)) = 1 - \frac{2}{2^n} .$$

З іншого боку,

$$\mathbb{P}(A_1 \cap A_2) = \mathbb{P}(G = 1) = \frac{n}{2^n} .$$

Прирівнюючи, маємо:

$$\frac{1+n}{2^n} \cdot \left(1 - \frac{2}{2^n}\right) = \frac{n}{2^n} ,$$

або

$$n + 1 = 2^{n-1} .$$

Це твердження істинне тоді й тільки тоді, коли  $n = 3$ . Таким чином, події  $A_1$  та  $A_2$  незалежні тоді й тільки тоді, коли  $n = 3$ .  $\square$

**Визначення 3.4.7.** Події  $A$  і  $B$  називають *умовно незалежними* (conditionally independent) за умови настання події  $C$ ,  $A \perp\!\!\!\perp B \mid C$ , якщо

$$\mathbb{P}(A \cap B \mid C) = \mathbb{P}(A \mid C)\mathbb{P}(B \mid C) .$$

□

**Приклад 3.4.8.** Події можуть бути умовно незалежними за умови події  $C$ , але не за умови події  $C^c$ .

Наприклад, нехай викладачів в університеті можна поділити на добрих і поганих. Добрий викладач поставить відмінно, якщо студент інтенсивно працюватиме. Поганий викладач поставить оцінку зі стелі незалежно від зусиль студента. Нехай  $G$  = «викладач добрий»,  $W$  = «студент інтенсивно працює»,  $A$  = «студенту поставили відмінно». Тоді  $\mathbb{P}(W \cap A | G^c) = \mathbb{P}(W | G^c) \mathbb{P}(A | G^c)$ , оскільки поганий викладач ставить оцінки незалежно від зусиль студента, але, очевидно, відповідна рівність не має місця для обумовлення подію  $G$ . □

**Приклад 3.4.9.** Із умовної незалежності не випливає незалежність безумовна.

Нехай є дві монетки — правильна і така, що герб випадає з імовірністю  $3/4$ . Випадковим чином обирають одну з них і підкидають  $n$  разів, і також невідомо, яку саме монетку обрано. Нехай  $F$  = «обрано правильну монетку»,  $A_i$  = «за  $i$ -им разом випав герб».

Тоді  $\mathbb{P}(A_1 \cap \dots \cap A_n | F) = \mathbb{P}(A_1 | F) \cdot \dots \cdot \mathbb{P}(A_n | F) = 1/2^n$ , оскільки підкиди монетки не впливають одне на одне. Так само  $\mathbb{P}(A_1 \cap \dots \cap A_n | F^c) = \mathbb{P}(A_1 | F^c) \cdot \dots \cdot \mathbb{P}(A_n | F^c) = 3^n/4^n$ . При цьому події  $A_i$ ,  $i = 1, \dots, n$ , не є безумовно незалежні, оскільки спостереження за послідовністю викидів змінюють апостеріорну ймовірність того, що монетка є правильною, а відтак імовірність кожного наступного герба змінюється. □

**Приклад 3.4.10.** Із безумовної незалежності не випливає незалежність умовна.

Нехай у деякого відлюдника є тільки двоє друзів, що можуть йому зателефонувати. Щодня вони незалежно один від одного вирішують, зателефонувати чи ні. Нехай  $A$  = «перший друг зателефонує наступного понеділка»,  $B$  = «інший друг зателефонує наступного понеділка»,  $\mathbb{P}(A) > 0$ ,  $\mathbb{P}(B) > 0$ .

Хоча ці події є безумовно незалежні, якщо обумовити їх подію  $C$  = «у понеділок надійшов тільки один дзвінок»,  $A$  і  $B$  перестають бути незалежними: дзвінок від одного друга виключає можливість дзвінка від другого. Іншими словами,  $\mathbb{P}(A | C) > 0$ , хоча  $\mathbb{P}(A | B, C) = 0$ . □

## 4. Дискретні випадкові величини. Схема Бернуллі

### 4.1. Поняття про випадкову величину

Поняття випадкової величини є фундаментальним для теорії ймовірностей і статистики. За великим рахунком, для кожного експерименту можна визначити деякий простір елементарних подій  $\Omega$ , задати на ньому  $\sigma$ -алгебру  $\mathcal{A}$  та ймовірнісну міру  $\mathbb{P}$ , і потім обчислювати ймовірності відповідних подій.

Проте на практиці в більшості випадків нас цікавить не сам результат експерименту (деякий елемент  $\omega \in \Omega$ ), а деяка чисрова функція від цього результату. Наприклад, нас може цікавити аналіз доходів працівників деякої галузі промисловості, і щоб не проводити опитування всіх мільйонів працівників (що на практиці є неможливим), ми можемо утворити деяку вибірку — випадковим чином відібрати працівників так, щоб інформація про їхні доходи була репрезентативна щодо доходів усієї популяції. Тоді сухо формально можна стверджувати, що результатом експерименту є множина *працівників*, проте для нас це зовсім нецікаво: нас цікавить множина *доходів* цих працівників. Ми хочемо мати можливість обчислювати ймовірності подій, сформульованих у термінах відповідних числових показників, напр., імовірність того, що дохід перевищує 10 000 грн або що.

Отже, неформально можна казати, що випадкова величина — це функція  $X : \Omega \rightarrow \mathbb{R}$ , що кожній елементарній події  $\omega \in \Omega$  зіставляє деяке значення  $X(\omega) \in \mathbb{R}$ , яке відображає суть досліджуваного явища (Рис. 4.1.1). Оскільки простір елементарних подій часто може бути доволі складним чи багатовимірним, а елементи  $\omega \in \Omega$  можуть бути довільні і необов'язково числові, використання випадкових величин дає змогу аналізувати відповідні явища у зручніший спосіб. Фактично, ми абстрагуємося від простору  $\Omega$ , і його початкова природа втрачає для нас своє значення.

Випадкові величини, як правило, позначаємо великими літерами латинського алфавіту ( $X, Y, Z$  тощо), а значення, яких вони набувають — маленькими ( $x, y, z$  тощо).

На одному й тому ж просторі  $\Omega$  можна визначати різні випадкові величини.

**Приклад 4.1.1.** Розгляньмо можливі випадкові величини:

1. Нехай викидують дві правильні шестигранні гральні кісточки. Простором елементарних подій є  $\Omega = \{(x, y) \in \{1, 2, 3, 4, 5, 6\}^2\}$ . Тоді можливими випадковими величинами є  $X$  = «сума випалих чисел»:  $X((1, 1)) = 2, X((1, 2)) = X((2, 1)) = 3$ , тощо, або  $Y$  = «сума випалих чисел є парним числом»,  $Y((1, 1)) = Y((1, 3)) = \dots = Y((6, 6)) = 1, Y((1, 2)) = Y((1, 4)) = \dots = Y((6, 5)) = 0$ .
2. Нехай оцінюють імовірність, що деякий пристрій пропрацює певну кількість часу. Простором елементарних подій  $\Omega$  є всі пристрої. Тоді  $X$  = «кількість часу нормальної роботи», і  $X(\omega) = t, t \in (0; \infty)$ . Іншою можливою величиною є  $Y$  = «вартість обслуговування протягом

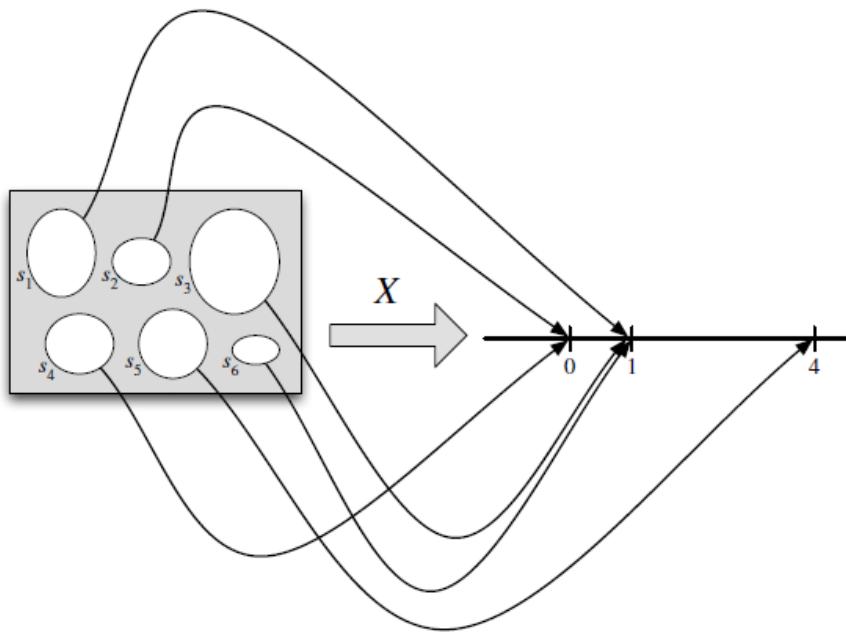


Рис. 4.1.1.: Ілюстрація випадкової величини  $X$  як функції, що відображає 6 елементів простору елементарних подій у числа 0, 1 і 4 ([1], Рис. 3.1)

часу  $t$ », де, скажімо,

$$Y(\omega) = 2 \left( 1 - 0.5e^{-0.2X(\omega)} \right) = Y(\omega) = 2 \left( 1 - 0.5e^{-0.2t} \right).$$

□

Варто зазначити, що хоча сама по собі функція  $X$  є детермінованою, «випадковість» у випадковій величині постає з випадкової природи експерименту. Результат експерименту  $\omega \in \Omega$  невідомий заздалегідь, і до його завершення можна хіба що оцінити ймовірність настання того чи того результату. Після завершення експерименту його результат зафікований, і випадкова величина набуває конкретного значення  $X(\omega)$ .

Даймо формальне визначення випадкової величини, яке узагальнює вищевикладену інтуїцію.

**Визначення 4.1.2.** Нехай маємо два вимірні простори  $(\Omega, \mathcal{A})$  і  $(\Omega', \mathcal{A}')$ . Функцію  $X : \Omega \rightarrow \Omega'$  називають *вимірною відносно  $\mathcal{A}$*  (measurable  $\mathcal{A}$  function), якщо для будь-якого  $A' \in \mathcal{A}'$  справедливо

$$X^{-1}(A') \equiv \{\omega \in \Omega : X(\omega) \in A'\} \in \mathcal{A},$$

тобто що *прообразом* (inverse image) кожної вимірної множини в одному просторі є деяка вимірна множина в іншому просторі.

Якщо з контексту очевидно, відносно якої  $\sigma$ -алгебри функція є вимірною, то просто кажуть, що вона вимірна. □

У загальному випадку не всі функції обов'язково повинні бути вимірними, проте всі функції, які можуть виринуті на практиці, є вимірними.

**Приклад 4.1.3.** Розгляньмо простір  $\Omega = \{1, 2, 3, 4\}$ . На цьому просторі визначмо дві  $\sigma$ -алгебри<sup>1</sup>:

$$\begin{aligned}\mathcal{A}_1 &= \{\emptyset, \{1\}, \{3\}, \{1, 3\}, \{2, 4\}, \{1, 2, 4\}, \{2, 3, 4\}, \{1, 2, 3, 4\}\} , \\ \mathcal{A}_2 &= \{\emptyset, \{1\}, \{4\}, \{1, 4\}, \{2, 3\}, \{1, 2, 3\}, \{2, 3, 4\}, \{1, 2, 3, 4\}\} .\end{aligned}$$

Можна перевірити, що ці множини замкнені відносно заперечень і (зліченних) об'єднань.

Нехай  $f : \Omega \rightarrow \mathbb{R}$  така, що  $f(x) = (-1)^x$ . Тоді ця функція може набувати всього двох значень, 1 і  $-1$ . Усі інші дійсні числа не мають своїх прообразів, тобто, наприклад,  $f^{-1}(2) = \emptyset$ .

Розгляньмо прообрази 1 і  $-1$ :

$$f^{-1}(1) = \{2, 4\} , f^{-1}(-1) = \{1, 3\} .$$

Оскільки  $\{2, 4\} \in \mathcal{A}_1$  і  $\{1, 3\} \in \mathcal{A}_1$ , проте  $\{2, 4\} \notin \mathcal{A}_2$ ,  $\{1, 3\} \notin \mathcal{A}_2$ , можемо зробити висновок, що функція  $f$  вимірна відносно  $\mathcal{A}_1$ , але не відносно  $\mathcal{A}_2$ .  $\square$

**Зауваження 4.1.4.** Будь-яка функція  $X : \Omega \rightarrow \mathbb{R}$ , де  $\Omega$  — дискретний імовірнісний простір, є вимірною, адже  $\sigma$ -алгебра  $2^\Omega$  для такого простору включає всі його можливі підмножини.  $\square$

**Визначення 4.1.5.** Дійснозначну вимірну функцію  $X : \Omega \rightarrow \mathbb{R}$  називають *випадковою величиною* (random variable).

При цьому за замовчуванням кладуть, що  $\sigma$ -алгеброю для простору  $\mathbb{R}$  є Борелева  $\sigma$ -алгебра  $\mathcal{B}$ .  $\square$

Таким чином, маємо ситуацію, коли з початкового абстрактного ймовірнісного простору  $(\Omega, \mathcal{A}, \mathbb{P})$  ми переходимо до нового простору  $(\mathbb{R}, \mathcal{B})$ . Нам потрібно навчитися обчислювати ймовірності подій у цьому новому просторі, але для цього потрібно з'ясувати, яку ймовірнісну міру на ньому використати.

**Визначення 4.1.6.** Функція (випадкова величина)  $X$  задає на вимірному просторі  $(\mathbb{R}, \mathcal{B})$  ймовірнісну міру

$$\mathbb{P}_X(A) \equiv \mathbb{P}_X(X \in A) = \mathbb{P}(\{\omega \in \Omega : X(\omega) \in A\}) = \mathbb{P}(X^{-1}(A)) , \quad A \in \mathcal{B} , \quad (4.1.1)$$

яку називають *розділом випадкової величини X* (distribution of a random variable). Також кажуть, що  $\mathbb{P}_X$  є *образом міри*  $\mathbb{P}$  під дією функції  $X$  ( $\mathbb{P}_X$  is a measure induced by  $X$ ).  $\square$

**Зауваження 4.1.7.** Із Визначення 4.1.6 стає зрозумілим, навіщо ми вимагали, щоб випадкова величина була вимірною функцією: обчислення ймовірнісних тверджень відносно випадкових величин зводиться до обчислення ймовірностей деяких підмножин  $\Omega$ . Вимагаючи, щоб випадкова величина була вимірною, ми гарантуємо, що такі ймовірності можна буде обчислити.  $\square$

<sup>1</sup> Як можна бачити, це не  $\sigma$ -алгебри, які ми звикли брати, тобто це не множини всіх підмножин простору  $\Omega$ .

Можемо довести, що розподіл випадкової величини справді є ймовірнісною мірою.

**Твердження 4.1.8.** Розподіл  $\mathbb{P}_X$  випадкової величини  $X : (\Omega, \mathcal{A}) \rightarrow (\mathbb{R}, \mathcal{B})$  є ймовірнісною мірою на  $(\mathbb{R}, \mathcal{B})$  в розумінні Визначення 2.1.5.

*Доведення.* Потрібно перевірити виконання трьох аксіом імовірнісної міри.

Очевидно, що розподіл є невід'ємним, адже (4.1.1) визначає його через імовірнісну міру, яка сама є невід'ємною.

Перевірмо,  $\mathbb{P}_X(\mathbb{R}) = 1$ :

$$\mathbb{P}_X(\mathbb{R}) = \mathbb{P}_X(X \in \mathbb{R}) = \mathbb{P}(\{\omega \in \Omega : X(\omega) \in \mathbb{R}\}) = \mathbb{P}(\Omega) = 1.$$

Залишилося показати  $\sigma$ -адитивність. Нехай  $A_1, A_2, \dots$  — деякі несумісні події з  $\mathcal{B}$ . Тоді можемо використати  $\sigma$ -адитивність початкової міри  $\mathbb{P}$ :

$$\begin{aligned} \mathbb{P}_X\left(\bigcup_{i=1}^{\infty} A_i\right) &= \mathbb{P}_X\left(X \in \bigcup_{i=1}^{\infty} A_i\right) = \mathbb{P}\left(\left\{\omega \in \Omega : X(\omega) \in \bigcup_{i=1}^{\infty} A_i\right\}\right) \\ &= \mathbb{P}\left(\bigcup_{i=1}^{\infty} \{\omega \in \Omega : X(\omega) \in A_i\}\right) = \sum_{i=1}^n \mathbb{P}(\{\omega \in \Omega : X(\omega) \in A_i\}) = \sum_{i=1}^n \mathbb{P}_X(X \in A_i). \end{aligned}$$

□

Відомо, що для будь-якої функції  $f : \mathbb{R} \rightarrow \mathbb{R}$  *носієм* є множина точок, у яких вона не дорівнює 0. За аналогією з носієм функції можна визначити поняття носія випадкової величини.

**Визначення 4.1.9.** *Носієм* (support) випадкової величини  $X$  називають множину  $\text{supp}(X) = S$ , яка є запереченням найбільшої відкритої множини  $O$  такої, що  $\mathbb{P}_X(O) = 0$ . Іншими словами, носій — це найменша замкнена множина, у якій «сконцентровано» всю ймовірність розподілу  $\mathbb{P}_X$ .

□

**Визначення 4.1.10.** Якщо випадкові величини  $X$  і  $Y$  задають на  $(\mathbb{R}, \mathcal{B})$  один і той самий розподіл, тобто  $\mathbb{P}_X(A) = \mathbb{P}_Y(A)$  для всіх  $A \in \mathcal{B}$ , то кажуть, що вони *рівні за розподілом* (equal in distribution):  $X \stackrel{d}{=} Y$ .

□

## 4.2. Дискретні випадкові величини

У Розд. 2.3 ми розглядали спосіб побудови ймовірнісної міри на дискретному просторі. Розподіли, які є образами дискретних імовірнісних мір, також називають дискретними, як і відповідні випадкові величини. У цьому випадку, оскільки початковий (дискретний) простір є зліченим, то випадкова величина може набувати також не більш ніж зліченну кількість значень. Як виявляється, саме це і є основним для визначення поняття дискретної випадкової величини та дискретного розподілу.

**Визначення 4.2.1.** Випадкову величину  $X$  називають *дискретною*, як і відповідний їй розподіл, якщо носій  $\text{supp}(X)$  є не більш ніж зліченим.

□

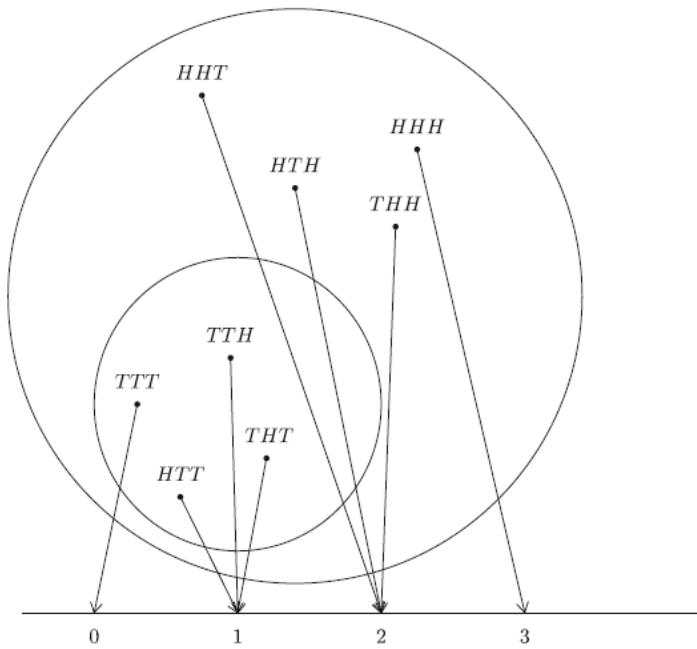


Рис. 4.2.1.: Ілюстрація до Прикладу 4.2.2 ([2], Рис. 2.8)

Відтак, хоча формально ми пишемо  $X : \Omega \rightarrow \mathbb{R}$ , потрібно пам'ятати, що насправді областю значень функції  $X$  є зліченна підмножина дійсних чисел (наприклад, усі натуральні числа тощо):  $X : \Omega \rightarrow \{x_1, x_2, \dots\}$ ,  $x_i \in \mathbb{R}$ . Тоді носієм випадкової величини  $X$  є множина  $\text{supp}(X) = \{x_1, x_2, \dots\}$ , у тому розумінні, що  $\mathbb{P}_X(\mathbb{R} \setminus \{x_1, x_2, \dots\}) = 0$ .

**Приклад 4.2.2.** Розгляньмо підкидування правильних монеток тричі поспіль та випадкову величину  $X$  = «сума випалих гербів». Позначивши герби за 1, а числа — за 0, маємо, що ймовірнісним простором у цьому випадку є  $(\Omega = \{000, 001, \dots, 111\}, \mathcal{A} = 2^\Omega, \mathbb{P})$ , де  $\mathbb{P}$  — дискретна ймовірнісна міра із Теореми 2.3.1, а кожна елементарна подія має однакову ймовірність  $1/8$ . Функція  $X$  на буває таких значень:  $X(000) = 0$ ,  $X(001) = X(010) = X(100) = 1$  тощо (Рис. 4.2.1). Тобто  $X$  має носій  $\text{supp}(X) = \{0, 1, 2, 3\}$ .

Тоді ймовірність події «випало не більше від 1 герба» можна обчислити так:

$$\mathbb{P}_X(X \leq 1) = \mathbb{P}(\{\omega \in \Omega : X(\omega) \leq 1\}) = \mathbb{P}(\{000, 001, 010, 100\}) = \frac{1}{8} + \frac{1}{8} + \frac{1}{8} + \frac{1}{8} = \frac{1}{2}.$$

□

У Лекції 2 ми відзначали, що ймовірнісну міру на деякій  $\sigma$ -алгебрі можна задати, визначивши деяку функцію на множинах із певного невеликого класу, наприклад, усіх пів інтервалів із  $\mathbb{R}$ . Для зручної роботи з розподілами випадкових величин було б також доречно мати можливість оперувати деякими функціями, які давали б змогу обчислювати ймовірності різних подій, не використовуючи щоразу формальне Визначення 4.1.5. Такими функціями є *функції розподілу* (distribution functions), які детально розглядаємо в наступних лекціях. Проте для дискретних випадкових величин кориснішим є поняття функції ймовірності, яке випливає з Теореми 2.3.1.

**Визначення 4.2.3.** Для опису розподілу  $\mathbb{P}_X$  дискретної випадкової величини  $X : \Omega \rightarrow \mathbb{R}$  з носієм  $\text{supp}(X) = \{x_1, x_2, \dots\}$  використовують *функцію ймовірності* (probability mass function)  $p_X(x) \equiv \mathbb{P}_X(X = x)$  таку, що:

$$p_X(x) \begin{cases} \neq 0, & x \in \{x_1, x_2, \dots\} \\ = 0, & x \notin \{x_1, x_2, \dots\} \end{cases}, \quad \sum_i p_X(x_i) = 1. \quad (4.2.1)$$

Тоді, за Теоремою 2.3.1, функція ймовірності визначатиме міру  $\mathbb{P}_X$  — розподіл випадкової величини. Зокрема,

$$\mathbb{P}_X(A) = \sum_{x_i \in A} p_X(x_i). \quad (4.2.2) \quad \square$$

**Приклад 4.2.4.** Розгляньмо експеримент із підкидуванням правильної монетки двічі поспіль. Імовірнісним простором у цьому випадку є  $(\Omega, 2^\Omega, \mathbb{P})$ , де  $\Omega = \{00, 01, 10, 11\}$ ,  $\mathbb{P}$  — дискретна ймовірнісна міра, де кожна  $\omega \in \Omega$  має однакову ймовірність  $1/4$ .

1. Нехай  $X$  = «число випалих гербів». Носієм цієї величини є  $\text{supp}(X) = \{0, 1, 2\}$ , а її функція ймовірності має вид

$$p_X(x) = \begin{cases} \frac{1}{4}, & x = 0 \\ \frac{1}{2}, & x = 1 \\ \frac{1}{4}, & x = 2 \\ 0, & x \neq 0, 1, 2 \end{cases}.$$

Використовуючи цю функцію, можна обчислювати ймовірності різних подій, абстрагувавшись від початкового простору елементарних подій. Наприклад,

$$\mathbb{P}_X(X \leq 1) = \mathbb{P}_X(X = 0) + \mathbb{P}_X(X = 1) = \frac{1}{4} + \frac{1}{2} = \frac{3}{4},$$

або ж

$$\mathbb{P}_X(X \leq 1) = 1 - \mathbb{P}_X(X = 2) = 1 - \frac{1}{4} = \frac{3}{4}.$$

2. Нехай  $Y$  = «число випалих чисел». Вочевидь,  $Y = 2 - X$ , і до того ж  $p_Y(z) = p_X(z)$  для всіх  $z = 0, 1, 2$ , зокрема тому що

$$\begin{aligned} \mathbb{P}_Y(Y = y) &= \mathbb{P}(\{\omega : Y(\omega) = y\}) \\ &= \mathbb{P}(\{\omega : 2 - X(\omega) = y\}) \\ &= \mathbb{P}(\{\omega : X(\omega) = 2 - y\}) = \mathbb{P}_X(X = 2 - y). \end{aligned}$$

Іншими словами,  $X \stackrel{d}{=} Y$ , тобто різні випадкові величини можуть мати одинаковий розподіл.

3. Нехай  $I = 1$ , якщо перший підкид завершився гербом, і  $0$  — у протилежному випадку, тобто

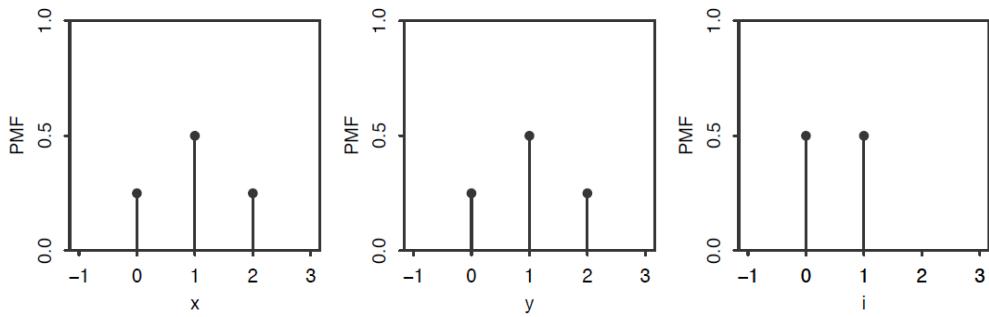


Рис. 4.2.2.: Приклади функцій імовірності для випадкових величин із Прикладу 4.2.4 (зліва направо:  $p_X(x)$ ,  $p_Y(y)$ ,  $p_I(i)$ ) ([1], Рис. 3.3)

$I(10) = I(11) = 1$ ,  $I(00) = I(01) = 0$ . Функція ймовірності має вид

$$p_I(x) = \begin{cases} \frac{1}{2}, & x = 0 \\ \frac{1}{2}, & x = 1 \\ 0, & x \neq 0, 1 \end{cases}.$$

Функції ймовірності вказаних випадкових величин наведено на Рис. 4.2.2. □

**Зауваження 4.2.5.** Для того, щоб постійно не уточнювати, що функція ймовірності дорівнює нулю за межами носія випадкової величини, у дальншому викладі будемо просто вказувати носій та значення функції ймовірності на ньому.

Також будемо казати про носій розподілу, розуміючи під цим носій будь-якої випадкової величини, яка має цей розподіл. □

Як можна бачити навіть із простих прикладів, один і той самий розподіл може бути використано для опису випадкових величин різного характеру. Потрібно розуміти, що в процесі розгляду дискретних (а потім і неперервних) випадкових величин ми часто використовуємо прості експерименти на кшталт підкидання монеток чи витягання кульок із ваз, але ці ситуації є ізоморфними іншим задачами прикладного характеру.

### 4.3. Сподівання та дисперсія дискретної випадкової величини

Сподівання<sup>2</sup> будь-якої випадкової величини є формалізацією поняття *середнього значення* (mean) випадкової величини. Загальну теорію сподівань ми розглянемо в наступних лекціях, а поки що зосередимося на простому випадку сподівання дискретної випадкової величини.

<sup>2</sup>Інколи кажуть «математичне сподівання».

Як відомо, *арифметичне середнє* (arithmetic mean) деяких чисел  $x_1, \dots, x_n$  дорівнює

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i .$$

Узагальнення арифметичного середнього є *зважене середнє* (weighted mean):

$$w(x) = \sum_{i=1}^n x_i w_i ,$$

де  $w_i \in [0; 1]$  — *вагові коефіцієнти* (weights) такі, що  $\sum_i w_i = 1$ . У випадку арифметичного середнього, очевидно,  $w_i = \frac{1}{n}$  для всіх  $i$ . Зважене середнє є гнучкішим поняттям, адже дає змогу надавати різної ваги різним значенням.

Зокрема, для дискретної випадкової величини природним є вибір вагових коефіцієнтів як імовірностей, із якими випадкова величина набуває своїх значень.

**Визначення 4.3.1.** *Сподіванням* (expectation) дискретної випадкової величини  $X$  є зважене середнє

$$\mathbb{E}[X] = \sum_{x \in \text{supp}(X)} x \cdot \mathbb{P}_X(X = x) . \quad (4.3.1)$$

Сподівання може бути нескінченно великим, а якщо одночасно

$$\mathbb{E}[X] = \sum_{\substack{x \in \text{supp}(X) \\ x > 0}} x \cdot \mathbb{P}_X(X = x) = \infty , \quad \mathbb{E}[X] = \sum_{\substack{x \in \text{supp}(X) \\ x < 0}} (-x) \cdot \mathbb{P}_X(X = x) = \infty ,$$

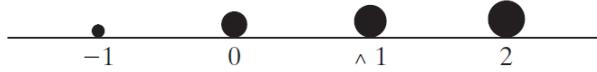
то кажуть, що сподівання *не існує* (does not exist).  $\square$

**Зауваження 4.3.2.** Надзвичайно важливо усвідомлювати, що  $\mathbb{E}[X]$ , якщо воно існує, є сталою (дійсним числом  $\mathbb{E}[X] \in \mathbb{R}$  або нескінченно великим значенням). У той же час  $X : \Omega \rightarrow \mathbb{R}$ , тобто випадкова величина є (випадковою) *функцією*. Зокрема, це означає, що сподівання не несе в собі жодної випадковості: це є фіксоване значення, яке заздалегідь відоме і може бути обчислено.  $\square$

Часто сподівання  $\mathbb{E}[X]$  позначають через  $\mu_X$ , але ми цього намагатимемося уникати, щоб не виникало плутанини з позначенням міри.

**Зауваження 4.3.3.** Поняття сподівання подібне до поняття центру мас у фізиці. Нехай маємо дискретну випадкову величину  $X$  із функцією ймовірності  $p(x_i)$ ,  $i \geq 1$ . Якщо уявити невагомий стрижень, на якому розміщено кулі масою  $p(x_i)$  в точках  $x_i$ ,  $i \geq 1$ , то точка, в якій стрижень перебуватиме в стані рівноваги, має називу центру мас і відповідає сподіванню  $\mathbb{E}[X]$  (Рис. 4.3.1).  $\square$

**Приклад 4.3.4.** Нехай страхова компанія виплачує 10 тис. грн за багаж, утрачений під час авіаперельоту. Історичні дані свідчать, що страховий випадок настає для 1 із 200 проданих полісів. Яка повинна бути вартість одного полісу?



$$p(-1) = .10, \quad p(0) = .25, \quad p(1) = .30, \quad p(2) = .35$$

$$\wedge = .9$$

Рис. 4.3.1.: Ілюстрація до Замітки 4.3.3, позначка  $\wedge$  відповідає центру мас ([8], Рис. 4.4)

Випадкова величина  $X$  = «втрати страхової компанії» в даному випадку має таку функцію ймовірності:

$$p_X(x) = \begin{cases} \frac{1}{200}, & x = -10,000 \\ 1 - \frac{1}{200}, & x = 0 \end{cases}.$$

Тоді сподівані втрати компанії становлять

$$\mathbb{E}[X] = -\frac{10,000}{200} + 0 \cdot \left(1 - \frac{1}{200}\right) = -50.$$

Таким чином, для покриття видатків для страхової компанії вигідно визначити вартість полісу в деяку суму більше 50 грн.  $\square$

Можна показати, що сподівання має властивість лінійності.

**Твердження 4.3.5.** Для будь-яких випадкових величин<sup>3</sup>  $X$  і  $Y$ , визначених на одному й тому ж імовірнісному просторі  $(\Omega, \mathcal{A}, \mathbb{P})$ , та будь-яких чисел  $a, b \in \mathbb{R}$  справедливо таке:

$$\mathbb{E}[aX + bY] = a\mathbb{E}[X] + b\mathbb{E}[Y]. \quad (4.3.2)$$

*Доведення.* Доведімо цю властивість для дискретних випадкових величини.

Нехай  $\text{supp}(X) = \{x_1, x_2, \dots\}$  — носій випадкової величини  $X$ . Тоді, за (4.1.1),

$$\mathbb{P}_X(X = x_i) = \mathbb{P}(\{\omega : X(\omega) = x_i\}),$$

а відтак

$$\begin{aligned} \mathbb{E}[X] &= \sum_{x \in \text{supp}(X)} x \mathbb{P}_X(X = x) = \sum_{x \in \text{supp}(X)} x \mathbb{P}(\{\omega : X(\omega) = x\}) \\ &= \sum_{x \in \text{supp}(X)} x \sum_{\omega : X(\omega) = x} \mathbb{P}(\{\omega\}) \\ &= \sum_{x \in \text{supp}(X)} \sum_{\omega : X(\omega) = x} X(\omega) \mathbb{P}(\{\omega\}) = \sum_{\omega \in \Omega} X(\omega) \mathbb{P}(\{\omega\}). \end{aligned}$$

Аналогічно можна подати

$$\mathbb{E}[Y] = \sum_{\omega \in \Omega} Y(\omega) \mathbb{P}(\{\omega\}).$$

<sup>3</sup>Будь-яких, не тільки дискретних, але тут ми доведемо для дискретних.

Тоді,

$$\begin{aligned}
 \mathbb{E}[aX + bY] &= \sum_{\omega \in \Omega} (aX + bY)(\omega) \mathbb{P}(\{\omega\}) \\
 &= \sum_{\omega \in \Omega} (aX(\omega) + bY(\omega)) \mathbb{P}(\{\omega\}) \\
 &= a \sum_{\omega \in \Omega} X(\omega) \mathbb{P}(\{\omega\}) + b \sum_{\omega \in \Omega} Y(\omega) \mathbb{P}(\{\omega\}) = a\mathbb{E}[X] + b\mathbb{E}[Y] .
 \end{aligned}$$

□

Сподівання також можна обчислити для деякої функції від випадкової величини. Загальний випадок розглядатимемо в наступних лекціях, а поки що зосередимося на дискретних величинах.

**Теорема 4.3.6** (Закон несвідомого статистика для дискретних величин (Law of unconscious statistician for discrete variables)). Нехай  $X$  — деяка дискретна випадкова величина з носієм  $\text{supp}(X) = \{x_1, x_2, \dots\}$  і функцією ймовірності  $p_X(x) = \mathbb{P}_X(X = x)$ , і нехай  $g : \mathbb{R} \rightarrow \mathbb{R}$  — деяка функція. Тоді

$$\mathbb{E}[g(X)] = \sum_{x \in \text{supp}(X)} g(x) \mathbb{P}_X(X = x) . \quad (4.3.3)$$

*Доведення.* Застосовуючи поелементно функцію  $g$ , дістаємо носій величини  $g(X)$ :  $\text{supp}(g(X)) = \{g(x_1), g(x_2), \dots\}$ . Варто зазначити, що деякі образи елементів носія величини  $X$  можуть бути одинакові ( $g(x_i) = g(x_j)$  для деяких  $i \neq j$ ).

Тоді, за (4.1.1),

$$\mathbb{P}_{g(X)}(g(X) = x'_i) = \mathbb{P}(\{\omega : g(X(\omega)) = x'_i\}) ,$$

а відтак

$$\begin{aligned}
 \mathbb{E}[g(X)] &= \sum_{x' \in \text{supp}(g(X))} x' \mathbb{P}_{g(X)}(g(X) = x') = \sum_{x' \in \text{supp}(g(X))} x' \mathbb{P}(\{\omega : g(X(\omega)) = x'\}) \\
 &= \sum_{x' \in \text{supp}(g(X))} x' \sum_{\omega : g(X(\omega)) = x'} \mathbb{P}(\{\omega\}) \\
 &= \sum_{x' \in \text{supp}(g(X))} \sum_{\omega : g(X(\omega)) = x'} g(X(\omega)) \mathbb{P}(\{\omega\}) \\
 &= \sum_{\omega \in \Omega} g(X(\omega)) \mathbb{P}(\{\omega\}) \\
 &= \sum_{x \in \text{supp}(X)} \sum_{\omega : X(\omega) = x} g(x) \mathbb{P}(\{\omega\}) \\
 &= \sum_{x \in \text{supp}(X)} g(x) \sum_{\omega : X(\omega) = x} \mathbb{P}(\{\omega\}) \\
 &= \sum_{x \in \text{supp}(X)} g(x) \mathbb{P}(\{\omega : X(\omega) = x\}) \\
 &= \sum_{x \in \text{supp}(X)} g(x) \mathbb{P}_X(X = x) .
 \end{aligned}$$

□

**Приклад 4.3.7.** Нехай  $X$  — випадкова величина, функція ймовірності якої має вид:

$$p_X(x) = \begin{cases} 0.2, & x = -1 \\ 0.5, & x = 0 \\ 0.3, & x = 1 \end{cases}.$$

Тоді

$$\mathbb{E}[X^2] = \sum_{x \in \text{supp}(X)} x^2 p_X(x) = (-1)^2 \cdot 0.2 + 0^2 \cdot 0.5 + 1^2 \cdot 0.3 = 0.5.$$

Варто звернути увагу, що

$$0.5 = \mathbb{E}[X^2] \neq (\mathbb{E}[X])^2 = (0.1)^2 = 0.01.$$

□

Одного лише сподівання для опису розподілу випадкової величини явно недостатньо. Так, наприклад, випадкові величини  $X$  і  $Y$  із функціями ймовірності

$$p_X(x) = \begin{cases} \frac{5}{8}, & x = -1 \\ \frac{1}{8}, & x = 1 \\ \frac{2}{8}, & x = 2 \end{cases}, \quad p_Y(y) = \begin{cases} \frac{5}{8}, & y = -10 \\ \frac{1}{8}, & y = 10 \\ \frac{2}{8}, & y = 20 \end{cases}$$

мають однакове сподівання,  $\mathbb{E}[X] = \mathbb{E}[Y] = 0$ , хоча, очевидно, їхні розподіли суттєво різняться.

Ми можемо застосувати Теорему 4.3.6 для визначення ще однієї надзвичайно важливої числової характеристики випадкової величини — її дисперсії.

**Визначення 4.3.8.** *Дисперсією* (variance) дискретної випадкової величини  $X$  є сподівання квадратів відхилень її значень від її сподівання:

$$\text{Var}(X) = \mathbb{E}[(X - \mathbb{E}[X])^2] = \sum_{x \in \text{supp}(X)} (x - \mathbb{E}[X])^2 \cdot \mathbb{P}_X(X = x). \quad (4.3.4)$$

Додатне значення квадратового кореня з дисперсії має назву *середньоквадратичного відхилення* (standard deviation):

$$\text{sd}(X) = \sqrt{\text{Var}(X)} = \sqrt{\sum_{x \in \text{supp}(X)} (x - \mathbb{E}[X])^2 \cdot \mathbb{P}_X(X = x)}. \quad (4.3.5)$$

□

Часто дисперсію позначають через  $\sigma_X^2$ , і тоді середньоквадратичне відхилення має позначення  $\sigma_X$ .

Дисперсія показує, наскільки (в середньому) далеко стоять від  $\mathbb{E}[X]$  значення величини  $X$ , тобто наскільки сильно значення величини  $X$  *розділено* відносно її сподівання. Оскільки одиниці виміру дисперсії є квадратами від одиниць виміру самої величини, на практиці застосовують середньоквадратичне відхилення, щоб можна було коректно порівнювати сумірні величини.

**Зауваження 4.3.9.** Використання квадратів у визначені дисперсії потрібно для того, щоб усі відстані від сподівання були додатними. Зокрема, за властивостями лінійності сподівання,

$$\mathbb{E}[X - \mathbb{E}[X]] = \mathbb{E}[X] - \mathbb{E}[\mathbb{E}[X]] = \mathbb{E}[X] - \mathbb{E}[X] = 0,$$

оскільки додатні відхилення компенсують від'ємні відхилення.

Альтернативним способом визначити дисперсію могло б бути сподівання виду  $\mathbb{E}[|X - \mathbb{E}[X]|]$ , адже модуль також перетворює від'ємні значення в додатні. Проте модуль є недиференційовним у точці 0, а тому його використання ускладнює аналіз властивостей випадкових величин.  $\square$

На практиці для обчислення дисперсії використовують альтернативну формулу її запису.

**Твердження 4.3.10.** Для будь-якої випадкової величини  $X$ <sup>4</sup>

$$\text{Var}(X) = \mathbb{E}[X^2] - (\mathbb{E}[X])^2. \quad (4.3.6)$$

*Доведення.* За визначенням дисперсії (4.3.4),

$$\begin{aligned} \text{Var}(X) &= \mathbb{E}[(X - \mathbb{E}[X])^2] \\ &= \mathbb{E}[X^2 - 2X\mathbb{E}[X] + (\mathbb{E}[X])^2] = \mathbb{E}[X^2] - 2\mathbb{E}[X]\mathbb{E}[X] + (\mathbb{E}[X])^2 = \mathbb{E}[X^2] - (\mathbb{E}[X])^2. \end{aligned}$$

$\square$

Наразі не будемо розглядати властивостей дисперсії: повернемося до відповідного аналізу, коли будемо розглядати неперервні випадкові величини. Відзначмо тут тільки, що дисперсія *не є* лінійною: для деяких констант  $a, b \in \mathbb{R}$  та деякої випадкової величини  $X$

$$\text{Var}(aX + b) = \mathbb{E}[(aX + b - \mathbb{E}[aX + b])^2] = \mathbb{E}[a^2(X - \mathbb{E}[X])^2] = a^2\text{Var}(X).$$

**Приклад 4.3.11.** Розгляньмо для правильної гральної кісточки випадкову величину  $X = \text{«число, яке випало»}$ . Її числові характеристики такі:

$$\begin{aligned} \mathbb{E}[X] &= \sum_{i=1}^6 i \cdot \frac{1}{6} = \frac{21}{6} = \frac{7}{2}, \\ \mathbb{E}[X^2] &= \sum_{i=1}^6 i^2 \cdot \frac{1}{6} = \frac{91}{6}, \\ \text{Var}(X) &= \mathbb{E}[X^2] - (\mathbb{E}[X])^2 = \frac{91}{6} - \frac{49}{4} = \frac{35}{12}, \\ \text{sd}(X) &= \sqrt{\frac{35}{12}} \approx 1.71. \end{aligned}$$

$\square$

Оскільки дисперсія завжди є невід'ємною (сума квадратів, помножених на ймовірності), то з (4.3.6) випливає, що

$$\mathbb{E}[X^2] \geq (\mathbb{E}[X])^2. \quad (4.3.7)$$

<sup>4</sup>Будь-якої, не тільки дискретної.

**Приклад 4.3.12.** Покладімо, що люди народжуються в кожний  $i$ -ий день року,  $i = 1, 2, \dots, 365$ , незалежно один від одного з імовірностю  $p_i$ ,  $\sum_{i=1}^{365} p_i = 1$ . Нехай  $A_{s,t}$  = «особи  $s$  і  $t$  народилися в один день». Підрахуймо ймовірність  $\mathbb{P}(A_{1,2})$ .

Нехай  $B_i$  = «особи 1 і 2 народилися в день  $i$ ». Через незалежність маємо  $\mathbb{P}(B_i) = p_i \cdot p_i = p_i^2$ . Оскільки  $A_{1,2}$  є об'єднанням  $B_i$ ,  $i = 1, \dots, 365$ , маємо:

$$\mathbb{P}(A_{1,2}) = \sum_{i=1}^{365} p_i^2.$$

Підрахуймо тепер імовірність  $\mathbb{P}(A_{1,2} | A_{1,3})$ . Нехай  $C_i$  = «особи 1, 2 і 3 народилися в день  $i$ ». Через незалежність маємо  $\mathbb{P}(C_i) = p_i^3$ . Оскільки  $A_{1,2} \cap A_{1,3}$  є об'єднанням  $C_i$ ,  $i = 1, \dots, 365$ , маємо:

$$\mathbb{P}(A_{1,2} | A_{1,3}) = \frac{\mathbb{P}(A_{1,2} \cap A_{1,3})}{\mathbb{P}(A_{1,2})} = \frac{\sum_{i=1}^{365} p_i^3}{\sum_{i=1}^{365} p_i^2},$$

адже подія  $A_{1,2} \cap A_{1,3}$  є об'єднанням несумісних подій, кожна з яких відповідає факту народження *трьох* осіб в один і той же день  $i$ ,  $i = 1, \dots, 365$ .

Розгляньмо випадкову величину  $X$ , яка дорівнює  $p_i$  з імовірністю  $p_i$ . Тоді

$$\mathbb{E}[X] = \sum_{i=1}^{365} p_i \cdot p_i = \sum_{i=1}^{365} p_i^2, \quad \mathbb{E}[X^2] = \sum_{i=1}^{365} p_i \cdot p_i^2 = \sum_{i=1}^{365} p_i^3.$$

Виходячи з (4.3.7), маємо:

$$\mathbb{E}[X^2] \geq (\mathbb{E}[X])^2 \Leftrightarrow \sum_{i=1}^{365} p_i^3 \geq \left( \sum_{i=1}^{365} p_i^2 \right)^2 \Leftrightarrow \mathbb{P}(A_{1,2} | A_{1,3}) \geq \mathbb{P}(A_{1,2}),$$

іншими словами, імовірність народитися в один і той самий день, за умови, що в цей самий день народилися дві особи, більше чи дорівнює безумовній імовірності народитися в один день. Інтуїтивно це зрозуміло, адже якщо відомо, що в деякий день народилися двоє людей, то це може свідчити, що відповідний день у певному сенсі є популярним, а тому імовірність народження третьої особи в цей же день підвищується.  $\square$

## 4.4. Дискретний рівномірний розподіл

У певному сенсі найбільш очевидним дискретним розподілом є рівномірний, який формалізує поняття класичної ймовірності з Визначення 1.3.1.

**Визначення 4.4.1.** Випадкова величина  $X$  має *рівномірний розподіл* (uniform distribution), якщо її функція ймовірності має вид:

$$p_X(x) = \frac{1}{|\text{supp}(X)|}. \quad (4.4.1)$$

Якщо випадкова величина  $X$  має рівномірний розподіл, позначають  $X \sim U(\text{supp}(X))$ .  $\square$

Очевидно, що ця функція ймовірності задовольняє умови з (4.2.1);

- вона невід'ємна;
- сума її значень по всіх елементах носія дорівнює 1:

$$\sum_{x \in \text{supp}(X)} \frac{1}{|\text{supp}(X)|} = \frac{|\text{supp}(X)|}{|\text{supp}(X)|} = 1.$$

**Приклад 4.4.2.** Нехай у капелюсі містяться 100 пронумерованих від 1 до 100 шматочків паперу. Послідовно витягають 5 шматочків (незалежно один від одного).

Нехай  $X_j$  = «номер на  $j$ -му витягнутому шматочку»,  $j = 1, \dots, 5$ . За побудовою  $X_j \sim U(\{1, \dots, 100\})$ ,  $j = 1, \dots, 5$ . Варто зазначити, що цей результат справедливий незалежно від того, чи вибір із повтореннями, чи ні. Найпростіше побачити цей факт — уявити, що замість п'яти послідовних витягів відбувається один *одночасний* витяг, але п'ятьма різними людьми. Цілком очевидно, що суть задачі не змінюється, але в такій інтерпретації стає зрозумілим, що ймовірність витягнути будь-який шматочок будь-ким із п'яти осіб є однакова і дорівнює 0.01.

Імовірність події «шматочок із номером 100 витягнуто принаймні один раз» дорівнює

$$\begin{aligned} \mathbb{P}_X(X_1 = 100 \cup \dots \cup X_5 = 100) &= 1 - \mathbb{P}(X_1 \neq 100 \cap \dots \cap X_5 \neq 100) \\ &= 1 - \mathbb{P}(X_1 \neq 100) \cdot \dots \cdot \mathbb{P}(X_5 \neq 100) \\ &= 1 - (1 - \mathbb{P}(X_1 = 100)) \cdot \dots \cdot (1 - \mathbb{P}(X_5 = 100)) \\ &= 1 - \left(\frac{99}{100}\right)^5 \approx 0.049. \end{aligned}$$

□

## 4.5. Схема Бернуллі

### 4.5.1. Розподіл Бернуллі

Розгляд найбільш поширених і важливих дискретних розподілів почнімо з розподілу випадкової величини, носій якої складається всього з двох значень — 0 і 1. Нехай деякий експеримент може завершитися успіхом із деякою ймовірністю  $p \in [0; 1]$ , або ж невдачею з імовірністю  $q \equiv 1 - p$ . Такий експеримент називають *випробуванням Бернуллі*<sup>5</sup> (Bernoulli trial). Розгляньмо випадкову величину  $X$  таку, що  $X = 1$ , якщо мав місце успіх, і  $X = 0$  — у протилежному випадку. Така величина має розподіл Бернуллі.

**Визначення 4.5.1.** Випадкова величина  $X$  має *розподіл Бернуллі* (Bernoulli distribution) із параметром  $p$ , якщо її функція ймовірності має вид:

$$p_X(1) = p, \quad p_X(0) = q = 1 - p,$$

або, що те ж саме

$$p_X(x) = p^x(1 - p)^{1-x}, \quad x \in \{0, 1\}. \quad (4.5.1)$$

<sup>5</sup>Назване так на честь швейцарського математика Якова Бернуллі (Jacob Bernoulli, 1654–1705).

Якщо випадкова величина  $X$  має розподіл Бернуллі, позначають  $X \sim \text{Bern}(p)$ .  $\square$

**Зауваження 4.5.2.** Насправді, оскільки розподіл Бернуллі визначає його параметр  $p$ , можемо говорити про *сімейство* розподілів Бернуллі, у якому для кожного  $p$  існує свій розподіл  $\text{Bern}(p)$ .  $\square$

**Твердження 4.5.3.** Якщо  $X \sim \text{Bern}(p)$ , то  $\mathbb{E}[X] = p$ , а  $\text{Var}(X) = p(1 - p)$ .

*Доведення.* Випадкова величина з розподілом Бернуллі має всього два значення, 1 і 0, із імовірностями  $p$  і  $1 - p$ , відповідно, тому

$$\mathbb{E}[X] = 1 \cdot p + 0 \cdot (1 - p) = p.$$

Для підрахунку дисперсії за (4.3.4) спочатку обчислімо сподівання від квадрату випадкової величини. Оскільки піднесення 0 та 1 до квадрату не змінює їх значення, маємо:

$$\mathbb{E}[X^2] = 1^2 \cdot p + 0^2 \cdot (1 - p) = p.$$

Тоді

$$\text{Var}(X) = \mathbb{E}[X^2] - (\mathbb{E}[X])^2 = p - p^2 = p(1 - p).$$

$\square$

Таким чином, як можна бачити, оскільки  $\text{Var}(X) = p(1 - p)$  досягає свого максимуму за  $p = 0.5$ , можемо зробити висновок, що найбільша невизначеність (найбільший розкид значень) спостерігається для величини Бернуллі з параметром  $p = 0.5$ , а найменший — з параметрами  $p = 0, p = 1$  (у цих випадках дисперсія взагалі нульова).

#### 4.5.2. Біномний розподіл

Розгляньмо набір послідовних незалежних випробувань Бернуллі. Такий набір має назву *схеми Бернуллі* (Bernoulli scheme).

**Визначення 4.5.4.** Нехай маємо  $n$  послідовних *незалежних* випробувань Бернуллі, у кожному з яких імовірність успіху дорівнює  $p$ . Тоді, випадкова величина  $X$  = «загальне число успіхів» має *біномний розподіл* (Binomial distribution) із параметрами  $n$  і  $p$ , що позначають через  $X \sim \text{Binom}(n, p)$ .  $\square$

Знайдімо функцію ймовірності такої випадкової величини.

**Твердження 4.5.5.** Якщо випадкова величина  $X \sim \text{Binom}(n, p)$ , то її функція ймовірності має вид:

$$p_X(k) = \mathbb{P}_X(X = k) = \binom{n}{k} p^k (1 - p)^{n - k}, \quad k = 0, 1, \dots, n. \quad (4.5.2)$$

*Доведення.* Щоб визначити функцію ймовірності, потрібно зрозуміти, на якому ймовірнісному просторі визначено відповідну випадкову величину.

Нескладно бачити, що  $\Omega = \{0, 1\}^n$ , тобто кожна елементарна подія — вектор довжини  $n$  із нулів (невдач) і одиниць (успіхів). На ньому, як завжди для дискретних просторів,  $\sigma$ -алгеброю є  $2^\Omega$ . Імовірність кожної елементарної події  $\omega \in \Omega$  можна дістати, врахувавши незалежність випробувань Бернуллі. Так, якщо  $\omega$  містить точно  $k$  одиниць і  $n - k$  нулів, її ймовірність буде  $\mathbb{P}(\{\omega\}) = p^k(1 - p)^{n-k}$ , незалежно від порядку слідування успіхів і невдач.

Кількість усіх елементарних подій, у яких міститься точно  $k$  успіхів, еквівалентна кількості варіантів вибору  $k$  елементів із  $n$  можливих, тобто дорівнює  $\binom{n}{k}$ .

За визначенням (4.1.1) маємо:

$$\begin{aligned}\mathbb{P}_X(X = k) &= \mathbb{P}(\{\omega : X(\omega) = k\}) = \sum_{\omega: X(\omega)=k} \mathbb{P}(\{\omega\}) \\ &= \sum_{i=1}^{\binom{n}{k}} \mathbb{P}(\{\omega_{k,i}\}) \\ &= \binom{n}{k} p^k (1 - p)^{n-k}.\end{aligned}$$

Залишилось тільки показати, що ця функція справді є функцією ймовірності. Вона очевидно невід'ємна, тому потрібно показати, що сума всіх імовірностей дорівнює 1. Цей результат безпосередньо випливає з біномної теореми:

$$(a + b)^n = \sum_{k=0}^n \binom{n}{k} a^k b^{n-k}.$$

У нашому випадку  $a = p$ ,  $b = 1 - p$ . □

**Зауваження 4.5.6.** Вочевидь,  $X \sim \text{Bern}(p)$  і  $Y \sim \text{Binom}(1, p)$  мають одинаковий розподіл. У цьому розумінні розподіл Бернуллі є частковим випадком біномного розподілу. □

На Рис. 4.5.1 зображено декілька функцій імовірності для біномного розподілу з різними параметрами  $n$  і  $p$ . Як можна бачити, коли  $p \neq 1/2$ , розподіл є *скошеним* (skewed), тобто не є симетричним відносно середнього значення. Також варто звернути увагу, що, як і підказує інтуїція, що більші значення параметру  $p$  для одного й того ж  $n$ , то частіше величина  $X$  набуває більших значень.

**Приклад 4.5.7.** Підкидають 5 монеток з імовірністю випаду герба  $p$ . Нехай  $X = \text{«число випалих гербів»}$ . Ця випадкова змінна матиме біномний розподіл  $X \sim \text{Binom}(5, p)$ . Зокрема,

$$\mathbb{P}_X(X = 2) = \binom{5}{2} p^2 (1 - p)^3 = 10p^2 (1 - p)^3,$$

а, скажімо,

$$\mathbb{P}_X(X \geq 4) = \mathbb{P}_X(X = 4) + \mathbb{P}_X(X = 5) = \binom{5}{4} p^4 (1 - p)^1 + \binom{5}{5} p^5 (1 - p)^0 = 5p^4 (1 - p) + p^5.$$

□

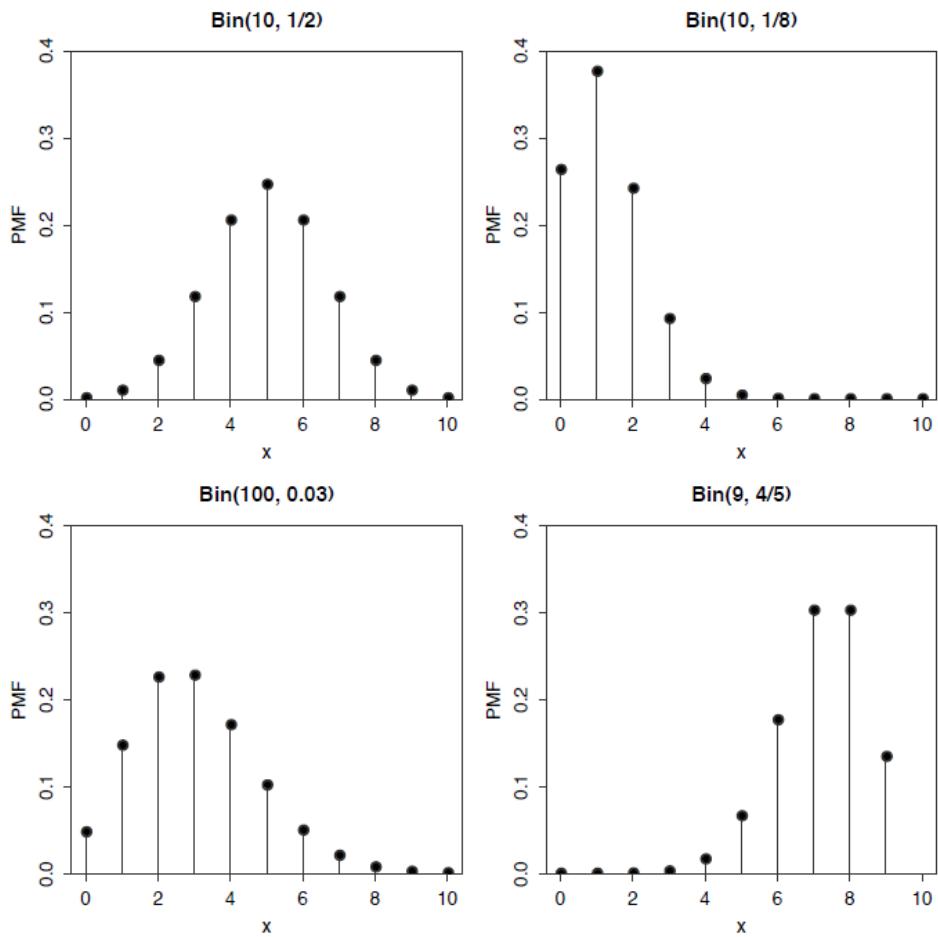


Рис. 4.5.1.: Приклади функцій імовірності для біномного розподілу ([1], Рис. 3.6)

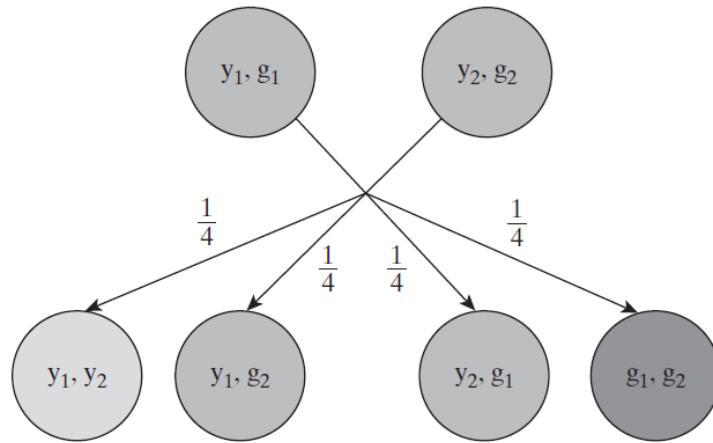


Рис. 4.5.2.: Схрещування двох гібридних рослин ([8], Рис. 4.6)

**Приклад 4.5.8.** Нехай деяку фенотипну рису організму, напр. колір насіння, визначає пара генів — домінантний  $d$  і рецесивний  $r$ . Особини з парами генів  $dd$ ,  $rd$  та  $dr$  мають домінантний фенотип, а особини з парою генів  $rr$  — рецесивний. Нашадок успадковує по одному гену від кожного з батьків з однаковою ймовірністю.

Нехай деяка пара батьків із гібридними генами має  $n = 4$  нашадків. Чому дорівнює ймовірність, що 3 з них матимуть домінантний фенотип?

На Рис. 4.5.2 наведено відповідний процес успадкування у контексті генів, що відповідають за колір насіння — домінантний жовтий  $y$  і рецесивний зелений  $g$ . Як можна бачити, імовірність того, що нашадок двох гібридних батьків матиме домінантний фенотип, дорівнює  $p = 1/4 + 1/4 + 1/4 = 3/4$ . Усього нашадків, за умовою, є 4. Таким чином, число нашадків із домінантним фенотипом  $X$  має біномний розподіл  $X \sim \text{Binom}(4, 3/4)$ , а тому

$$\mathbb{P}_X(X = 3) = \binom{4}{3} \left(\frac{3}{4}\right)^3 \left(\frac{1}{4}\right)^1 = \frac{27}{64}.$$
□

**Твердження 4.5.9.** Нехай  $X \sim \text{Binom}(n, p)$ . Тоді випадкова величина  $Y = n - X \sim \text{Binom}(n, 1 - p)$ .

*Доведення.* Для кожного  $k = 0, 1, \dots, n$  маємо

$$\begin{aligned} \mathbb{P}_Y(Y = k) &= \mathbb{P}(\{\omega : Y(\omega) = k\}) = \mathbb{P}(\{\omega : X(\omega) = n - k\}) = \mathbb{P}_X(X = n - k) \\ &= \binom{n}{n - k} p^{n-k} (1 - p)^k = \binom{n}{k} (1 - p)^k p^{n-k}, \end{aligned}$$

де ми використали Визначення 4.1.6 та формулу функції ймовірності біномного розподілу (4.5.2). □

Відповідний результат очевидний, адже величина  $Y$  просто рахує кількість невдач, що є повністю дзеркальною задачею.

**Твердження 4.5.10.** Якщо  $X \sim \text{Binom}(n, p)$ , то  $\mathbb{E}[X] = np$ , а  $\text{Var}(X) = np(1 - p)$ .

*Доведення.* За визначенням сподівання та з урахуванням (4.5.2) маємо:

$$\begin{aligned}
 \mathbb{E}[X] &= \sum_{k=0}^n k \cdot \binom{n}{k} p^k (1-p)^{n-k} = \sum_{k=1}^n k \cdot \binom{n}{k} p^k (1-p)^{n-k} \\
 &= n \sum_{k=1}^n \binom{n-1}{k-1} p^k (1-p)^{n-k} \\
 &= np \sum_{k=1}^n \binom{n-1}{k-1} p^{k-1} (1-p)^{n-k} \\
 &= np \sum_{j=0}^{n-1} \binom{n-1}{j} p^j (1-p)^{n-j-1} = np \cdot 1 = np ,
 \end{aligned}$$

де ми використали тотожність  $k \cdot \binom{n}{k} = n \binom{n-1}{k-1}$ .

Також ми скористалися тим фактом, що

$$\sum_{j=0}^{n-1} \binom{n-1}{j} p^j (1-p)^{n-j-1} = 1 ,$$

оскільки це сума ймовірностей для величини з розподілом Binom  $(n-1, p)$ , а тому дорівнює 1.

Для обчислення дисперсії спочатку обчислімо сподівання  $X^2$ , яке розпишемо так:

$$\mathbb{E}[X^2] = \mathbb{E}[X] + \mathbb{E}[X(X-1)] = np + \mathbb{E}[X(X-1)] ,$$

де

$$\begin{aligned}
 \mathbb{E}[X(X-1)] &= \sum_{k=0}^n k(k-1) \cdot \binom{n}{k} p^k (1-p)^{n-k} \\
 &= \sum_{k=2}^n k(k-1) \cdot \frac{n!}{k!(n-k)!} p^k (1-p)^{n-k} \\
 &= \sum_{k=2}^n \frac{n!}{(k-2)!(n-k)!} p^k (1-p)^{n-k} \\
 &= n(n-1) \sum_{k=2}^n \frac{(n-2)!}{(k-2)!(n-k)!} p^k (1-p)^{n-k} \\
 &= n(n-1)p^2 \sum_{k=2}^n \frac{(n-2)!}{(k-2)!((n-2)-(k-2))!} p^{k-2} (1-p)^{(n-2)-(k-2)} \\
 &= n(n-1)p^2 \sum_{j=0}^{n-2} \frac{(n-2)!}{j!((n-2)-j)!} p^j (1-p)^{n-2-j} \\
 &= n(n-1)p^2 \sum_{j=0}^{n-2} \binom{n-2}{j} p^j (1-p)^{n-2-j} \\
 &= n(n-1)p^2 \cdot 1 = n(n-1)p^2 ,
 \end{aligned}$$

оскільки відповідна сума є сумаю всіх імовірностей для величини з розподілом Binom ( $n - 2, p$ ), а тому дорівнює 1.

Нарешті,

$$\text{Var}(X) = \mathbb{E}[X^2] - (\mathbb{E}[X])^2 = (np + n(n-1)p^2) - (np)^2 = np - np^2 = np(1-p) .$$

□

**Зауваження 4.5.11.** Прийом, використаний у доведенні Твердження 4.5.10, є дуже поширеним і потребує окремої уваги. У теорії ймовірностей ми часто намагаємося «помітити», що деякі вирази можна звести до суми<sup>6</sup> значень функції ймовірності деякого розподілу, про яку заздалегідь відомо, що вона дорівнює 1. □

**Приклад 4.5.12.** Продовжмо розгляд Прикладу 4.5.8. Ми встановили, що число нащадків із домінантним фенотипом  $X \sim \text{Binom}(4, 3/4)$ . Згідно з Твердженням 4.5.10,

$$\mathbb{E}[X] = 4 \cdot \frac{3}{4} = 3 , \quad \text{Var}(X) = 4 \cdot \frac{3}{4} \cdot \left(1 - \frac{3}{4}\right) = \frac{3}{4} ,$$

тобто в середньому варто сподіватися на 3 нащадків із домінантним фенотипом, а середнім відхиленням від цього сподівання буде  $\text{sd}(X) = \sqrt{\frac{3}{4}} \approx 0.87$ . □

---

<sup>6</sup>Або в загальному випадку інтегралу, про що мова в наступних лекціях.

## 5. Деякі поширені дискретні розподіли

У попередній лекції ми розглянули визначення дискретної випадкової величини та пов'язані з нею поняття розподілу та сподівання, а також почали знайомитися з розповсюдженими на практиці дискретними розподілами. У цій лекції ми продовжимо наше знайомство з дискретними розподілами.

### 5.1. Розподіл Пуассона

Розгляньмо дуже поширений розподіл, за допомогою якого на практиці моделюють кількість успіхів із-посеред послідовних випробувань, що сталися за певний проміжок часу, коли кількість випробувань дуже велика, а ймовірність успіху — мала.

**Визначення 5.1.1.** Випадкова величина  $X$  має *розподіл Пуассона* (Poisson distribution)<sup>1</sup> із параметром  $\lambda > 0$ ,  $X \sim \text{Pois}(\lambda)$ , якщо її функція ймовірності має вид

$$p_X(k) = \mathbb{P}_X(X = k) = \frac{e^{-\lambda} \lambda^k}{k!}, \quad k = 0, 1, 2, \dots. \quad (5.1.1)$$

□

Те, що така функція є справді функцією ймовірності в тому розумінні, що вона відповідає Визначенняю 4.2.3, можна бачити з того, що вона невід'ємна, а, враховуючи розвинення в ряд Тейлора експоненційної функції

$$e^\lambda = \frac{\lambda^k}{k!},$$

маємо

$$\sum_{k=0}^{\infty} p_X(k) = \sum_{k=0}^{\infty} \frac{e^{-\lambda} \lambda^k}{k!} = e^{-\lambda} \sum_{k=0}^{\infty} \frac{\lambda^k}{k!} = e^{-\lambda} e^{\lambda} = 1.$$

Параметр  $\lambda$  можна інтерпретувати як частоту появи деяких подій («успіхів»), які стаються з малою ймовірністю. Як правило, якого визначають емпірично.

**Приклад 5.1.2.** Деякі приклади випадкових величин, які доречно моделювати за допомогою розподілу Пуассона:

- кількість електронних листів, отримуваних за годину: хоча потенційно електронні листи може надіслати велике число осіб (число випробувань велике), імовірність надсилання листа окремо взятою особою саме протягом встановленої години є низька. Тоді, наприклад, можна покласти  $\lambda = 20$  листів на годину;

<sup>1</sup>Уперше запропонував французький математик Симеон Дені Пуассон (Siméon Denis Poisson, 1781–1840) у 1837 р. у роботі «Дослідження ймовірностей судових рішень у кримінальних та цивільних справах».

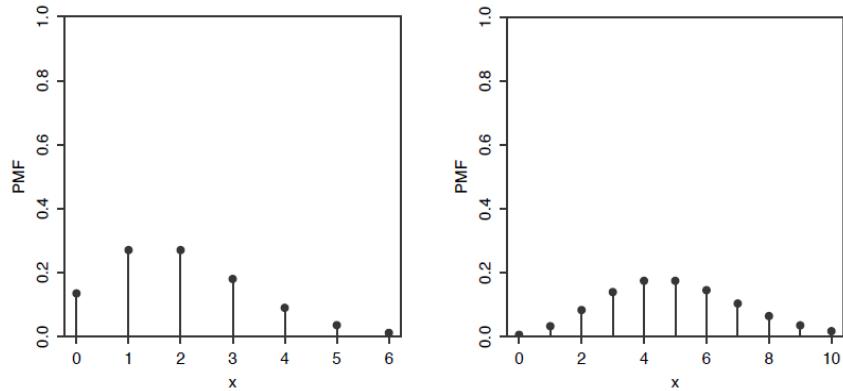


Рис. 5.1.1.: Приклади функцiй iмовiрностi для розподiлiв Пуассона Pois (2) i Pois (5), ([1], Рис. 4.7)

- кiлькiсть шматочкiв шоколаду в печивi: якiщо роздiлити печиво на велике чiсло маленьких частинок (число випробувань велике), iмовiрнiсть появi шоколадного шматочку в деякiй окремо взятiй частинцi буде дуже мала. Тодi, наприклад, можна покласти  $\lambda = 10$  шматочкiв шоколаду на одне печиво;
- чiсло землетрусiв на рiк у деякому регiонi планети: iмовiрнiсть виникнення землетрусу в конкретнiй локацiї в конкретний момент часу дуже мала, але протягом року чiсло пар «локацiя-час» дуже велике. Тодi, наприклад, можна покласти  $\lambda = 2$  землетруси на рiк.

□

**Приклад 5.1.3.** Нехай чiсло типографських помилок на окремо взятiй сторiнцi книжки має розподiл Пуассона Pois (0.5). Тодi iмовiрнiсть хоча б однiєї помилки дорiвнює

$$\mathbb{P}_X (X \geq 1) = 1 - \mathbb{P}_X (X = 0) = 1 - \frac{e^{-0.5} 0.5^0}{0!} \approx 0.393.$$

□

На Рис. 5.1.1 зображенi функцiї iмовiрностi для розподiлу Пуассона з параметрами  $\lambda = 2$  i  $\lambda = 5$ . Як можна бачити, зi збiльшенням значення параметру  $\lambda$  функцiя iмовiрностi стає бiльше симетричною.

Як зазначено вище, розподiл Пуассона дoreчно застосовувати для моделювання явищ, якi стаються нечасто. Проте це не означає, що такi явища мають розподiл Пуассона. Розподiл Пуассона є всього лише апроксимацiєю, яку зручно використовувати на практицi.

Так, зокрема, розподiл Пуассона можна розглядати як границю бiномного розподiлу, коли чiслу випробувань прямує до нескiнченностi, а iмовiрнiсть успiху — до нуля.

**Теорема 5.1.4.** Нехай  $X \sim \text{Binom}(n, p)$ ,  $n \rightarrow \infty$ ,  $p_n \rightarrow 0$ , але  $np_n \rightarrow \lambda$ , тобто залишається фiксованим. Тодi функцiя розподiлу  $X$  прямує до функцiї розподiлу  $Y \sim \text{Pois}(\lambda)$ .

*Доведення.* Функція ймовірності біномного розподілу з Твердження 4.5.5 має вид

$$\mathbb{P}_X(X = k) = \binom{n}{k} p_n^k (1 - p_n)^{n-k} = \frac{n(n-1)\dots(n-k+1)}{k!} p_n^k (1 - p_n)^{n-k}.$$

Візьмімо границю:

$$\begin{aligned} \lim_{\substack{n \rightarrow \infty \\ p_n \rightarrow 0 \\ np_n \rightarrow \lambda}} \frac{n(n-1)\dots(n-k+1)}{k!} p_n^k (1 - p_n)^{n-k} &= \lim_{\substack{n \rightarrow \infty \\ p_n \rightarrow 0 \\ np_n \rightarrow \lambda}} \frac{n(n-1)\dots(n-k+1)}{k!} \left(\frac{\lambda}{n}\right)^k \left(1 - \frac{\lambda}{n}\right)^{n-k} \\ &= \lim_{\substack{n \rightarrow \infty \\ p_n \rightarrow 0 \\ np_n \rightarrow \lambda}} \frac{n}{n} \cdot \frac{n-1}{n} \cdot \dots \cdot \frac{n-k+1}{n} \cdot \frac{1}{k!} \lambda^k \left(1 - \frac{\lambda}{n}\right)^{n-k} \\ &= \lim_{\substack{n \rightarrow \infty \\ p_n \rightarrow 0 \\ np_n \rightarrow \lambda}} 1 \cdot \frac{\lambda^k}{k!} \left(1 - \frac{\lambda}{n}\right)^n \left(1 - \frac{\lambda}{n}\right)^{-k} \\ &= \frac{\lambda^k}{k!} \cdot e^{-\lambda} \cdot 1 \\ &= \frac{e^{-\lambda} \lambda^k}{k!}, \end{aligned}$$

де ми використали наслідок із другої важливої границі:

$$\lim_{n \rightarrow \infty} \left(1 - \frac{\lambda}{n}\right)^n = e^{-\lambda}.$$

□

Теорема 5.1.4 застосовна, коли число випробувань  $n$  велике, імовірність успіху  $p$  мала, але їх добуток  $np$  є не дуже великим і не дуже малим, тобто збільшення числа випробувань і зменшення ймовірності компенсують одне одне. До таких прикладів належать, наприклад:

- число осіб у громаді, які доживуть до 100 років;
- число неправильно набраних телефонних номерів на день;
- число пакетів собачого корму, проданих у деякому магазині за день;
- число клієнтів поштового відділку на день;
- число вакансій, що відкриваються за рік у судовій системі;
- число  $\alpha$ -частинок, випромінюваних протягом фіксованого періоду часу.

У кожній із цих ситуацій відповідні випадкові величини мають біномний розподіл, де випробуваннями є окремі особи в громаді, покупці зоомагазину, посади в судовій системі тощо.

**Приклад 5.1.5.** Нехай  $X \sim \text{Binom}(25, 1/16)$ . Тоді значення функції ймовірності цієї величини в точці  $x = 2$  дорівнює

$$p_X(2) = \binom{25}{2} \left(\frac{1}{16}\right)^2 \left(\frac{15}{16}\right)^{23} \approx 0.266.$$

Використовуючи апроксимацію з Теореми 5.1.4, маємо для  $Y \sim \text{Pois}(25/16)$ :

$$p_Y(2) = \frac{e^{-\frac{25}{16}} \left(\frac{25}{16}\right)^2}{2!} \approx 0.256 .$$

Похибка апроксимації складає приблизно 3.7%, що зовсім непогано, враховуючи мале значення  $n = 25$ .

Якщо збільшити  $n$  до 50, маємо

$$p_X(2) = \binom{50}{2} \left(\frac{1}{16}\right)^2 \left(\frac{15}{16}\right)^{48} \approx 0.216 , \quad p_Y(2) = \frac{e^{-\frac{50}{16}} \left(\frac{50}{16}\right)^2}{2!} \approx 0.213 ,$$

і похибка апроксимація вже падає до 0.1%.  $\square$

**Приклад 5.1.6.** Нехай  $X$  = «число відвідувачів вебсайту на день». Щодня відвідати сайт незалежно одне від одного вирішує мільйон користувачів, з імовірністю відвідання  $p = 2 \times 10^{-6}$ . Використовуючи апроксимацію з Теореми 5.1.4, можемо обчислити ймовірність відвідання сайту щонайменше трьома користувачами:

$$\begin{aligned} \mathbb{P}_X(X \geq 3) &= 1 - \mathbb{P}_X(X < 3) \approx 1 - (\mathbb{P}_Y(0) + \mathbb{P}_Y(1) + \mathbb{P}_Y(2)) \\ &= 1 - \left( e^{-2} + e^{-2} \cdot 2 + \frac{e^{-2} \cdot 2^2}{2!} \right) = 1 - 5 \cdot e^{-2} \approx 0.3233 . \end{aligned}$$

$\square$

Дуже часто розподіл Пуассона використовують для опису явищ, які стаються в деякі випадкові моменти часу. Тоді загальне число таких явищ протягом деякого періоду часу матиме приблизний розподіл Пуассона. Неформально це можна зрозуміти, якщо розглядати кожний «момент часу» як випробування, де «успіхом» є настання певної події.

Як правило, час моделюють як незліченний числовий проміжок, тому «моментів часу» є незліченна кількість. Щоб формалізувати відповідні інтуїтивні уявлення, розгляньмо таку модель. Нехай для деякого  $\lambda > 0$  справедливі такі твердження:

1. Імовірність, що в інтервалі довжиною  $h$  станеться точно одна подія, дорівнює  $\lambda h + o(h)^2$ .
2. Імовірність, що в інтервалі довжиною  $h$  станеться більше однієї події, дорівнює  $o(h)$ .
3. Імовірності появи подій у неперетинаних інтервалах є незалежні одна від одної.

Іншими словами, для дуже малих  $h$  імовірність настання однієї події дорівнює приблизно  $\lambda h$ , а двох і більше подій — приблизно нулю.

Розгляньмо число  $N(h)$  подій, що стаються в деякому інтервалі довжиною  $h$ :  $I = [0; h]$ . Для

<sup>2</sup>Позначення  $o(h)$  використовують для будь-якої функції  $f(h)$  такої, що  $\lim_{h \rightarrow 0} \frac{f(h)}{h} = 0$ , тобто для будь-якої функції, яка прямує до нуля швидше за  $h$ .

підрахунку  $\mathbb{P}_N(N(h) = k)$  розбиймо інтервал на  $n$  неперетинаних підінтервалів довжини  $\frac{h}{n}$ . Тоді

$$\begin{aligned}\mathbb{P}_N(N(h) = k) &= \mathbb{P}(\text{«}k \text{ інтервалів містять точно по 1 подій, а решта } (n - k) \text{ містять 0 подій}\text{»}) \\ &+ \mathbb{P}_X(N(h) = k \cap \text{«принаймні 1 підінтервал містить 2 і більше подій»}) \\ &\equiv \mathbb{P}(A) + \mathbb{P}(B) .\end{aligned}$$

Розгляньмо подію  $B$ . Оскільки вона є перетином двох подій, очевидно вона не може перевищувати однієї з них, а тому за властивістю монотонності ймовірності маємо

$$\begin{aligned}\mathbb{P}(B) &\leq \mathbb{P}(\text{«принаймні 1 підінтервал містить 2 і більше подій»}) \\ &= \mathbb{P}\left(\bigcup_{i=1}^n \text{«саме } i\text{-ий інтервал містить 2 і більше подій»}\right) .\end{aligned}$$

Можемо застосувати нерівність Була (2.2.2) та припущення нашої моделі:

$$\begin{aligned}\mathbb{P}\left(\bigcup_{i=1}^n \text{«саме } i\text{-ий інтервал містить 2 і більше подій»}\right) \\ \leq \sum_{i=1}^n \mathbb{P}(\text{«саме } i\text{-ий інтервал містить 2 і більше подій»}) \\ = \sum_{i=1}^n o\left(\frac{h}{n}\right) = n \cdot o\left(\frac{h}{n}\right) .\end{aligned}$$

Оскільки

$$\lim_{n \rightarrow \infty} n \cdot o(h/n) = \lim_{n \rightarrow \infty} h \frac{o(h/n)}{h/n} = h \lim_{t \equiv h/n \rightarrow 0} \frac{o(t)}{t} = 0 ,$$

за визначенням  $o(t)$ . Таким чином,

$$\lim_{n \rightarrow \infty} \mathbb{P}(B) = 0 .$$

Імовірність події  $A$  можна обчислити за допомогою біномного розподілу Binom  $(n, \lambda \frac{h}{n} + o(\frac{h}{n}))$ :

$$\mathbb{P}(A) = \binom{n}{k} \left(\lambda \frac{h}{n} + o\left(\frac{h}{n}\right)\right)^k \left(1 - \lambda \frac{h}{n} - o\left(\frac{h}{n}\right)\right)^{n-k} .$$

До того ж

$$np_n = n \left(\lambda \frac{h}{n} + o\left(\frac{h}{n}\right)\right) = \lambda h + h \left(\frac{o(h/n)}{h/n}\right) \xrightarrow{n \rightarrow \infty} \lambda h ,$$

а отже можемо застосувати результат Теореми 5.1.4:

$$\lim_{n \rightarrow \infty} \mathbb{P}(A) = \frac{e^{-\lambda h} (\lambda h)^k}{k!} ,$$

а відтак і

$$\lim_{n \rightarrow \infty} \mathbb{P}_N(N(h) = k) = \frac{e^{-\lambda h} (\lambda h)^k}{k!} .$$

Іншими словами, число подій, які стаються в часовому інтервалі довжиною  $h$ , має розподіл Пуассона з параметром  $\lambda h$ , якщо кількість цих інтервалів дуже велика. Серед випадкових величин,

які можна описувати розподілом Пуассона, такі:

- число землетрусів за певний період часу;
- число воєн на рік;
- число електронів, випущених нагрітих катодом, за певний період часу;
- число смертей серед клієнтів страхової компанії за певний період часу.

**Приклад 5.1.7.** Нехай землетруси на західному узбережжі США стаються в кількості 2 випадки на тиждень. Тоді ймовірність появи 3 землетрусів протягом наступних 2 тижнів становить

$$\begin{aligned}\mathbb{P}_N(N(2) \geq 3) &= 1 - \mathbb{P}_N(N(2) = 0) - \mathbb{P}_N(N(2) = 1) - \mathbb{P}_N(N(2) = 2) \\ &= 1 - e^{-4} - 4e^{-4} - \frac{4^2}{2!}e^{-4} \approx 0.762.\end{aligned}$$

□

Поки що ми розглядали тільки ситуацію, коли випробування могли завершитися успіхом з однаковою ймовірністю. Насправді, розподіл Пуассона також можна використовувати як апроксимацію розподілу суми випадкових величин, навіть якщо ймовірності подій різні.

**Твердження 5.1.8** (Пуассонівська парадигма (Poisson paradigm)). Нехай  $A_1, \dots, A_n$  мають імовірності  $\mathbb{P}(A_j) = p_j$ ,  $j = 1, \dots, n$ , де  $n$  велике, а  $p_j$  малі. Події  $A_j$ ,  $j = 1, \dots, n$ , повинні бути незалежні або *слабко залежні* (weakly dependent). Тоді  $X = \sum_{j=1}^n \mathbb{1}\{A_j\}$ , тобто число подій, які сталися, має розподіл, який можна апроксимувати розподілом Пуассона з параметром  $\lambda = \sum_{j=1}^n p_j$ .

Ми не будемо доводити цього твердження, оскільки для цього потрібно розглядати додатковий матеріал.

Також можна довести, що якість апроксимації задає така нерівність:

$$|\mathbb{P}_X(X \in B) - \mathbb{P}_N(N \in B)| \leq \min \left\{ 1, \frac{1}{\lambda} \right\} \cdot \sum_{j=1}^n p_j^2,$$

де  $N \sim \text{Pois}(\lambda)$ . Іншими словами, апроксимація буде якісною, якщо сума квадратів імовірностей дуже мала (або принаймні мала в порівнянні з  $\lambda$ ).

Ми не будемо розглядати поняття слабкої залежності, але для ілюстрації цього поняття розглянемо ситуацію, коли кожний із  $n$  чоловіків повинен випадковим чином вибрати свого капелюха з непрозорого мішка з  $n$  капелюхів, кожен із яких належить одному з чоловіків. Нехай  $E_i = \text{«}i\text{-ий чоловік витяг свого капелюха}\text{»}$ . Очевидно,  $\mathbb{P}(E_i) = 1/n$ , а  $\mathbb{P}(E_i | E_j) = 1/(n-1)$ ,  $i \neq j$ . Тобто події  $E_i$  і  $E_j$  не є незалежними, хоча якщо  $n$  дуже велике, різниця між  $1/n$  і  $1/(n+1)$  буде незначною, тому ці події є слабко залежними.

Наочанок підрахуймо сподівання та дисперсію розподілу Пуассона.

**Твердження 5.1.9.** Якщо  $X \sim \text{Pois}(\lambda)$ , то  $\mathbb{E}[X] = \text{Var}(X) = \lambda$ .

*Доведення.* За визначенням сподівання:

$$\begin{aligned}
 \mathbb{E}[X] &= \sum_{k=0}^{\infty} k \cdot \frac{e^{-\lambda} \lambda^k}{k!} \\
 &= \sum_{k=1}^{\infty} k \cdot \frac{e^{-\lambda} \lambda^k}{k!} \\
 &= \lambda e^{-\lambda} \sum_{k=1}^{\infty} \frac{\lambda^{k-1}}{(k-1)!} \\
 &= \lambda e^{-\lambda} \sum_{j=0}^{\infty} \frac{\lambda^j}{j!} \\
 &= \lambda e^{-\lambda} e^{\lambda} = \lambda.
 \end{aligned}$$

Для підрахунку дисперсії спочатку підрахуймо  $\mathbb{E}[X^2]$ :

$$\begin{aligned}
 \mathbb{E}[X^2] &= \sum_{k=0}^{\infty} k^2 \cdot \frac{e^{-\lambda} \lambda^k}{k!} \\
 &= \sum_{k=1}^{\infty} k^2 \cdot \frac{e^{-\lambda} \lambda^k}{k!} \\
 &= \lambda \sum_{k=1}^{\infty} k \cdot \frac{e^{-\lambda} \lambda^{k-1}}{(k-1)!} \\
 &= \lambda \sum_{j=0}^{\infty} (j+1) \cdot \frac{e^{-\lambda} \lambda^j}{j!} \\
 &= \lambda \left( \sum_{j=0}^{\infty} j \cdot \frac{e^{-\lambda} \lambda^j}{j!} + \sum_{j=0}^{\infty} \frac{e^{-\lambda} \lambda^j}{j!} \right) \\
 &= \lambda(\lambda + 1),
 \end{aligned}$$

оскільки перша сума в дужках є сподіванням величини з розподілом  $\text{Pois}(\lambda)$ , тобто дорівнює  $\lambda$ , а друга сума є сумаю всіх значень функції ймовірності величини з розподілом  $\text{Pois}(\lambda)$ , тобто дорівнює 1.

Відтак, згідно з (4.3.6),

$$\text{Var}(X) = \mathbb{E}[X^2] - (\mathbb{E}[X])^2 = \lambda(\lambda + 1) - \lambda^2 = \lambda.$$

□

## 5.2. Геометричний розподіл

Зі схемою Бернуллі також пов'язані декілька інших дискретних розподілів.

**Визначення 5.2.1.** Нехай маємо послідовні *незалежні* випробування Бернуллі, у кожному з яких імовірність успіху дорівнює  $p$ . Тоді, випадкова величина  $X$  = «кількість випробувань до

появи першого успіху» має *геометричний розподіл* (geometric distribution) із параметром  $p$ , що позначають через  $X \sim \text{Geom}(p)$ .  $\square$

Знайдімо функцію ймовірності такої випадкової величини.

**Твердження 5.2.2.** Якщо випадкова величина  $X \sim \text{Geom}(p)$ , то її функція ймовірності має вид:

$$p_X(k) = \mathbb{P}_X(X = k) = (1 - p)^k p, \quad k = 0, 1, \dots. \quad (5.2.1)$$

**Доведення.** Щоб визначити функцію ймовірності, потрібно зрозуміти, на якому ймовірнісному просторі визначено відповідну випадкову величину.

Розгляньмо послідовність з  $n$  випробувань. Як і в доведенні Твердження 4.5.5,  $\Omega = \{0, 1\}^n$ , тобто кожна елементарна подія — вектор довжини  $n$  із нулів (невдач) і одиниць (успіхів). На ньому  $\sigma$ -алгеброю є  $2^\Omega$ . Імовірність кожної елементарної події  $\omega \in \Omega$ , у якій містяться точно  $k$  успіхів, як було зазначено в доведенні Твердження 4.5.5, є  $\mathbb{P}(\{\omega\}) = p^k(1 - p)^{n-k}$ .

Розгляньмо всі  $\omega$ , у яких перші  $k$ ,  $0 \leq k < n$ , випробувань завершились невдачею, а  $(k+1)$ -е — успіхом:  $\omega_1 = \dots = \omega_k = 0, \omega_{k+1} = 1, \omega_{k+2} \in \{0, 1\}, \dots, \omega_n \in \{0, 1\}$ . Решта випробувань можуть бути як 0, так і 1. Із обговорення біномного розподілу зрозуміло, що існує  $\binom{n-k-1}{j}$  таких  $\omega$ , у яких перший успіх стався на  $(k+1)$ -ому випробуванні, а після цього було точно  $j$  успіхів.

Таким чином,

$$\begin{aligned} \mathbb{P}_X(X = k) &= \mathbb{P}(\{\omega : X(\omega) = k\}) \\ &= \mathbb{P}(\{\omega : \omega_1 = \dots = \omega_k = 0, \omega_{k+1} = 1, \omega_{k+2} \in \{0, 1\}, \dots, \omega_n \in \{0, 1\}\}) \\ &= (1 - p)^k p \sum_{j=0}^{n-k-1} \binom{n - k - 1}{j} p^j (1 - p)^{n-k-1-j} \\ &= (1 - p)^k p, \end{aligned}$$

оскільки сума дорівнює 1 як сума значень функції ймовірності для випадкової величини з розподілом  $\text{Binom}(n - k - 1, p)$ .

Як можна бачити, функція ймовірності не залежить від  $n$ , тому можна казати, що відповідна функція ймовірності справедлива для довільної послідовності випробувань.

Залишилось тільки показати, що ця функція справді є функцією ймовірності. Вона очевидно невід'ємна, тому потрібно показати, що сума всіх імовірностей дорівнює 1:

$$\sum_{k=0}^{\infty} (1 - p)^k p = p \sum_{k=0}^{\infty} (1 - p)^k = p \cdot \frac{1}{1 - (1 - p)} = 1,$$

де ми використали формулу суми нескінченно спадної геометричної прогресії.  $\square$

Із доведення Твердження 5.2.2 стає зрозумілим, чому геометричний розподіл має таку назву: ймовірності утворюють нескінченно спадну геометричну прогресію.

На Рис. 5.2.1 зображено функцію ймовірності для геометричного розподілу з параметром  $p = 0.5$ . Як можна бачити, що більше значення цього параметру, то сильніше функція ймовірності прямує до нуля.

**Зауваження 5.2.3.** У літературі не існує єдиного бачення, який саме розподіл визначати як геометричний. Деякі автори подають геометричний розподіл як розподіл випадкової величини

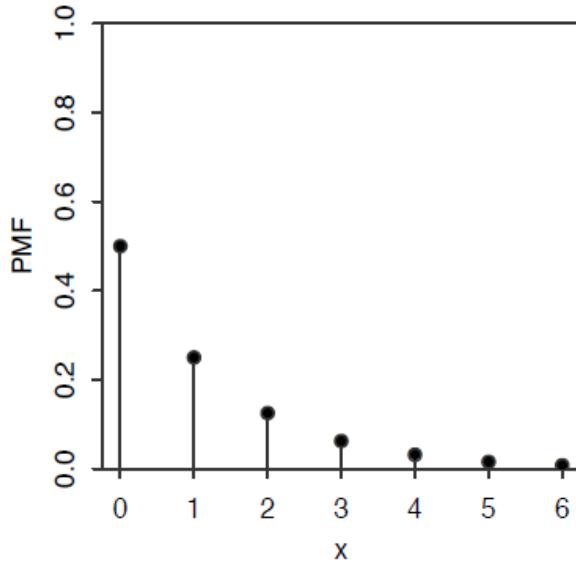


Рис. 5.2.1.: Функція розподілу для  $X \sim \text{Geom}(0.5)$  ([1], Рис. 4.5)

$Y$  = «кількість випробувань до появи першого успіху, включно з самим успіхом». Ми будемо вважати, що така величина має окремий розподіл:  $Y \sim \text{FS}(p)$ <sup>3</sup>. Очевидно, що якщо  $X \sim \text{Geom}(p)$ , то  $X + 1 \sim \text{FS}(p)$ .  $\square$

**Приклад 5.2.4.** Нехай у вазі містяться  $w$  білих і  $b$  чорних кульок. Кульки послідовно і випадковим чином виймають із вази до появи першої білої кульки. Кожну вийняту кульку повертають у вазу перед витяганням наступної. Тоді число витягів  $X$ , потрібних для появи першої білої кульки, має геометричний розподіл  $X \sim \text{Geom}\left(\frac{w}{w+b}\right)$ , а ймовірність того, що для витягу білої кульки знадобиться щонайменше  $k$  витягань, дорівнює:

$$\begin{aligned}
 \mathbb{P}_X(X \geq k) &= \frac{w}{w+b} \sum_{i=k}^{\infty} \left(\frac{b}{w+b}\right)^i \\
 &= \frac{w}{w+b} \left(\frac{b}{w+b}\right)^k \cdot \sum_{j=0}^{\infty} \left(\frac{b}{w+b}\right)^j \\
 &= \frac{w}{w+b} \left(\frac{b}{w+b}\right)^k \cdot \frac{1}{1 - \frac{b}{w+b}} \\
 &= \left(\frac{b}{w+b}\right)^k.
 \end{aligned}$$

$\square$

<sup>3</sup>За першими літерами словосполучення «перший успіх» англійською — «first success».

Іншими словами, для величини  $X \sim \text{Geom}(p)$  справедлива властивість

$$\mathbb{P}_X(X \geq k) = (1 - p)^k.$$

Більше того, умовна ймовірність того, що  $X \geq k + j$  за умови, що всі перші  $j$  випробувань завершилися невдачами, дорівнює

$$\begin{aligned}\mathbb{P}_X(X \geq k + j \mid X \geq j) &= \frac{\mathbb{P}_X((X \geq k + j) \cap (X \geq j))}{\mathbb{P}_X(X \geq j)} \\ &= \frac{\mathbb{P}_X(X \geq k + j)}{\mathbb{P}_X(X \geq j)} \\ &= \frac{(1 - p)^{k+j}}{(1 - p)^j} = (1 - p)^k.\end{aligned}$$

Таким чином, геометричний розподіл є розподілом *без пам'яти* (memoryless), тобто тривалість очікування першого успіху, за умови, що він до цього ще не встиг настати, не залежить від попередньої історії.

**Твердження 5.2.5.** Якщо  $X \sim \text{Geom}(p)$ , то  $\mathbb{E}[X] = \frac{1-p}{p}$ , а  $\text{Var}(X) = \frac{1-p}{p^2}$ .

*Доведення.* За визначенням сподівання та з урахуванням (5.2.1) маємо:

$$\begin{aligned}\mathbb{E}[X] &= \sum_{k=0}^{\infty} k \cdot (1 - p)^k p = \sum_{k=1}^{\infty} k \cdot (1 - p)^k p \\ &= \sum_{k=1}^{\infty} (k - 1 + 1)(1 - p)^k p \\ &= \sum_{k=1}^{\infty} (k - 1)(1 - p)^k p + \sum_{k=1}^{\infty} (1 - p)^k p + p - p \\ &= \sum_{j=0}^{\infty} j \cdot (1 - p)^{j+1} p + \sum_{k=0}^{\infty} (1 - p)^k p - p \\ &= (1 - p)\mathbb{E}[X] + 1 - p,\end{aligned}$$

оскільки перша сума є добутком  $(1 - p)$  і сподівання величини з розподілом  $\text{Geom}(p)$ , а друга сума є сумою всіх значень функції ймовірності величини з розподілом  $\text{Geom}(p)$ , тобто 1.

Отже, ми дістали рівність, із якої можна виразити

$$\mathbb{E}[X] = \frac{1 - p}{p}.$$

Для підрахунку дисперсії знайдімо

$$\begin{aligned}
 \mathbb{E}[X^2] &= \sum_{k=0}^{\infty} k^2 \cdot (1-p)^k p = \sum_{k=1}^{\infty} k^2 \cdot (1-p)^k p \\
 &= \sum_{k=1}^{\infty} (k-1+1)^2 \cdot (1-p)^k p \\
 &= \sum_{k=1}^{\infty} (k-1)^2 (1-p)^k p + 2 \sum_{k=1}^{\infty} (k-1)(1-p)^k p + \sum_{k=1}^{\infty} (1-p)^k p + p - p \\
 &= \sum_{j=0}^{\infty} j^2 \cdot (1-p)^{j+1} p + 2 \sum_{j=0}^{\infty} j \cdot (1-p)^{j+1} p + \sum_{k=0}^{\infty} (1-p)^k p - p \\
 &= (1-p)\mathbb{E}[X^2] + 2(1-p)\mathbb{E}[X] + 1 - p \\
 &= (1-p)\mathbb{E}[X^2] + 2\frac{(1-p)^2}{p} + 1 - p.
 \end{aligned}$$

Отже, ми дістали рівність, із якої можна виразити

$$\mathbb{E}[X^2] = \frac{(1-p)(2-p)}{p^2}.$$

Нарешті,

$$\text{Var}(X) = \mathbb{E}[X^2] - (\mathbb{E}[X])^2 = \frac{(1-p)(2-p)}{p^2} - \frac{(1-p)^2}{p^2} = \frac{1-p}{p^2}.$$

□

**Зауваження 5.2.6.** Якщо  $Y \sim \text{FS}(p)$ , а, відповідно,  $X = Y - 1 \sim \text{Geom}(p)$ , то

$$\mathbb{E}[Y] = \mathbb{E}[X+1] = \mathbb{E}[X] + 1 = \frac{1-p}{p} + 1 = \frac{1}{p}.$$

□

### 5.3. Від'ємний біномний розподіл

У повній аналогії з біномним розподілом як узагальненням розподілу Бернуллі можна увести від'ємний біномний розподіл як узагальнення розподілу геометричного.

**Визначення 5.3.1.** Нехай маємо послідовні незалежні випробування Бернуллі, у кожному з яких імовірність успіху дорівнює  $p$ . Тоді, випадкова величина  $X$  = «загальне число невдач до  $r$ -того успіху» має *від'ємний біномний розподіл* (negative Binomial distribution) із параметрами  $r$  і  $p$ , що позначають через  $X \sim \text{NegBinom}(r, p)$ . □

Знайдімо функцію ймовірності такої випадкової величини.

**Твердження 5.3.2.** Якщо випадкова величина  $X \sim \text{NegBinom}(r, p)$ , то її функція ймовірності має вид:

$$p_X(j) = \mathbb{P}_X(X = j) = \binom{j+r-1}{r-1} p^r (1-p)^j, \quad j = 0, 1, \dots. \quad (5.3.1)$$

**Доведення.** Щоб визначити функцію ймовірності, потрібно зрозуміти, на якому ймовірнісному просторі визначено відповідну випадкову величину.

Розгляньмо послідовність з  $n$  випробувань. Як і в доведенні Твердження 5.2.2,  $\Omega = \{0, 1\}^n$ , тобто кожна елементарна подія — вектор довжини  $n$  із нулів (невдач) і одиниць (успіхів). На ньому  $\sigma$ -алгеброю є  $2^\Omega$ . Імовірність кожної елементарної події  $\omega \in \Omega$ , у якій містяться точно  $k$  успіхів, як було зазначено в доведенні Твердження 5.2.2, є  $\mathbb{P}(\{\omega\}) = p^k(1-p)^{n-k}$ .

Подія  $X = j$ , тобто що сталося  $j$  невдач до  $r$ -ого успіху, дорівнює

$$\mathbb{P}_X(X = j) = \mathbb{P}\left(\left\{\omega : \sum_{i=1}^{j+r} \omega_i = r, \omega_{j+r+1} \in \{0, 1\}, \dots, \omega_n \in \{0, 1\}\right\}\right) \equiv \mathbb{P}(B_j),$$

тобто ймовірності множини всіх таких елементарних подій  $\omega \in B_j$ , у яких перші  $r+j$  випробувань яких містять  $r$  успіхів і  $j$  невдач, а решта випробувань — довільні. Із обговорення біномного розподілу зрозуміло, що існує  $\binom{j+r-1}{r-1}$  можливих варіантів розташування успіхів і невдач серед перших  $r+j$  випробувань (оскільки останній випад обов'язково має бути успішним, його позиція зафіксована і тому не враховується).

Розгляньмо деяку конкретну  $\omega' \in B_j$ . Позначмо через  $B_j(\omega')$  множину всіх таких елементарних подій із  $B_j$ , де перші  $r+j$  випробувань завершилися точно так, як і перші  $r+j$  випробувань із  $\omega'$ . Також позначмо через  $B_{j,m}(\omega')$  множину таких елементарних подій із  $B_j(\omega')$ , у яких серед останніх  $n-r-j$  випробувань міститься точно  $m$  успіхів. Тоді її ймовірність дорівнюватиме

$$\mathbb{P}(B_{j,m}(\omega')) = p^r (1-p)^j \cdot \binom{n-r-j}{m} p^m (1-p)^{n-r-j-m},$$

оскільки для кожного  $m$  існує  $\binom{n-r-j}{m}$  варіантів реалізації всіх інших випробувань, де  $m$  — кількість успіхів серед  $n-r-j$  випробувань. Тоді

$$\mathbb{P}(B_j(\omega')) = p^r (1-p)^j \sum_{m=0}^{n-r-j} \binom{n-r-j}{m} p^m (1-p)^{n-r-j-m} = p^r (1-p)^j,$$

оскільки маємо суму значень функції ймовірності для розподілу  $\text{Binom}(n-r-j, p)$ , яка дорівнює 1.

Нарешті,

$$\mathbb{P}(B_j) = \binom{j+r-1}{r-1} p^r (1-p)^j,$$

з урахуванням вищепереданих аргументів про кількість можливих варіантів розташування успіхів і невдач серед перших  $r+j$  випробувань.

Отже, ми показали, що для  $n$  випробувань

$$\mathbb{P}_X(X = j) = \mathbb{P}(B_j) = \binom{j+r-1}{r-1} p^r (1-p)^j, \quad 0 \leq j \leq n-r.$$

Як можна бачити, функція ймовірності не залежить від  $n$ , тому можна казати, що відповідна функція ймовірності справедлива для довільної послідовності випробувань.

Залишилось тільки показати, що ця функція справді є функцією ймовірності. Вона очевидно невід'ємна, тому потрібно показати, що сума всіх імовірностей дорівнює 1.

Для цього ми використаємо без доведення узагальнений біном Ньютона:

$$(a+b)^r = \sum_{j=0}^{\infty} \binom{r}{j} a^{r-j} b^j, \quad |a| > |b|, r \in \mathbb{R},$$

де біномні коефіцієнти визначено так:

$$\binom{r}{j} = \begin{cases} \frac{r(r-1)\dots(r-j+1)}{j!}, & r \geq j \\ 0, & r < j \end{cases},$$

що повністю відповідає стандартному визначенням для натуральних  $r \in \mathbb{N}$ , але також має коректне визначення для довільних  $r \in \mathbb{R}$ , у тому числі від'ємних<sup>4</sup>.

Ураховуючи вищена ведений результат, справедливою є така властивість:

$$\begin{aligned} \binom{j+r-1}{r-1} &= \frac{(j+r-1)(j+r-2)\dots r}{j!} \\ &= (-1)^j \frac{(-r)(-r-1)\dots(-r-(j-1))}{j!} \\ &= (-1)^j \binom{-r}{j}. \end{aligned}$$

Нарешті,

$$\begin{aligned} \sum_{j=0}^{\infty} \binom{j+r-1}{r-1} p^r (1-p)^j &= p^r \sum_{j=0}^{\infty} \binom{j+r-1}{r-1} (1-p)^j \\ &= p^r \sum_{j=0}^{\infty} \binom{-r}{j} (-1)^j (1-p)^j \\ &= p^r \sum_{j=0}^{\infty} \binom{-r}{j} 1^{r-j} (p-1)^j \\ &= p^r (1+p-1)^{-r} \\ &= p^r p^{-r} = 1. \end{aligned}$$

□

На Рис. 5.3.1 зображені функції ймовірності для від'ємного біномного розподілу з  $p = 0.5$  та різними значеннями параметра  $r$ .

**Зауваження 5.3.3.** У літературі не існує єдиного бачення, який саме розподіл визначати як від'ємний біномний. Деякі автори кажуть, що від'ємний біномний розподіл має випадкова величина  $Y = \text{«загальна кількість випробувань, потрібна для одержання } r \text{ успіхів»}$ . Тоді функція

<sup>4</sup>Звідси й назва — від'ємний біномний розподіл.

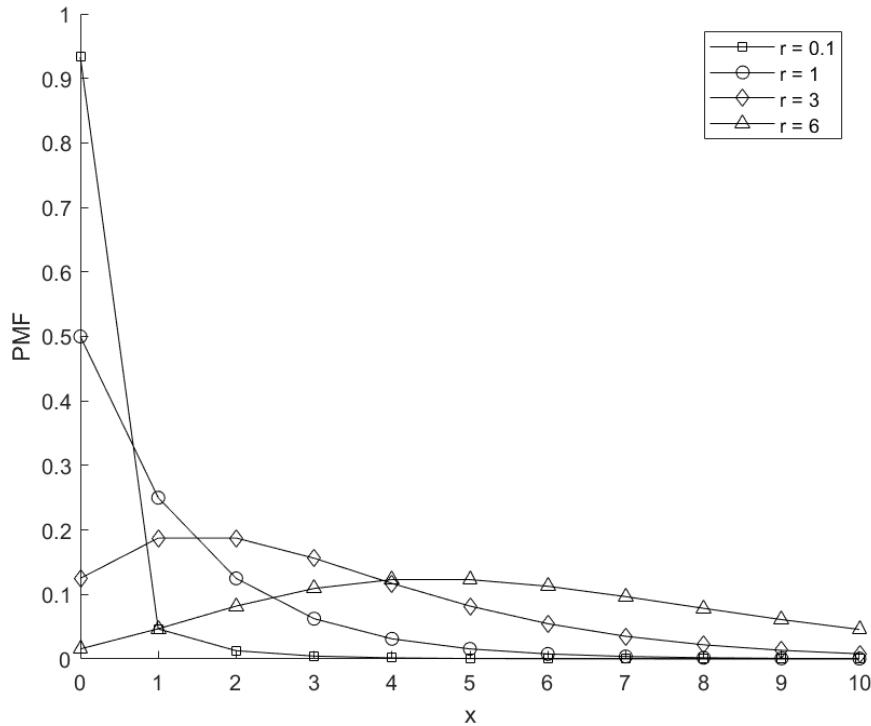


Рис. 5.3.1.: Приклади функцiй розподiлу для  $X \sim \text{NegBinom}(r, 0.5)$  для рiзних значень  $r$

ймовiрностi має вид

$$\mathbb{P}_Y(Y = k) = \binom{k-1}{r-1} p^r (1-p)^{k-r}, k = r, r+1, \dots.$$

□

**Приклад 5.3.4** (Задача Банаха про сiрники<sup>5</sup>). Математик уявив із собою 2 пачки з  $N$  сiрниками кожна, по однiй у кожнiй руцi. Щоразу запалюючи свою трубку, вiн виймає сiрник із однiєї з 2 пачок з однаковою ймовiрнiстю. Нехай на якомусь етапi математик помiчає, що одна пачка повнiстю спорожнiла. Чому дорiвнює ймовiрнiсть, що в iншiй пачцi залишилось  $k$  сiрникiв?

Моделювати вiдповiдну ситуацiю дoreчно за допомогою вiд'ємного бiномного розподiлу. Розгляньмо спочатку ситуацiю, що першою спорожнiла пачка у правiй руцi. Тодi за успiх покладемо витяг сiрника з правої пачки, а за невдачу — iз лiвої. Якщо права пачка виявилася порожньою пiд час чергового витягання, то це означає, що математик вибрав її  $N+1$  разiв, тобто мало мiсце  $N+1$  успiхiв. Тодi в лiвiй пачцi залишилось  $k = N - m$  сiрникiв, а отже сталося  $m = N - k$  невдач.

Вiдповiдно, iмовiрнiсть подiї  $A_k$  = «права пачка порожня, а в лiвiй пачцi залишилося  $k$  сiрникiв»

<sup>5</sup>Ця задача виросла з iронiчного коментаря польського математика Хуго Штайнхауза (Hugo Steinhaus, 1887–1972) стосовно iншого вiдомого польського математика Стефана Банаха (Stefan Banach, 1892–1945) та його звичку курити трубку.

можна обчислити як  $\mathbb{P}_Y(Y = N - k)$  для  $Y \sim \text{NegBinom}(N + 1, \frac{1}{2})$ :

$$\mathbb{P}(A_k) = \binom{(N - k) + (N + 1) - 1}{(N + 1) - 1} \left(\frac{1}{2}\right)^{N+1} \left(\frac{1}{2}\right)^{N-k} = \binom{2N - k}{N} \left(\frac{1}{2}\right)^{2N-k+1}.$$

Симетрична ситуація справедлива, якщо спорожніла пачка в лівій руці, а тому загальна ймовірність, що в одній із пачок лишилось  $X = k$  сірників, дорівнює

$$\mathbb{P}_X(X = k) = 2\mathbb{P}(A_k) = \binom{2N - k}{N} \left(\frac{1}{2}\right)^{2N-k}.$$

□

Зафіксуємо без доведення (його наведено в Розд. 5.6) вирази для сподівання та дисперсії від'ємного біномного розподілу.

**Твердження 5.3.5.** Якщо  $X \sim \text{NegBinom}(r, p)$ , то  $\mathbb{E}[X] = r \frac{1-p}{p}$ , а  $\text{Var}(X) = r \frac{1-p}{p^2}$ .

## 5.4. Гіпергеометричний розподіл

Нехай маємо вазу, у якій міститься  $w$  білих та  $b$  чорних кульок, і нехай витягають  $n$  кульок із повтореннями. Тоді випадкова величина  $X = \text{«кількість білих кульок серед витягнутих»}$  має біномний розподіл:  $X \sim \text{Binom}\left(n, \frac{w}{w+b}\right)$ . Якщо ж кульки витягають без повторень, то випадкова величина  $X$  матиме інший розподіл.

**Визначення 5.4.1.** Нехай маємо вазу, у якій міститься  $w$  білих та  $b$  чорних кульок. Тоді, випадкова величина  $X = \text{«кількість білих кульок серед } n \text{ витягнутих без повторень»}$  має *гіпергеометричний розподіл* (hypergeometric distribution) із параметрами  $w$ ,  $b$  і  $n$ , що позначають через  $X \sim \text{HGeom}(w, b, n)$ .

У загальному випадку під  $w$  і  $b$  розуміють число *бажаних* і *небажаних* результатів, а  $n$  — загальне число випробувань. □

Знайдімо функцію ймовірності такої випадкової величини.

**Твердження 5.4.2.** Якщо випадкова величина  $X \sim \text{HGeom}(w, b, n)$ , то функція її розподілу має вид<sup>6</sup>:

$$\mathbb{P}_X(X = k) = \frac{\binom{w}{k} \binom{b}{n-k}}{\binom{w+b}{n}}, \quad k = 0, 1, \dots, n. \quad (5.4.1)$$

*Доведення.* За (4.1.1) маємо:

$$\mathbb{P}_X(X = k) = \mathbb{P}(\{\omega : X(\omega) = k\}) = \sum_{\omega: X(\omega)=k} \mathbb{P}(\{\omega\}).$$

<sup>6</sup>Варто звернути увагу, що біномний коефіцієнт дорівнює 0, якщо  $k > w$  чи  $n - k > b$ , тому ми зазначили, що  $k$  пробігає всі значення від 0 до  $n$ , хоча, очевидно,  $k$  не може бути більшим від  $w$  чи меншим від  $n - b$ .

Кожна зі згаданих  $\omega$  — це деяка послідовність із  $n$  «витягнутих кульок» (результатів випробувань), серед яких є точно  $k$  «білих кульок» (успішних результатів). Кількість усіх таких послідовностей еквівалентна кількості варіантів вибору  $k$  елементів із  $w$  можливих, помножених на кількість вибору решти  $n - k$  елементів із  $b$  можливих, тобто  $\binom{w}{k} \binom{b}{n-k}$ .

Імовірність кожної окремої послідовності  $\omega$  з точно  $k$  «білими кульками» однакова і дорівнює  $\frac{1}{\binom{w+b}{n}}$ , оскільки всього існує  $\binom{w+b}{n}$  способів витягнути  $n$  кульок з-поміж  $w + b$  наявних без повторень. Отже,

$$\mathbb{P}_X(X = k) = \sum_{\omega: X(\omega) = k} \mathbb{P}(\{\omega\}) = \frac{\binom{w}{k} \binom{b}{n-k}}{\binom{w+b}{n}}.$$

Залишилось тільки показати, що сума всіх імовірностей дорівнює 1. Цей результат безпосередньо випливає з тотожності Вандермонда<sup>7</sup>:

$$\binom{w+b}{n} = \sum_{j=0}^n \binom{w}{j} \cdot \binom{b}{n-j}. \quad (5.4.2)$$

Цю тотожність ми довели на практичних заняттях.  $\square$

У загальному випадку гіпергеометричний розподіл застосовний до ситуацій, коли об'єкти мають два типи міток, а мітки принаймні одного типу присвоюють у випадковий спосіб. Тоді випадкова величина  $X$  відповідає числу об'єктів з мітками обох типів.

У нашому попередньому контексті такими мітками є колір кульки та той факт, відібрано її чи ні. Розгляньмо ще декілька прикладів задач, ізоморфних розглянутій вище.

**Приклад 5.4.3** (Кільцювання оленів). У лісі проживає  $N$  оленів. У деякий день  $m$  із них відловлюють, кільцюють і випускають назад у ліс. У деякий інший день відловлюють  $n$  нових оленів. Покладімо, що в новій групі оленів імовірність появі закільцюваних тварин така сама, як і для незакільцюваних (тобто закільцювані тварини не уникають дослідників навмисно).

Тоді число закільцюваних оленів у новій групі становить  $X \sim \text{HGeom}(m, N - m, n)$ .  $\square$

**Приклад 5.4.4.** Нехай випадковим чином вибирати 5 карт із добре перетасованої колоди. Тоді число тузів у кожній такій комбінації складає  $X \sim \text{HGeom}(4, 48, 5)$ , оскільки тузи відіграють роль білих кульок, а всі інші карти — чорних (загальне число — 52 карти).

Тоді, наприклад,

$$\mathbb{P}_X(X = 3) = \frac{\binom{4}{3} \binom{48}{2}}{\binom{52}{5}} \approx 0.0017,$$

тобто ймовірність витягу трьох тузів надзвичайно низька.  $\square$

<sup>7</sup>Названа так на честь французького математика Александра-Теофіля Вандермонда (Alexandre-Théophile Vandermonde, 1735–1796), проте вперше про нього є згадка в далекому 1303 р. у працях китайського математика Чжу Шицьзе.

**Твердження 5.4.5.** Випадкові величини  $X \sim \text{HGeom}(w, b, n)$  і  $Y \sim \text{HGeom}(n, w + b - n, w)$  рівні за розподілом.

*Доведення.*

$$\mathbb{P}_X(X = k) = \frac{\binom{w}{k} \binom{b}{n-k}}{\binom{w+b}{n}} = \frac{w!b!n!(w+b-n)!}{k!(w+b)!(w-k)!(n-k)!(b-n+k)!},$$

$$\mathbb{P}_Y(Y = k) = \frac{\binom{n}{k} \binom{w+b-n}{w-k}}{\binom{w+b}{w}} = \frac{w!b!n!(w+b-n)!}{k!(w+b)!(w-k)!(n-k)!(b-n+k)!},$$

тобто два вирази є ідентичними.  $\square$

Результат Твердження 5.4.5 стає очевидним, якщо помітити, що перша величина відповідає числу білих кульок у множині  $n$  кульок, витягнутих із  $w+b$  можливих, а друга величина відповідає числу *витягнутих* кульок із множини *білих* кульок. Іншими словами, обидві величини описують один і той самий набір кульок, тільки два типи міток — колір і властивість бути витягнутою — змінено місцями.

**Твердження 5.4.6.** Якщо  $X \sim \text{HGeom}(w, b, n)$ , то  $\mathbb{E}[X] = \frac{nw}{w+b}$ , а  $\text{Var}(X) = \frac{nw(b+w-n)}{(w+b)^2(w+b-1)}$ .

*Доведення.* За визначенням сподівання та з урахуванням (5.4.1) маємо:

$$\mathbb{E}[X] = \sum_{k=0}^n k \cdot \frac{\binom{w}{k} \binom{b}{n-k}}{\binom{w+b}{n}}.$$

Використовуючи комбінаторну тотожність  $\binom{n}{m} = \frac{n}{m} \binom{n-1}{m-1}$ , маємо:

$$\begin{aligned} \mathbb{E}[X] &= \sum_{k=0}^n k \cdot \frac{\frac{w}{k} \binom{w-1}{k-1} \binom{b}{(n-1)-(k-1)}}{\frac{n}{n} \binom{(w-1)+b}{n-1}} \\ &= \frac{nw}{w+b} \sum_{k=0}^n k \cdot \frac{\binom{w-1}{k-1} \binom{b}{(n-1)-(k-1)}}{\binom{(w-1)+b}{n-1}} \\ &= \frac{nw}{w+b} \sum_{j=0}^{n-1} \frac{\binom{w-1}{j} \binom{b}{(n-1)-j}}{\binom{(w-1)+b}{n-1}} \\ &= \frac{nw}{w+b}, \end{aligned}$$

оскільки сума, що залишилася — це сума всіх значень функції ймовірності для величини з розподiлом HGeom  $(w - 1, b, n - 1)$ , тобто 1.

Для обчислення дисперсії використаймо пiдхiд iз доведення Твердження 4.5.10.

$$\mathbb{E}[X^2] = \mathbb{E}[X] + \mathbb{E}[X(X - 1)] = \frac{nw}{w + b} + \mathbb{E}[X(X - 1)] ,$$

де

$$\begin{aligned} \mathbb{E}[X(X - 1)] &= \sum_{k=0}^n k(k - 1) \cdot \frac{\frac{w}{k} \binom{w-1}{k-1} \binom{b}{(n-1)-(k-1)}}{\frac{n}{w+b} \binom{(w-1)+b}{n-1}} \\ &= \sum_{k=2}^n k(k - 1) \cdot \frac{\frac{w}{k} \binom{w-1}{k-1} \binom{b}{(n-1)-(k-1)}}{\frac{n}{w+b} \binom{(w-1)+b}{n-1}} \\ &= \sum_{k=2}^n (k - 1) \cdot \frac{\frac{w(w-1)}{k-1} \binom{w-2}{k-2} \binom{b}{(n-1)-(k-1)}}{\frac{n}{w+b} \frac{w+b-1}{n-1} \binom{(w-2)+b}{n-2}} \\ &= \frac{n(n-1)w(w-1)}{(w+b)(w+b-1)} \sum_{j=0}^{n-2} \frac{\binom{w-2}{j} \binom{b}{(n-2)-j}}{\binom{(w-2)+b}{n-2}} \\ &= \frac{n(n-1)w(w-1)}{(w+b)(w+b-1)} , \end{aligned}$$

оскiльки вiдповiдна сума є сумаю всiх значень функцiї ймовiрностi для випадкової величини з розподiлом HGeom  $(w - 2, b, n - 2)$ , а тому дорiвнює 1.

Нарештi,

$$\text{Var}(X) = \mathbb{E}[X^2] - (\mathbb{E}[X])^2 = \left( \frac{nw}{w+b} + \frac{n(n-1)w(w-1)}{(w+b)(w+b-1)} \right) - \left( \frac{nw}{w+b} \right)^2 = \frac{nwb(w+b-n)}{(w+b)^2(w+b-1)} .$$

□

Як розподiл Пуассона є граничним для бiномного, так i бiномний є граничним — для гiпергеометричного.

**Твердження 5.4.7.** Якщо  $X \sim \text{HGeom}(w, b, n)$ , i  $N = w + b \rightarrow \infty$ , a  $\frac{w}{w+b} \rightarrow p$  залишається сталою, то функцiя розподiлу  $p_X(x)$  прямує до функцiї розподiлу величини  $Y \sim \text{Binom}(n, p)$ .

*Доведення.* Розпишімо функцію ймовірності  $X$ :

$$\begin{aligned}
 \mathbb{P}_X(X = k) &= \frac{\binom{w}{k} \binom{b}{n-k}}{\binom{w+b}{n}} \\
 &= \binom{n}{k} \frac{\binom{w+b-n}{w-k}}{\binom{w+b}{w}} \\
 &= \binom{n}{k} \frac{w!}{(w-k)!} \frac{b!}{(b-n+k)!} \frac{(w+b-n)!}{(w+b)!} \\
 &= \binom{n}{k} \frac{w(w-1)\dots(w-k+1)b(b-1)\dots(b-n+k+1)}{(w+b)(w+b-1)\dots(w+b-n+1)}.
 \end{aligned}$$

Для обчислення границі відповідного виразу за  $w+b \rightarrow \infty$ ,  $\frac{w}{w+b} \rightarrow p$ , поділімо кожну дужку під знаком границі на  $w+b$ . З урахуванням того, що для деякого натурального  $m \in \mathbb{N}$ :

$$\begin{aligned}
 \lim_{\substack{w+b \rightarrow \infty \\ \frac{w}{w+b} \rightarrow p}} \frac{w-m}{w+b} &= p-0=p, \\
 \lim_{\substack{w+b \rightarrow \infty \\ \frac{w}{w+b} \rightarrow p}} \frac{b-m}{w+b} &= \lim_{\substack{w+b \rightarrow \infty \\ \frac{w}{w+b} \rightarrow p}} \frac{(w+b)-w-m}{w+b} = 1-p-0=1-p, \\
 \lim_{\substack{w+b \rightarrow \infty \\ \frac{w}{w+b} \rightarrow p}} \frac{w+b-m}{w+b} &= 1-0=1,
 \end{aligned}$$

маємо:

$$\lim_{\substack{w+b \rightarrow \infty \\ \frac{w}{w+b} \rightarrow p}} \binom{n}{k} \frac{w(w-1)\dots(w-k+1)b(b-1)\dots(b-n+k+1)}{(w+b)(w+b-1)\dots(w+b-n+1)} = \binom{n}{k} \frac{p^k(1-p)^k}{1^k} = \binom{n}{k} p^k(1-p)^k.$$

□

Нескладно бачити, чому так відбувається. Оскільки біномний розподіл пов'язаний із витягуванням кульок із повтореннями, а гіпергеометричний — без, зі збільшенням загального числа кульок до нескінченності ці два способи витягувань стають однаковими.

На практиці результат цього твердження дає змогу використовувати біномний розподіл замість гіпергеометричного як його апроксимацію, коли  $N = w+b$  достатньо велике у порівнянні з  $n$ .

**Приклад 5.4.8.** Нехай  $X \sim \text{HGeom}(70, 90, 25)$ . Тоді значення функції ймовірності цієї величини в т.  $x = 10$  дорівнює

$$p_X(10) = \frac{\binom{70}{10} \binom{90}{25-10}}{\binom{70+90}{25}} \approx 0.166,$$

що вимагає проведення великої кількості непростих обчислень.

З іншого боку,  $\frac{w}{w+b} = \frac{7}{16}$ , і тому

$$p_X(10) \approx \binom{25}{10} \left(\frac{7}{16}\right)^{10} \left(\frac{9}{16}\right)^{15} = 0.15.$$

Наближене значення доволі близьке до точного, хоча похибка апроксимації складає приблизно 10.7%.  $\square$

## 5.5. Умовна функція ймовірності

У Лекції 3 ми розглядали поняття умовної ймовірності  $\mathbb{P}(A | B)$ , тобто ймовірності події  $A$  за умови, що сталася подія  $B$ . Аналогічне поняття можна увести для випадкових величин, тобто можна обчислювати ймовірність, що випадкова величина  $X$  набуває деяких значень, за умови, що інша випадкова величина  $Y$  дорівнює певному значенню. Уведімо відповідне поняття для дискретних випадкових величин, а загальний випадок будемо розглядати в наступних лекціях.

**Визначення 5.5.1.** Для будь-яких дискретних випадкових величин  $X$  і  $Z$  функцію  $p_{X|Z}(x | z) \equiv \mathbb{P}_{X|Z}(X = x | Z = z)$  називають *умовною функцією ймовірності* (conditional probability mass function) величини  $X$  за умови  $Z = z$ .  $\square$

Обчислення умовної функції розподілу можна виконати, використовуючи Визначення 4.1.6:

$$\mathbb{P}_{X|Z}(X = x | Z = z) = \mathbb{P}(\{\omega : X(\omega) = x\} \mid \{\omega : Z(\omega) = z\}) = \frac{\mathbb{P}(\{\omega : (X(\omega) = x) \cap (Z(\omega) = z)\})}{\mathbb{P}(\{\omega : Z(\omega) = z\})}. \quad (5.5.1)$$

Зокрема, справедливе правило Бееса з Теореми 3.3.1:

$$\mathbb{P}_{X|Z}(X = x | Z = z) = \frac{\mathbb{P}_{Z|X}(Z = z | X = x) \mathbb{P}_X(X = x)}{\mathbb{P}_Z(Z = z)}. \quad (5.5.2)$$

**Приклад 5.5.2** (Точний тест Фішера). Нехай деякий дослідник бажає дослідити питання, чи схильні до деякої хвороби чоловіки більше від жінок. Для цього формують випадкову вибірку з  $n$  жінок і  $m$  чоловіків, і кожного з цієї вибірки тестиють на наявність хвороби. Для спрощення вважатимемо, що тест ніколи не помилляється. Тоді, очевидно, випадкові величини  $X$  і  $Y$ , які відповідають кількості хворих жінок і чоловіків, відповідно, мають такі розподіли:

$$X \sim \text{Binom}(n, p_1), \quad Y \sim \text{Binom}(m, p_2),$$

де  $p_1$  і  $p_2$  — параметри розподілів (імовірності захворюваності серед жінок і чоловіків, відповідно), які досліднику невідомі. На основі даних вибірки дослідник прагне оцінити, чи  $p_1 = p_2$ . Оцінку дослідник робитиме за допомогою одного з методів статистики, відомого як *точний тест Фішера* (Fisher's exact test).

Розгляньмо Табл. 5.5.1, у якій рядки відповідають стану здоров'я, стовпці — статі особи, а комірки — кількості спостережень із відповідними характеристиками. Такі таблиці називають *таблицями спряженості* (contingency tables). У таблиці спряженості використано такі позначення:

Табл. 5.5.1.: Таблиця спряженості для Прикладу 5.5.2

	Жінки	Чоловіки	Разом
Хворі	$x$	$r - x$	$r$
Здорові	$n - x$	$m - r + x$	$n + m - r$
Разом	$n$	$m$	$n + m$

- $x$ : кількість хворих жінок у даній вибірці (тобто реалізоване значення випадкової величини  $X$ );
- $r = X + Y$  — загальне число хворих.

Процедуру виконання тесту Фішера ми розглядати не будемо, оскільки це предмет вивчення навчальної дисципліни «Математична статистика». Коротко його суть полягає в тому, що значення  $x$ , якого величина  $X$  набуває у даній вибірці, не може занадто сильно відхилятися від сподівання  $\mathbb{E}[X]$ , якщо насправді  $p_1 = p_2$ . У протилежному випадку можна буде зробити висновок, що  $p_1 \neq p_2$ .

Таким чином, для виконання тесту спершу потрібно знайти функцію ймовірності для величини  $X$ . Оскільки параметр  $p_1$  невідомий, ми натомість шукатимемо *умовну* функцію розподілу:

$$\begin{aligned} \mathbb{P}_{X|X+Y}(X = x | X + Y = r) &= \frac{\mathbb{P}_{X+Y|X}(X + Y = r | X = x) \mathbb{P}_X(X = x)}{\mathbb{P}_{X+Y}(X + Y = r)} \\ &= \frac{\mathbb{P}_Y(Y = r - x) \mathbb{P}_X(X = x)}{\mathbb{P}_{X+Y}(X + Y = r)}. \end{aligned}$$

Перехід  $\mathbb{P}_{X|X+Y}(X = x | X + Y = r) = \mathbb{P}_Y(Y = r - x)$  можна здійснити з урахуванням незалежності  $X$  і  $Y$ .

За умови  $p_1 = p_2 = p$ , сума величин  $X + Y$  має біномний розподіл  $X + Y \sim \text{Binom}(n + m, p)$ , адже вона показує число успіхів із  $n + m$  випробувань з однаковою ймовірністю. Отже,

$$\mathbb{P}_{X|X+Y}(X = x | X + Y = r) = \frac{\binom{m}{r-x} p^{r-x} (1-p)^{m-r+x} \cdot \binom{n}{x} p^x (1-p)^{n-x}}{\binom{n+m}{r} p^r (1-p)^{n+m-r}} = \frac{\binom{n}{x} \binom{m}{r-x}}{\binom{n+m}{r}}.$$

Відтак,  $X | X + Y \sim \text{HGeom}(n, m, r)$ , а параметр  $p$  взагалі випав із формули функції ймовірності.  $\square$

## 5.6. Доведення окремих тверджень

Доведімо вирази для сподівання та дисперсії від'ємного біномного розподілу з Твердження 5.3.5.

*Доведення.* Для доведення використаймо прийом із Заваження 4.5.11:

$$\begin{aligned}
 \mathbb{E}[X] &= \sum_{j=0}^{\infty} j \cdot \binom{j+r-1}{r-1} p^r (1-p)^j = r \frac{1-p}{p} \sum_{j=1}^{\infty} \frac{p}{1-p} \frac{j}{r} \cdot \binom{j+r-1}{r-1} p^r (1-p)^j \\
 &= r \frac{1-p}{p} \sum_{j=1}^{\infty} \frac{p}{1-p} \frac{j}{r} \cdot \binom{j+r-1}{j} p^r (1-p)^j \\
 &= r \frac{1-p}{p} \sum_{j=1}^{\infty} \binom{j+r-1}{j-1} p^{r+1} (1-p)^{j-1} \\
 &= r \frac{1-p}{p} \sum_{k=0}^{\infty} \binom{k+(r+1)-1}{k} p^{r+1} (1-p)^k \\
 &= r \frac{1-p}{p} \sum_{k=0}^{\infty} \binom{k+(r+1)-1}{(r+1)-1} p^{r+1} (1-p)^k \\
 &= r \frac{1-p}{p} ,
 \end{aligned}$$

оскільки сума всіх імовірностей для випадкової величини з розподілом NegBinom( $r+1, p$ ) дорівнює 1, а

$$\frac{j}{r} \cdot \binom{j+r-1}{j} = \frac{j}{r} \cdot \frac{(j+r-1)!}{j!(r-1)!} = \frac{(j+r-1)!}{(j-1)!r!} = \binom{j+r-1}{j-1} .$$

Для обчислення дисперсії спочатку підрахуймо сподівання від квадрата випадкової величини:

$$\begin{aligned}
 \mathbb{E}[X^2] &= \sum_{j=0}^{\infty} j^2 \cdot \binom{j+r-1}{r-1} p^r (1-p)^j = \sum_{j=1}^{\infty} j^2 \cdot \frac{(j+r-1)!}{j!(r-1)!} p^r (1-p)^j \\
 &= \sum_{j=1}^{\infty} j \cdot \frac{(j+r-1)!}{(j-1)!(r-1)!} p^r (1-p)^j \\
 &= \sum_{m=0}^{\infty} (m+1) \cdot \frac{(m+r)!}{m!(r-1)!} p^r (1-p)^{m+1} \\
 &= \sum_{m=0}^{\infty} m \cdot \frac{(m+r)!}{m!(r-1)!} p^r (1-p)^{m+1} + \sum_{m=0}^{\infty} \frac{(m+r)!}{m!(r-1)!} p^r (1-p)^{m+1} .
 \end{aligned}$$

Першу суму обчислімо так:

$$\begin{aligned}
 \sum_{m=0}^{\infty} m \cdot \frac{(m+r)!}{m!(r-1)!} p^r (1-p)^{m+1} &= \sum_{m=0}^{\infty} \frac{(m+r)!}{(m-1)!(r-1)!} p^r (1-p)^{m+1} \\
 &= \sum_{s=0}^{\infty} \frac{(s+r+1)!}{s!(r-1)!} p^r (1-p)^{s+2} \\
 &= \sum_{s=0}^{\infty} \frac{(s+(r+2)-1)!}{s!(r-1)!} p^r (1-p)^{s+2} \\
 &= r(r+1) \frac{(1-p)^2}{p^2} \sum_{s=0}^{\infty} \frac{(s+(r+2)-1)!}{s!(r-1)!} \cdot p^{r+2} (1-p)^s \\
 &= r(r+1) \frac{(1-p)^2}{p^2} \sum_{s=0}^{\infty} \frac{(s+(r+2)-1)!}{s!((r+2)-1)!} \cdot p^{r+2} (1-p)^s \\
 &= r(r+1) \frac{(1-p)^2}{p^2} \sum_{s=0}^{\infty} \binom{s+(r+2)-1}{(r+2)-1} \cdot p^{r+2} (1-p)^s \\
 &= r(r+1) \frac{(1-p)^2}{p^2},
 \end{aligned}$$

оскільки сума всіх імовірностей для випадкової величини з розподілом  $\text{NegBinom}(r+2, p)$  дорівнює 1.

Другу суму обчислімо так:

$$\begin{aligned}
 \sum_{m=1}^{\infty} \frac{(m+r)!}{m!(r-1)!} p^r (1-p)^{m+1} &= r \frac{1-p}{p} \sum_{m=1}^{\infty} \frac{(m+(r+1)-1)!}{m!((r+1)-1)!} p^{r+1} (1-p)^m \\
 &= r \frac{1-p}{p} \sum_{m=0}^{\infty} \binom{m+(r+1)-1}{(r+1)-1} p^{r+1} (1-p)^m \\
 &= r \frac{1-p}{p},
 \end{aligned}$$

оскільки сума всіх імовірностей для випадкової величини з розподілом  $\text{NegBinom}(r+1, p)$  дорівнює 1.

Остаточно, маємо:

$$\text{Var}(X) = \mathbb{E}[X^2] - (\mathbb{E}[X])^2 = r(r+1) \frac{(1-p)^2}{p^2} + r \frac{1-p}{p} - \left(r \frac{1-p}{p}\right)^2 = r \frac{1-p}{p^2}.$$

□

Частина II.  
Випадкові величини

## 6. Неперервні випадкові величини

Як було зазначено під час уведення Визначення 4.1.5, випадковою величиною є будь-яка дійсно-значна вимірна функція  $X : \Omega \rightarrow \mathbb{R}$ , що діє з деякого ймовірнісного простору  $(\Omega, \mathcal{A}, \mathbb{P})$  на Борелів простір  $(\mathbb{R}, \mathcal{B})$ . При цьому випадкова величина задає нову міру на Борелевій  $\sigma$ -алгебрі — розподіл  $\mathbb{P}_X$  — який, згідно з (4.1.1), має вид

$$\mathbb{P}_X(X \in A) = \mathbb{P}(X^{-1}(A)) = \mathbb{P}(\{\omega \in \Omega : X(\omega) \in A\}) , \quad A \in \mathcal{B} ,$$

де

$$X^{-1}(A) = \{\omega \in \Omega : X(\omega) \in A\}$$

є прообразом множини  $A$ .

Як ми зазначали раніше, працювати з визначенням розподілу безпосередньо надзвичайно важко, адже постійно обчислювати ймовірності подій за участю випадкової величини, повертаючись до міри на початковому ймовірнісному просторі  $(\Omega, \mathcal{A}, \mathbb{P})$ , незручно. Під час огляду дискретних випадкових величин ми з'ясували, що розподіл випадкової величини можна компактно описати за допомогою функції ймовірності, яка кожному елементу зі (зліченного) носія величини  $X$  зіставляє його ймовірність  $\mathbb{P}_X(X = x)$ . Проте якщо носій *незліченний*, то сума перетворюється в інтеграл  $\mathbb{P}_X(X \in A) = \int_A \mathbb{P}_X(X = x) dx$ , а відтак  $\mathbb{P}_X(X = a) = \int_a^a \mathbb{P}_X(X = x) dx = 0$ . Отже функція ймовірності в цьому випадку є безкорисною.

У загальному випадку для опису будь-якої випадкової величини нам знадобиться інша функція — функція розподілу.

### 6.1. Функції розподілу

Як було зазначено в Зауваженні 1.5.21, Борелеву  $\sigma$ -алгебру, серед іншого, породжують промені виду  $(-\infty; a]$ ,  $a \in \mathbb{R}$ . Іншими словами, будь-яку Борелеву множину  $A \in \mathcal{B}$  можна утворити зліченними запереченнями й об'єднаннями променів. Згідно з властивістю  $\sigma$ -адитивності будь-якої міри, якщо ми знаємо  $\mathbb{P}((-\infty; a])$  для всіх  $a \in \mathbb{R}$ , то ми можемо обчислити ймовірність будь-якої множини  $A \in \mathcal{B}$  як (зліченну) суму ймовірностей відповідних променів. Це і є функція розподілу.

**Визначення 6.1.1.** *Функцією розподілу* (distribution function, інколи cumulative distribution function, CDF) випадкової величини  $X$  є функція

$$F_X(x) \equiv \mathbb{P}_X(X \leq x) = \mathbb{P}_X((-\infty; x]) , \quad x \in \mathbb{R} . \quad (6.1.1)$$

□

Розгляньмо декілька властивостей функції розподілу, які випливають безпосередньо з її визначення.

**Твердження 6.1.2.** Функція розподілу  $F_X(x)$  має такі властивості:

(i) є неспадною:  $F_X(x) \leq F_X(y)$ , якщо  $x \leq y$ ;

(ii) є неперервною справа<sup>1</sup>:  $\lim_{x \downarrow c} F_X(x) = F_X(c)$ ;

(iii) є нормованою:

$$\lim_{x \rightarrow -\infty} F_X(x) = 0, \\ \lim_{x \rightarrow \infty} F_X(x) = 1.$$

*Доведення.* (i) Нехай  $x \leq y$ . Тоді в силу властивості монотонності міри з Теореми 2.2.1,

$$F_X(x) = \mathbb{P}_X((-\infty; x]) \leq \mathbb{P}_X((-\infty; y]) = F_X(y);$$

(ii) оскільки за Теоремою 2.2.8 будь-яка міра є неперервною, у тому числі неперервною зверху, можемо розглянути спадну послідовність променів  $I_n = (-\infty; c + h_n]$ ,  $h_n \rightarrow 0$  (наприклад,  $h_n = 1/n$ ). Очевидно, що

$$\lim_{n \rightarrow \infty} I_n = (-\infty; c],$$

а тоді

$$\lim_{n \rightarrow \infty} \mathbb{P}_X(I_n) = \mathbb{P}_X\left(\lim_{n \rightarrow \infty} I_n\right) = \mathbb{P}_X((-\infty; c]),$$

тобто

$$\lim_{n \rightarrow \infty} F_X(c + h_n) = \lim_{x \downarrow c} F_X(x) = F_X(c);$$

(iii) для доведення  $\lim_{x \rightarrow -\infty} F_X(x) = 0$  можна скористатися спадною послідовністю  $I_n = (-\infty; x - h_n]$ ,  $h_n \rightarrow \infty$  (наприклад,  $h_n = n$ ), оскільки

$$\lim_{n \rightarrow \infty} I_n = \emptyset,$$

а тоді

$$\lim_{n \rightarrow \infty} F_X(x - h_n) = \lim_{n \rightarrow \infty} \mathbb{P}_X(I_n) = \mathbb{P}_X\left(\lim_{n \rightarrow \infty} I_n\right) = 0.$$

Схожим чином можна показати  $\lim_{x \rightarrow \infty} F_X(x) = 1$ , розглянувши зростаючу послідовність  $I_n = (-\infty; x + h_n]$ ,  $h_n \rightarrow \infty$ , для якої  $\lim_{n \rightarrow \infty} I_n = \Omega$ , а  $\mathbb{P}_X(\Omega) = 1$ .

□

Ілюстрацію відповідних властивостей наведено на Рис. 6.1.1 для функції розподілу величини Бернуллі  $X \sim \text{Bern}(0.7)$ . Як можна бачити, значення  $F(x)$  дорівнює 0 для  $x < 0$ , оскільки ймовірність, що  $X$  набуде від'ємних значень, дорівнює 0. На інтервалі  $0 \leq x < 1$  функція  $F(x) = 0.3$ , оскільки на цьому інтервалі подія  $X \leq x$  еквівалентна події  $X = 0$ . Нарешті, для всіх  $x \geq 1$  функція  $F(x) = 1$ , оскільки  $X \leq 1$  стається з одиничною ймовірністю. Ми бачимо, що в точках  $x = 0$  і  $x = 1$  функція розподілу має *стрибок* (jump)<sup>2</sup>, тобто вона не є неперервною, проте є неперервною справа.

<sup>1</sup>Функція *неперервна справа* (continuous from the right) в точці  $c$ , якщо  $\lim_{x \rightarrow c} f(x) \equiv \lim_{x \downarrow c} f(x)$  інше.

<sup>2</sup>Він же розрив I роду.

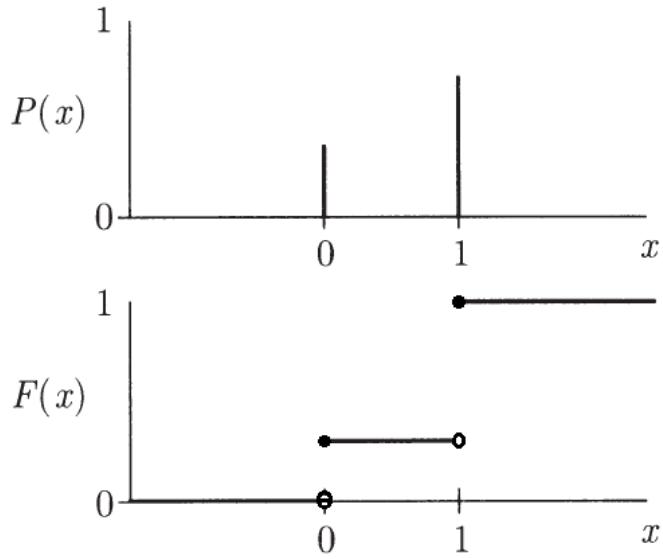


Рис. 6.1.1.: Функція ймовірності  $P(x)$  та функція розподілу  $F(x)$  для випадкової величини  $X \sim \text{Bern}(0.7)$  ([9], Рис. 4.5.2)

**Зауваження 6.1.3.** Функція розподілу  $F_X$  є неперервною справа, але вона не є неперервною зліва.

Повторюючи кроки з доведення властивості (ii), маємо таке. За Теоремою 2.2.8 будь-яка міра є неперервною, у тому числі неперервною *знизу*, а отже можемо розглянути зростаючу послідовність променів  $I_n = (-\infty; c - \frac{1}{n}]$ . Тоді

$$\lim_{n \rightarrow \infty} \mathbb{P}_X(I_n) = \mathbb{P}_X\left(\lim_{n \rightarrow \infty} I_n\right) ,$$

а

$$\lim_{n \rightarrow \infty} F_X\left(c - \frac{1}{n}\right) = \lim_{x \uparrow c} F_X(x) \equiv \lim_{\substack{x \rightarrow c \\ x < c}} F_X(x) \equiv F(c-) .$$

Проте в цьому випадку

$$\lim_{n \rightarrow \infty} I_n = (-\infty; c) ,$$

оскільки  $c \notin (-\infty; c - 1/n]$  для кожного  $n$ .

Отже,

$$F(c-) = \mathbb{P}_X\left(\lim_{n \rightarrow \infty} I_n\right) = \mathbb{P}_X((-\infty; c)) = \mathbb{P}_X(X < c) .$$

Тому  $F_X(c-) \neq F_X(c)$  у загальному випадку. У кожній такій точці  $c$  функція  $F$  має стрибок.

Відповідні міркування чимось схожі на спроби показати в Прикладі 1.5.7, чому алгебра пів інтервалів не є  $\sigma$ -алгеброю.  $\square$

**Приклад 6.1.4.** Розгляньмо *індикаторну величину* (indicator random variable) для деякої події  $A$  з імовірнісного простору  $(\Omega, \mathcal{A}, \mathbb{P})$ :

$$\mathbb{1}_A(\omega) = \begin{cases} 1, & \omega \in A \\ 0, & \omega \notin A \end{cases} .$$

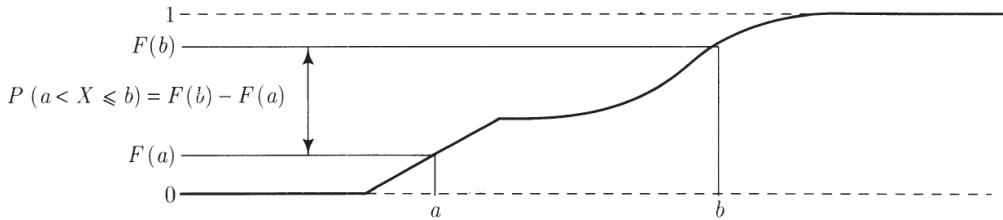


Рис. 6.1.2.: Ілюстрація до Твердження 6.1.5 ([9], Рис. 4.5.1)

Прообрази променів під дією цієї функції такі:

$$(\mathbb{1}_A(\omega))^{-1}((-\infty; x]) = \begin{cases} \emptyset, & x < 0 \\ A^c, & x \in [0; 1) \\ \Omega, & x \geq 1 \end{cases}.$$

Відповідно, функція розподілу, за визначенням, дорівнює:

$$F_{\mathbb{1}_A(\omega)}(x) = \begin{cases} 0, & x < 0 \\ \mathbb{P}(A^c), & x \in [0; 1) \\ 1, & x \geq 1 \end{cases}.$$

□

Доведімо деякі правила обчислення ймовірностей на основі функцій розподілу.

**Твердження 6.1.5.** Нехай  $F : \mathbb{R} \rightarrow \mathbb{R}$  — функція розподілу деякої випадкової величини  $X$ . Тоді:

- (i)  $\mathbb{P}_X(X \geq x) = 1 - F(x-);$
- (ii)  $\mathbb{P}_X((a; b]) = F(b) - F(a), a, b \in \mathbb{R};$
- (iii)  $\mathbb{P}_X(X = x) = F(x) - F(x-).$

*Доведення.* (i) Оскільки  $\mathbb{P}_X(\mathbb{R}) = 1$ , за властивостями ймовірнісної міри маємо:

$$\mathbb{P}_X(X \geq x) = 1 - \mathbb{P}_X(X < x) = 1 - F(x-);$$

(ii) це випливає з

$$\mathbb{P}_X((a; b]) = \mathbb{P}_X((-\infty; b]) - \mathbb{P}_X((-\infty; a]) = F(b) - F(a);$$

(iii) це випливає з

$$\mathbb{P}(X = x) = \mathbb{P}_X(X \leq x) - \mathbb{P}_X(X < x) = F(x) - F(x-).$$

□

Ілюстрації Твердження 6.1.5, зокрема, властивості (ii) наведено на Рис. 6.1.2.

Виявляється, що не тільки деякий розподіл випадкової величини можна описати за допомогою функції розподілу, а що й будь-яка функція розподілу *однозначно* задає розподіл деякої випадкової величини.

**Теорема 6.1.6.** Нехай  $F : \mathbb{R} \rightarrow \mathbb{R}$  — деяка функція, яка задовольняє умови (i)–(ii) з Твердження 6.1.2. Тоді в Борелевому просторі  $(\mathbb{R}, \mathcal{B})$  існує єдина міра  $\nu$  така, що  $\nu((a; b]) = F(b) - F(a)$ ,  $a, b \in \mathbb{R}$ .

Якщо ця функція додатково задовольняє умову (iii) з Твердження 6.1.2, то міра  $\nu$  є ймовірнісною мірою, тобто існує випадкова величина  $X : (\Omega, \mathcal{A}, \mathbb{P}) \rightarrow (\mathbb{R}, \mathcal{B})$  на деякому ймовірнісному просторі така, що  $F(x) = \mathbb{P}_X(X \leq x)$ .

Часткове доведення цієї теореми можна знайти в Розд. 6.5.

Цілком очевидно, що якщо покласти  $F(x) = x$  на пів інтервалі  $(0; 1]$  (тоді  $F(x) = 0$  для  $x \leq 0$ , і  $F(x) = 1$  для  $x > 1$ , інакше не будуть виконуватися властивості функцію розподілу), то ми дістанемо міру Лебега на  $(0; 1]$ :  $\lambda((a; b]) = F(b) - F(a) = b - a$ , яку ми визначили в Розд. 2.4.

**Зауваження 6.1.7.** Замість математично строгої, але багатослівної фрази «функція розподілу  $F_X$ , яка однозначно задає розподіл  $\mathbb{P}_X$  для деякої випадкової величини  $X$ », будемо просто казати — «функція розподілу  $F_X$  величини  $X$ ».

Також опускатимемо індексування  $F_X$  і писатимемо просто  $F$ , якщо за контекстом зрозуміло, про яку саме випадкову величину мова.  $\square$

Властивість (iii) з Твердження 6.1.5 дає змогу збудувати функцію розподілу для будь-якої дискретної випадкової величини. По-перше, кількість стрибків у функції розподілу може бути хіба що зліченна.

**Твердження 6.1.8.** У будь-якої функції розподілу  $F(x)$  є не більш ніж зліченна кількість стрибків.

Доведення цього твердження можна знайти в Розд. 6.5.

**Приклад 6.1.9.** Нехай носій дискретної випадкової величини  $X$  складається зі зліченного числа точок:  $\text{supp}(X) = \{x_1, x_2, \dots\}$ ,  $x_i < x_{i+1}$ ,  $i = 1, 2, \dots$ . Функція ймовірності дорівнює  $p_X(x_i)$  для кожної точки  $x_i$ ,  $i = 1, 2, \dots$ , та 0 — для всіх інших точок.

Із Твердження 6.1.5 випливає, що в кожній точці  $x_i$  функція розподілу  $F_X$  повинна мати стрибок  $F_X(x_i) - F_X(x_i-)$ , а в усіх інших точках — бути неперервною (і сталою):  $F_X(x) = F_X(x-)$ ,  $x \neq x_i$ ,  $i = 1, 2, \dots$

Таким чином, для побудови функції розподілу в деякій точці  $x$  на основі функції ймовірності  $p_X$  потрібно просумувати всі значення функції ймовірності для елементів її носія, що не перевищують  $x$ :

$$F_X(x) = \sum_{\substack{y \in \text{supp}(X) \\ y \leq x}} p_X(y). \quad (6.1.2)$$

Нехай  $X \sim \text{Binom}(4, 0.5)$ . На Рис. 6.1.3 зображені функція ймовірності і функція розподілу для  $X$ . Так, наприклад,

$$F_X(1.5) = p_X(0) + p_X(1) = \left(\frac{1}{2}\right)^4 + 4 \left(\frac{1}{2}\right)^4 = \frac{5}{16},$$

а, скажімо,

$$p_X(2) = F_X(2) - F_X(1) = \frac{11}{16} - \frac{5}{16} = \frac{6}{16} = \frac{3}{8}.$$

$\square$

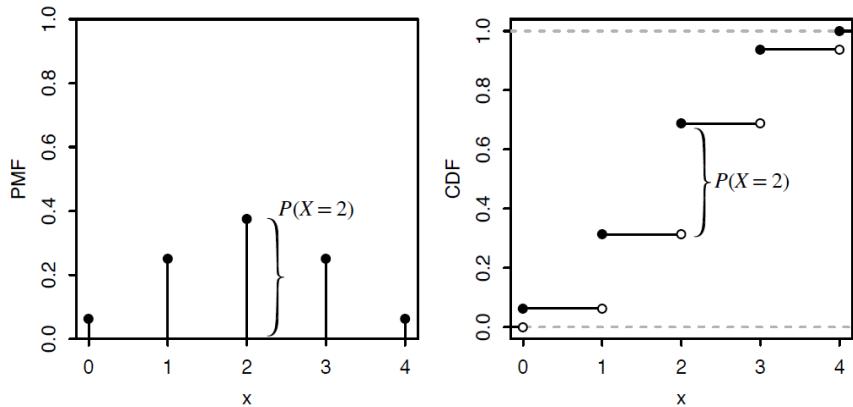


Рис. 6.1.3.: Функція ймовірності та функція розподілу випадкової величини  $X \sim \text{Binom}(4, 0.5)$  ([1], Рис. 3.8)

## 6.2. Зв'язок інтегралу та міри

У попередніх лекціях під час розгляду дискретних випадкових величин ми з'ясували, що за побудовою ймовірнісної міри на дискретному ймовірнісному просторі  $(\Omega, \mathcal{A}, \mathbb{P})$ , імовірністю деякої події  $A \in \mathcal{A}$  є сума ймовірностей елементарних подій, що їй належать:  $\mathbb{P}(A) = \sum_{\omega \in A} \mathbb{P}(\omega)$ . Тоді для деякої дискретної випадкової величини  $X : \Omega \rightarrow \mathbb{R}$  ймовірність події  $X \in A$  можна було обчислити через розподіл  $\mathbb{P}_X$  як суму значень відповідної функції ймовірності:

$$\mathbb{P}_X(A) = \sum_{x_i \in A} p_X(x_i).$$

Виявляється, ці міркування можна узагальнити на довільний імовірнісний простір і довільні випадкові величини. Як відомо, узагальненням поняття суми є поняття *інтегралу* (integral). Оскільки ми працюємо в парадигмі мір, то нас цікавить інтеграл *за мірою* (integral with respect to a measure). Зокрема, якщо такою мірою є міра Лебега  $\lambda$ , то інтеграл має назву *інтегралу Лебега* (Lebesgue integral). Ми не будемо постійно це уточнювати, а використовуватиме поняття інтегралу саме в цьому розумінні, якщо не буде зазначено інакше.

Окрім цього, у Лекції 4 ми увели поняття сподівання дискретної випадкової величини (Визначення 4.3.1) як зважене середнє значень із її носія, де ваговими коефіцієнтами є значення функції ймовірності для відповідних елементів носія:

$$\mathbb{E}[X] = \sum_{x \in \text{supp}(X)} x \cdot \mathbb{P}_X(X = x).$$

Відповідне поняття також можна узагальнити на випадок довільних випадкових величин, і оскільки узагальненням суми є інтеграл, то сподівання довільної випадкової величини *також* доречно визначити як *інтеграл* за її ж розподілом.

Детальне виведення інтегралу деякої функції з усіма доведеннями ми робити не будемо, оскільки це роблять у курсі «Математичного аналізу». Ми тільки згадаємо основні кроки побудови інтегралу деякої випадкової величини<sup>3</sup>.

<sup>3</sup>Усі твердження будуть справедливі для інтегралу будь-якої вимірної функції.

### 6.2.1. Крок 1. Індикаторні величини

Найпростішою випадковою величиною є *індикаторна* (indicator variable):

$$\mathbb{1}_A(x) = \begin{cases} 1, & x \in A \\ 0, & x \notin A \end{cases}.$$

Як і у випадку з розподілом випадкової величини, даватимемо собі право зловживати позначенням і застосовувати позначення  $\mathbb{1}\{x \in A\}$  замість  $\mathbb{1}_A(x)$ . Наприклад,  $\mathbb{1}\{x < 0\} \equiv \mathbb{1}_{\{x:x<0\}}(x)$ .

Тоді інтеграл індикаторної функції *визначає* як

$$\int \mathbb{1}_A d\mathbb{P} = \mathbb{P}(A), \quad (6.2.1)$$

Часто для спрощення запису писатимемо

$$\int_A f d\mathbb{P} = \int f \cdot \mathbb{1}_A d\mathbb{P}.$$

**Зауваження 6.2.1.** Нескладно бачити, що індикаторна величина є нічим іншим, як *величиною Бернуллі*. Справді, індикаторна величина набуває тільки двох значень, а тому

$$\begin{aligned} \mathbb{P}_X(\{1\}) &= \mathbb{P}(\{\omega : \mathbb{1}\{\omega \in A\} = 1\}) = \mathbb{P}(A), \\ \mathbb{P}_X(\{0\}) &= \mathbb{P}(\{\omega : \mathbb{1}\{\omega \in A\} = 0\}) = \mathbb{P}(A^c), \end{aligned}$$

а відтак  $X = \mathbb{1}_A$  породжує розподіл Бернуллі з Визначення 4.5.1:  $X \sim \text{Bern}(\mathbb{P}(A))$ . □

### 6.2.2. Крок 2. Прості випадкові величини

**Визначення 6.2.2.** Випадкову величину  $X : (\Omega, \mathcal{A}, \mathbb{P}) \rightarrow (\mathbb{R}, \mathcal{B})$  називають *простою* (simple), якщо її можна подати як

$$X(\omega) = \begin{cases} a_1, & \omega \in A_1 \\ a_2, & \omega \in A_2 \\ \vdots \\ a_n, & \omega \in A_n \end{cases},$$

або в компактнішому виді<sup>4</sup> як *скінченну* суму

$$X(\omega) = \sum_{i=1}^n a_i \mathbb{1}\{\omega \in A_i\}, \quad (6.2.2)$$

де  $A_i$  — деякі попарно несумісні події, а  $a_i > 0$ ,  $i = 1, \dots, n$ . □

<sup>4</sup> Альтернативно можна було б записати  $X(\omega) = \prod_{i=1}^n a_i^{\mathbb{1}\{\omega \in A_i\}}$ .

Тоді інтеграл простої функції визначають як

$$\int X d\mathbb{P} = \sum_{i=1}^n a_i \mathbb{P}(A_i) , \quad (6.2.3)$$

де прийнято вважати, що  $0 \cdot \infty = \infty \cdot 0 = 0$ .

**Приклад 6.2.3.** Розглянемо ймовірнісний простір  $((0; 1], \mathcal{B}(0; 1], \lambda)$  та задаймо (прості) випадкові величини так:

$$X(\omega) = \begin{cases} 5, & \omega > \frac{1}{3} \\ 3, & \omega \leq \frac{1}{3} \end{cases} , \quad Y(\omega) = \begin{cases} 2, & \omega \in \mathbb{Q} \\ 4, & \omega = \frac{1}{\sqrt{2}} \\ 6, & \omega \leq \frac{1}{4}, \omega \notin \mathbb{Q} \\ 8, & \text{інакше} \end{cases} .$$

Тоді інтегралами цих простих функцій будуть

$$\begin{aligned} \int X d\lambda &= 5 \cdot \lambda\left(\left(\frac{1}{3}; 1\right]\right) + 3 \cdot \lambda\left(\left(0; \frac{1}{3}\right]\right) = \frac{10}{3} + \frac{3}{3} = \frac{13}{3} , \\ \int Y d\lambda &= 2 \cdot \lambda(\mathbb{Q}) + 4 \cdot \lambda\left(\left\{\frac{1}{\sqrt{2}}\right\}\right) + 6 \cdot \lambda\left(\left\{\omega : \omega \leq \frac{1}{4}, \omega \notin \mathbb{Q}\right\}\right) + 8 \cdot \lambda\left(\left\{\omega : \omega > \frac{1}{4}, \omega \notin \mathbb{Q}\right\}\right) \\ &= 2 \cdot 0 + 4 \cdot 0 + \frac{6}{4} + \frac{24}{4} = \frac{15}{2} . \end{aligned}$$

□

**Зауваження 6.2.4.** Подання (6.2.2) у загальному випадку не є однозначним, проте можна довести, що визначення інтегралу є коректним, у тому сенсі, що інтеграл простої функції  $f$  дорівнюватиме одному й тому ж значенню, незалежно від варіантів подання  $f$ . □

### 6.2.3. Крок 3. Невід'ємні випадкові величини

Нехай маємо загальну невід'ємну випадкову величину  $X$ , тобто таку, що  $X \geq 0$  майже напевно. Кваліфікатор «майже напевно», відповідно до Визначення 2.5.5, означає, що величина може набувати від'ємних значень на множині міри нуль (прообразами від'ємних чисел є множини міри нуль). Із дільшого викладу стане зрозуміло, що для підрахунку інтегралу множини міри нуль не становлять жодної цінності, тому ми даємо собі право послабити вимоги до цієї величини на множинах міри нуль.

Тоді інтеграл визначають так:

$$\int X d\mathbb{P} = \sup \left\{ \int Y d\mathbb{P}, Y \text{ проста}, Y \leq X \right\} . \quad (6.2.4)$$

Неформально, інтеграл — це границя інтегралів простих функцій, які не перевищують  $X$ . При цьому значення цього інтегралу може бути нескінченно великим.

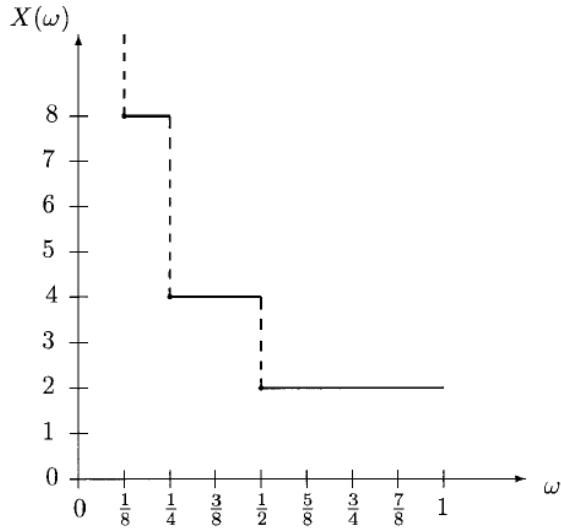


Рис. 6.2.1.: Функція ймовірності для випадкової величини з Прикладу 6.2.5 ([10], Рис. 4.2.1)

**Приклад 6.2.5.** Нехай  $X : ((0; 1], \mathcal{B}(0; 1], \lambda) \rightarrow \mathbb{R}$  така, що  $X(\omega) = 2^n$ ,  $2^{-n} \leq \omega < 2^{-(n-1)}$ ,  $n \in \mathbb{N}$  (Рис. 6.2.1). Безперечно, така величина є *ступінчастою* (step random variable), проте вона не є простою, адже набуває зліченного, а не скінченного, числа значень.

Тим не менше, інтеграл можна обчислити за визначенням. Нехай  $Y = 2^k$ ,  $2^{-k} \leq \omega < 2^{-(k-1)}$ ,  $k = 1, 2, \dots, N$ . Така величина буде простою, і до того ж не перевищуватиме  $X$ . Нескладно бачити, що й інтеграл  $Y$  також не перевищуватиме інтегралу  $X$ :

$$\int X d\lambda \geq \int Y d\lambda = \sum_{k=1}^N 2^k \cdot \lambda([2^{-k}; 2^{-(k-1)})) = \sum_{k=1}^N 2^k 2^{-k} = N$$

для будь-якого  $N \in \mathbb{N}$ . Отже, супремум нескінченний, тобто  $\int X d\lambda = \infty$ . □

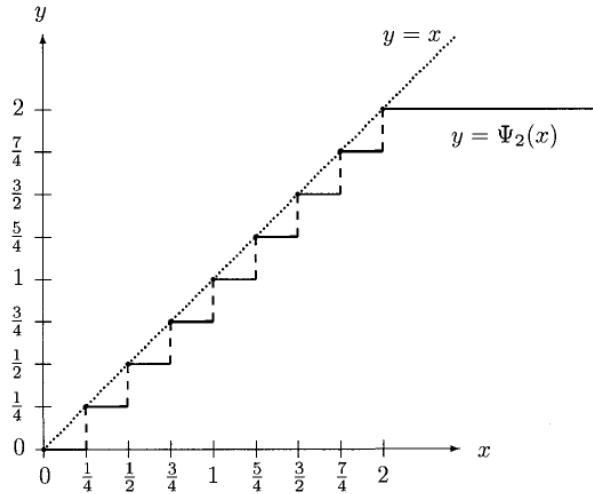
Звісно, постійно обчислювати інтеграл через супремум є задачею нетривіальною, тому на практиці використовують таку теорему.

**Теорема 6.2.6** (Теорема про монотонну збіжність (Monotone convergence theorem)). Розгляньмо деяку неспадну послідовність невід'ємних випадкових величин  $X_1, X_2, \dots$  таких, що  $X_1(\omega) \leq X_2(\omega) \leq \dots$  для всіх  $\omega$ , і  $\lim_{n \rightarrow \infty} X_n = X$  майже напевно<sup>5</sup>. Тоді

$$\lim_{n \rightarrow \infty} \int X_n d\mathbb{P} = \int \lim_{n \rightarrow \infty} X_n d\mathbb{P} = \int X d\mathbb{P}.$$

**Зауваження 6.2.7.** Умова, що послідовність випадкових величин повинна бути неспадною, є ключовою. Нехай на ймовірнісному просторі  $((0; 1], \mathcal{B}(0; 1], \lambda)$  визначено  $X_n = n \cdot \mathbb{1}_{\{\omega \in (0; 1/n)\}}$ .

<sup>5</sup>Тобто існує така вимірна множина  $N$  міри нуль,  $\mathbb{P}(N) = 0$ , що  $\lim_{n \rightarrow \infty} X_n(\omega) = X(\omega) \forall \omega \in N^c$ .

Рис. 6.2.2.: Графік функції  $\Psi_2(x)$  ([10], Рис. 4.2.4)

Тоді  $X_n \xrightarrow{n \rightarrow \infty} 0$ , але

$$\int X_n d\lambda = n \cdot \frac{1}{n} = 1$$

для всіх  $n$ .  $\square$

Нарешті, використання Теореми 6.2.6 віправдано, адже можемо показати, що будь-яку невід'ємну випадкову величину  $X$  можна подати як границю монотонно неспадної послідовності простих величин.

**Твердження 6.2.8.** Нехай  $X$  — деяка невід'ємна випадкова величина. Нехай

$$X_n = \Psi_n(X) = \min \{n, 2^{-n} \cdot \lfloor 2^n x \rfloor\},$$

де  $\lfloor r \rfloor$  — найменше ціле, що не перевищує  $r$ . Іншими словами,  $\Psi_n(x)$  являє собою округлене донизу значення  $x$ , обрізане по  $n$ . Наприклад, на Рис. 6.2.2 наведено графік  $\Psi_2(x)$ .

Тоді  $X_n \geq 0$ , а  $\lim_{n \rightarrow \infty} X_n = X$ .

*Доведення.* Нескладно бачити, що  $\Psi_1(x) \leq \Psi_2(x) \leq \dots$  для всіх  $x$  за побудовою. А для кожного фіксованого  $x \geq 0$  маємо

$$\lim_{n \rightarrow \infty} \Psi_n(x) = \lim_{n \rightarrow \infty} \min \{n, 2^{-n} \cdot \lfloor 2^n x \rfloor\} = \lim_{n \rightarrow \infty} \frac{\lfloor 2^n x \rfloor}{2^n} = \lim_{n \rightarrow \infty} \frac{2^n x - \{2^n x\}}{2^n} = x,$$

де  $\{x\}$  — дробова частина числа  $x$ .  $\square$

Відтак Теорему 6.2.6 можна застосовувати до будь-якої невід'ємної випадкової величини  $X$ .

#### 6.2.4. Крок 4. Довільні випадкові величини

Нарешті, довільну випадкову величину можна розписати як суму двох невід'ємних величин:

$$X = X^+ - X^-,$$

де

$$\begin{aligned} X^+(\omega) &= \max \{X(\omega), 0\} , \\ X^-(\omega) &= \max \{-X(\omega), 0\} . \end{aligned}$$

Тоді інтеграл визначають так:

$$\int X d\mathbb{P} = \int X^+ d\mathbb{P} - \int X^- d\mathbb{P} . \quad (6.2.5)$$

Існує така класифікація сподівань:

- якщо одночасно  $\int X^+ d\mathbb{P} < \infty$  і  $\int X^- d\mathbb{P} < \infty$ , то кажуть, що випадкова величина  $X$  інтегровна (integrable);
- якщо тільки один із  $\int X^+ d\mathbb{P}$ ,  $\int X^- d\mathbb{P}$  скінчений, то інтеграл існує, але є нескінченним;
- якщо ж одночасно  $\int X^+ d\mathbb{P} = \infty$  і  $\int X^- d\mathbb{P} = \infty$ , то кажуть, що інтеграл не існує (does not exist).

Якщо помітити, що  $X^+ + X^- = |X|$ , то умовою інтегровності випадкової величини є

$$\int X^+ d\mathbb{P} + \int X^- d\mathbb{P} = \int |X| d\mathbb{P} < \infty . \quad (6.2.6)$$

**Визначення 6.2.9.** Простір усіх інтегровних вимірних функцій  $f : (\Omega, \mathcal{A}, \mathbb{P}) \rightarrow \mathbb{R}$  називають простором  $L^1(\Omega, \mathcal{A}, \mathbb{P})$ , або просто  $L^1$ .  $\square$

Використовуючи уведене таким чином поняття інтегралу, можна розглянути деякі його властивості, які доводять у рамках інших дисциплін.

**Твердження 6.2.10.** Нехай  $X, Y \in L^1(\Omega, \mathcal{A}, \mathbb{P})$ . Тоді:

- якщо  $X \geq 0$  майже напевно, то  $\int X d\mathbb{P} \geq 0$ ;
- лінійність (linearity): для всіх  $a, b \in \mathbb{R}$ ,  $\int (aX + bY) d\mathbb{P} = a \int X d\mathbb{P} + b \int Y d\mathbb{P}$ ;
- якщо  $X \leq Y$  майже напевно, то  $\int X d\mathbb{P} \leq \int Y d\mathbb{P}$ ;
- якщо  $X = Y$  майже напевно, то  $\int X d\mathbb{P} = \int Y d\mathbb{P}$ ;
- $|\int X d\mathbb{P}| \leq \int |X| d\mathbb{P}$ .

Із властивості лінійності (ii) з Твердження 6.2.10, зокрема, стає зрозуміло, чому ми постійно уточнюємо, що деяка величина  $X$  має певну властивість майже напевно. Нехай властивість  $X$  не виконується для деякої множини  $N$  міри нуль. Тоді за властивістю лінійності

$$\begin{aligned} \int X d\mathbb{P} &= \int X(\omega) \cdot \mathbb{1}_{\{\omega \in N\}} d\mathbb{P} + \int X(\omega) \cdot \mathbb{1}_{\{\omega \notin N\}} d\mathbb{P} \\ &= \int_N X(\omega) d\mathbb{P} + \int_{N^c} X(\omega) d\mathbb{P} = \int_{N^c} X(\omega) d\mathbb{P} , \end{aligned}$$

оскільки  $\mathbb{P}(N) = 0$ .

**Приклад 6.2.11.** Будь-яку дискретну величину  $X$ , яка набуває значень  $a_1, a_2, \dots$  на множинах  $A_i \equiv \{\omega : X(\omega) = a_i\}$ ,  $i = 1, 2, \dots$ , можна подати як зліченну суму

$$X(\omega) = \sum_{i=1}^{\infty} a_i \mathbb{1}_{\{\omega \in A_i\}} .$$

Це схоже на форму запису простої функції, але такою не є, адже тут у загальному випадку маємо не скінченну суму, а ряд. Проте інтеграл у цьому випадку все одно матиме форму

$$\int X d\mathbb{P} = \sum_{i=1}^{\infty} a_i \mathbb{P}(A_i) ,$$

оскільки, з урахуванням Теореми 6.2.6,

$$\begin{aligned} \int X d\mathbb{P} &= \int X^+ d\mathbb{P} - \int X^- d\mathbb{P} \\ &= \int \sum_{i=1}^{\infty} a_i \mathbb{1}_{\{A_i \cap \{\omega : X(\omega) \geq 0\}\}} d\mathbb{P} - \int \sum_{i=1}^{\infty} (-a_i) \mathbb{1}_{\{A_i \cap \{\omega : X(\omega) \leq 0\}\}} d\mathbb{P} \\ &= \sum_{i=1}^{\infty} \int a_i \mathbb{1}_{\{A_i \cap \{\omega : X(\omega) \geq 0\}\}} d\mathbb{P} - \sum_{i=1}^{\infty} \int (-a_i) \mathbb{1}_{\{A_i \cap \{\omega : X(\omega) \leq 0\}\}} d\mathbb{P} \\ &= \sum_{i=1}^{\infty} a_i \mathbb{P}(A_i \cap \{\omega : X(\omega) \geq 0\}) + \sum_{i=1}^{\infty} a_i \mathbb{P}(A_i \cap \{\omega : X(\omega) \leq 0\}) \\ &= \sum_{i=1}^{\infty} a_i \mathbb{P}(A_i) . \end{aligned}$$

□

### 6.3. Абсолютно неперервні випадкові величини та щільність розподілу

У Розділі 6.1 ми з'ясували, що ймовірності подій за участю випадкової величини  $X$  можна обчислити безпосередньо в термінах функції розподілу  $F_X$  замість того, щоб використовувати апарат міри  $\mathbb{P}$  на просторі  $(\Omega, \mathcal{A}, \mathbb{P})$ . На практиці винятково корисними є такі випадкові величини, розподіли яких є абсолютно неперервними відносно міри Лебега.

**Визначення 6.3.1.** Нехай на вимірному просторі  $(\Omega, \mathcal{A})$  визначено дві міри,  $\mu$  і  $\nu$ . Тоді міра  $\nu$  є *абсолютно неперервною відносно  $\mu$*  (absolutely continuous with respect to  $\mu$ ), що позначають як  $\nu \ll \mu$ , якщо

$$\forall A \in \mathcal{A}, \quad \mu(A) = 0 \Rightarrow \nu(A) = 0 .$$

□

У теорії міри доводять фундаментальний результат, за яким будь-яку міру  $\nu$ , абсолютно неперервну відносно  $\mu$ , можна подати як *інтеграл* від деякої функції за мірою  $\mu$ .

**Теорема 6.3.2** (Теорема Радона-Нікодіма (Radon–Nikodym theorem)<sup>6</sup>). Якщо на вимірному просторі  $(\Omega, \mathcal{A})$  визначено дві  $\sigma$ -скінченні міри  $\mu$  і  $\nu$ ,  $\nu \ll \mu$ , тоді існує невід'ємна вимірна функція  $f : \Omega \rightarrow [0; \infty)$  така, що

$$\nu(A) = \int_A f d\mu, \quad \forall A \in \mathcal{A}. \quad (6.3.1)$$

Функцію  $f$  із Теореми 6.3.2 називають *похідною Радона-Нікодіма* (Radon-Nikodym derivative). Безпосереднє застосування цієї теореми до випадкових величин таке.

**Визначення 6.3.3.** Випадкова величина  $X$  є **абсолютно неперервна** (absolutely continuous), якщо її розподіл абсолютно неперервний відносно міри Лебега:

$$\mathbb{P}_X(A) = \int_A f_X d\lambda, \quad \forall A \in \mathcal{B}.$$

□

З урахуванням (6.1.1), маємо:

$$F_X(x) = \mathbb{P}_X((-\infty; x]) = \int_{(-\infty; x]} f_X d\lambda. \quad (6.3.2)$$

**Визначення 6.3.4.** Функція  $f_X$  із (6.3.2) має назву *щільноти розподілу* (probability density function, PDF). □

Варто зазначити, що термін абсолютно неперервний тут доречний також і тому, що функція розподілу  $F_X$ , яку можна подати через щільність (6.3.2), є абсолютно неперервною в класичному розумінні, тобто

$$\forall \varepsilon > 0 \quad \exists \delta > 0 \quad \forall [a_i; b_i] \subset \mathbb{R}, i = 1, \dots, k, \quad \sum_{i=1}^k (b_i - a_i) < \varepsilon \Rightarrow \sum_{i=1}^k |F_X(b_i) - F_X(a_i)| < \varepsilon.$$

Ми цього доводити не будемо.

Як відомо з курсу «Математичний аналіз», якщо функція інтегровна за Ріманом<sup>7</sup> (Riemann integrable), то вона інтегровна за Лебегом. Усі корисні на практиці щільноті розподілу є інтегровні за Ріманом, а тому для таких абсолютно неперервних величин справедливо

$$F_X(x) = \int_{(-\infty; x]} f_X d\lambda = \int_{-\infty}^x f_X(t) dt,$$

де останній інтеграл є стандартним інтегралом Рімана. Іншими словами,

$$f_X(x) = F'_X(x), \quad (6.3.3)$$

<sup>6</sup>Теорему названо на честь австрійського математика Йоганна Карла Августа Радона, 1887–1956), який уперше довів цю теорему для простору  $\mathbb{R}^n$  у 1913 р., та польського математика Отто Марчина Нікодіма (Otto Marcin Nikodym, 1887–1974), який узагальнив її для будь-якого вимірного простору в 1930 р.

<sup>7</sup>Георг Фридрих Бернгард Ріман (Georg Friedrich Bernhard Riemann, 1826–1866) — відомий німецький математик.

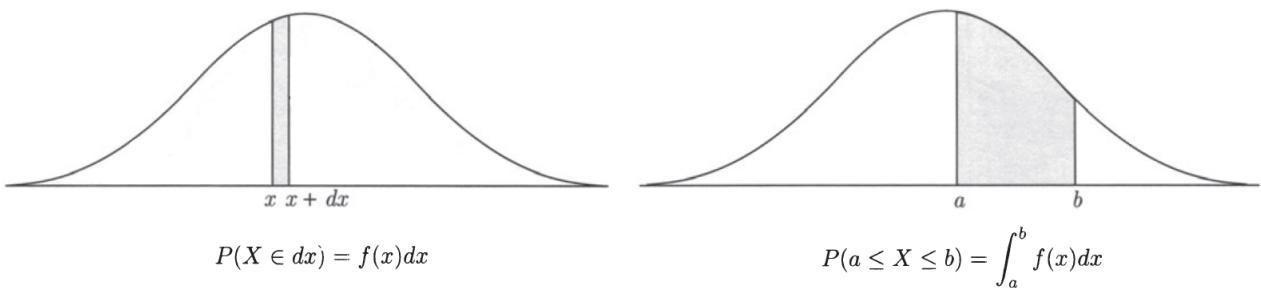


Рис. 6.3.1.: Ілюстрація обчислення ймовірностей за допомогою щільностей розподілів ([9, с. 263])

для всіх точок  $x$ , у яких функція  $F_X$  диференційовна.

**Зауваження 6.3.5.** Існують, звісно, і такі щільності розподілу, які не є інтегровні за Ріманом. Наприклад, функція Діріхле, обмежена пів інтервалом  $(0; 1]$ ,

$$f(x) = \mathbb{1}\{x \in \mathbb{Q} \cap (0; 1]\}$$

є щільністю для деякого (мало цікавого) розподілу. Але, як відомо, вона не є інтегровною за Ріманом, хоч  $\int f d\lambda = 0$  як інтеграл простої функції:

$$\int f d\lambda = 1 \cdot \lambda(\mathbb{Q} \cap (0; 1]) + 0 \cdot \lambda((\mathbb{R} \setminus \mathbb{Q}) \cap (0; 1]) = 1 \cdot 0 + 0 \cdot 1 = 0.$$

□

Далі в цьому курсі будемо вважати, що щільності розподілу абсолютно неперервних випадкових величин інтегровні за Ріманом.

Оскільки абсолютно неперервні функції розподілу є неперервні, то для них справедливо узагальнення властивостей із Твердження 6.1.5.

**Твердження 6.3.6.** Нехай  $F : \mathbb{R} \rightarrow \mathbb{R}$  — функція розподілу деякої абсолютно неперервної випадкової величини  $X$ . Тоді:

- (i)  $\mathbb{P}_X(X \geq x) = 1 - F(x);$
- (ii)  $\mathbb{P}_X((a; b]) = \mathbb{P}_X([a; b)) = \mathbb{P}_X((a; b)) = \mathbb{P}_X([a; b]) = F(b) - F(a) = \int_a^b f(x) dx, a, b \in \mathbb{R};$
- (iii)  $\mathbb{P}_X(X = x) = 0.$

*Доведення.* Усі ці результати очевидні через те, що  $F(x) = F(x-)$  для всіх  $x \in \mathbb{R}$ . □

Як стане зрозуміло нижче, щільність розподілу дуже подібна до функції ймовірності з Визначення 4.2.3, проте її інтерпретація трішки інша. Для розподілів із неперервною функцією розподілу  $\mathbb{P}_X(X = x) = 0$ , проте щільність розподілу дас змогу визначити імовірність потрапляння в деякий нескінченно малий інтервал:  $\mathbb{P}_X(X \in (x; x + dx)) = f(x)dx$  (Рис. 6.3.1). Варто пам'ятати, що  $f(x)$  сама по собі не є ймовірністю (вона навіть може перевищувати 1).

Для абсолютно неперервних випадкових величин можна уточнити поняття носія.

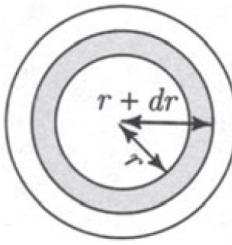


Рис. 6.3.2.: Ілюстрація до Прикладу 6.3.8 ([9, с. 270])

**Визначення 6.3.7.** Носієм (support) абсолютно неперервної випадкової величини  $X$  зі щільністю  $f_X(x)$  називають множину  $\text{supp}(X) = \{x \in \mathbb{R} : f_X(x) > 0\}$ .  $\square$

**Приклад 6.3.8.** Розглянемо випадкову величину  $R$  = «відстань від центра одиничного кола до випадкової точки, що належить одиничному кругу» (Рис. 6.3.2). Імовірність потрапляння випадкової точки  $x$  у круг радіуса  $r$ , згідно з геометричною інтерпретацією ймовірності, дорівнює відношенню його площини до площини всього простору (одиничного круга):  $\mathbb{P}_R(R \leq r) = \frac{\pi r^2}{\pi} = r^2$ . Імовірність потрапляння точки в кільце між кругами радіусів  $r_1 < r_2 \leq 1$  дорівнює

$$\mathbb{P}_R(r_1 \leq R \leq r_2) = r_2^2 - r_1^2.$$

Щільність розподілу дає ймовірність потрапляння точки в кільце нескінченно малого радіуса  $dr$ :

$$f_R(r)dr = \mathbb{P}_R(r \leq R \leq r + dr) = (r + dr)^2 - r^2 = 2rdr + (dr)^2.$$

Нехтуючи доданком  $(dr)^2$ , маємо:

$$f_R(r) = 2r \cdot \mathbb{1}\{0 < r \leq 1\}.$$

Носієм  $R$  є  $\text{supp}(R) = (0; 1]$ .  $\square$

Будь-яка щільність має такі ж властивості, як і функція ймовірності: вона невід'ємна, а сума (інтеграл) усіх значень дорівнює 1.

**Твердження 6.3.9.** Будь-яка функція  $f : \mathbb{R} \rightarrow [0; \infty)$  є щільністю для деякого розподілу тоді й тільки тоді, коли

$$\int_{-\infty}^{\infty} f(t) dt = F(\infty) = 1.$$

*Доведення.* Щільність розподілу невід'ємна за визначенням, а

$$\int_{-\infty}^{\infty} f(t) dt = F(\infty) = 1$$

за Твердженням 6.1.2.

Доведімо достатність умов. Оскільки інтеграл функції  $f$  дорівнює 1, вона є інтегровною. Тоді ця функція задає деякий розподіл:

$$\mathbb{P}_f(A) = \int_A f d\lambda, \quad A \in \mathcal{B}.$$

Можна формально покласти  $X : (\mathbb{R}, \mathcal{B}, \mathbb{P}_f) \rightarrow (\mathbb{R}, \mathcal{B})$  як  $X(\omega) = \omega$ . Тоді, за Визначенням 4.1.6,

$$\mathbb{P}_X(A) = \mathbb{P}_f(\{\omega : X(\omega) \in A\}) = \mathbb{P}_f(\{\omega \in A\}) = \mathbb{P}_f(A) = \int_A f d\lambda.$$

Отже розподіл величини  $X$  має щільність  $f$ .  $\square$

**Приклад 6.3.10.** Нехай випадкова величина  $X$  має щільність

$$f(x) = C(4x - 2x^2) \cdot \mathbb{1}\{0 < x < 2\}.$$

Значення константи  $C$  можна встановити з того факту, що інтеграл щільності дорівнює 1:

$$\int_{-\infty}^{\infty} C(4x - 2x^2) \cdot \mathbb{1}\{0 < x < 2\} dx = \int_0^2 C(4x - 2x^2) dx = C \left( 2x^2 - \frac{2x^3}{3} \right) \Big|_0^2 = 1,$$

звідки  $C = 3/8$ .  $\square$

**Зауваження 6.3.11.** Варто зазначити, що всі щільності розподілу еквівалентні з точністю до рівності майже напевно. Справді, за Твердженням 6.2.10, якщо функції  $f = g$  майже напевно, то  $\int_A f d\lambda = \int_A g d\lambda$ . Таким чином, наприклад

$$\int_A f d\lambda = \int_A (f + \mathbb{1}\{x \in \mathbb{Q} \cap (0; 1]\}) d\lambda,$$

хоча функція справа взагалі не є інтегровною за Ріманом.  $\square$

На перший погляд може здатися, що дискретні і абсолютно неперервні випадкові величини являють собою різні види випадкових величин. Насправді, дискретні випадкові величини також є абсолютно неперервними.

**Зауваження 6.3.12.** Як би це не було на перший погляд парадоксально, але дискретні випадкові величини є частковим випадком абсолютно неперервних величин. Єдина різниця полягає в тому, що замість міри Лебега потрібно використати лічну міру з Прикладу 2.1.4. Але оскільки лічна міра для незліченного простору (яким є  $\mathbb{R}$ ) не є  $\sigma$ -скінченною (тому для неї незастосовна Теорема 6.3.2), її потрібно трішки модифікувати для нашого випадку.

Нехай випадкова величина  $X : (\Omega, \mathcal{A}, \mathbb{P}) \rightarrow (\mathbb{R}, \mathcal{B}, \#)$  є дискретною з носієм  $X_0 = \{x_1, x_2, \dots\}$ . Тоді  $\#$  визначмо як

$$\#(B) = |X_0 \cap B|, \quad B \in \mathcal{B}.$$

Можна показати, що це є міра в розумінні Визначення 2.1.2, і що вона  $\sigma$ -скінчена.

Покажімо, що розподіл  $\mathbb{P}_X$  величини  $X$  є абсолютно неперервним відносно лічної міри. Для цього потрібно показати, що  $\#(B) = 0 \Rightarrow \mathbb{P}_X(B) = 0$ . Справді, якщо  $\#(B) = 0$ , то це означає, що  $X_0 \cap B = \emptyset$ , тобто  $B \subseteq X_0^c$ . Відтак,

$$\mathbb{P}_X(X \in B) = \mathbb{P}(\{\omega : X(\omega) \in B\}) \leq \mathbb{P}(\{\omega : X(\omega) \in X_0^c\}) = 1 - \mathbb{P}(\{\omega : X(\omega) \in X_0\}) = 1 - 1 = 0.$$

Отже, за Теоремою 6.3.2,

$$\mathbb{P}_X(X \in B) = \int_B f d\#,$$

для деякої невід'ємної функції  $f$ . Ключовим результатом є те, що функція  $f$  — це і є функція ймовірності  $p_X$ . Справді, за (4.2.2),

$$\mathbb{P}_X(X \in B) = \sum_{x \in (X_0 \cap B)} p_X(x).$$

Вищенаведений інтеграл є нічим іншим, як інтегралом від дискретної величини  $f : (\mathbb{R}, \mathcal{B}, \#) \rightarrow (\mathbb{R}, \mathcal{B})$  такої, що

$$f(x) = \begin{cases} p_X(x), & x \in X_0 \\ 0, & x \notin X_0 \end{cases}.$$

Як і в Прикладі 6.2.11,

$$\int_B f d\# = \int f \cdot \mathbb{1}_B d\# = 0 \cdot \#(X_0^c \cap B) + \sum_{x \in (X_0 \cap B)} f(x) \#(\{x\}) = \sum_{x \in (X_0 \cap B)} p_X(x).$$

Серед іншого,

$$\mathbb{P}_X(X = y) = \int_{\{y\}} f d\# = \sum_{x=y} p_X(x) = p_X(y),$$

що і є визначенням функції ймовірності.  $\square$

Отже, теорія дискретних величин є складовою частиною загальнішої теорії абсолютно неперервних величин. Більше того, загальна теорія також дає змогу працювати з випадковими величинами, які не є *ні дискретними, ні абсолютно неперервними*.

**Приклад 6.3.13** (Суміш дискретної і абсолютно неперервної величини). Нехай тривають підкидання монетки, яка випадає гербом з імовірністю  $p$ . Якщо випадає число, то гравець втрачає 1 грн, а якщо герб — то з інтервалу  $[1; 2]$  випадковим чином вибирають число, яке становить його виграш.

Нехай величина  $X$  = «виграш після одного підкидання».

У цьому випадку ймовірнісний простір є декартовим добутком  $\Omega = \Omega_1 \times \Omega_2$ , де  $\Omega_1 = \{\text{Герб, Число}\}$ , а  $Y : \Omega_2 \rightarrow \mathbb{R}$  така, що  $f_Y(y) = 1 \cdot \mathbb{1}\{y \in [1; 2]\}$  (тобто всі значення з  $[1; 2]$  рівні ймовірні).

Тоді

$$X(\omega_1, \omega_2) = \begin{cases} -1 & \omega_1 = \text{Число} \\ Y(\omega_2) & \omega_1 = \text{Герб} \end{cases}.$$

Якщо  $x \in [-1; 1]$ , то

$$F_X(x) = \mathbb{P}_X(X \leq x) = \mathbb{P}_X(X = -1) = 1 - p.$$

Якщо  $x \in [1; 2)$ , то

$$\begin{aligned}
 F_X(x) &= \mathbb{P}_X((X = -1) \cup (X \in [1; x])) = \mathbb{P}_X(X = -1) + \mathbb{P}_X(X \in [1; x]) \\
 &= 1 - p + \mathbb{P}((\omega_1 = \text{Герб}) \cap \{\omega_2 : Y(\omega_2) \leq x\}) \\
 &= 1 - p + p \cdot \mathbb{P}_Y([1; x]) \\
 &= 1 - p + p \cdot \int_{y \in [1; x]} 1 d\lambda \\
 &= 1 - p + p \cdot \int_1^x dy \\
 &= 1 - 2p + px.
 \end{aligned}$$

Нарешті,  $F_X(x) = 0$ , якщо  $x < -1$  (оскільки  $x$  ніколи не буває менша від  $-1$ ), і  $F_X(x) = 1$ , якщо  $x \geq 2$  (оскільки  $x$  завжди буває менша від  $2$ ).

Отже,

$$F_X(x) = \begin{cases} 0, & x < -1 \\ 1 - p, & x \in [-1; 1) \\ 1 - 2p + px, & x \in [1; 2) \\ 1, & x \geq 2 \end{cases}.$$

Варто звернути увагу, що ця функція не є неперервною, але вона також і не є функцією розподілу деякої дискретної величини, оскільки не є ступінчатою.  $\square$

Щоб існувала щільність розподілу, функція розподілу повинна бути саме абсолютно неперервною, а не просто неперервною.

**Приклад 6.3.14.** Розгляньмо множину Кантора з Прикладу 2.5.4 — заперечення відкритої множини

$$U = \bigcup_{i=1}^{\infty} U_i,$$

де  $U_i = \bigcup_{j=1}^{2^i-1} J_{i,j}$  — множина всіх інтервалів  $J_{i,j}$ , вирізаних після кроку  $i$ .

Розгляньмо множину

$$U_0 = (-\infty; 0) \cup U \cup (1; \infty).$$

Уведімо позначення  $J_{i,0} = (-\infty; 0)$  та  $J_{i,2^i} = (1, \infty)$ . Тоді

$$U_0 = \bigcup_{i=1}^{\infty} \bigcup_{j=0}^{2^i-1} J_{i,j}.$$

Нехай  $c_{i,j} = j/2^i$ ,  $i \geq 1$ ,  $0 \leq j \leq 2^i$ . Тоді нехай  $F_0 : U_0 \rightarrow \mathbb{R}$  така, що  $F_0(x) = c_{i,j}$ , якщо  $x \in J_{i,j}$ .

Множину значень, на якій функція стала, називають *множиною рівня* (level set). За побудовою, інтервали  $J_{i,j}$  є інтервалами рівня, оскільки  $F_0 = c_{i,j}$  на них. Частину графіку функції  $F_0$  подано на Рис. 6.3.3, де зображені множини рівня  $J_{i,j}$ ,  $1 \leq i \leq 4$ . Повний графік функції можна уявити як нескінчений набір усіх більшого числа всіх менших множин рівня, які в підсумку заповнюють собою увесь простір між зображеннями на графіку інтервалами.

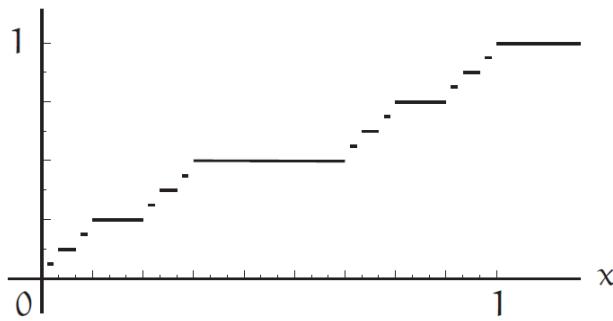


Рис. 6.3.3.: Графік множин рівня функції  $F_0(x)$  із Прикладу 6.3.14 після чотирьох ітерацій побудови множини Кантора ([7], Рис. 12.4.1)

Очевидно, що функція  $F_0$  неспадна і зростає від 0 до 1. Щоб завершити побудову функції розподілу  $F$ , ми повинні уточнити, чому дорівнює  $F$  у точках із множини кантора  $C = U_0^c$ :

$$F(x) = \inf \{F_0(t) : x < t, t \in U_0\}.$$

Така функція розподілу відповідає так званому *розподілу Кантора*. Можна довести, що така функція буде неперервною (хоча це інтуїтивно зрозуміло з її графіка). Проте така функція не є абсолютно неперервною: похідна  $F'$  існує і дорівнює 0 на всіх  $J_{i,j}$ , але оскільки  $\lambda(C) = 0$ , то виходить, що  $F' = 0$  майже напевно. Відповідно, вона не може бути щільністю розподілу<sup>8</sup>.  $\square$

На практиці найчастіше використовують абсолютно неперервні величини, тому для спрощення термінології такі величини часто називають просто неперервними.

## 6.4. Сподівання та дисперсія як інтеграли

У Розділі 4.3 ми розглядали поняття сподівання та дисперсії випадкових величин. Як ми з'ясували вище, сподівання дискретної величини як зважене середнє має своїм узагальненням сподівання будь-якої величини як інтеграл.

**Визначення 6.4.1.** *Сподіванням* (expectation) будь-якої випадкової величини  $X$  є її інтеграл:

$$\mathbb{E}[X] = \int X d\mathbb{P}. \quad (6.4.1)$$

$\square$

У Зauważенні 6.3.12 ми побачили, що інтегралом за лічною мірою від деякої функції (у нашому випадку, функції ймовірності) є сума (ряд) її значень. Відтак формула обчислення сподівання як зваженого середнього (4.3.1) є частковим випадком (6.4.1).

Оскільки на практиці ми стикаємося з доволі різними ймовірнісними мірами, за якими потрібно інтегрувати випадкові величини, було б непогано мати спосіб обчислення інтегралів шляхом зведення їх до інтегралів деякого відомого виду.

<sup>8</sup>Для кожного  $A \in \mathcal{A}$ ,  $\mathbb{P}_X(A) = 0$ , незалежно від того,  $\lambda(A) = 0$  чи ні.

**Теорема 6.4.2** (Закон несвідомого статистика (Law of unconscious statistician)). Нехай маємо деяку випадкову величину  $X : (\Omega, \mathcal{A}, \mathbb{P}) \rightarrow (\mathbb{R}, \mathcal{B})$ , яка породжує розподіл  $\mathbb{P}_X$ , абсолютно неперервний відносно деякої міри  $\mu$  зі щільністю  $f_X$ . Нехай  $g : \mathbb{R} \rightarrow \mathbb{R}$  — деяка вимірна функція, так що  $g(X) : \Omega \rightarrow \mathbb{R}$  також є випадковою величиною. Тоді

$$\int g(X(\omega)) d\mathbb{P} = \int g(x) d\mathbb{P}_X = \int g(x) \cdot f_X(x) d\mu, \quad (6.4.2)$$

у тому розумінні, що якщо існує один інтеграл, то існують усі інші, і вони дорівнюють одне одному.

Доведення цього закону здійснюють у курсі «Математичного аналізу». Альтернативне доведення можна знайти в Розд. 6.5.

**Зауваження 6.4.3.** Теорема 6.4.2 дає змогу звести обчислення інтегралу в деякому абстрактному просторі до обчислення рядів і звичних інтегралів Рімана.

Для дискретних величин розподіл завжди абсолютно неперервний відносно лічної міри, а тому

$$\mathbb{E}[g(X)] = \int g(X(\omega)) d\mathbb{P} = \int g(x) \cdot p_X(x) d\# = \sum_{x \in \text{supp}(X)} g(x) \cdot p_X(x), \quad (6.4.3)$$

де  $p_X$  — функція ймовірності. Це відповідає формулі (4.3.3).

Для (абсолютно) неперервних величин розподіл абсолютно неперервний відносно міри Лебега, а тому<sup>9</sup>

$$\mathbb{E}[g(X)] = \int g(X(\omega)) d\mathbb{P} = \int g(x) \cdot f_X(x) d\lambda = \int_{-\infty}^{\infty} g(x) \cdot f_X(x) dx, \quad (6.4.4)$$

де  $f_X$  — щільність розподілу.  $\square$

**Зауваження 6.4.4.** Окремі автори визначають сподівання не як інтеграл Лебега за мірою, а як інтеграл Стилт'єса (Stieltjes integral) за функцією розподілу  $F_X$ :

$$\mathbb{E}[g(X)] = \int g(x) d\mathbb{P}_X = \int g(x) dF_X(x),$$

який потім можна розписати, врахувавши, що  $dF_X(x) = f_X(x)dx$ . Проте для нашого способу уведення формул (6.4.3)–(6.4.4) розгляд таких інтегралів необов'язковий.  $\square$

У Визначенні 4.3.8 було уведено поняття дисперсії дискретної випадкової величини. У формулі (4.3.4) фігурують тільки сподівання. Отже, відповідне визначення залишається справедливим для випадкових величин у загальному випадку.

**Приклад 6.4.5.** Продовжмо розгляд Прикладу 6.3.8.

<sup>9</sup>Ми розглядаємо тільки щільності, інтегровні за Ріманом, хоча, як було зазначено раніше, існують і (малоцікаві) неінтегровні за Ріманом щільності.

Підрахуймо сподівання та дисперсію випадкової величини  $R$ :

$$\begin{aligned}\mathbb{E}[R] &= \int_{-\infty}^{\infty} r f_R(r) dr = \int_0^1 2r^2 dr = \frac{2}{3}, \\ \mathbb{E}[R^2] &= \int_{-\infty}^{\infty} r^2 f_R(r) dr = \int_0^1 2r^3 dr = \frac{1}{2}, \\ \text{Var}(R) &= \mathbb{E}[R^2] - (\mathbb{E}[R])^2 = \frac{1}{18}.\end{aligned}$$

Отже, середня відстань випадкової точки круга до його центра дорівнює  $2/3$ , а середній розкид відносно сподівання дорівнює  $\sqrt{1/18} \approx 0.236$ .  $\square$

Із властивостей лінійності інтегралу випливають властивості сподівання та дисперсії, які ми раніше доводили в контексті дискретних випадкових величин.

**Твердження 6.4.6.** Для сподівання  $\mathbb{E}[X]$  та дисперсії  $\text{Var}(X)$  деякої випадкової величини  $X$ , якщо вони існують, справедливі такі властивості:

- $\mathbb{E}[aX + bY] = a\mathbb{E}[X] + b\mathbb{E}[Y]$ ;
- $\text{Var}(aX + b) = a^2\text{Var}(X)$ ;
- якщо  $\text{Var}(X) = 0$ , то  $X = c$ ,  $c \in \mathbb{R}$ , майже напевно.

Безперечно, існують такі розподіли, сподівання яких є нескінченим, а є й такі, сподівання яких узагалі не існує.

**Приклад 6.4.7.** Нехай випадкова величина  $X$  має щільність

$$f(x) = \frac{1}{(1+x)^2} \cdot \mathbb{1}\{x > 0\}.$$

Тоді її сподівання дорівнює

$$\mathbb{E}[X] = \int_{-\infty}^{\infty} \frac{x}{(1+x)^2} \mathbb{1}\{x > 0\} dx = \int_0^{\infty} \left( \frac{1}{1+x} - \frac{1}{(1+x)^2} \right) dx = \left( \ln(1+x) + \frac{1}{1+x} \right) \Big|_0^{\infty} = \infty.$$

$\square$

**Приклад 6.4.8.** Розгляньмо відомий на практиці *стандартний розподіл Коши* (standard Cauchy distribution)<sup>10</sup>, який має щільність

$$f(x) = \frac{1}{\pi(1+x^2)}, \quad x \in \mathbb{R}.$$

Тоді

$$\mathbb{E}[X] = \frac{1}{\pi} \int_{-\infty}^{\infty} \frac{x}{1+x^2} dx = \frac{1}{2\pi} \ln(1+x^2) \Big|_{-\infty}^{\infty} = \frac{1}{2\pi}(\infty - \infty),$$

тобто маємо невизначеність, а тому сподівання взагалі не існує.

$\square$

<sup>10</sup>Названий на честь французького математика Огюстена-Луї Коши (Augustin-Louis Cauchy, 1789–1857).

## 6.5. Доведення окремих тверджень

Доведімо Теорему 6.1.6.

*Доведення.* Частину з існуванням міри ми доводити не будемо, оскільки педагогічно в цьому немає нічого цікавого. Доведення базується на застосуванні Теореми 2.4.1, для чого потрібно перевірити, що  $\nu$  має всі властивості міри на алгебрі зі скінченних об'єднань і заперечень пів інтервалів. Це технічна вправа, яка не сильно відрізняється від перевірки властивостей міри для довжин пів інтервалів, яку ми виконували під час уведення міри Лебега. (Власне кажучи, поклавши  $F(x) = x$ , ми і отримаємо міру Лебега.)

Що ж стосується існування змінної  $X$ , то з  $\lim_{x \rightarrow -\infty} F(x) = 0$  випливає  $\nu(-\infty; x] = F(x) - \lim_{x \rightarrow -\infty} F(x) = F(x)$ , а з  $\lim_{x \rightarrow \infty} F(x) = 1$  випливає  $\nu(\mathbb{R}) = 1$ . Отже, достатньо покласти  $(\Omega, \mathcal{A}, \mathbb{P}) = (\mathbb{R}, \mathcal{B}, \nu)$  і визначити випадкову величину  $X : (\mathbb{R}, \mathcal{B}, \nu) \rightarrow (\mathbb{R}, \mathcal{B})$  як  $X(\omega) = \omega$ . Тоді

$$\mathbb{P}_X(X \leq x) = \nu(\{\omega \in \mathbb{R} : \omega \leq x\}) = F(x).$$

□

Доведімо закон несвідомого статистика з Теореми 6.4.2.

*Доведення.* Доведення виконуватимемо за кроками побудови інтегралу.

**Крок 1. Індикаторні величини** Нехай  $g(x) = \mathbb{1}\{x \in A\}$ ,  $A \in \mathcal{B}$ . Оскільки  $X(\omega) = x \in A$  означає, що  $\omega \in X^{-1}(A)$ , маємо, що  $g(X(\omega)) = \mathbb{1}\{\omega \in X^{-1}(A)\}$ . Тоді з урахуванням (6.2.1) маємо:

$$\begin{aligned} \int g(X(\omega)) d\mathbb{P} &= \int \mathbb{1}\{\omega \in X^{-1}(A)\} d\mathbb{P} = \int_{X^{-1}(A)} 1 d\mathbb{P} = \mathbb{P}(X^{-1}(A)) \\ &= \mathbb{P}_X(A) = \int_A 1 d\mathbb{P}_X = \int \mathbb{1}\{x \in A\} d\mathbb{P}_X = \int g(x) d\mathbb{P}_X. \end{aligned}$$

Друга частина рівності випливає безпосередньо з Теореми 6.3.2:

$$\int g(x) d\mathbb{P}_X = \int_A d\mathbb{P}_X = \mathbb{P}_X(A) = \int_A f_X(x) d\mu = \int g(x) \cdot f_X(x) d\mu.$$

**Крок 2. Прості випадкові величини** Нехай  $g(x)$  — проста функція:

$$g(x) = \sum_{i=1}^n a_i \mathbb{1}\{x \in A_i\}, \quad a_i > 0, i = 1, \dots, n.$$

За аналогією з попереднім випадком,

$$g(X(\omega)) = \sum_{i=1}^n a_i \mathbb{1}\{\omega \in X^{-1}(A_i)\}, \quad a_i > 0, i = 1, \dots, n.$$

Тоді, використовуючи (6.2.3), властивість лінійності інтегралу, та з урахуванням попереднього

кроку маємо:

$$\begin{aligned} \int g(X(\omega)) d\mathbb{P} &= \sum_{i=1}^n a_i \int \mathbb{1}\{\omega \in X^{-1}(A_i)\} d\mathbb{P} \\ &= \sum_{i=1}^n a_i \int \mathbb{1}\{x \in A_i\} d\mathbb{P}_X = \int \left( \sum_{i=1}^n a_i \mathbb{1}\{x \in A_i\} \right) d\mathbb{P}_X = \int g(x) d\mathbb{P}_X . \end{aligned}$$

Що стосується другої частини рівності, то

$$\begin{aligned} \int g(x) d\mathbb{P}_X &= \int \left( \sum_{i=1}^n a_i \mathbb{1}\{x \in A_i\} \right) d\mathbb{P}_X \\ &= \sum_{i=1}^n a_i \int \mathbb{1}\{x \in A_i\} d\mathbb{P}_X = \sum_{i=1}^n a_i \int_{A_i} d\mathbb{P}_X = \sum_{i=1}^n a_i \mathbb{P}_X(A_i) \\ &= \sum_{i=1}^n a_i \int_{A_i} f_X(x) d\mu = \sum_{i=1}^n a_i \int f_X(x) \cdot \mathbb{1}\{x \in A_i\} d\mu = \int f_X(x) \sum_{i=1}^n a_i \mathbb{1}\{x \in A_i\} d\mu \\ &= \int g(x) \cdot f_X(x) d\mu . \end{aligned}$$

**Крок 3. Невід'ємні випадкові величини** Якщо  $g(x)$  — довільна невід'ємна величина, то відповідно до Твердження 6.2.8 існує неспадна послідовність  $\{g_n(x)\}$  невід'ємних функцій, кожна з яких є простою:

$$g_n(x) = \sum_{i=1}^{r_n} a_{ni} \mathbb{1}\{x \in A_{ni}\} , \quad a_{ni} \geq 0 , \quad i = 1, \dots, r_n ,$$

а  $\lim_{n \rightarrow \infty} g_n(x) = g(x)$  майже напевно. За Теоремою 6.2.6,

$$\int g(x) d\mathbb{P}_X = \lim_{n \rightarrow \infty} \int g_n(x) d\mathbb{P}_X .$$

Аналогічно попереднім випадкам,

$$g_n(X(\omega)) = \sum_{i=1}^{r_n} a_{ni} \mathbb{1}\{\omega \in X^{-1}(A_{ni})\} ,$$

і до того ж

$$\lim_{n \rightarrow \infty} g_n(X(\omega)) = g(X(\omega))$$

майже напевно. Відтак, за Теоремою 6.2.6,

$$\int g(X(\omega)) d\mathbb{P} = \lim_{n \rightarrow \infty} \int g_n(X(\omega)) d\mathbb{P} .$$

Більше того, із попереднього кроку безпосередньо випливає, що для кожного  $n$

$$\int g_n(X(\omega)) d\mathbb{P} = \int g_n(x) d\mathbb{P}_X .$$

Разом усі ці твердження дають бажаний результат:

$$\int g(X(\omega)) d\mathbb{P} = \int g(x) d\mathbb{P}_X .$$

Що стосується другої частини рівності, то для кожного  $g_n(x)$  справедливо, за попереднім кроком:

$$\int g_n(x) d\mathbb{P}_X = \int g_n(x) \cdot f_X(x) d\mu .$$

З іншого боку, якщо  $g_n(x) \rightarrow_{n \rightarrow \infty} g(x)$  майже напевно, то і  $g_n(x) \cdot f_X(x) \rightarrow_{n \rightarrow \infty} g(x) \cdot f_X(x)$  майже напевно. Звідси випливає

$$\int g(x) d\mathbb{P}_X = \lim_{n \rightarrow \infty} \int g_n(x) d\mathbb{P}_X = \lim_{n \rightarrow \infty} \int g_n(x) \cdot f_X(x) d\mu = \int g(x) \cdot f_X(x) d\mu .$$

**Крок 4. Довільні випадкові величини** Нарешті, нехай  $g(x) = g^+(x) - g^-(x)$ . Тоді, відповідно, і  $g(X(\omega)) = g^+(X(\omega)) - g^-(X(\omega))$ . Якщо інтеграл  $\int g(X(\omega)) d\mathbb{P}$  існує<sup>11</sup>, то принаймні один із інтегралів

$$\int g^+(X(\omega)) d\mathbb{P} , \quad \int g^-(X(\omega)) d\mathbb{P}$$

скінчений. За попереднім кроком,

$$\int g^+(X(\omega)) d\mathbb{P} = \int g^+(x) d\mathbb{P}_X , \quad \int g^-(X(\omega)) d\mathbb{P}_X = \int g^-(x) d\mathbb{P}_X ,$$

і тому принаймні один із інтегралів

$$\int g^+(x) d\mathbb{P}_X , \quad \int g^-(x) d\mathbb{P}_X$$

скінчений. Тоді

$$\begin{aligned} \int g(X(\omega)) d\mathbb{P} &= \int g^+(X(\omega)) d\mathbb{P} - \int g^-(X(\omega)) d\mathbb{P} \\ &= \int g^+(x) d\mathbb{P}_X - \int g^-(x) d\mathbb{P}_X \\ &= \int g(x) d\mathbb{P}_X . \end{aligned}$$

Аналогічного результата можна досягти, поклавши спочатку існування інтегралу  $\int g(x) d\mathbb{P}_X$ . Таким чином, існування одного інтегралу має наслідком існування іншого, і вони дорівнюють один одному.

Абсолютно аналогічно можна довести і другу частину рівності.  $\square$

<sup>11</sup>При цьому він може бути нескінчений.

## 7. Функції від випадкових величин. Деякі неперервні розподіли

У минулій лекції ми розглянули визначення випадкової величини в найзагальнішому випадку та розглянули функції та щільності розподілів, за допомогою яких можна компактно описувати розподіли випадкових величин. У поточній лекції ми розглянемо декілька поширеніших на практиці неперервних розподілів та дізнаємося, як утворити розподіл деякої функції від випадкової величини.

### 7.1. Стандартний рівномірний розподіл

Найпростіший неперервний розподіл, який розглянемо — це *рівномірний розподіл* (uniform distribution) на  $(0; 1]$ . Точніше, цей розподіл ми вже розглядали, і він має назву міри Лебега<sup>1</sup>. У цьому сенсі рівномірний розподіл є безпосереднім аналогом класичної ймовірності у випадку незліченного простору.

**Визначення 7.1.1.** Випадкова величина  $X$  має *стандартний* (standard) рівномірний розподіл,  $X \sim U((0; 1])$ , якщо її щільність розподілу дорівнює

$$f_X(x) = \begin{cases} 1, & x \in (0; 1] \\ 0, & x \notin (0; 1] \end{cases} = \mathbb{1}\{x \in (0; 1]\} . \quad (7.1.1)$$

□

Функція рівномірного розподілу має такий вид:

$$F_X(x) = \int_{(-\infty; x]} \mathbb{1}\{x \in (0; 1]\} d\lambda = \int_{-\infty}^x \mathbb{1}\{t \in (0; 1]\} dt = \begin{cases} 0, & x < 0 \\ x, & 0 \leq x < 1 \\ 1, & x \geq 1 \end{cases} . \quad (7.1.2)$$

Графіки щільності та функції стандартного рівномірного розподілу наведено на Рис. 7.1.1.

<sup>1</sup>Оскільки це абсолютно неперервний розподіл, значення щільності в одній точці ні на що не впливає, тому можна сказати, що рівномірний розподіл задано на замкненому інтервалі  $[0; 1]$  або ж на відкритому інтервалі  $(0; 1)$ . Проте ми будемо формально класифікувати, що носієм є пів інтервал із тих міркувань, що саме на такому просторі ми визначили міру Лебега.

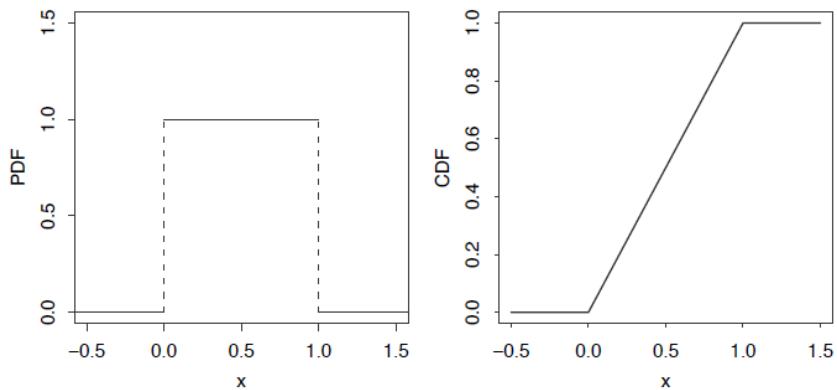


Рис. 7.1.1.: Щільність та функція стандартного рівномірного розподілу на  $(0; 1]$  ([1], Рис. 5.6)

**Приклад 7.1.2.** Нехай  $X \sim U((0; 1])$ . Тоді

$$\begin{aligned}\mathbb{P}_X(X \leq 0.3) &= \int_0^{0.3} 1 dx = F_X(0.3) = 0.3, \\ \mathbb{P}_X(X \geq 0.6) &= 1 - \int_0^{0.6} 1 dx = 1 - F_X(0.6) = 0.4, \\ \mathbb{P}_X(0.3 \leq X \leq 0.8) &= F_X(0.8) - F_X(0.3) = 0.5.\end{aligned}$$

□

Підрахуймо сподівання та дисперсію стандартного рівномірного розподілу.

**Твердження 7.1.3.** Якщо  $X \sim U((0; 1])$ , то  $\mathbb{E}[X] = \frac{1}{2}$ , а  $\text{Var}(X) = \frac{1}{12}$ .

*Доведення.* Обчислення здійснюватимемо за формулами (6.4.4) та (4.3.6):

$$\begin{aligned}\mathbb{E}[X] &= \int_{-\infty}^{\infty} x \cdot \mathbb{1}\{x \in (0; 1]\} dx = \int_0^1 x \cdot 1 dx = \frac{1}{2}, \\ \mathbb{E}[X^2] &= \int_{-\infty}^{\infty} x^2 \cdot \mathbb{1}\{x \in (0; 1]\} dx = \int_0^1 x^2 \cdot 1 dx = \frac{1}{3}, \\ \text{Var}(X) &= \mathbb{E}[X^2] - (\mathbb{E}[X])^2 = \frac{1}{3} - \frac{1}{4} = \frac{1}{12}.\end{aligned}$$

Це інтуїтивно зрозуміло, адже якщо щільність є сталою на інтервалі  $(0; 1]$ , то «середнім значенням» повинна бути середина інтервалу.

## 7.2. Стандартний нормальний розподіл

Без перебільшення, *найбільш важливим розподілом у теорії ймовірностей та статистиці є стандартний нормальний розподіл* (standard normal distribution)<sup>2</sup>. Його важливість стане зро-

<sup>2</sup>Нормальний розподіл було вперше уведено французьким математиком Авраамом де Муавром (Abraham de Moivre, 1667–1754) у 1733 р. як спосіб апроксимації ймовірностей для біномного розподілу з дуже великою кількістю випробувань  $n$ .

зумілою, коли ми розглянемо центральну граничну теорему, яка дає теоретичне обґрунтування емпіричному спостереженню, що значна кількість випадкових явищ має приблизно нормальній розподіл, принаймні асимптотично. Корисність розподілу величезна, адже ціла низка реальних явищ у навколошньому світі має нормальній розподіл, наприклад, зріст людини, швидкість молекули газу в будь-якому напрямку, похибка вимірювання деякої фізичної величини тощо.

Часто цей розподіл також називають *Гауссевим розподілом* (Gaussian distribution), оскільки німецький математик Карл Фридрих Гаусс (Karl Friedrich Gauss, 1777–1855) використовував його для прогнозування розташування астрономічних об'єктів. Із середини XIX ст. статистики почали помічати, що гістограми для більшості наборів даних мають форму, яка відповідає щільності Гауссової розподілу, і тому за порадою британського статистики Карла Пірсона (Karl Pearson, 1857–1936) цей розподіл стали називати нормальним.

**Визначення 7.2.1.** Випадкова величина  $Z$ <sup>3</sup> має стандартний нормальній розподіл,  $Z \sim N(0, 1)$ , якщо її щільність розподілу має вид

$$\phi(z) = \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}}, \quad z \in \mathbb{R}. \quad (7.2.1)$$

□

Позначення для цієї щільноти через  $\phi$  замість  $f$  обумовлено історично.

Як можна бачити, щільність є парною функцією:  $\phi(-z) = \phi(z)$ , і до того ж

$$\phi'(z) = \left( \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}} \right)' = -\frac{z}{\sqrt{2\pi}} e^{-\frac{z^2}{2}} = -z\phi(z).$$

**Твердження 7.2.2.** Функція (7.2.1) відповідає вимогам до щільностей розподілу, викладеним у Твердженні 6.3.9.

**Доведення.** Очевидно, що щільність є додатною для всіх  $z \in \mathbb{R}$ . Тому потрібно показати, що її інтеграл дорівнює 1. Підінтегральна функція доволі складна для інтегрування, тому використаємо такий прийом із переходом до полярних координат  $x = r \cos \theta$ ,  $y = r \sin \theta$ :

$$\begin{aligned} \left( \int_{-\infty}^{\infty} e^{-\frac{x^2}{2}} dx \right) \cdot \left( \int_{-\infty}^{\infty} e^{-\frac{y^2}{2}} dy \right) &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} e^{-\frac{x^2+y^2}{2}} dx dy \\ &= \int_0^{2\pi} \int_0^{\infty} e^{-\frac{r^2}{2}} r dr d\theta \\ &= \int_0^{2\pi} \int_0^{\infty} e^{-u} du d\theta \\ &= \int_0^{2\pi} - \left( \lim_{u \rightarrow \infty} e^{-u} - e^0 \right) dr d\theta = \int_0^{2\pi} d\theta = 2\pi. \end{aligned}$$

Звідси випливає результат

$$\int_{-\infty}^{\infty} e^{-\frac{z^2}{2}} dz = \sqrt{2\pi},$$

<sup>3</sup>Таке позначення стандартної нормальної величини замість  $X$  обумовлено історично, і ми його дотримуватимемося.

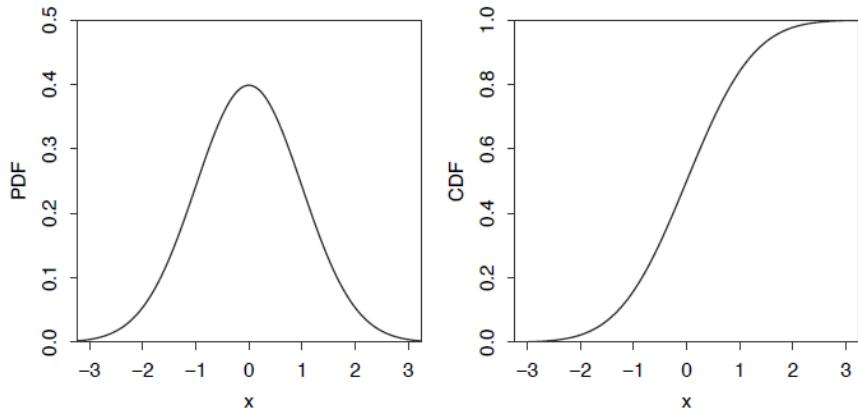


Рис. 7.2.1.: Щільність та функція стандартного нормального розподілу на  $(0; 1]$  ([1], Рис. 5.11)

а отже

$$\int_{-\infty}^{\infty} \phi(z) dz = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-\frac{z^2}{2}} dz = 1.$$

□

Варто зазначити, що функція стандартного нормального розподілу

$$\Phi(z) = \int_{-\infty}^z \phi(t) dt = \int_{-\infty}^z \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}} dt \quad (7.2.2)$$

не має аналітичного виразу. Її значення можна підрахувати тільки чисельно<sup>4</sup>. Позначення цієї функції через  $\Phi$  замість  $F$  також обумовлено історично.

Графіки щільності та функції стандартного нормального розподілу наведено на Рис. 7.2.1. Нескладно бачити, і це можна перевірити аналітично, що  $\Phi$  — строго зростаюча функція.

Для функції розподілу справедливі деякі корисні властивості.

**Твердження 7.2.3.** Функція стандартного нормального розподілу  $\Phi$  має такі властивості:

- (i)  $\Phi(z) = 1 - \Phi(-z)$ ;
- (ii)  $\Phi^{-1}(p) = -\Phi^{-1}(1 - p)$ ,  $p \in (0; 1)$ .

*Доведення.* (i) Використовуючи той факт, що  $\phi$  — парна функція, маємо

$$\Phi(-z) = \int_{-\infty}^{-z} \phi(t) dt = \int_z^{\infty} \phi(u) du = 1 - \int_{-\infty}^z \phi(u) du = 1 - \Phi(z);$$

(ii) оскільки  $\Phi$  — строго зростаюча функція, обернена до неї функція існує і також є строго зростаюча. Нехай  $p = \Phi(z)$ , тоді

$$z = \Phi^{-1}(p).$$

З іншого боку,

$$\Phi^{-1}(1 - p) = \Phi^{-1}(1 - \Phi(z)) = \Phi^{-1}(\Phi(-z)) = -z.$$

<sup>4</sup>До появи потужних і дешевих комп'ютерів значення цієї функції табулювали, проте нині використання таких таблиць подібне на застосування логаритмічної лінійки для підрахунку добутків великих чисел.

Оскільки  $z = -(-z)$ , маємо  $\Phi^{-1}(p) = -\Phi^{-1}(1 - p)$ .

□

**Приклад 7.2.4.** Нехай мережею потрібно передати бінарне повідомлення — деяке двійкове число  $x \in \{0, 1\}$ . Нехай розподіл завади в каналі на основі практичних спостережень визначили як стандартний нормальній:  $N \sim N(0, 1)$ . Тоді приймач на вході отримує значення  $R = x + N$ . Він декодує повідомлення так: якщо  $R \geq 0.5$ , вважають, що було послано 1, а якщо  $R < 0.5$ , що 0. Розгляньмо ймовірності помилок під час пересилання такого повідомлення.

Сигнал 1 може бути неправильно розпізнано як 0 у випадку, коли  $x = 1$ , а  $x + N = 1 + N < 0.5$ . Тоді

$$\mathbb{P}(\text{«помилка при розпізнаванні 1»}) = \mathbb{P}_N(1 + N < 0.5) = \Phi(-0.5) \approx 0.3085.$$

У схожий спосіб сигнал 0 може бути неправильно розпізнано як 1 у випадку, коли  $x = 0$ , а  $x + N = 0 + N \geq 0.5$ . Тоді

$$\mathbb{P}(\text{«помилка при розпізнаванні 0»}) = \mathbb{P}_N(0 + N \geq 0.5) = 1 - \Phi(0.5) = \Phi(-0.5) \approx 0.3085.$$

Для зменшення ймовірності помилки повідомлення кодують так, що замість 0 посилають  $-2$ , а замість 1 посилають 2. У цьому випадку матимемо відповідники ймовірностей помилок:

$$\mathbb{P}(\text{«помилка при розпізнаванні 1»}) = \mathbb{P}_N(2 + N < 0.5) = \Phi(-1.5) \approx 0.0668,$$

$$\mathbb{P}(\text{«помилка при розпізнаванні 0»}) = \mathbb{P}_N(-2 + N \geq 0.5) = 1 - \Phi(2.5) \approx 0.0062.$$

Як можна бачити, за такого кодування ймовірності помилок суттєво менші. □

Підрахуймо сподівання та дисперсію стандартного нормальногорозподілу.

**Твердження 7.2.5.** Якщо  $Z \sim N(0, 1)$ , то  $\mathbb{E}[Z] = 0$ , а  $\text{Var}(Z) = 1$ .

*Доведення.* Спочатку підрахуймо сподівання:

$$\begin{aligned} \mathbb{E}[Z] &= \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} z e^{-\frac{z^2}{2}} dz = -\frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-\frac{z^2}{2}} d\left(-\frac{z^2}{2}\right) \\ &= -\frac{1}{\sqrt{2\pi}} \left( e^{-\frac{z^2}{2}} \Big|_{-\infty}^{\infty} \right) = 0. \end{aligned}$$

Варто зазначити, що ми могли б помітити, що маємо інтеграл від непарної функції за симетричним проміжком. Проте оскільки наш проміжок нескінченно великий, ми повинні коректно перевірити, що сподівання існує. Наприклад, функція  $f(x) = 1/x$  також непарна, проте цілком очевидно, що її інтеграл не існує в тому розумінні, що інтеграли  $f^+$  та  $f^-$  одночасно нескінченно великі.

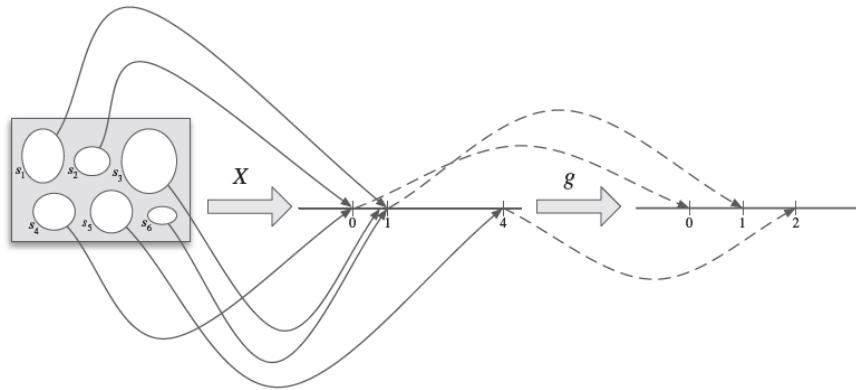


Рис. 7.3.1.: Ілюстрація функції  $g(\cdot) = \sqrt{\cdot}$ , застосованої до випадкової величини  $X$ , яка відображає 6 елементів простору елементарних подій у числа 0, 1 і 4 ([1], Рис. 3.9)

Із (4.3.6) безпосередньо випливає

$$\begin{aligned}
 \text{Var}(Z) &= \mathbb{E}[Z^2] - (\mathbb{E}[Z])^2 = \mathbb{E}[Z^2] = \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} z^2 e^{-\frac{z^2}{2}} dz \\
 &= \int_0^{\infty} \frac{2}{\sqrt{2\pi}} z^2 e^{-\frac{z^2}{2}} dz \\
 &\quad \left[ u = z, dv = ze^{-\frac{z^2}{2}} dz \right] \\
 &\quad \left[ du = dz, v = -e^{-\frac{z^2}{2}} \right] \\
 &= \frac{2}{\sqrt{2\pi}} \left( -ze^{-\frac{z^2}{2}} \Big|_0^{\infty} + \int_0^{\infty} e^{-\frac{z^2}{2}} dz \right) \\
 &= \frac{2}{\sqrt{2\pi}} \left( 0 + \frac{\sqrt{2\pi}}{2} \right) \\
 &= 1,
 \end{aligned}$$

де границя  $\lim_{z \rightarrow \infty} ze^{-\frac{z^2}{2}} = 0$  хоча б за правилом Л'Опітала (L'Hôpital's rule), а останній інтеграл є половиною інтегралу щільності розподілу.  $\square$

Тепер стає зрозуміло, чому в позначенні стандартного нормального розподілу  $N(0, 1)$  використано числа 0 і 1.

### 7.3. Функції від випадкових величин

На практиці скрізь виникають ситуації, коли потрібно оперувати випадковими величинами, які є функціями від інших випадкових величин, наприклад, ступенями випадкових величин, їх сумами чи добутками, лінійними комбінаціями тощо (Рис. 7.3.1).

Для коректного уведення поняття функції від випадкової величини спочатку потрібно з'ясувати, які функції є вимірними, адже нам потрібно, щоб результат застосування функції залишався вимірним. Згідно з Визначенням 4.1.2, функція  $X : (\Omega, \mathcal{A}) \rightarrow (\Omega', \mathcal{A}')$  є вимірною, якщо  $X^{-1}(A') \subseteq \mathcal{A}$

для будь-якого  $A' \in \mathcal{A}'$ , тобто що будь-який прообраз вимірної множини є вимірною множиною, де прообраз ми визначили як

$$X^{-1}(A') \equiv \{\omega \in \Omega : X(\omega) \in A'\} \in \mathcal{A}$$

Можна розглянути поняття прообразу не окремої множини, а цілого класу множин  $\mathcal{C}$ :

$$X^{-1}(\mathcal{C}) = \{A \subseteq \Omega : A = X^{-1}(B), B \in \mathcal{C}\}, \quad (7.3.1)$$

тобто це клас усіх прообразів відповідних множин.

Розгляньмо твердження, яке дає змогу значно спростити перевірку вимірності функцій.

**Теорема 7.3.1.** Нехай деяка функція  $X$  діє з вимірного простору  $(\Omega, \mathcal{A})$  у вимірний простір  $(\Omega', \mathcal{A}')$ .

Тоді якщо  $\mathcal{A}'$  породжено множинами з деякого класу  $\mathcal{C}'$ :  $\mathcal{A}' = \mathcal{A}(\mathcal{C}')$ , то функція  $X$  є вимірною, якщо  $X^{-1}(A') \in \mathcal{A}$  для будь-якої множини  $A' \in \mathcal{C}'$ .

Доведення цієї теореми наведено в Розд. 7.6.

Практичне застосування Теореми 7.3.1 полягає в тому, що вимірність будь-якої функції достатньо перевірити для множин із класу, який породжує Борелеву  $\sigma$ -алгебру, що зробити значно простіше, ніж перевіряти цю властивість для всіх можливих множин із цієї  $\sigma$ -алгебри.

Також можна показати, що композиція вимірних функцій є вимірною функцією.

**Твердження 7.3.2.** Нехай деяка функція  $X : (\Omega, \mathcal{A}) \rightarrow (\Omega', \mathcal{A}')$  є вимірною і нехай деяка функція  $g : (\Omega', \mathcal{A}') \rightarrow (\Omega'', \mathcal{A}'')$  також є вимірною. Тоді вимірною є їх композиція  $g \circ X : \Omega \rightarrow \Omega''$ , де  $(g \circ X)(\omega) \equiv g(X(\omega))$ ,  $\omega \in \Omega$ .

*Доведення.* Оскільки  $g$  вимірна, то для кожної  $A'' \in \mathcal{A}''$  справедливо  $A' = g^{-1}(A'') \in \mathcal{A}'$ . За властивостями композицій,  $(g(X))^{-1}(A'') = X^{-1}(g^{-1}(A'')) = X^{-1}(A')$ . Але оскільки  $X$  сама по собі є вимірною, і  $A' \in \mathcal{A}'$ , то і  $X^{-1}(A') \in \mathcal{A}$ , а відтак і  $(g(X))^{-1}(A'') \in \mathcal{A}$ .  $\square$

**Зауваження 7.3.3.** Безпосереднім наслідком Твердження 7.3.2 є той факт, що вимірна функція від випадкової величини є випадковою величиною.  $\square$

**Визначення 7.3.4.** Вимірну функцію  $f : (\mathbb{R}, \mathcal{B}) \rightarrow (\mathbb{R}, \mathcal{B})$  називають *Борелевою функцією* (Borel function).  $\square$

Окрім функцій від однієї випадкової величини також корисно розглядати функції від *декількох* випадкових величин, наприклад, їх суми чи добутку. Сукупність випадкових величин має назву випадкового вектора.

**Визначення 7.3.5.** Вимірну функцію  $\mathbf{X} : \Omega \rightarrow \mathbb{R}^k$  називають *випадковим вектором* (random vector).

За замовчуванням кладуть, що  $\sigma$ -алгеброю для простору  $\mathbb{R}^k$  є Борелева  $\sigma$ -алгебра  $\mathcal{B}^k$ .  $\square$

Виявляється, що координатами будь-якого випадкового вектора є випадкові величини.

**Твердження 7.3.6.** Нехай маємо функцію  $\mathbf{X} = (X_1, \dots, X_k)^\top : (\Omega, \mathcal{A}) \rightarrow (\mathbb{R}^k, \mathcal{B}^k)$ . Функція  $\mathbf{X}$  є випадковим вектором тоді й тільки тоді, коли  $X_i$  є випадковими величинами,  $i = 1, \dots, k$ .

Доведення цього твердження наведено в Розд. 7.6.

За аналогією з одновимірним випадком, вимірну функцію  $f : (\mathbb{R}^n, \mathcal{B}^n) \rightarrow (\mathbb{R}^m, \mathcal{B}^m)$  також називають Борелевою. Твердження 7.3.2 дає змогу утворювати випадкові вектори шляхом застосування Борелевих функцій до інших випадкових векторів.

Отже, Теорема 7.3.1 у поєднанні з Твердженням 7.3.2 уможливлює утворення нових випадкових величин шляхом застосування функцій до випадкових величин, вимірність яких можна перевірити для класу множин, що породжують Борелеву  $\sigma$ -алгебру.

З'ясуємо, які функції є вимірними (Борелевими).

**Твердження 7.3.7.** Індикаторна функція  $X(\omega) = \mathbb{1}\{\omega \in A\} : (\Omega, \mathcal{A}) \rightarrow (\mathbb{R}, \mathcal{B})$ ,  $A \in \mathcal{A}$  є вимірною.

*Доведення.* За Теоремою 7.3.1 вимірність індикаторної функції достатньо перевірити для промінів виду  $(-\infty; a]$ ,  $a \in \mathbb{R}$ :

$$(\mathbb{1}\{x \in A\})^{-1}((-\infty; a]) = \begin{cases} \Omega, & a \geq 1 \\ A^c, & a \in [0; 1) \\ \emptyset, & a < 0 \end{cases}.$$

Усі ці три множини є вимірними.  $\square$

**Твердження 7.3.8.** Нехай деяка функція  $X : (\mathbb{R}^k, \mathcal{B}^k) \rightarrow (\mathbb{R}^m, \mathcal{B}^m)$  є неперервною. Тоді вона є Борелевою.

*Доведення.* Однією з властивостей неперервних функцій є той факт, що прообразом будь-якої відкритої множини є відкрита множина<sup>5</sup>. Борелева  $\sigma$ -алгебра  $\mathcal{B}^m$ , за визначенням, породжена класом  $m$ -вимірних відкритих множин  $\mathcal{O}^m$ . Іншими словами, для будь-якої  $O^m \in \mathcal{O}^m$  справедливо  $X^{-1}(O^m) \in \mathcal{O}^k$ . Але Борелева  $\sigma$ -алгебра  $\mathcal{B}^k$  також породжена множиною відкритих множин  $\mathcal{O}^k$ . Відтак,  $X^{-1}(O^m) \in \mathcal{B}^k$ . За Теоремою 7.3.1,  $X$  є вимірною.  $\square$

**Зауваження 7.3.9.** Із Тверджень 7.3.6 і 7.3.8 також випливає, що будь-яка *кусково-неперервна* (piecewise continuous) функція також є вимірною.

Будь-яку кусково-неперервну функцію

$$f(x) = \begin{cases} f_1(x), & x \in A_1 \\ \vdots & \\ f_k(x), & x \in A_k \end{cases},$$

де кожна  $f_i$  неперервна,  $i = 1, \dots, k$ , можна подати як суму

$$f(x) = \sum_{i=1}^k f_i(x) \cdot \mathbb{1}\{x \in A_i\}.$$

<sup>5</sup>Власне, це є визначенням неперервної функції для найбільш загальних топологічних просторів, на яких навіть не визначено поняття міри.

За Твердженнями 7.3.7 та 7.3.8 індикаторні та неперервні функції є вимірними.

Добуток  $f_i(x) \cdot \mathbb{1}\{x \in A_i\}$  є вимірною функцією, оскільки це неперервна функція від випадкового вектора  $(f_i, \mathbb{1}_{A_i})^\top$ , який сам по собі є вимірний згідно з Твердженням 7.3.6.

Аналогічно вимірною є сума вимірних функцій.

Тому й лінійна комбінація  $f(x)$  індикаторних і неперервних функцій є вимірною.  $\square$

**Зауваження 7.3.10.** Отже, результатом застосування будь-якої неперервної функції до випадкового вектору є випадковий вектор. У тому числі, якщо  $\mathbf{X} = (X_1, \dots, X_k)^\top$  і  $\mathbf{Y} = (Y_1, \dots, Y_k)^\top$  є випадковими векторами в  $\mathbb{R}^k$ , а  $a, b \in \mathbb{R}^k$ , то випадковими векторами будуть  $a\mathbf{X} + b$ ,  $\mathbf{X} + \mathbf{Y}$ ,  $\mathbf{X}^\top \cdot \mathbf{Y}$ ,  $X_1/X_2$  (якщо  $X_2 \neq 0$ , оскільки ця функція є кусково-неперервною),  $\mathbf{X}^\top \mathbf{X}$ ,  $|\mathbf{X}|$ ,  $\sum_{i=1}^k X_i$ ,  $\prod_{i=1}^k X_i$ ,  $\max\{X_1, \dots, X_k\}$  тощо.  $\square$

Насамкінець доведімо, що послідовність випадкових величин збігається до випадкової величини, якщо відповідна границя існує. Для цього спочатку згадаймо визначення границі та пов'язані з ним визначення для числових послідовностей.

**Визначення 7.3.11.** *Нижньою межею* (lower bound) деякої множини  $A$  є число  $a \in \mathbb{R}$  таке, що  $\forall x \in A \ x \geq a$ . Відповідно, *верхньою межею* (upper bound) деякої множини  $A$  є число  $b \in \mathbb{R}$  таке, що  $\forall x \in A \ x \leq b$ .

*Точною нижньою межею* (greatest lower bound, infimum) деякої множини  $A$  є найбільша з її нижніх меж, тобто число  $\inf A \in \mathbb{R}$  таке, що для всіх нижніх меж  $a$  множини  $A$  виконується  $a \leq \inf A$ . Відповідно, *точною верхньою межею* (least upper bound, supremum) деякої множини  $A$  є найменша з її верхніх меж, тобто число  $\sup A \in \mathbb{R}$  таке, що для всіх верхніх меж  $b$  множини  $A$  виконується  $\sup A \leq b$ .

Нехай маємо деяку числову послідовність  $\{x_1, x_2, \dots\}$ . *Нижньою границею* (limit inferior) цієї послідовності є

$$\liminf_{n \rightarrow \infty} x_n = \lim_{k \rightarrow \infty} \inf_{n \geq k} \{x_n\} . \quad (7.3.2)$$

Аналогічно, *верхньою границею* (limit superior) цієї послідовності є

$$\limsup_{n \rightarrow \infty} x_n = \lim_{k \rightarrow \infty} \sup_{n \geq k} \{x_n\} . \quad (7.3.3)$$

$\square$

У курсі «Математичний аналіз» доводять, що

$$\liminf_{n \rightarrow \infty} x_n = \sup_{k \in \mathbb{N}} \left\{ \inf_{n \geq k} \{x_n\} \right\} ,$$

$$\limsup_{n \rightarrow \infty} x_n = \inf_{k \in \mathbb{N}} \left\{ \sup_{n \geq k} \{x_n\} \right\} .$$

Також доводять, що границя послідовності існує, якщо її нижня та верхня границя дорівнюють одна одній:

$$\lim_{n \rightarrow \infty} x_n = \liminf_{n \rightarrow \infty} x_n = \limsup_{n \rightarrow \infty} x_n .$$

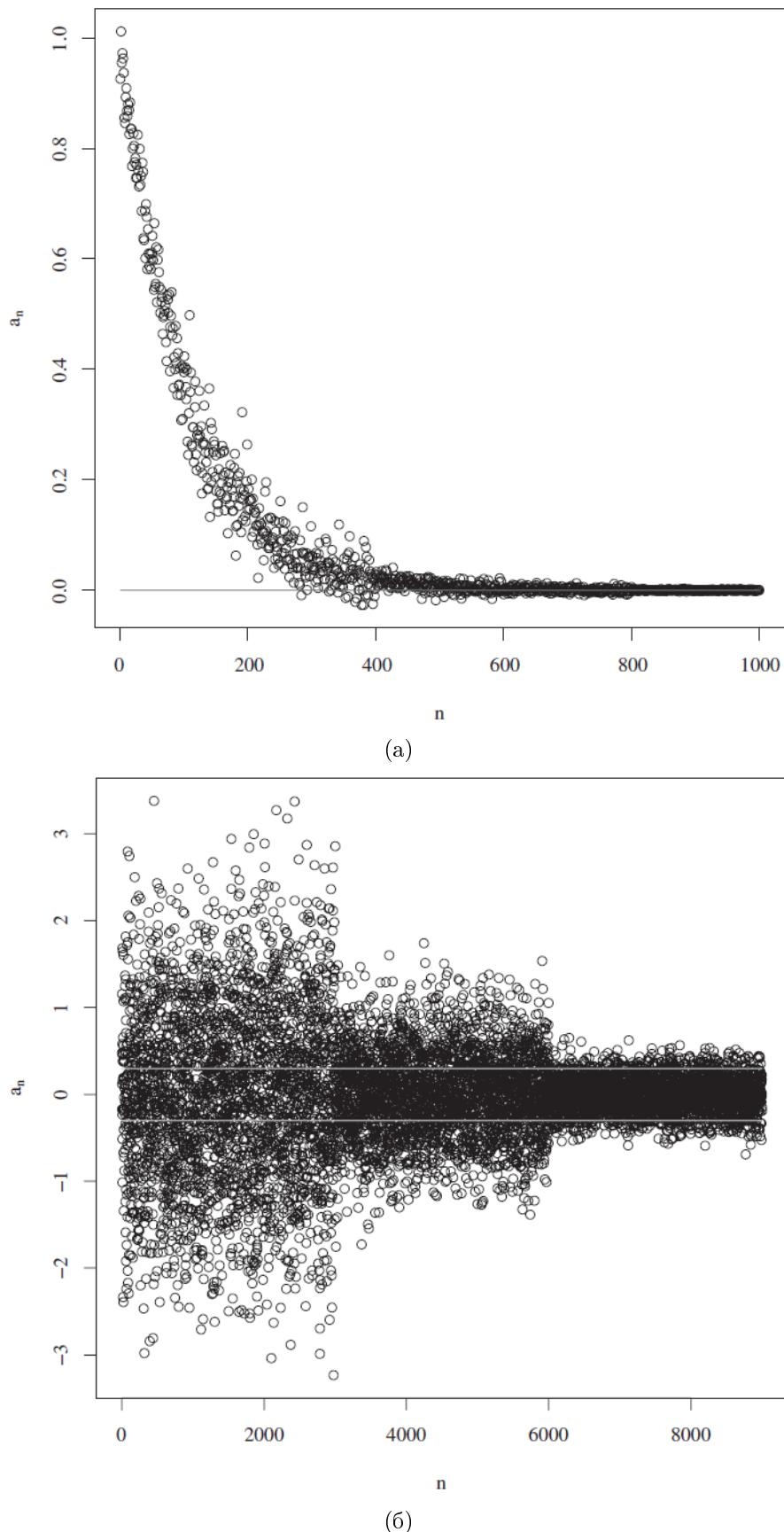


Рис. 7.3.2.: Дві числові послідовності: (а) границя існує (суцільна лінія); (б) границі не існує, оскільки нижня й верхня границі (нижня й верхня суцільні лінії відповідно) не дорівнюють одній ([11], Рис. 1.5–1.6)

Цей факт добре проілюстровано на Рис. 7.3.2.

Аналогічні властивості справедливі для випадкових величин.

**Твердження 7.3.12.** Нехай  $X_n, n \geq 1$  — деякі випадкові величини. Тоді випадковими величинами є<sup>6</sup>:

$$\begin{aligned} (\inf_n X_n)(\omega) &= \inf_n X_n(\omega), \\ (\sup_n X_n)(\omega) &= \sup_n X_n(\omega), \\ (\liminf_{n \rightarrow \infty} X_n)(\omega) &= \liminf_n X_n(\omega), \\ (\limsup_{n \rightarrow \infty} X_n)(\omega) &= \limsup_n X_n(\omega). \end{aligned}$$

Зокрема, якщо  $\liminf_{n \rightarrow \infty} X_n(\omega) = \limsup_{n \rightarrow \infty} X_n(\omega)$ , то випадковою величиною також є функція  $\lim_{n \rightarrow \infty} X_n$ .

*Доведення.* Як відомо з Теореми 7.3.1, вимірність функції достатньо перевірити на деякому класі множин, що породжують  $\mathcal{B}$ , зокрема на класі променів  $\mathcal{C} = \{(-\infty, a], a \in \mathbb{R}\}$ .

Так, для деякого  $a \in \mathbb{R}$  прообраз відповідного променя під дією супремуму має вид:

$$\left\{ \omega : \sup_n X_n(\omega) \leq a \right\} = \bigcap_{n=1}^{\infty} \{ \omega : X_n(\omega) \leq a \} \in \mathcal{A}.$$

Відтак,  $\sup_n X_n$  є випадковою величиною.

Оскільки  $\inf_n X_n = -\sup_n (-X_n)$ , то в силу вимірності множення на константу  $\inf_n X_n$  також є випадковою величиною.

Оскільки  $\liminf_{n \rightarrow \infty} X_n = \sup_{k \in \mathbb{N}} \left\{ \inf_{n \geq k} \{X_n\} \right\}$ , ми маємо композицію двох вимірних функцій, а відтак це також вимірна функція.

Аналогічні міркування справедливі для  $\limsup_{n \rightarrow \infty} X_n$ . □

## 7.4. Розподіл функції від випадкової величини

Після розгляду класу функцій, які можна застосовувати до випадкових величин, не порушуючи їх вимірності, доречно перейти до вивчення способів обчислення розподілів, які мають відповідні новоутворені випадкові величини.

У загальному випадку функцію розподілу для будь-якої випадкової величини можна дістати за Визначенням 6.1.1. А саме, якщо  $X : (\Omega, \mathcal{A}, \mathbb{P}) \rightarrow \mathbb{R}$  і  $g : \mathbb{R} \rightarrow \mathbb{R}$  є Борелевою, то  $Y = g(X)$  має функцію розподілу

$$\begin{aligned} F_Y(y) &= \mathbb{P}_Y(Y \leq y) = \mathbb{P}_X(g(X) \leq y) = \mathbb{P}_X(X \in g^{-1}((-\infty; y])) \\ &= \mathbb{P}(\{\omega : X(\omega) \in g^{-1}((-\infty; y])\}), \end{aligned} \tag{7.4.1}$$

де  $g^{-1}((-\infty; y]) = \{x \in \mathbb{R} : g(x) \leq y\}$ .

<sup>6</sup>Зовсім строго кажучи, ми повинні вимагати, щоб імовірнісний простір  $(\Omega, \mathcal{A}, \mathbb{P})$  був *повний* (complete) у тому розумінні, що будь-яка підмножина множини міри нуль має міру нуль. У цьому курсі ми вважаємо, що всі ймовірнісні простори повні.

**Приклад 7.4.1** (Розподіл  $\chi_1^2$ ). Нехай  $X \sim N(0, 1)$ , а  $g(x) = x^2$ . Тоді

$$\mathbb{P}_Y(Y \leq y) = \mathbb{P}_X(X^2 \leq y) = \begin{cases} 0, & y < 0 \\ \mathbb{P}_X(X \in [-\sqrt{y}; \sqrt{y}]), & y \geq 0 \end{cases},$$

оскільки прообразом  $(-\infty; y]$  може бути або  $\emptyset$  для від'ємних  $y$ , або  $[-\sqrt{y}; \sqrt{y}]$  — для невід'ємних. Тоді, для  $y \geq 0$  маємо

$$\mathbb{P}_Y(Y \leq y) = \Phi(\sqrt{y}) - \Phi(-\sqrt{y}) = 2\Phi(\sqrt{y}) - 1.$$

Остаточно

$$F_Y(y) = (2\Phi(\sqrt{y}) - 1) \cdot \mathbb{1}\{y \geq 0\}.$$

Щільність розподілу величини  $Y$  можна дістати як похідну. Для  $y < 0$ , очевидно,  $f_Y(y) = 0$ . Для  $y > 0$  маємо

$$f_Y(y) = (2\Phi(\sqrt{y}) - 1)' = 2\phi(\sqrt{y}) \cdot \frac{1}{2\sqrt{y}} = \frac{1}{\sqrt{2\pi y}} e^{-y/2}.$$

У точці 0 функція розподілу недиференційовна, тому можемо покласти, наприклад,  $f_Y(0) = 0$ . Оскільки всі щільності визнанено з точністю до множини міри нуль, ми могли б із таким же успіхом покласти  $f(0) = 100$  або що, проте вибір  $f_Y(0) = 0$  є типовим у цій ситуації.

Остаточно маємо

$$f_Y(y) = \frac{1}{\sqrt{2\pi y}} e^{-y/2} \cdot \mathbb{1}\{y > 0\}.$$

Можна пересвідчитися, що це справді щільність, якщо помітити, що

$$\frac{1}{\sqrt{2\pi}} \int_0^\infty \frac{1}{\sqrt{2\pi y}} e^{-y/2} dy = \frac{\sqrt{2}}{\sqrt{2\pi}} \int_0^\infty t^{-1/2} e^{-t} dt = \frac{1}{\sqrt{\pi}} \cdot \sqrt{\pi} = 1$$

оскільки відповідний інтеграл є значенням гамма-функції в точці 0.5, а загальновідомо, що  $\Gamma(0.5) = \sqrt{\pi}$ .

Ця щільність відповідає так званому *розподілу  $\chi^2$  з одним ступенем свободи*:  $Y \sim \chi_1^2$ . Детальніше розподіл  $\chi^2$  ми розглянемо в наступних лекціях.  $\square$

Підхід із використанням (7.4.1) загальний і працює для будь-яких величин, проте для величин, які мають щільність (у тому числі відносно лічної міри), існують альтернативні способи.

#### 7.4.1. Розподіл функції від дискретної випадкової величини

Для дискретної випадкової величини  $X$ , згідно з (6.1.2), функція розподілу має вид

$$F_X(x) = \sum_{\substack{y \in \text{supp}(X) \\ y \leq x}} p_X(y),$$

а відтак формула (7.4.1) набуває такого виду:

$$\mathbb{P}_Y(Y = y) = \mathbb{P}_X(g(X) = y) = \sum_{x: g(x)=y} \mathbb{P}_X(X = x). \quad (7.4.2)$$

**Приклад 7.4.2.** Розгляньмо спрощений варіант випадкового процесу, відомого як *випадкове блукання* (random walk).

Нехай деяка частинка покроково рухається вздовж числової осі, на кожному кроці пересувуючись на 1 вправо або вліво з однаковими ймовірностями. Усі кроки незалежні один від одного. На початку руху частинка перебуває в точці 0. Нехай  $Y$  — точка, у якій зупинилася частинка після  $n$  кроків.

Знайдімо функцію ймовірності  $p_Y$ . Якщо кожний крок розглядати як випробування Бернуллі, де успіхом уважати рух управо, то число кроків управо  $X$ , які встигне зробити частинка, має розподіл  $\text{Binom}(n, 0.5)$ . Якщо  $X = j$ , то це означає, що частинка зробила  $j$  кроків управо і  $n - j$  кроків уліво, а тому кінцева точка її перебування є  $Y = j - (n - j) = 2j - n$ . Зокрема, це значить, що частинка за парну кількість кроків може описанитися тільки в парній позиції, а за непарну — тільки в непарній.

Отже,  $Y = 2X - n$ . Це взаємно однозначне відображення, і тому функція ймовірності має простий вид:

$$\mathbb{P}_y(Y = k) = \mathbb{P}_X(2X - n = k) = \mathbb{P}_X\left(X = \frac{n+k}{2}\right) = \binom{n}{\frac{n+k}{2}} 0.5^n,$$

де  $\text{supp}(Y) = \{k \in \mathbb{Z} \cap [-n; n] : n + k \text{ парне}\}$ .

Розгляньмо тепер випадкову величину  $D$  — відстань точки від 0 після  $n$  кроків,  $n$  — парне. Тоді  $D = |Y|$ . Це відображення не є взаємно однозначним, а тому за формулою (7.4.2) маємо:

$$\begin{aligned} \mathbb{P}_D(D = 0) &= \mathbb{P}_Y(Y = 0) = \binom{n}{\frac{n}{2}} 0.5^n, \\ \mathbb{P}_D(D = k) &= \sum_{y:|y|=k} \mathbb{P}_Y(Y = y) = \mathbb{P}_Y(Y = k) + \mathbb{P}_Y(Y = -k) = 2 \binom{n}{\frac{n+k}{2}} 0.5^n. \end{aligned} \quad \square$$

**Зауваження 7.4.3.** Не варто плутати застосування функції до випадкової величини,  $g(X)$ , і застосування цієї ж функції до функції ймовірності  $g(p_X(x))$ . Формула (7.4.2) дає чітке правило підрахунку  $p_Y$ . Зокрема, зовсім некоректно стверджувати, що для здобуття функції ймовірності  $Y = 2X$  потрібно помножити  $p_X(x)$  на 2. Це позбавлено сенсу, адже в цьому випадку сума всіх значень функції ймовірності перевищуватиме 1. На Рис. 7.4.1 наведено графіки функції ймовірності деякої дискретної величини  $X$  та функції ймовірності величини  $2X$ .

□

#### 7.4.2. Розподіл функції від абсолютно неперервної випадкової величини

Для абсолютно неперервних випадкових величин в окремих випадках справедлива так звана формула заміни змінних.

**Теорема 7.4.4** (Формула заміни змінних (Change of variables formula)). Нехай  $X$  — абсолютно неперервна величина зі щільністю  $f_X$ , неперервною на  $\text{supp}(X)$ . Нехай  $Y = g(X)$ , де  $g$  взаємно однозначна на  $\text{supp}(X)$ , а функція  $g^{-1}(y)$  неперервно диференційовна на  $g(\text{supp}(X))$ . Тоді

$$f_Y(y) = f_X(g^{-1}(y)) \cdot \left| \frac{d}{dy} g^{-1}(y) \right| \cdot \mathbb{1}_{\{y \in g(\text{supp}(X))\}}. \quad (7.4.3)$$

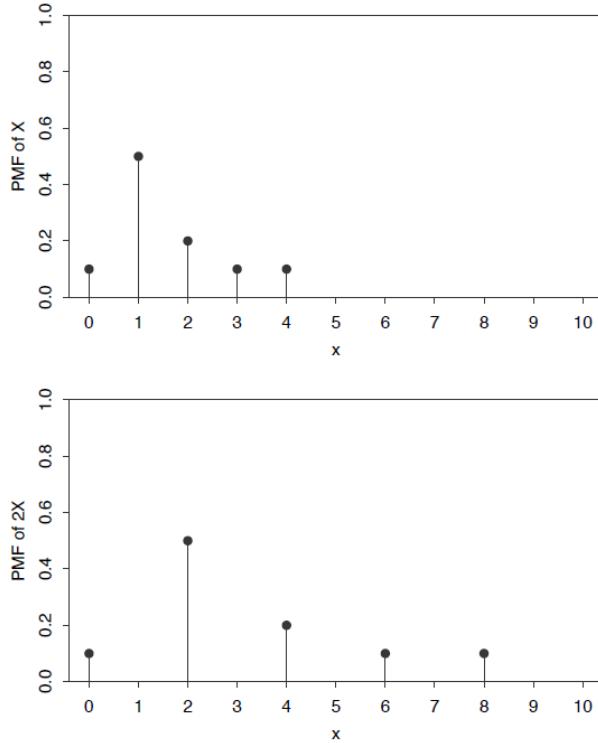


Рис. 7.4.1.: Ілюстрацiя до Зauваження 7.4.3 ([1], Рис. 3.11)

*Доведення.* Оскiльки  $g$  взаємно однозначна на  $\text{supp}(X)$ , вона повинна бути строго монотонна на  $\text{supp}(X)$ . Якщо вона строго зростаюча, то з (7.4.1) випливає, що для  $y \in g(\text{supp}(X))$

$$F_Y(y) = \mathbb{P}_Y(g(X) \leq y) = \mathbb{P}_X(X \leq g^{-1}(y)) = F_X(g^{-1}(y)).$$

Оскiльки  $g^{-1}(y)$  неперервно диференцiйовна на  $g(\text{supp}(X))$ ,

$$f_Y(y) = \frac{d}{dy}F_Y(y) = \frac{d}{dy}F_X(g^{-1}(y)) = \frac{d}{dy} \left( \int_{-\infty}^{g^{-1}(y)} f_X(x) dx \right) = f_X(g^{-1}(y)) \cdot \frac{d}{dy}g^{-1}(y).$$

Якщо ж  $g$  строго спадна, то для  $y \in g(\text{supp}(X))$ :

$$F_Y(y) = \mathbb{P}_Y(g(X) \leq y) = \mathbb{P}_X(X \geq g^{-1}(y)) = 1 - F_X(g^{-1}(y)) ,$$

$$f_Y(y) = \frac{d}{dy}F_Y(y) = -\frac{d}{dy}F_X(g^{-1}(y)) = -\frac{d}{dy} \left( \int_{-\infty}^{g^{-1}(y)} f_X(x) dx \right) = f_X(g^{-1}(y)) \cdot \left( -\frac{d}{dy}g^{-1}(y) \right) .$$

Варто звернути увагу, що для спадної  $g$  справедливо  $\frac{d}{dy}g^{-1}(y) < 0$ , тому вищенаведений вираз коректний ( $f_Y(y)$  невiд'ємна).

За межами носiя щiльнiсть дорiвнює 0, а в точках недиференцiйовностi вiдповiдної функцiї покладемо  $f_Y$  рiвною 0. Остаточно маємо

$$f_Y(y) = f_X(g^{-1}(y)) \cdot \left| \frac{d}{dy}g^{-1}(y) \right| \cdot \mathbb{1}\{y \in g(\text{supp}(X))\} .$$

□

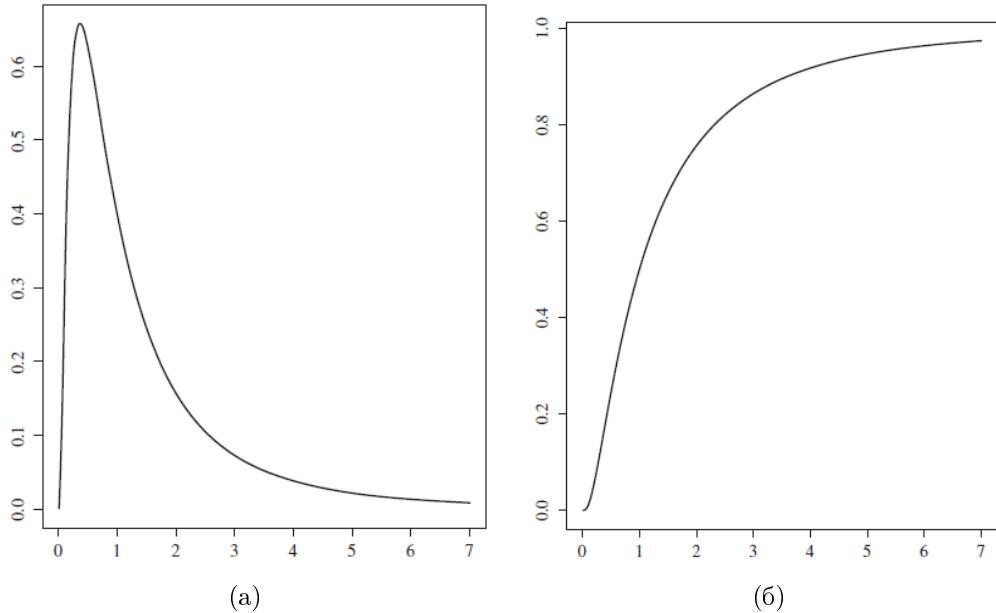


Рис. 7.4.2.: Графіки для стандартного логнормального розподілу: (а) щільність розподілу; (б) функція розподілу ([11], Рис. 2.9)

**Приклад 7.4.5** (Логнормальний розподіл (Log-normal distribution)). Нехай  $X \sim N(0, 1)$ ,  $Y = g(x) = e^X$ , тобто  $X = \ln Y$ . Тоді кажуть, що величина  $Y$  має *логнормальний* (log-normal) розподіл. Його щільність на основі (7.4.3) дорівнює

$$f_Y(y) = f_X(\ln y) \cdot \left| \frac{d}{dy} \ln y \right| \cdot \mathbb{1} \{y \in e^{\mathbb{R}}\} = \phi(\ln y) \cdot \frac{1}{y} \cdot \mathbb{1} \{y > 0\} .$$

На Рис. 7.4.2 зображені щільність та функцію розподілу для логнормального розподілу, що відповідає стандартному нормальному розподілу.

Сподівання логнормальної випадкової величини дорівнює

$$\begin{aligned} \mathbb{E}[Y] &= \int_0^\infty y \cdot \phi(\ln y) \cdot \frac{1}{y} dy \\ &= \int_0^\infty \frac{1}{\sqrt{2\pi}} e^{-\frac{\ln^2 y}{2}} dy \\ &= \int_{-\infty}^\infty \frac{1}{\sqrt{2\pi}} e^{-\frac{u^2}{2}} e^u du \\ &= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^\infty e^{-\frac{1}{2}(u^2 - 2u + 1 - 1)} du \\ &= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^\infty e^{-\frac{(u-1)^2}{2} + \frac{1}{2}} du \\ &= e^{\frac{1}{2}} \cdot \frac{1}{\sqrt{2\pi}} \int_{-\infty}^\infty e^{-\frac{v^2}{2}} dv = e^{\frac{1}{2}}, \end{aligned}$$

де ми використали той факт, що інтеграл щільності стандартного нормального розподілу є 1. □

У випадках, коли функція  $g$  не є взаємно однозначною, інколи можливо розбити носій величини  $X$  на неперетинанні підмножини так, що на кожній із них  $g$  буде взаємно однозначною.

**Твердження 7.4.6.** Нехай  $X$  — абсолютно неперервна величина зі щільністю  $f_X$ , неперервною на  $\text{supp}(X)$ . Нехай  $Y = g(X)$ . Нехай існує розбиття  $A_0, A_1, \dots, A_k$  носія  $\text{supp}(X)$  таке, що  $\mathbb{P}_X(X \in A_0) = 0$ .

Нехай на  $A_1, \dots, A_k$  визначено функції  $g_1(x), \dots, g_k(x)$  відповідно, такі що:

- $g(x) = g_i(x)$  для  $x \in A_i$ ,  $i = 1, \dots, k$ ;
- $g_i(x)$  взаємно однозначна на  $A_i$ ,  $i = 1, \dots, k$ ;
- множина  $B = \{y : g_i(x) = y, x \in A_i\}$  однаакова для всіх  $i = 1, \dots, k$ ;
- функції  $g_i^{-1}(y)$  неперервно диференційовні на  $B$ ,  $i = 1, \dots, k$ .

Тоді

$$f_Y(y) = \mathbb{1}\{y \in B\} \cdot \sum_{i=1}^k f_X(g_i^{-1}(y)) \cdot \left| \frac{d}{dy} g_i^{-1}(y) \right|. \quad (7.4.4)$$

**Приклад 7.4.7.** Застосування Твердження 7.4.6 проілюструймо на вже розглянутому Прикладі 7.4.1.

Покладімо  $A_0 = \{0\}$ ,  $A_1 = (-\infty; 0)$ ,  $A_2 = (0; \infty)$ . На множинах  $A_1$  і  $A_2$  функція  $g(x) = x^2$  є строго монотонною, а  $g_1^{-1}(y) = -\sqrt{y}$ ,  $g_2^{-1}(y) = \sqrt{y}$ .

Тоді

$$f_Y(y) = \mathbb{1}\{y > 0\} \cdot \left( f_X(-\sqrt{y}) \left| -\frac{1}{2\sqrt{y}} \right| + f_X(\sqrt{y}) \left| \frac{1}{2\sqrt{y}} \right| \right) = \mathbb{1}\{y > 0\} \cdot \phi(\sqrt{y}) \cdot \frac{1}{\sqrt{y}}. \quad \square$$

## 7.5. Сім'ї розподілів відносно зсуву-масштабування

На практиці поширеними є випадкові величини, які можна утворювати одну з одної шляхом *перетворення зсуву та масштабування* (location-scale transformation):

$$Y = aX + b, \quad a, b \in \mathbb{R}, \quad a > 0. \quad (7.5.1)$$

Параметр  $b$  відповідає за зсув, а параметр  $a$  — за масштабування.

Сукупність розподілів, які пов'язано відповідним чином, становить *сім'ю відносно зсуву-масштабування* (location-scale family). Можна казати, що ця сім'я розподілів замкнена відносно операції (7.5.1).

Оскільки для такої сім'ї виконуються всі умови Теореми 7.4.4, можемо записати формулу щільності величини  $Y$ :

$$f_Y(y) = f_X\left(\frac{y-b}{a}\right) \cdot \frac{1}{a} \cdot \mathbb{1}\{y \in a \cdot \text{supp}(X) + b\}. \quad (7.5.2)$$

**Приклад 7.5.1.** Функція і щільність логістичного розподілу (logistic distribution) мають вид

$$\begin{aligned} F(x) &= \frac{e^{\frac{x-\mu}{\sigma}}}{1 + e^{\frac{x-\mu}{\sigma}}} , \\ f(x) = F'(x) &= \frac{e^{\frac{x-\mu}{\sigma}}}{\sigma \left(1 + e^{\frac{x-\mu}{\sigma}}\right)^2} , \end{aligned} \quad (7.5.3)$$

де  $\mu \in \mathbb{R}$ ,  $\sigma > 0$ .

Цей розподіл належить до сім'ї розподілів відносно зсуву-масштабування. Це можна перевірити з використанням (7.5.2) безпосередньо. Нехай  $X$  має логістичний розподіл, тоді  $Y = aX + b$  має таку щільність розподілу:

$$f_Y(y) = f_X\left(\frac{y-b}{a}\right) \cdot \frac{1}{a} = \frac{e^{\frac{y-b}{a}-\mu}}{a\sigma\left(1 + e^{\frac{y-b}{a}-\mu}\right)^2} = \frac{e^{\frac{y-(a\mu+b)}{a\sigma}}}{a\sigma\left(1 + e^{\frac{y-(a\mu+b)}{a\sigma}}\right)^2} ,$$

що відповідає логістичному розподілу з параметрами  $a\mu + b$  і  $a\sigma$ .  $\square$

### 7.5.1. Рівномірний розподіл

Прикладом сім'ї розподілів відносно зсуву-масштабування є рівномірні розподіли.

Нехай  $U \sim \text{U}((0; 1))$ . Розгляньмо випадкову величину  $X = a + (b - a)U$ . Згідно з (7.5.2), щільність такої величини дорівнює

$$f_Y(y) = f_X\left(\frac{y-a}{b-a}\right) \cdot \frac{1}{b-a} \cdot \mathbb{1}\{y \in a + (b-a) \cdot \text{supp}(X)\} = \frac{1}{b-a} \cdot \mathbb{1}\{y \in (a; b]\} .$$

**Визначення 7.5.2.** Випадкова величина  $X$  має *рівномірний розподіл* (uniform distribution) на  $(a; b]$ ,  $X \sim \text{U}((a; b])$ , якщо її щільність розподілу дорівнює

$$f(x) = \frac{1}{b-a} \cdot \mathbb{1}\{x \in (a; b]\} . \quad (7.5.4)$$

$\square$

Функція рівномірного розподілу має такий вид:

$$F_X(x) = \int_{(-\infty; x]} \mathbb{1}\{x \in (a; b]\} d\lambda = \int_{-\infty}^x \mathbb{1}\{t \in (a; b]\} dt = \begin{cases} 0, & x < a \\ \frac{x-a}{b-a}, & a \leq x < b \\ 1, & x \geq b \end{cases} . \quad (7.5.5)$$

Отже рівномірний розподіл є нічим іншим, як *нормованою* мірою Лебега, тобто для  $x \in (a; b)$  маємо

$$\mathbb{P}_X((a; x]) = F_X(x) - F_X(a) = \frac{x-a}{b-a} = \frac{\lambda((a; x])}{\lambda((a; b])} .$$

**Зауваження 7.5.3.** Варто пам'ятати, що застосування функцій до випадкових величин не є тотожним застосуванню функцій до їхніх щільностей.

Так, наприклад, якщо  $X \sim U((0; 1])$ , то  $3X + 1 \sim U((1; 4])$ , і  $f_{3X+1}(x) = \frac{1}{3} \cdot \mathbb{1}\{x \in (1; 4]\}$ , що зовсім не дорівнює  $3f_X(x) + 1 = 3 \cdot \mathbb{1}\{(0; 1]\} + 1$ .  $\square$

**Твердження 7.5.4.** Якщо  $X \sim U((a; b])$ , то  $\mathbb{E}[X] = \frac{a+b}{2}$ , а  $\text{Var}(X) = \frac{(b-a)^2}{12}$ .

*Доведення.* Сподівання та дисперсію рівномірного розподілу можна обчислити за визначенням, а можна скористатися тим, що  $X = a + (b - a)U$ , де  $U \sim U((0; 1])$ , та Твердженням 6.4.6:

$$\begin{aligned}\mathbb{E}[X] &= \mathbb{E}[a + (b - a)U] = a + (b - a)\mathbb{E}[U] = a + \frac{b - a}{2} = \frac{a + b}{2}, \\ \text{Var}(X) &= \text{Var}(a + (b - a)U) = (b - a)^2 \text{Var}(U) = \frac{(b - a)^2}{12}.\end{aligned}$$

 $\square$ 

**Приклад 7.5.5.** Автобуси прибувають на зупинку що 15 хвилин: о 7 ранку, о 7:15, о 7:30 і т.д. Нехай пасажир приходить на зупинку в деякий момент часу  $X$ , рівномірно розподілений на часовому проміжку від 7:00 до 7:30. З'ясуймо, чому дорівнює ймовірність очікування на автобус менше 5 хвилин та більше 10 хвилин.

Оскільки за описом  $X \sim U((0; 30])$ , пасажир ждатиме автобуса менше 5 хвилин, тільки якщо він приде на зупинку між 7:10 і 7:15 або між 7:25 і 7:30. Тоді

$$\begin{aligned}\mathbb{P}(\text{час очікування менше 5 хвилин}) &= \mathbb{P}_X(10 < X < 15) + \mathbb{P}_X(25 < X < 30) \\ &= \int_{10}^{15} \frac{1}{30} dx + \int_{25}^{30} \frac{1}{30} dx = \frac{1}{3}.\end{aligned}$$

За цією ж логікою, очікування більше 10 хвилин означає приуття на зупинку між 7:00 і 7:05 або між 7:15 і 7:20:

$$\begin{aligned}\mathbb{P}(\text{час очікування більше 10 хвилин}) &= \mathbb{P}_X(0 < X < 5) + \mathbb{P}_X(15 < X < 20) \\ &= \int_0^5 \frac{1}{30} dx + \int_{15}^{20} \frac{1}{30} dx = \frac{1}{3}.\end{aligned}\quad \square$$

## 7.5.2. Нормальний розподіл

Безпосереднім узагальненням стандартного нормального розподілу, розглянутого в Розділі 7.2, є нормальний розподіл зі сподіванням  $\mu$ <sup>7</sup> та дисперсією  $\sigma^2$ .

**Визначення 7.5.6.** Випадкова величина  $X$  має *нормальний розподіл* (normal distribution) зі сподіванням  $\mu$  та дисперсією  $\sigma^2$  ( $\sigma > 0$ ),  $X \sim N(\mu, \sigma^2)$ , якщо  $X = \mu + \sigma Z$ , де  $Z \sim N(0, 1)$ .  $\square$

<sup>7</sup>Позначення  $\mu$  в цьому контексті є загальноприйнятым, тому ми його також використовуватимемо, сподіваючись, що з позначенням міри його сплутати буде неможливо.

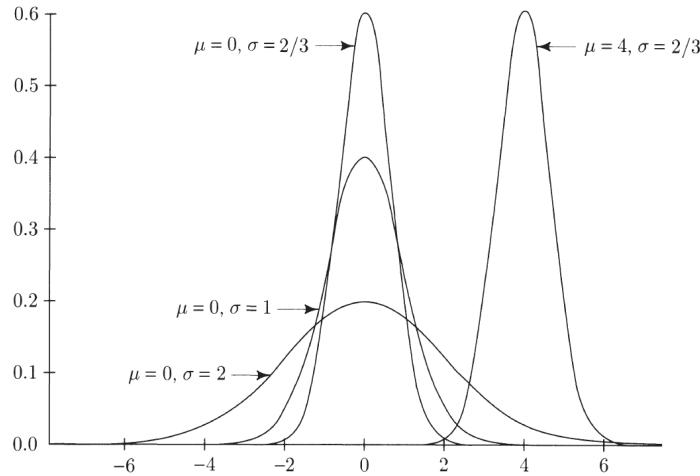


Рис. 7.5.1.: Щільності нормального розподілу з різними параметрами  $\mu$  і  $\sigma$  ([9, с. 268])

Зв'язок між функціями та щільностями стандартного нормального розподілу і довільного нормального розподілу визначає таке твердження.

**Твердження 7.5.7.** Нехай  $X \sim N(\mu, \sigma^2)$ . Тоді

$$\begin{aligned} F(x) &= \Phi\left(\frac{x-\mu}{\sigma}\right), \\ f(x) &= \frac{1}{\sigma} \cdot \phi\left(\frac{x-\mu}{\sigma}\right) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}}. \end{aligned} \quad (7.5.6)$$

*Доведення.* За визначенням функції розподілу, маємо:

$$F(x) = \mathbb{P}_X (X \leq x) = \mathbb{P}_X \left( \frac{X-\mu}{\sigma} \leq \frac{x-\mu}{\sigma} \right) = \Phi\left(\frac{x-\mu}{\sigma}\right).$$

Оскільки щільність є похідною функції розподілу, маємо:

$$f(x) = F'(x) = \frac{1}{\sigma} \phi\left(\frac{x-\mu}{\sigma}\right).$$

Цей результат також є прикладом застосування (7.5.2). □

Результат цього твердження гранично корисний на практиці, адже всі обчислення ймовірностей, пов'язані з використанням нормального розподілу з деякими параметрами  $\mu \neq 0$  чи  $\sigma^2 \neq 1$  зводять до обчислення ймовірностей із використанням стандартного нормального розподілу. Це пов'язано з тим, що обчислення функції нормального розподілу здійснюють чисельно, і для максимальної оптимізації таких обчислень на комп'ютері алгоритми обчислення функції стандартного нормального розподілу вписано якомога ефективніше<sup>8</sup>.

Графіки щільностей нормальних розподілів із різними параметрами наведено на Рис. 7.5.1.

<sup>8</sup>Наприклад, у мові Python, зокрема в бібліотеці `scipy`, функцією нормального розподілу є `scipy.stats.norm.cdf`. У коді чітко видно, що для її обчислення викликають функцію `scipy.special.ndtr`, яка, у свою чергу, викликає функцію `erf`, яка тісно пов'язана з функцією стандартного нормального розподілу і яку обчислюють через відповідну поліноміальну апроксимацію.

**Приклад 7.5.8.** Експерт із питань установлення батьківства стверджує в суді, що тривалість (у днях) вагітності у людей має приблизно нормальній розподіл із параметрами  $\mu = 270$  і  $\sigma^2 = 100$ . Відповідач у суді довів, що він перебував за межами країни в період, який почався за 290 днів та завершився за 240 днів до дати народження дитини.

Якщо покласти, що відповідач — справді батько дитини, імовірність, що мати могла народити дитину від цього батька, становить

$$\begin{aligned}\mathbb{P}_X((X > 290) \cup (X < 240)) &= \mathbb{P}_X(X > 290) + \mathbb{P}_X(X < 240) \\ &= \mathbb{P}_X\left(\frac{X - 270}{10} > 2\right) + \mathbb{P}_X\left(\frac{X - 270}{10} < -3\right) \\ &= 1 - \Phi(2) + \Phi(-3) \approx 0.0241,\end{aligned}$$

тобто дуже мала.  $\square$

**Зауваження 7.5.9.** Застосування перетворення (7.5.1) до нормальної величини  $X \sim N(\mu, \sigma^2)$  знову дає нормальну величину  $Y \sim N(a\mu + b, a^2\sigma^2)$ , що безпосередньо випливає з (7.5.2):

$$f_Y(y) = f_X\left(\frac{y - b}{a}\right) \cdot \frac{1}{a} = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(\frac{y-b}{a}-\mu)^2}{2\sigma^2}} \cdot \frac{1}{a} = \frac{1}{\sqrt{2\pi}a\sigma} e^{-\frac{(y-(a\mu+b))^2}{2a^2\sigma^2}}. \quad \square$$

**Твердження 7.5.10.** Якщо  $X \sim N(\mu, \sigma^2)$ , то  $\mathbb{E}[X] = \mu$ , а  $\text{Var}(X) = \sigma^2$ .

*Доведення.* Сподівання та дисперсію рівномірного розподілу можна обчислити за визначенням, а можна скористатися тим, що  $X = \mu + \sigma Z$ , де  $Z \sim N(0, 1)$ , та Твердженням 6.4.6:

$$\begin{aligned}\mathbb{E}[X] &= \mathbb{E}[\mu + \sigma Z] = \mu, \\ \text{Var}(X) &= \text{Var}(\mu + \sigma Z) = \sigma^2.\end{aligned} \quad \square$$

**Зауваження 7.5.11.** У загальному випадку до будь-якої величини  $X$  зі сподіванням  $\mu$  та дисперсією  $\sigma^2$  можна застосувати процедуру стандартизації:

$$Z = \frac{X - \mu}{\sigma}. \quad (7.5.7)$$

Нескладно бачити, що в цьому випадку

$$\begin{aligned}\mathbb{E}[Z] &= \mathbb{E}\left[\frac{X - \mu}{\sigma}\right] = 0, \\ \text{Var}(Z) &= \text{Var}\left(\frac{X - \mu}{\sigma}\right) = \frac{1}{\sigma^2} \text{Var}(X) = 1.\end{aligned}$$

Таку стандартизовану величину часто називають *Z-оцінкою* (Z-score).  $\square$

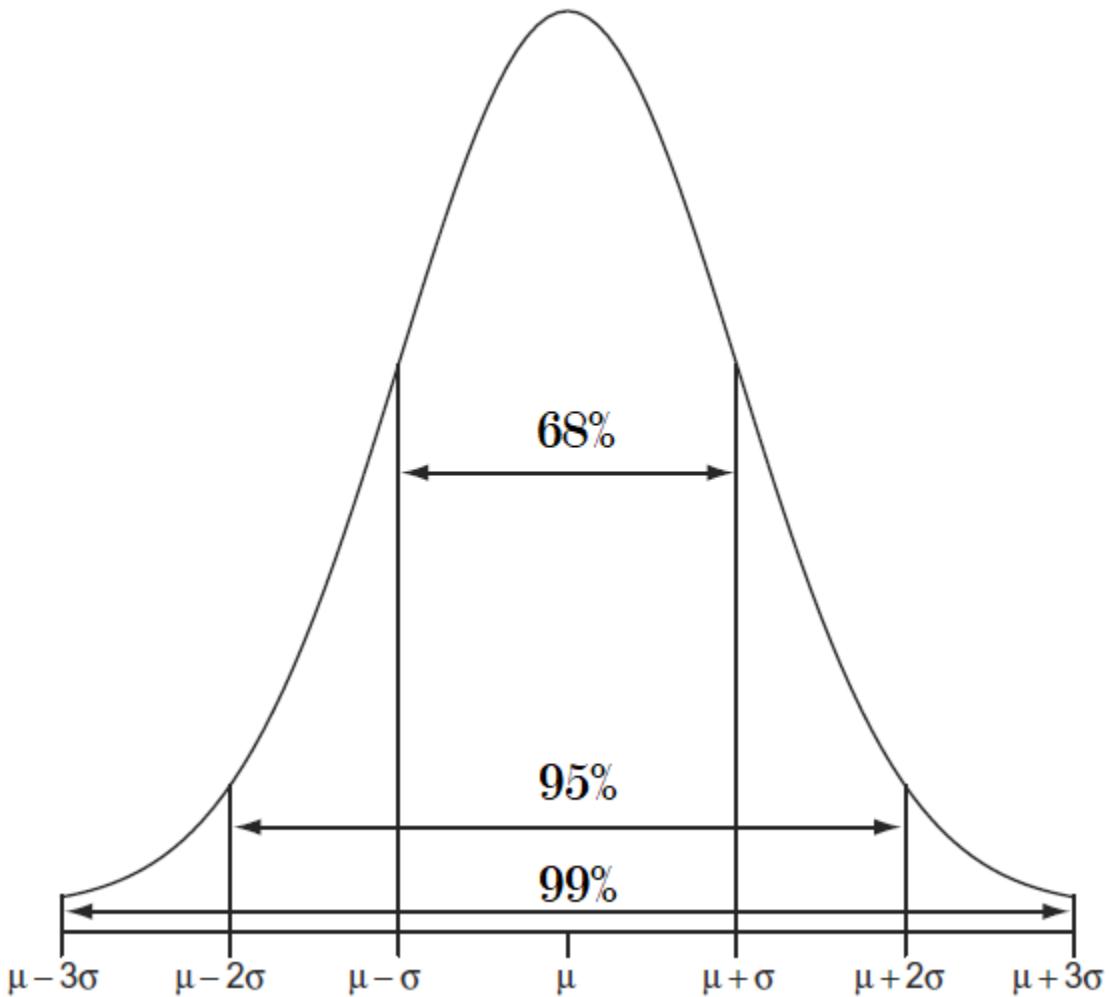


Рис. 7.5.2.: Ілюстрація ймовірностей потрапляння випадкової нормальної величини в проміжки відстанню 1, 2 і 3 середньоквадратичні відхилення від сподівання ([2], Рис. 6.11)

Для нормального розподілу справедливі такі підрахунки (Рис. 7.5.2):

$$\begin{aligned}
 \mathbb{P}_X (|X - \mu| \leq \sigma) &= \mathbb{P}_X \left( \left| \frac{X - \mu}{\sigma} \right| \leq 1 \right) = \mathbb{P}_Z (-1 \leq Z \leq 1) \\
 &= \Phi(1) - \Phi(-1) = 2\Phi(1) - 1 \approx 0.68, \\
 \mathbb{P}_X (|X - \mu| \leq 2\sigma) &= \mathbb{P}_X \left( \left| \frac{X - \mu}{\sigma} \right| \leq 2 \right) = \mathbb{P}_Z (-2 \leq Z \leq 2) \\
 &= \Phi(2) - \Phi(-2) = 2\Phi(2) - 1 \approx 0.95, \\
 \mathbb{P}_X (|X - \mu| \leq 3\sigma) &= \mathbb{P}_X \left( \left| \frac{X - \mu}{\sigma} \right| \leq 3 \right) = \mathbb{P}_Z (-3 \leq Z \leq 3) \\
 &= \Phi(3) - \Phi(-3) = 2\Phi(3) - 1 \approx 0.997.
 \end{aligned}$$

Ці значення можна використовувати для швидкого приблизного підрахунку відповідних імо-

вірностей. Зокрема, під час великого числа повторень деякого експерименту, результатом якого є стандартна нормальнна величина, 99.7% усіх результатів лежатимуть на відстані до трьох середньоквадратичних відхилень від сподівання (на проміжку  $[-3\sigma; 3\sigma]$ ). Часто відповідне правило називають «правилом трьох сигм».

## 7.6. Доведення окремих тверджень

Доведімо критерій вимірності, викладений у Теоремі 7.3.1. Для цього спочатку розглянемо декілька лем.

**Лема 7.6.1.** Нехай деяка функція  $X$  діє з вимірного простору  $(\Omega, \mathcal{A})$  у вимірний простір  $(\Omega', \mathcal{A}')$ . Тоді справедливі такі твердження:

- (i)  $X^{-1}((A')^c) = (X^{-1}(A'))^c$  для деякої  $A' \in \mathcal{A}'$ ;
- (ii)  $X^{-1}(\bigcup_{i \in I} A'_i) = \bigcup_{i \in I} X^{-1}(A'_i)$  для деяких  $A'_i \in \mathcal{A}'$ ,  $i \in I$ , де  $I$  — деяка зліченна множина індексів;
- (iii) якщо  $A' \subseteq B'$ ,  $A', B' \in \mathcal{A}'$ , то  $X^{-1}(A') \subseteq X^{-1}(B')$ .

*Доведення.* (i) За визначенням прообразу,

$$X^{-1}((A')^c) = \{\omega : X(\omega) \in (A')^c\} = \{\omega : X(\omega) \notin A'\} .$$

З іншого боку,

$$(X^{-1}(A'))^c = (\{\omega : X(\omega) \in A'\})^c = \{\omega : X(\omega) \notin A'\} .$$

Як можна бачити, це одна й та сама множина;

- (ii) за визначенням прообразу,

$$X^{-1}\left(\bigcup_{i \in I} A'_i\right) = \left\{\omega : X(\omega) \in \bigcup_{i \in I} A'_i\right\} = \bigcup_{i \in I} \{\omega : X(\omega) \in A'_i\} .$$

З іншого боку,

$$\bigcup_{i \in I} X^{-1}(A'_i) = \bigcup_{i \in I} \{\omega : X(\omega) \in A'_i\} .$$

Знову ж таки, як можна бачити, це одна й та сама множина;

- (iii) за визначенням прообразу,

$$X^{-1}(A') = \{\omega : X(\omega) \in A'\} .$$

Якщо  $A' \subseteq B'$ , то це означає, що  $X(\omega) \in A'$  має наслідком  $X(\omega) \in B'$ , а відтак

$$X^{-1}(A') = \{\omega : X(\omega) \in A'\} \subseteq \{\omega : X(\omega) \in B'\} = X^{-1}(B') .$$

□

**Лема 7.6.2.** Нехай деяка функція  $X$  діє з вимірного простору  $(\Omega, \mathcal{A})$  у вимірний простір  $(\Omega', \mathcal{A}')$ .

Нехай  $\mathcal{C} = X^{-1}(\mathcal{A}')$ . Тоді  $\mathcal{C}$  є  $\sigma$ -алгеброю.

*Доведення.* Доведімо всі три властивості  $\sigma$ -алгебр із Визначення 1.5.8:

- (i) потрібно показати, що  $\Omega \in X^{-1}(\mathcal{A}')$ . За (7.3.1), до прообразу  $\mathcal{A}'$  входять усі множини, обра-  
зами яких є множини з  $\mathcal{A}'$ , у тому числі  $\Omega' \in \mathcal{A}'$ . Прообразом  $\Omega'$  є

$$X^{-1}(\Omega') = \{\omega : X(\omega) \in \Omega'\} ,$$

тобто будь-які елементарні події  $\omega \in \Omega$ . Таким чином,  $\Omega \in X^{-1}(\mathcal{A}')$ ;

- (ii) нехай  $A' \in \mathcal{A}'$ . Оскільки  $\mathcal{A}'$  є  $\sigma$ -алгеброю, то разом із  $A'$  до неї належить також  $(A')^c$ . Тоді  $X^{-1}(A') \in \mathcal{C}$  і  $X^{-1}((A')^c) \in \mathcal{C}$ . За Лемою 7.6.1,  $X^{-1}((A')^c) = (X^{-1}(A'))^c$ . Таким чином, маємо, що одночасно  $X^{-1}(A') \in \mathcal{C}$  і  $(X^{-1}(A'))^c \in \mathcal{C}$ ;
- (iii) нехай  $A'_i \in \mathcal{A}'$ ,  $i \in I$ , де  $I$  — деяка зліченна множина індексів. Тоді  $X^{-1}(A'_i) \in \mathcal{C}$ ,  $i \in I$ . Потрібно показати, що  $\bigcup_{i \in I} X^{-1}(A'_i) \in \mathcal{C}$ . За Лемою 7.6.1,  $\bigcup_{i \in I} X^{-1}(A'_i) = X^{-1}(\bigcup_{i \in I} A'_i)$ . Але оскільки  $\mathcal{A}'$  є  $\sigma$ -алгеброю, то  $\bigcup_{i \in I} A'_i \in \mathcal{A}'$ , а відтак  $X^{-1}(\bigcup_{i \in I} A'_i) \in \mathcal{C}$ .

□

**Лема 7.6.3.** Нехай деяка функція  $X$  діє з вимірного простору  $(\Omega, \mathcal{A})$  у вимірний простір  $(\Omega', \mathcal{A}')$ .

Нехай  $\mathcal{C}'$  — клас підмножин  $\Omega'$  такий, що прообрази всіх його елементів вимірні:

$$\mathcal{C}' = \{A' \subseteq \Omega' : X^{-1}(A') \in \mathcal{A}\} .$$

Тоді  $\mathcal{C}'$  є  $\sigma$ -алгеброю.

*Доведення.* Доведення дуже схоже на доведення Леми 7.6.2. □

**Лема 7.6.4.** Нехай деяка функція  $X$  діє з вимірного простору  $(\Omega, \mathcal{A})$  у вимірний простір  $(\Omega', \mathcal{A}')$ .

Нехай  $\mathcal{C}'$  — клас множин, що породжує  $\mathcal{A}'$ :  $\mathcal{A}(\mathcal{C}') = \mathcal{A}'$ . Тоді прообразом  $\mathcal{A}'$  є  $\sigma$ -алгебра, поро-  
джена прообразами множин із  $\mathcal{C}'$ :  $\mathcal{A}(X^{-1}(\mathcal{C}')) = X^{-1}(\mathcal{A}')$ .

*Доведення.* Для доведення рівності доведемо включення в обидва боки.

Спочатку доведімо  $\mathcal{A}(X^{-1}(\mathcal{C}')) \subseteq X^{-1}(\mathcal{A}')$ . Дійсно, оскільки  $\mathcal{C}' \subseteq \mathcal{A}'$ , то за Лемою 7.6.1 маємо, що  $X^{-1}(\mathcal{C}') \subseteq X^{-1}(\mathcal{A}')$ . За Лемою 7.6.2,  $X^{-1}(\mathcal{A}')$  є  $\sigma$ -алгеброю, а тому, за Лемою 1.5.15,

$$\mathcal{A}(X^{-1}(\mathcal{C}')) \subseteq \mathcal{A}(X^{-1}(\mathcal{A}')) = X^{-1}(\mathcal{A}') .$$

Тепер доведімо  $\mathcal{A}(X^{-1}(\mathcal{C}')) \supseteq X^{-1}(\mathcal{A}')$ . Розглянемо клас множин

$$\mathcal{C}^* = \{A' \subseteq \Omega' : X^{-1}(A') \in \mathcal{A}(X^{-1}(\mathcal{C}'))\} .$$

За Лемою 7.6.3,  $\mathcal{C}^*$  є  $\sigma$ -алгеброю, а тому  $\mathcal{C}' \subseteq \mathcal{C}^*$  (адже  $\mathcal{C}' \subseteq \mathcal{A}(\mathcal{C}')$ , а  $\mathcal{A}(\mathcal{C}')$  — найменша з  $\sigma$ -алгебр, що містить  $\mathcal{C}'$ ). За Лемою 1.5.15,  $\mathcal{A}(\mathcal{C}') \subseteq \mathcal{C}^*$ , а відтак, за Лемою 7.6.1,  $X^{-1}(\mathcal{A}') \subseteq X^{-1}(\mathcal{C}^*)$ . З іншого боку,  $X^{-1}(\mathcal{C}^*) \subseteq \mathcal{A}(X^{-1}(\mathcal{C}'))$  за визначенням  $\mathcal{C}^*$ . Із цього випливає  $X^{-1}(\mathcal{A}') \subseteq \mathcal{A}(X^{-1}(\mathcal{C}'))$ .

Разом включення в обидва боки доводять рівність  $\mathcal{A}(X^{-1}(\mathcal{C}')) = X^{-1}(\mathcal{A}')$ . □

Тепер можемо довести Теорему 7.3.1.

*Доведення.* Нехай  $X : (\Omega, \mathcal{A}) \rightarrow (\Omega', \mathcal{A}')$ . Тоді  $X$  буде вимірною, якщо  $X^{-1}(\mathcal{A}') \subseteq \mathcal{A}$ .

Нехай  $\mathcal{C}'$  — клас множин, що породжує  $\mathcal{A}'$ :  $\mathcal{A}(\mathcal{C}') = \mathcal{A}'$ . Покладімо, що  $X^{-1}(\mathcal{C}') \subseteq \mathcal{A}$ . Тоді, за Лемою 1.5.15,  $\mathcal{A}(X^{-1}(\mathcal{C}')) \subseteq \mathcal{A}$ . За Лемою 7.6.4,  $\mathcal{A}(X^{-1}(\mathcal{C}')) = X^{-1}(\mathcal{A}')$ , тобто  $X^{-1}(\mathcal{A}') \subseteq \mathcal{A}$ . □

Доведімо тепер Твердження 7.3.6.

**Доведення.** Спочатку доведімо вимірність випадкового вектора на основі вимірності окремих випадкових функцій.

Як було зазначено в Лекції 1, Борелеву  $\sigma$ -алгебру  $\mathcal{B}^k$  породжують множини виду

$$S_x = \left\{ y \in \mathbb{R}^k : y_i \leq x_i, i = 1, \dots, n \right\}.$$

За Теоремою 7.3.1, достатньо перевірити вимірність випадкового вектора на множинах із цього класу. Тоді

$$\begin{aligned} \mathbf{X}^{-1}(S_x) &= \{ \omega : \mathbf{X}(\omega) \in S_x \} \\ &= \{ \omega : X_1(\omega) \leq x_1 \cap \dots \cap X_n(\omega) \leq x_n \} \\ &= \{ \omega : X_1(\omega) \leq x_1 \} \cap \dots \cap \{ \omega : X_n(\omega) \leq x_n \} \\ &= (X_1^{-1}((-\infty; x_1])) \cap \dots \cap (X_n^{-1}((-\infty; x_n])) . \end{aligned}$$

Якщо кожна  $X_i$  є вимірною, то всі  $X_i^{-1}((-\infty; x_i]) \in \mathcal{A}$ ,  $i = 1, \dots, k$ , а відтак і їх перетин також належить  $\mathcal{A}$ , тобто  $\mathbf{X}^{-1}(S_x) \in \mathcal{A}$ .

Тепер покладімо, що вимірним є випадковий вектор. Зафіксуємо  $x_j = x$  і спрямуймо всі інші координати в нескінченність:  $x_1 = \dots = x_{j-1} = x_{j+1} = \dots = x_k \rightarrow \infty$ . Тоді матимемо множину

$$S_{x_j} = \left\{ y \in \mathbb{R}^k : y_j \leq x, y_i \in \mathbb{R}, i \neq j \right\}.$$

Оскільки  $\mathbf{X}$  є вимірною, то

$$\mathbf{X}^{-1}(S_{x_j}) = \{ \omega : X_j(\omega) \leq x, X_i(\omega) \in \mathbb{R}, i \neq j \} = \{ \omega : X_j(\omega) \leq x \} \in \mathcal{A}.$$

Іншими словами,  $X_j$  є вимірною за Теоремою 7.3.1, адже вона є вимірною для всіх променів, що породжують  $\mathcal{B}$ . Відповідні міркування справедливі для будь-якого  $j = 1, \dots, k$ .  $\square$

## 8. Окремі застосування неперервних випадкових величин

### 8.1. Інтегральне перетворення ймовірностей

Один із важливих результатів, пов'язаних із неперервними випадковими величинами, є спосіб перетворення стандартного рівномірного розподілу у деякий довільний розподіл.

**Теорема 8.1.1** (Інтегральне перетворення ймовірностей (Probability integral transform)). Нехай випадкова величина  $X : (\Omega, \mathcal{A}, \mathbb{P}) \rightarrow \mathbb{R}$  має функцію розподілу  $F_X(x) = \mathbb{P}_X(X \leq x)$ . Тоді:

- (i) якщо  $F_X$  неперервна, то випадкова величина  $U = F_X(X)$  має розподіл  $U \sim \text{U}((0; 1))$ ;
- (ii) якщо  $U \sim \text{U}((0; 1))$ , то випадкова величина  $F_X^\dagger(U)$ , де

$$F_X^\dagger(u) = \inf \{x \in \mathbb{R} : F_X(x) \geq u\} , \quad (8.1.1)$$

має розподіл  $\mathbb{P}_X^1$ .

*Доведення.* (i) За визначенням функції розподілу,

$$F_U(u) = \mathbb{P}_U(U \leq u) = \mathbb{P}_X(F_X(X) \leq u) = \mathbb{P}_X(X \in F_X^{-1}((-\infty; u])) .$$

Оскільки  $F_X(X) \in [0; 1]$ , то маємо  $F_U(u) = 0$  для  $u \leq 0$  та  $F_U(u) = 1$  для  $u \geq 1$ .

Розгляньмо  $u \in (0; 1)$ . Тоді

$$F_X^{-1}((-\infty; u]) = \{x \in \mathbb{R} : F_X(x) \leq u\} .$$

Нехай  $x_{\max} = \max \{x \in \mathbb{R} : F_X(x) \leq u\}$ . Цей елемент існує, оскільки:

- функція  $F_X$  неперервна: якби вона мала розрив, то для деяких  $u$  прообразом була б порожня множина;
- ми поклали  $u < 1$ : цілком очевидно, що найбільшого елементу, для якого  $F_X(X) \leq 1 < u$ , не існує (він нескінченно великий).

Тоді в загальному випадку

$$F_X^{-1}((-\infty; u]) = (-\infty; x_{\max}] ,$$

оскільки функція  $F_X$  неспадна. Зрозуміло, що якщо функція  $F_X$  є строго зростаючою, вона є взаємно однозначною, а тому  $F_X^{-1}((-\infty; u]) = (-\infty; F_X^{-1}(u)]$ , тобто в цьому випадку  $x_{\max} = F_X^{-1}(u)$ .

Отже в підсумку маємо, що для  $u \in (0; 1)$

$$F_U(u) = \mathbb{P}_X(X \in (-\infty; x_{\max}]) = \mathbb{P}_X(X \leq x_{\max}) = F_X(x_{\max}) .$$

<sup>1</sup> Якщо функція  $F_X$  є строго зростаючою, вона є взаємно однозначною, а тому  $F_X^\dagger(u) = F_X^{-1}(u)$ . Вибір  $\inf$  замість  $\min$  обумовлено, зокрема, тим, що тоді  $F_X^\dagger(0) = -\infty$ .

Оскільки  $x_{\max} \in F_X^{-1}(u)$ <sup>2</sup> (або навіть  $x_{\max} = F_X^{-1}(u)$  для взаємно однозначної функції),  $F_X(x_{\max}) = u$  через неспадність функції, а відтак  $F_U(u) = u$ .

Отже

$$F_U(u) = \begin{cases} 0, & u < 0 \\ u, & 0 \leq u < 1 \\ 1, & u \geq 1 \end{cases}$$

тобто  $F_U(u)$  є нічим іншим, як функцією розподілу для стандартного рівномірного розподілу (7.1.2);

(ii) нехай  $Z = F_X^\dagger(U)$ ,  $U \sim \text{U}((0; 1])$ . Для кожного  $z \in \mathbb{R}$  маємо:

$$F_Z(z) = \mathbb{P}_Z(Z \leq z) = \mathbb{P}_U(F_X^\dagger(U) \leq z) = \mathbb{P}\left(\left\{\omega : F_X^\dagger(U(\omega)) \leq z\right\}\right).$$

Якщо нам вдасться показати, що

$$\left\{\omega : F_X^\dagger(U(\omega)) \leq z\right\} = \{\omega : U(\omega) \leq F_X(z)\},$$

то ми матимемо

$$F_Z(z) = \mathbb{P}(\{\omega : U(\omega) \leq F_X(z)\}) = \mathbb{P}_U(U \leq F_X(z)) = \begin{cases} 0, & F_X(z) < 0 \\ F_X(z), & F_X(z) \in [0; 1] \\ 1, & F_X(z) \geq 1 \end{cases},$$

тобто фактично що  $F_Z(z) = F_X(z)$ .

І справді,

$$\left\{\omega : F_X^\dagger(U(\omega)) \leq z\right\} = \{\omega : \inf\{x : F_X(x) \geq U(\omega)\} \leq z\} \subseteq \{\omega : U(\omega) \leq F_X(z)\},$$

оскільки якщо  $U(\omega) \leq F_X(x)$ , і найменший такий  $x \leq z$ , то  $U(\omega) \leq F_X(z)$  через неспадність функції  $F_X$ .

З іншого боку, якщо  $\{\omega : U(\omega) \leq F_X(z)\}$ , то  $z \in \{x : F_X(x) \geq U(\omega)\}$ , а відтак

$$z \geq \inf\{x : F_X(x) \geq U(\omega)\} = F_X^\dagger(U(\omega)),$$

тобто

$$\{\omega : U(\omega) \leq F_X(z)\} \subseteq \left\{\omega : F_X^\dagger(U(\omega)) \leq z\right\}.$$

Включення в обидва боки дає рівність відповідних множин.

□

Теорему 8.1.1 на практиці можна використовувати у двох випадках:

- для перетворення вибірки значень з невідомого розподілу у вибірку з рівномірним розподілом, потреба в чому виникає у деяких статистичних методах;

<sup>2</sup>Саме тому ми використали  $\max$ , а не  $\sup$ .

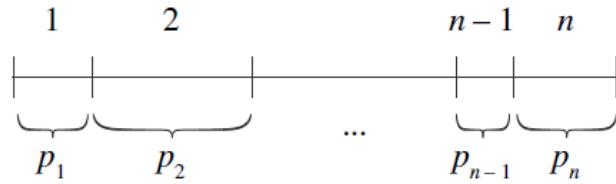


Рис. 8.1.1.: Ілюстрація до генерування значень дискретної випадкової величини ([1], Рис. 5.9)

- для генерування випадкових чисел, розподілених за певним законом. Для цього достатньо згенерувати рівномірно розподілені випадкові числа на інтервалі  $(0; 1]$ <sup>3</sup> і застосувати до них функцію  $F_X^\dagger$ , відповідну бажаному розподілу.

**Приклад 8.1.2** (Генерування випадкових чисел із дискретним розподілом). Нехай стоять задача згенерувати значення деякої випадкової величини  $X$  зі скінчненим носієм  $\text{supp}(X) = \{x_1, \dots, x_n\}$  та функцією ймовірності  $p_X(x_i) = p_i$ ,  $i = 1, \dots, n$ . Функція розподілу такої величини має вид

$$F_X(x) = \begin{cases} 0, & x < x_1 \\ p_1, & x \in [x_1; x_2) \\ p_1 + p_2, & x \in [x_2; x_3) \\ \vdots \\ \sum_{i=1}^{n-1} p_i, & x \in [x_{n-1}; x_n) \\ 1, & x \geq x_n \end{cases}.$$

Тоді функція (8.1.1) для даного прикладу матиме вид

$$F_X^\dagger(u) = \begin{cases} x_1, & u < p_1 \\ x_2, & u \in [p_1; p_1 + p_2) \\ x_3, & u \in [p_1 + p_2; p_1 + p_2 + p_3) \\ \vdots, \\ x_n, & u \geq \sum_{i=1}^{n-1} p_i \end{cases}.$$

Таким чином, спочатку генеруємо рівномірно розподілене на інтервалі  $(0; 1]$  випадкове число  $u$ , а потім застосовуємо до нього функцію  $F_X^\dagger(u)$ . Ілюстративно це показано на Рис. 8.1.1, де  $x_i = i$ , а інтервал  $(0; 1]$  розбито на підінтервали. Потрапляння числа  $u$  у відповідний підінтервал визначає індекс  $i$  значення  $x_i$  випадкової величини  $X$ .

Найчастіше саме цим результатом користуються, коли хочуть згенерувати величину з розподілом Бернуллі з параметром  $p$ . Спочатку генерують випадкове число  $x$  на інтервалі  $(0; 1]$ . Якщо  $x \leq p$ , то повертають одиницю, інакше повертають нуль.  $\square$

<sup>3</sup>У будь-якій системі програмування існують функції, які можуть це зробити — так звані генератори псевдовипадкових чисел (pseudorandom number generators). Наприклад, у мові Python для цього можна використати функцію `random` із модуля `numpy.random`.

**Приклад 8.1.3.** Розгляньмо логістичний розподіл із Прикладу 7.5.1.

Відповідні функцію розподілу та щільність можна переписати так:

$$F(x) = \frac{1}{1 + e^{-x}}, \quad f(x) = \frac{e^{-x}}{(1 + e^{-x})^2}.$$

Функція розподілу є взаємно однозначною, а відтак

$$F^\dagger(u) = F^{-1}(u) = \ln\left(\frac{u}{1-u}\right), \quad u \in (0; 1)^4.$$

Відповідно до Теореми 8.1.1, якщо  $U \sim U((0; 1])$ , то випадкова величина  $Y = \ln\left(\frac{U}{1-U}\right)$  має логістичний розподіл.

Цей факт можна проілюструвати за допомогою відповідної симуляції за *методом Монте-Карло* (Monte Carlo simulation)<sup>5</sup>, суть якого полягає у відтворенні деякого випадкового експерименту велику кількість разів і використання одержаних результатів для дослідження властивостей процесу, який їх породжує. У нашому випадку ми згенеруємо  $10^5$  випадкових чисел на інтервалі  $(0; 1)$ , застосуємо до них функцію  $F^{-1}$  та побудуємо відповідну гістограму (Лістинг 8.1.1). Як можна бачити з Рис. 8.1.2, гістограма дуже близька до аналітичного графіка щільності розподілу.

Лістинг 8.1.1: Код для Прикладу 8.1.3

```

import matplotlib.pyplot as plt
from matplotlib import rc
import numpy as np
from numpy.random import seed
from scipy.stats import uniform
5

rc('text', usetex=True)
rc('text.latex', unicode=True)
rc('text.latex', preamble=r'\usepackage[utf8]{inputenc}')
rc('text.latex', preamble=r'\usepackage[ukrainian]{babel}')

seed(100)
u = uniform.rvs(size=10**5)
x = np.sort(np.log(u/(1 - u)))
15

pdf = np.exp(x) / (1 + np.exp(x))**2

fig, ax1 = plt.subplots()

20
ax1.hist(x, bins=100, alpha=0.5)
ax1.set_ylabel("Гістограма", fontsize=15)
ax1.tick_params(axis='y', labelsize=10)
ax1.set_ylim(0)
ax1.yaxis.label.set_color("blue")
25

```

<sup>4</sup>Зверніть увагу, що для жодного  $x \in \mathbb{R}$  функція  $F(x)$  не набуває значень 0 або 1. Вона їх набуває тільки в граничних переходах, а тому функція  $F^\dagger$  визначена тільки на інтервалі  $(0; 1)$ .

<sup>5</sup>Відповідну назву запропонував угорсько-американський математик Джон фон Нойманн (John von Neumann, 1903–1957) для опису першої симуляції такого роду. Це сталося в процесі дослідження поведінки нейтронів у науковій лабораторії Лос-Аламос під час Другої світової війни. Ця задача була занадто складною, щоб розв'язати її аналітично, і було запропоновано моделювати відповідний експеримент на комп'ютері. Назва «Монте-Карло», вочевидь, походить від назви курорту в Монако, відомого своїми гральними закладами.

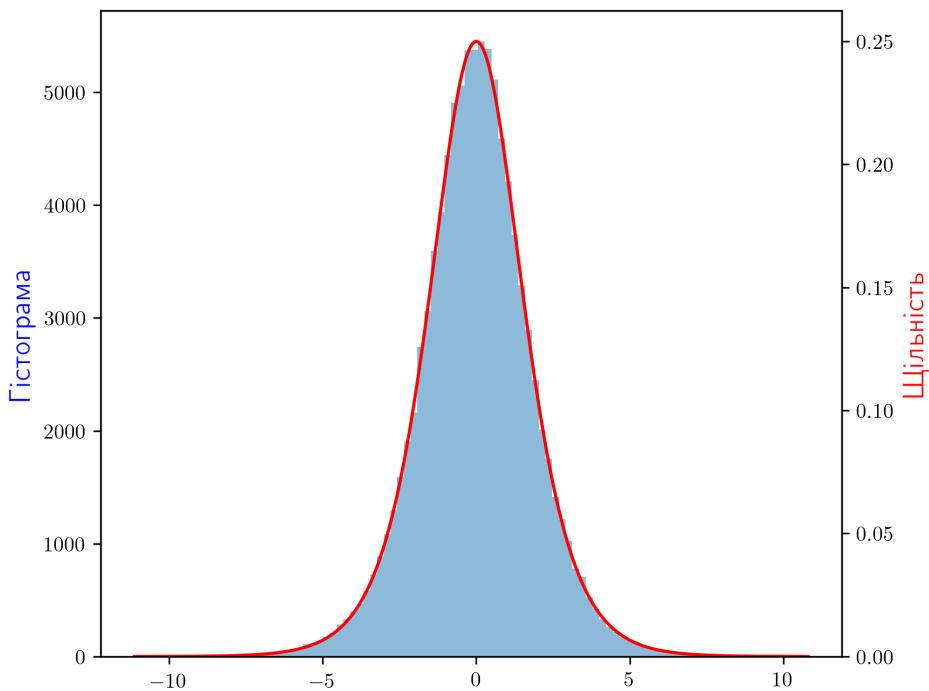


Рис. 8.1.2.: Графік щільності логістичного розподілу та гістограма  $10^5$  результатів застосування оберненої функції логістичного розподілу до випадкових чисел для Прикладу 8.1.3

```

ax2 = ax1.twinx()

ax2.plot(x, pdf, color="red")
ax2.set_ylabel("Щільність", fontsize=15)
50 ax2.tick_params(axis='y', labelsize=10)
ax2.set ylim(0)
ax2.yaxis.label.set_color("red")

fig.tight_layout()
55 plt.savefig("../images/Lecture 8/logistic.png", dpi=300)

```

□

## 8.2. Бета-розподіл та елементи Беєсівського виведення

Розгляньмо дуже цікавий розподіл та з'ясуємо, як його можна використовувати для оновлення інформації про розподіл деякого невідомого параметра.

**Визначення 8.2.1.** Випадкова величина  $X$  має *Бета-розподіл* (Beta distribution) із параметрами  $a > 0$ ,  $b > 0$ ,  $X \sim \text{Beta}(a, b)$ , якщо її щільність дорівнює

$$f(x) = \frac{1}{B(a, b)} x^{a-1} (1-x)^{b-1} \cdot \mathbb{1}_{\{0 < x < 1\}} , \quad (8.2.1)$$

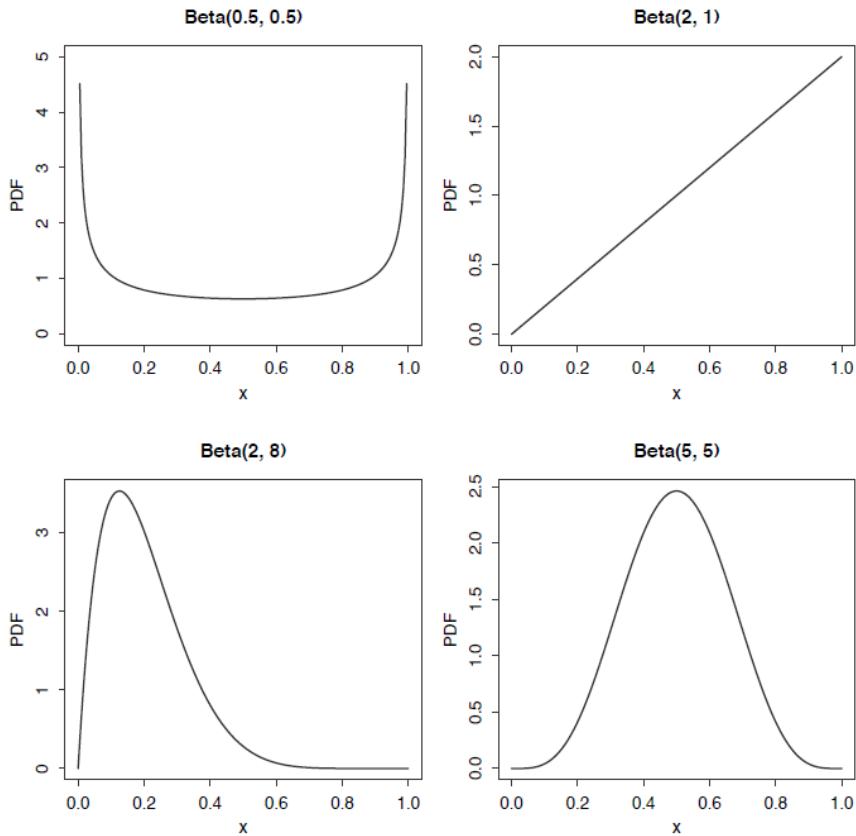


Рис. 8.2.1.: Щільності Бета-розподілу з різними параметрами  $a$  і  $b$  ([1], Рис. 8.4)

де  $B(a, b)$  — бета-функція виду

$$B(a, b) = \int_0^1 x^{a-1} (1-x)^{b-1} dx .$$

□

Інтеграл щільності Бета-розподілу тривіально дорівнює 1, адже значення бета-функції  $B(a, b)$  за визначенням і є відповідним інтегралом  $\int_0^1 x^{a-1} (1-x)^{b-1} dx$ . Функція розподілу для Бета-розподілу в загальному випадку не має простого аналітичного запису.

Як можна помітити, розподіл Beta(1, 1) тотожний стандартному рівномірному розподілу.

Графіки щільностей Бета-розподілів із різними параметрами наведено на Рис. 8.2.1. У вічі впадають дві загальні тенденції:

- якщо  $a < 1$ ,  $b < 1$ , то щільність опукла. Якщо ж  $a > 1$ ,  $b > 1$ , то щільність вгнута;
- якщо  $a = b$ , то щільність симетрична відносно 0.5. Якщо  $a > b$ , то щільність скошена вліво, а якщо  $a < b$ , то щільність скошена вправо.

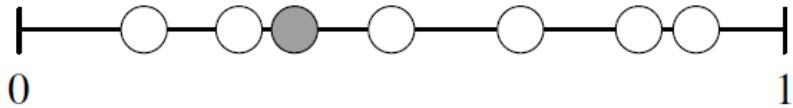


Рис. 8.2.2.: Ілюстрація до Зауваження 8.2.2 ([1], Рис. 8.5)

**Зауваження 8.2.2.** Як відомо з курсу «Математичний аналіз», для натуральних  $a$  і  $b$

$$B(a, b) = \frac{(a-1)!(b-1)!}{(a+b-1)!}.$$

Якщо трішки переписати цей результат, то можна помітити, що

$$\frac{(a-1)!(b-1)!}{(a+b-1)!} = \frac{(a-1)!(b-1)!}{(a+b-1) \cdot ((a-1)+(b-1))!} = \frac{1}{a+b-1} \cdot \frac{1}{\binom{(a-1)+(b-1)}{b-1}}.$$

Позначивши  $(a-1)+(b-1) = n$ ,  $b-1 = k$ , маємо

$$\binom{n}{k} B(n-k+1, k+1) = \int_0^1 \binom{n}{k} p^n (1-p)^{n-k} dp = \frac{1}{n+1}.$$

Цей результат можна вивести за допомогою експерименту, який запропонував Томас Беес. Нехай маємо  $n$  білих куль і одну сіру. Якщо ці кулі кинути випадково на інтервал  $(0; 1]$ , то координати цих куль будуть мати стандартний рівномірний розподіл (Рис. 8.2.2).

Нехай  $X = \text{«число білих куль зліва від сірої»}$ . Це є дискретна випадкова величина з носієм  $\text{supp}(X) = \{0, 1, \dots, n\}$ . Також нехай  $B = \text{«координата сірої кулі}$ ,  $B \sim U((0; 1])$ . За умови, що  $B = p$ , число білих куль зліва від  $p$  має біномний розподіл  $X \sim \text{Binom}(n, p)$ , оскільки кожна куля зліва — це успіх, який може статися з імовірністю  $p$ .

Розгляньмо випадкову величину  $X_k = \mathbb{1}\{X = k\} \sim \text{Bern}(\mathbb{P}_X(X = k))$ . Її сподівання дорівнює  $\mathbb{E}[X_k] = \mathbb{P}_X(X = k)$ . Згідно з законом ітерованих сподівань, який ми розглядалимо далі в цьому курсі,

$$\mathbb{P}_X(X = k) = \mathbb{E}[X_k] = \mathbb{E}[\mathbb{E}[X_k | B]] = \int_{-\infty}^{\infty} \mathbb{P}_X(X = k | B = p) \cdot f_B(p) dp = \int_0^1 \binom{n}{k} p^n (1-p)^{n-k} dp.$$

З іншого боку, розгляньмо ситуацію, коли всі  $n+1$  куль білі. Киньмо їх випадково на  $(0; 1]$ . Оберімо одну з куль навмисно та розфарбуймо її в сірий колір. Згідно з класичним визначенням імовірності, імовірність того, що такою кулею виявиться куля з номером  $k$ ,  $k = 0, 1, \dots, n$ , є  $\frac{1}{n+1}$ .

Із еквівалентності відповідних ситуацій і випливає шукана рівність

$$\int_0^1 \binom{n}{k} p^n (1-p)^{n-k} dp = \frac{1}{n+1}.$$

□

Завдяки своїй гнучкості, у статистиці Бета-розподіл часто використовують для моделювання імовірностей невідомих параметрів. Наприклад, нехай ми маємо деяку монетку, але не знаємо, із

якою ймовірністю  $p$  вона випаде гербом. Ми волімо з'ясувати цю ймовірність шляхом підкидання монетки  $n$  разів і аналізу результатів випадів. Згідно з *частотним підходом* до інтерпретації ймовірностей, ми можемо порахувати середню частку гербів із-посеред  $n$  випалих, яка для великих  $n$  буде близька до справжнього значення. Ми це вивчатимемо в наступних лекціях під назвою закону великих чисел.

У рамках великої галузі статистики, що має назву *Беєсівського виведення* (Bayesian inference), невідомий параметр  $p$  розглядають як *випадкову величину*, що має деякий розподіл. Спочатку цей розподіл є *апріорним* (prior), а з надходженням нової інформації його оновлюють і дістають *апостеріорний* (posterior) розподіл. Звісно, такий підхід працює в будь-яких схожих ситуаціях, а тут ми зосередимося на оцінюванні параметра  $p$  для монетки.

Розгляньмо випадок, коли апріорним розподілом є Бета-розподіл:  $p \sim \text{Beta}(a, b)$ , де константи  $a$  і  $b$  відомі. Нехай  $X = \text{«число гербів із-посеред } n \text{ підкидань монетки»}$ . Для деякого фіксованого  $p$  відомо, то  $X | p \sim \text{Binom}(n, p)$ . Тоді для пошуку апостеріорного розподілу можна застосувати Теорему Беєса 3.3.1, але замість функції ймовірності для  $p$  ми використаємо щільність, оскільки  $p$  є неперервною випадковою величиною. Конкретніше,

$$\mathbb{P}_{P|X}(p \in (x; x + dx) | X = k) = \frac{\mathbb{P}_X(X = k | p) \cdot \mathbb{P}_P(p \in (x; x + dx))}{\mathbb{P}_X(X = k)},$$

і, спрямувавши  $dx \rightarrow 0^6$ , маємо

$$\begin{aligned} f_{p|X}(p | X = k) &= \frac{\mathbb{P}_X(X = k | p) f_p(p)}{\mathbb{P}_X(X = k)} \\ &= \frac{\binom{n}{k} p^k (1-p)^{n-k} \cdot \frac{1}{B(a, b)} p^{a-1} (1-p)^{b-1}}{\mathbb{P}_X(X = k)}. \end{aligned}$$

Коли  $a = b = 1$ , ми показали в Завданні 8.2.2, що  $\mathbb{P}_X(X = k) = \frac{1}{n+1}$ . Якщо  $a$  і  $b$  — деякі натуральні числа, то розподіл  $X$  можна вивести за тією ж логікою. У найзагальнішому випадку розподіл  $X$  може бути доволі складним.

Однією з ключових особливостей під час виконання Беєсівського виведення є те, що для обчислення апостеріорного розподілу нам *непотрібно* знати  $\mathbb{P}_X(X = k)$ . Ми можемо взагалі об'єднати всі константи, які не залежать від  $p$ , в одній константі  $C$ :

$$f_{p|X}(p | X = k) = C \cdot p^k (1-p)^{n-k} \cdot p^{a-1} (1-p)^{b-1} = C \cdot p^{a+k-1} (1-p)^{b+n-k-1}.$$

Самостійного значення константа  $C$  не має, вона існує тільки для того, щоб інтеграл щільності дорівнював 1. Часто її опускають і просто пишуть

$$f_{p|X}(p | X = k) \propto p^{a+k-1} (1-p)^{b+n-k-1}.$$

У будь-якому випадку з зовнішнього вигляду відповідної щільності можна зробити висновок, що  $p | X = k \sim \text{Beta}(a+k, b+n-k)$ . Іншими словами, якщо апріорний розподіл є Бета-розподілом, а  $X$  має біномний розподіл за умови  $p$ , то апостеріорний розподіл параметру  $p$  залишається Бета-розподілом. Кажуть, що Бета-розподіл є *спряженим апріорним розподілом* (conjugate prior) для біномного розподілу.

<sup>6</sup>Ми розглянемо цю формулу детальніше в іншій лекції в контексті умовних сподівань.

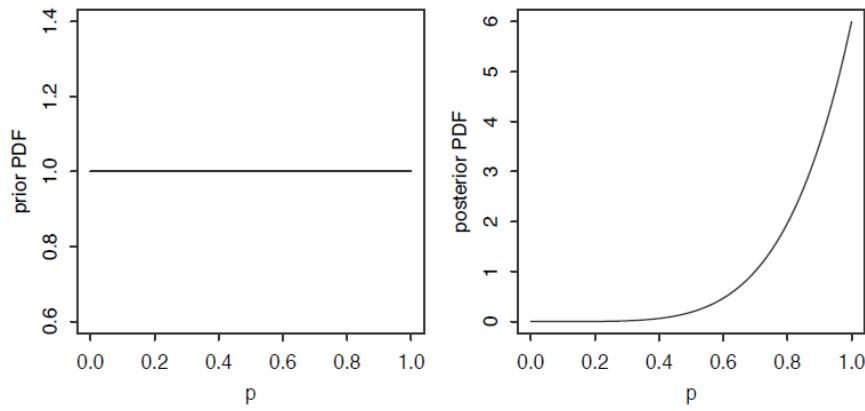


Рис. 8.2.3.: Ілюстрація до Зauważення 8.2.3 ([1], Рис. 8.6)

**Зauważення 8.2.3.** Варто помітити, що формула для оновлення розподілу параметру  $p$  є дуже простою. Ми просто додаємо загальне число успіхів  $k$  до першого параметра Бета-розподілу і загальне число невдач  $n - k$  до другого параметра Бета-розподілу.

Наприклад, нехай наш апіорний розподіл є стандартним рівномірним, тобто  $p \sim \text{Beta}(1, 1)$ . За такого підходу до моделювання ми не надаємо переваги жодному конкретному значенню  $p$ : до появи додаткових фактів вони всі для нас рівномірні. Тоді якщо ми підкинули монетку 5 разів, ми можемо оновити наші уявлення про ймовірність  $p$  шляхом додавання 5 до першого параметру відповідного Бета-розподілу:  $p | X = 5 \sim \text{Beta}(6, 1)$  (Рис. 8.2.3). □

Знайдімо сподівання та дисперсію Бета-розподілу.

**Твердження 8.2.4.** Якщо  $X \sim B(a, b)$ , то  $\mathbb{E}[X] = \frac{a}{a+b}$ , а  $\text{Var}(X) = \frac{ab}{(a+b)^2(a+b+1)}$ .

*Доведення.* За формулою обчислення сподівань маємо

$$\mathbb{E}[X] = \frac{1}{B(a, b)} \int_0^1 x \cdot x^{a-1} (1-x)^{b-1} dx = \frac{1}{B(a, b)} \int_0^1 x^a (1-x)^{b-1} dx = \frac{B(a+1, b)}{B(a, b)} = \frac{a}{a+b},$$

оскільки існує взаємозв'язок між Бета-функцією і Гамма-функцією:

$$B(a, b) = \frac{\Gamma(a)\Gamma(b)}{\Gamma(a+b)},$$

де

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

Звідси, враховуючи властивість  $\Gamma(x+1) = x\Gamma(x)$ , маємо

$$\mathbb{E}[X] = \frac{\Gamma(a+1)\Gamma(b)}{\Gamma(a+b+1)} \cdot \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} = \frac{a}{a+b}.$$

У схожий спосіб

$$\begin{aligned}
 \mathbb{E}[X^2] &= \frac{1}{B(a, b)} \int_0^1 x^2 \cdot x^{a-1} (1-x)^{b-1} dx \\
 &= \frac{1}{B(a, b)} \int_0^1 x^{a+1} (1-x)^{b-1} dx \\
 &= \frac{B(a+2, b)}{B(a, b)} \\
 &= \frac{B(a+2, b)}{B(a+1, b)} \cdot \frac{B(a+1, b)}{B(a, b)} \\
 &= \frac{(a+1)a}{(a+b+1)(a+b)}.
 \end{aligned}$$

Остаточно маємо

$$\text{Var}(X) = \mathbb{E}[X^2] - (\mathbb{E}[X])^2 = \frac{ab}{(a+b)^2(a+b+1)}.$$

□

**Приклад 8.2.5.** Нехай для лікування деякої хвороби виготовили нові ліки, і для перевірки їх ефективності влаштовують клінічні випробування за участю  $n$  пацієнтів. Кожний пацієнт може вилікуватися з імовірністю  $p$  незалежно один від одного. Ми хочемо з'ясувати, чому дорівнює значення  $p$ .

Згідно з Бессівським виведенням, ми вважатимемо, що  $p$  є випадковою величиною, апіорний розподіл якої  $p \sim \text{Beta}(1, 1) = U((0; 1])$ . Нехай після вживання ліків  $usi$  пацієнти одужали. Тоді, згідно з вищепереданими міркуваннями, ми повинні оновити параметри Бета-розподілу. Апостеріорний розподіл буде  $\text{Beta}(1+n, 1)$ .

Тоді апостеріорним сподіванням для  $p$  буде значення  $\mathbb{E}[p | X = n] = \frac{1+n}{(1+n)+1} = \frac{n+1}{n+2}$ . А, наприклад, імовірність того, що  $p > 0.5$ , можна обчислити як відповідний інтеграл:

$$\mathbb{P}_p \left( p \geq \frac{1}{2} \right) = \int_{\frac{1}{2}}^1 \frac{1}{B(1+n, 1)} p^{1+n-1} (1-p)^{1-1} dp = \int_{\frac{1}{2}}^1 (n+1)p^n dp = p^{n+1} \Big|_{\frac{1}{2}}^1 = 1 - \frac{1}{2^{n+1}}.$$

□

### 8.3. Порядкові статистики

Розгляньмо важливий приклад застосування вимірних функцій до випадкового вектора, який не потребує вивчення поняття розподілу такого вектора, про що буде мова в дальших лекціях.

Нехай маємо  $n$  випадкових величин  $X_1, \dots, X_n$ . Відповідно до Твердження 7.3.6, їх можна об'єднати у випадковий вектор  $\mathbf{X} = (X_1, \dots, X_n)^\top$ . Цілком очевидно, що довільна перестановка елементів цього вектора є вимірною функцією. Також вимірними є функції  $\min$  і  $\max$ . Поєднуючи ці функції в певному порядку, можемо визначити такі важливі випадкові величини.

**Визначення 8.3.1.** Нехай  $X_1, \dots, X_n$  — деякі випадкові величини. Тоді *порядковими статистиками*<sup>7</sup> (order statistics) будуть випадкові величини  $X_{(1)}, \dots, X_{(n)}$  такі, що

$$\begin{aligned} X_{(1)} &= \min \{X_1, \dots, X_n\} , \\ X_{(2)} &= \min \{\{X_1, \dots, X_n\} \setminus \{X_{(1)}\}\} , \\ &\vdots \\ X_{(n-1)} &= \min \{\{X_1, \dots, X_n\} \setminus \{X_{(1)}, \dots, X_{(n-2)}\}\} \\ X_{(n)} &= \min \{\{X_1, \dots, X_n\} \setminus \{X_{(1)}, \dots, X_{(n-1)}\}\} = \max \{X_1, \dots, X_n\} , \end{aligned}$$

тобто  $X_{(k)}$  — це  $k$ -а найменша величина, або ж  $k$ -а порядкова статистика.  $\square$

Зосередьмо свою увагу на незалежних<sup>8</sup> (абсолютно) неперервних величинах<sup>9</sup>, що мають однаковий розподіл  $\mathbb{P}_X$ . Розгляньмо спочатку розподіл максимальної величини  $X_{(n)}$ :

$$F_{X_{(n)}}(x) = \mathbb{P}_{X_{(n)}}(X_{(n)} \leq x) .$$

Для того, щоб найбільше значення було менше від деякого числа  $x$ , потрібно, щоб *усі значення* були менші від такого значення. Через незалежність величин та рівність їхніх розподілів, імовірність такої події є добутком імовірностей окремих подій:

$$F_{X_{(n)}}(x) = \mathbb{P}_X(X_1 \leq x) \cdot \dots \cdot \mathbb{P}_X(X_n \leq x) = (F_X(x))^n . \quad (8.3.1)$$

Щільність такого розподілу можна дістати як похідну<sup>10</sup>:

$$f_{X_{(n)}}(x) = n(F_X(x))^{n-1} f_X(x) . \quad (8.3.2)$$

У схожий спосіб маємо функцію розподілу мінімального значення, помітивши, що воно було *більше* від деякого числа  $x$ , якщо *всі значення* будуть більші від такого значення:

$$F_{X_{(1)}}(x) = \mathbb{P}_{X_{(1)}}(X_{(1)} \leq x) = 1 - \mathbb{P}_{X_{(1)}}(X_{(1)} \geq x) = 1 - (1 - F_X(x))^n . \quad (8.3.3)$$

Щільність такого розподілу можна дістати як похідну:

$$f_{X_{(1)}}(x) = n(1 - F_X(x))^{n-1} f_X(x) . \quad (8.3.4)$$

У загальному випадку маємо таке Твердження.

**Твердження 8.3.2.** Нехай  $X_1, \dots, X_n$  — незалежні неперервні величини з однаковим розподілом

<sup>7</sup>Статистикою в загальному випадку називають будь-яку функцію від даних у деякій вибірці.

<sup>8</sup>На даному етапі ми ще формально не визначили, що розуміти під незалежністю випадкових величин, проте інтуїтивно повинно бути зрозуміло, що якщо  $X$  та  $Y$  незалежні, то  $\mathbb{P}_{X,Y}((X \in A) \cap (Y \in B)) = \mathbb{P}_X(X \in A) \mathbb{P}_Y(Y \in B)$ .

<sup>9</sup>Для дискретних величин існує ненульова імовірність, що дві випадкові величини набудуть однакового значення, і наш аналіз трішки ускладниться.

<sup>10</sup>Пам'ятаючи, що в точках недиференційовності функції розподілу можна покласти довільні значення щільності.

$\mathbb{P}_X$ . Тоді функція і щільність розподілу порядкової статистики  $X_{(k)}$ ,  $k = 1, \dots, n$ , мають вид:

$$\begin{aligned} F_{X_{(k)}}(x) &= \sum_{j=k}^n \binom{n}{j} (F_X(x))^j (1 - F_X(x))^{n-j}, \\ f_{X_{(k)}}(x) &= n \binom{n-1}{k-1} f_X(x) (F_X(x))^{j-1} (1 - F_X(x))^{n-j}. \end{aligned} \quad (8.3.5)$$

**Доведення.** Подія  $X_{(k)} \leq x$  настає тоді, коли принаймні  $k$  величин із набору набувають значення, меншого від  $x$ . Нехай  $N = \text{«число величин із набору } X_1, \dots, X_n, \text{ значення яких менші від } x\text{»}$ . Імовірність потрапляння кожної окремої величини  $X_i$ ,  $i = 1, \dots, n$ , у промінь  $(-\infty; x]$  дорівнює  $\mathbb{P}_X(X_i \leq x) = F_X(x)$ , а відтак  $N \sim \text{Binom}(n, F_X(x))$ . Отже

$$\mathbb{P}_{X_{(k)}}(X_{(k)} \leq x) = \mathbb{P}_N(N \geq k) = \sum_{j=k}^n \binom{n}{j} (F_X(x))^j (1 - F_X(x))^{n-j}.$$

Щільність розподілу можна дістати шляхом диференціювання функції розподілу.  $\square$

**Приклад 8.3.3.** Нехає ми маємо набір незалежних випадкових величин  $U_1, \dots, U_n \sim \text{U}((0; 1))$ . Тоді для  $u \in (0; 1)$ ,  $f_U(u) = 1$ ,  $F_U(u) = u$ , а відтак щільність можна дістати з (8.3.5):

$$f_{U_{(k)}}(u) = n \binom{n-1}{k-1} x^{k-1} (1-x)^{n-k} \cdot \mathbb{1}\{u \in (0; 1)\},$$

що є нічим іншим, як щільністю Бета-розподілу з параметрами  $a = k$ ,  $b = n - k + 1$  (Рис. 8.3.1).

Звідси випливає, що

$$\mathbb{E}[U_{(k)}] = \frac{k}{k + (n - k + 1)} = \frac{k}{n + 1}.$$

Зокрема, якщо величин усього дві, то

$$\mathbb{E}[\min\{U_1, U_2\}] = \mathbb{E}[U_{(1)}] = \frac{1}{3}, \quad \mathbb{E}[\max\{U_1, U_2\}] = \mathbb{E}[U_{(2)}] = \frac{2}{3}.$$

$\square$

## 8.4. Експоненційний розподіл

Експоненційний розподіл є аналогом геометричного розподілу у неперервному випадку. Якщо величина з геометричним розподілом відповідає кількості невдач перед першим успіхом, то величина, що має експоненційний розподіл, відповідає часу до появи першого успіху. Успіхи стаються з інтенсивністю  $\lambda$  на одиницю часу.

**Визначення 8.4.1.** Випадкова величина  $X$  має *експоненційний розподіл* (exponential distribution),  $X \sim \text{Exp}(\lambda)$ ,  $\lambda > 0$ , якщо її щільність розподілу має вид

$$f_X(x) = \lambda e^{-\lambda x} \cdot \mathbb{1}\{x > 0\}. \quad (8.4.1)$$

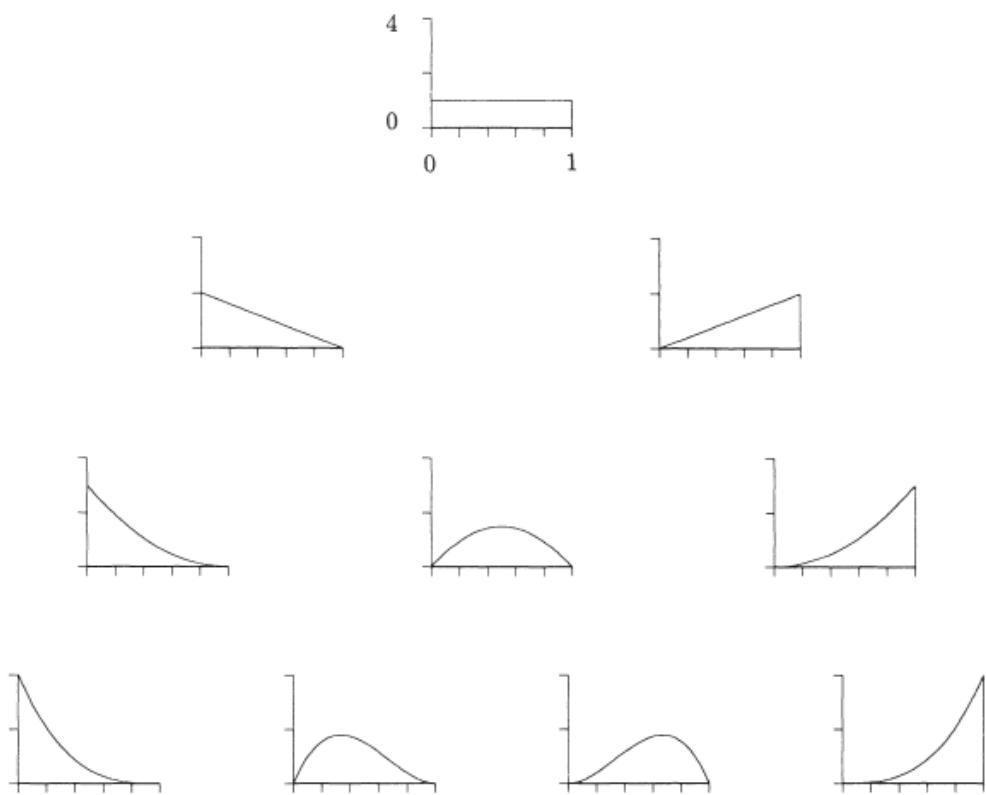
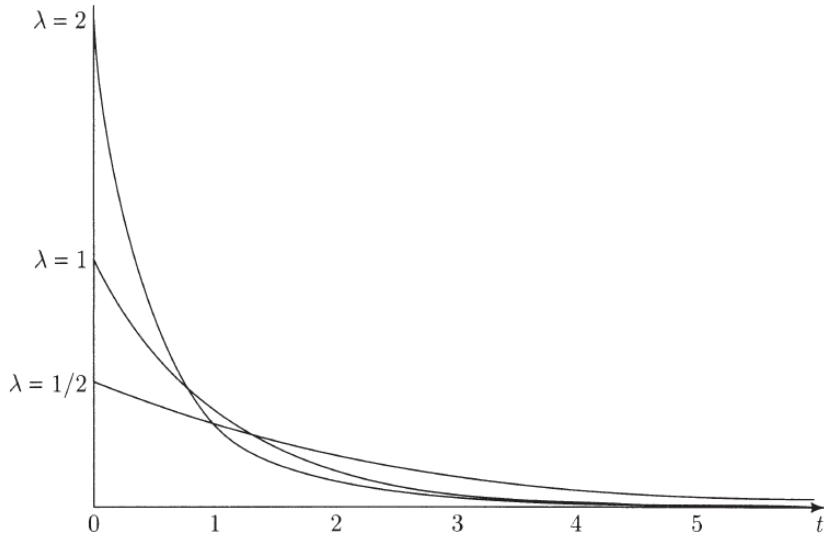


Рис. 8.3.1.: Ілюстрація до Прикладу 8.3.3: графіки щільностей розподілів порядкових статистик для стандартних рівномірних величин: кожний рядок  $n = 1, 2, 3, 4$  відповідає ситуації з  $n$  величинами ([9], Рис. 4.6.1)

Рис. 8.4.1.: Щільності експоненційного розподілу з різними параметрами  $\lambda$  ([9, с. 280])

□

Функція розподілу має вид

$$F_X(x) = \int_{-\infty}^x \lambda e^{-\lambda t} \cdot \mathbb{1}\{t > 0\} dt = \begin{cases} 0, & x < 0 \\ 1 - e^{-\lambda x}, & x \geq 0 \end{cases}.$$

Графіки щільностей експоненційного розподілу наведено на Рис. 8.4.1.

**Приклад 8.4.2.** Нехай тривалість  $X$  телефонної розмови (у хвилинах) має експоненційний розподіл із параметром  $\lambda = 0.1$ . У цьому випадку інтерпретація параметру  $\lambda$  така, що у послідовності телефонних дзвінків, де наступний починається відразу після завершення попереднього, завершення розмови стається (в середньому) 0.1 разів на хвилину, або 1 раз на 10 хвилин.

Тоді можемо оцінити ймовірності тривалостей щойно розпочатої розмови. Зокрема,

$$\begin{aligned} \mathbb{P}_X(X > 10) &= 1 - F(10) = 1 - (1 - e^{-0.1 \cdot 10}) = e^{-1} \approx 0.368, \\ \mathbb{P}_X(10 < X < 20) &= F(20) - F(10) = 1 - e^{-0.1 \cdot 20} - (1 - e^{-0.1 \cdot 10}) = e^{-1} - e^{-2} \approx 0.233. \end{aligned} \quad \square$$

Підрахуймо сподівання та дисперсію експоненційного розподілу.

**Твердження 8.4.3.** Якщо  $X \sim \text{Exp}(\lambda)$ , то  $\mathbb{E}[X] = \frac{1}{\lambda}$ , а  $\text{Var}(X) = \frac{1}{\lambda^2}$ .

*Доведення.* Сподівання шукатимемо за допомогою інтегрування частинами:

$$\begin{aligned}\mathbb{E}[X] &= \int_{-\infty}^{\infty} x \cdot \lambda e^{-\lambda x} \mathbb{1}\{x > 0\} dx = \int_0^{\infty} x \cdot \lambda e^{-\lambda x} dx \\ &\quad \left[ u = x, dv = \lambda e^{-\lambda x} dx \right] \\ &\quad \left[ du = dx, v = -e^{-\lambda x} \right] \\ &= -xe^{-\lambda x} \Big|_0^{\infty} + \int_0^{\infty} e^{-\lambda x} dx \\ &= 0 - \frac{1}{\lambda} e^{-\lambda x} \Big|_0^{\infty} = \frac{1}{\lambda}.\end{aligned}$$

Також застосування інтегрування частинами дає

$$\begin{aligned}\mathbb{E}[X^2] &= \int_{-\infty}^{\infty} x^2 \cdot \lambda e^{-\lambda x} \mathbb{1}\{x > 0\} dx = \int_0^{\infty} x^2 \cdot \lambda e^{-\lambda x} dx \\ &\quad \left[ u = x^2, dv = \lambda e^{-\lambda x} dx \right] \\ &\quad \left[ du = 2x dx, v = -e^{-\lambda x} \right] \\ &= -x^2 e^{-\lambda x} \Big|_0^{\infty} + 2 \int_0^{\infty} x e^{-\lambda x} dx \\ &= 0 + \frac{2}{\lambda} \mathbb{E}[X] = \frac{2}{\lambda^2}.\end{aligned}$$

Нарешті

$$\text{Var}(X) = \mathbb{E}[X^2] - (\mathbb{E}[X])^2 = \frac{2}{\lambda^2} - \frac{1}{\lambda^2} = \frac{1}{\lambda^2}.$$

□

Експоненційний розподіл, як і геометричний, є розподілом без пам'яти, тобто тривалість  $t$  очікування першого успіху, за умови, що протягом часу  $s$  він ішле не встиг настати, така сама, як тривалість очікування першого успіху «з самого початку»:

$$\mathbb{P}_X(X \geq s + t \mid X \geq s) = \mathbb{P}_X(X \geq t), \quad s, t \geq 0.$$

Справді:

$$\begin{aligned}\mathbb{P}_X(X \geq s + t \mid X \geq s) &= \frac{\mathbb{P}_X((X \geq s + t) \cap (X \geq s))}{\mathbb{P}_X(X \geq s)} = \frac{\mathbb{P}_X(X \geq s + t)}{\mathbb{P}_X(X \geq s)} \\ &= \frac{1 - (1 - e^{-\lambda(s+t)})}{1 - (1 - e^{-\lambda s})} = e^{-\lambda t} = \mathbb{P}(X \geq t).\end{aligned}$$

Більше того, експоненційний розподіл — *единий* додатний неперервний розподіл, який не має пам'яти.

**Твердження 8.4.4.** Нехай  $X$  — додатна неперервна випадкова величина без пам'яти. Тоді  $X \sim \text{Exp}(\lambda)$ .

*Доведення.* Розглянемо так звану *функцію надійності* (survival function)  $G(x) = 1 - F(x)$ , де  $F$  —

функція розподілу  $X$ . Згідно з вищеперечисленими міркуваннями, якщо розподіл не має пам'яти, то  $G(s+t) = G(s)G(t)$ , зокрема,  $G(2t) = G(t)^2$ ,  $G(3t) = G(2t+t) = G(t)^3$ , і т.д., тобто  $G(mt) = G(t)^m$ ,  $m = 1, 2, \dots$ .

Більше того, з  $G(t) = G(2 \cdot t/2) = G(t/2)^2$  випливає  $G(t/2) = G(t)^{1/2}$ , а відтак

$$G\left(\frac{t}{n}\right) = (G(t))^{\frac{1}{n}}, \quad n = 1, 2, \dots.$$

Отже,

$$G\left(\frac{m}{n}t\right) = \left(G\left(\frac{t}{n}\right)\right)^m = (G(t))^{\frac{m}{n}}, \quad n = 1, 2, \dots, \quad m = 1, 2, \dots,$$

тобто  $G(xt) = G(t)^x$ ,  $x \in \mathbb{Q}$ .

Нарешті, для будь-якого дійсного  $x \in \mathbb{R}$  можна знайти послідовність раціональних  $\{x_n\}$  таких, що  $x_n \rightarrow x$ . Ураховуючи неперервність  $G$ , матимемо:

$$G(xt) = \lim_{n \rightarrow \infty} G(x_n t) = \lim_{n \rightarrow \infty} G(t)^{x_n} = G(t)^x.$$

Поклавши  $t = 1$ ,  $\lambda = -\ln G(1) > 0$  маємо:

$$G(x) = G(1 \cdot x) = G(1)^x = \left(e^{-\lambda}\right)^x = e^{-\lambda x}.$$

Отже  $F(x) = 1 - e^{-\lambda x}$ , що відповідає експоненційному розподілу.  $\square$

Відтак експоненційний розподіл є єдиним можливим варіантом для опису явищ, де час очікування появи деякої події не має пам'яти, наприклад:

- атоми радіоактивних ізотопів, зокрема вуглецю-14, урану-235, стронцію-90 у випадкові моменти часу ралтово розпадаються з одночасним випроміненням радіації або певних частинок;
- на відміну від деякого *механічного* приладу, який із часом об'єктивно зношується, а тому очікувана тривалість його роботи в різні моменти часу буде різною, *електронні* компоненти не зношується з часом (принаймні в осяжній перспективі використання приладу), а виходять із ладу ралтово і непрогнозовано.

**Приклад 8.4.5.** Стронцій-90 — особливо токсичний побічний продукт ядерного вибуху з періодом напіврозпаду майже 28 років, тобто  $\mathbb{P}_T(T \leq 28) = \mathbb{P}_T(T > 28) = 0.5$ , де  $T$  — тривалість життя одного атому.

Із цієї інформації можна встановити параметр експоненційного розподілу  $\lambda$ :

$$0.5 = \mathbb{P}_T(T > 28) = e^{-28\lambda} \Rightarrow \lambda = \frac{\ln 2}{28} \approx 0.0248,$$

тобто на один рік у середньому розпадаються 0.0248 атома.

Середня тривалість життя такого атома за Твердженням 8.4.3 дорівнює  $1/\lambda \approx 40.4$  року.

Імовірність, що атом стронцію-90 проживе щонайменше 50 років, дорівнює

$$\mathbb{P}_T(T > 50) = e^{-50\lambda} = e^{-50 \cdot \frac{\ln 2}{28}} \approx 0.290.$$

Також цю ймовірність можна інтерпретувати як частку від початкової маси стронцію-90, який збережеться після 50 років (оскільки кількість атомів шалена, тому інтерпретація ймовірності як частоти доречна).

Нарешті, кількість років після ядерного вибуху, потрібна для розпаду 99% утвореного стронцію-90, становить

$$e^{-\lambda \cdot y} = 0.01 \Rightarrow y = \frac{\ln 10}{\lambda} = \frac{\ln 100}{\ln 2} \cdot 28 \approx 186 \text{ років}.$$

□

**Приклад 8.4.6.** Нехай очікувана тривалість  $X$  безперебійної роботи деякого транзистора має експоненційний розподіл. Середня тривалість життя такого транзистора — 100 робочих годин. Нехай транзистор успішно працює протягом 50 годин. Підрахуймо ймовірність виходу його з ладу протягом найближчої години.

З урахуванням властивості не мати пам'яти інформація про 50 годин є абсолютно іррелевантною:

$$\mathbb{P}_X(X < 51 | X > 50) = \mathbb{P}_X(X < 1) = 1 - 0.01e^{-0.01} \approx 0.01,$$

де ми використали формулу сподівання з Твердження 8.4.3, щоб установити значення параметру  $\lambda$ :  $100 = 1/\lambda \Rightarrow \lambda = 0.01$ . □

## 8.5. Процес Пуассона

Експоненційний розподіл пов'язаний із розподілом Пуассона, розглянутим у Розділі 5.1<sup>11</sup>. Розгляньмо деяку послідовність подій, які стаються у випадкові моменти часу<sup>12</sup>. Якщо послідовність має такі властивості:

- кількість подій  $N$  в інтервалі часу довжиною  $t$  має розподіл  $N \sim \text{Pois}(\lambda t)$ ;
- поява подій в одному часовому інтервалі не залежить від появи подій в іншому часовому інтервалі без спільного перетину;

то таку послідовність називають *процесом Пуассона* (Poisson process) зі швидкістю настання подій  $\lambda$ . Ми не будемо детально розглядати в цьому курсі властивості процесу Пуассона, а тільки обмежимося тим фактом, що час очікування появи наступної події після появи попередньої  $T \sim \text{Exp}(\lambda)$  для всіх подій.

Справді, розгляньмо розподіл часу очікування появи першої події,  $T_1$ . Звернімо увагу на той факт, що подія  $T_1 > t$  — це та сама подія, що «за період часу від 0 до  $t$  не сталося жодної події», тобто

$$\mathbb{P}_{T_1}(T_1 \leq t) = 1 - \mathbb{P}_{T_1}(T_1 > t) = 1 - \mathbb{P}_N(N = 0) = 1 - \frac{e^{-\lambda t}(\lambda t)^0}{0!} = 1 - e^{-\lambda t},$$

відтак  $T_1 \sim \text{Exp}(\lambda)$ .

Аналогічно можна розглянути *різницю* між часом настання другої події  $T_2$  та часом настання першої події  $T_1$ . Оскільки за умовами процесу Пуассона імовірність настання подій в одному

<sup>11</sup>Тому й параметри цих розподілів позначають однаковою літерою!

<sup>12</sup>Таку послідовність часто називають потоком подій.

інтервалі не залежить від імовірності настання події в іншому, розподіл величини  $T_2 - T_1$  нічим не відрізняється від розподілу величини  $T_1$ , тобто  $T_2 - T_1 \sim \text{Exp}(\lambda)$ . Відповідні міркування справедливі для різниць у часі між будь-якими сусідніми подіями.

**Зауваження 8.5.1.** Час очікування  $T_2$  є сумаю двох незалежних випадкових величин,  $T_1$  і  $T_2 - T_1$ . Розподіл  $T_2$  не є експоненційним, а так званим Гамма-розподілом (Gamma-distribution), про що буде мова в наступних лекціях.  $\square$

**Приклад 8.5.2.** У поштовому відділку працюють два клерки. Коли пан Іваненко заходить до відділку, він бачить, що обидва клерки зайняті обслуговуванням двох клієнтів, панів Петренка і Сидоренка відповідно. Пан Іваненко займає місце в черзі, яка одна для обох клерків. Покладімо, що час  $T$  роботи клерка з одним клієнтом має експоненційний розподіл із параметром  $\lambda$ .

Підрахуймо ймовірність події, що пан Іваненко покине поштовий відділок останнім. Коли пана Іваненка почне обслуговувати один із клерків, другий усе ще буде зайнятий (або з паном Петренком, або з паном Сидоренком, що непринципово). За властивістю експоненційного розподілу не мати пам'яти очікувана тривалість роботи кожного клерка зі своїм клієнтом становим на цей момент точно така сама, як і коли пан Іваненко тільки-но зайде до відділку. Можливі два варіанти розвитку ситуації: пан Іваненко завершить свої справи або першим, або другим. Нехай  $T_1$  — тривалість часу обслуговування пана Іваненка, а  $T_2$  — тривалість часу обслуговування іншого з двох панів, що залишилися. Тоді

$$\mathbb{P}(T_1 < T_2) + \mathbb{P}(T_2 < T_1) = 1,$$

але через симетрію задачі  $\mathbb{P}(T_1 < T_2) = \mathbb{P}(T_2 < T_1)$ , і тому ймовірність, що пан Іваненко вийде з відділку останнім, дорівнює 0.5.  $\square$

Для незалежних експоненційних величин  $X_1 \sim \text{Exp}(\lambda_1), \dots, X_n \sim \text{Exp}(\lambda_n)$  справедлива властивість, що їх мінімум також є експоненційною величиною. Це безпосередньо випливає з (8.3.3), з тією відмінністю, що розподіл величин може відрізнятися:

$$F_{X_{(1)}}(x) = 1 - (1 - F_{X_1}(x)) \cdot \dots \cdot (1 - F_{X_n}(x)) = 1 - (e^{-\lambda_1 x} \dots e^{-\lambda_n x}) = 1 - e^{-x \cdot \sum_{i=1}^n \lambda_i},$$

тобто

$$X_{(1)} \sim \text{Exp}\left(\sum_{i=1}^n \lambda_i\right). \quad (8.5.1)$$

Інтуїтивно ми можемо розглядати кожну величину як час очікування настання деякої події в певному процесі Пуассона, а всього таких процесів —  $n$  штук.

**Приклад 8.5.3.** Троє студентів о 13:00 почали одночасно готуватися до екзамену з «Теорії ймовірностей». Час підготовки кожного студента  $T_i$ ,  $i = 1, 2, 3$ , має експоненційний розподіл  $T_i \sim \text{Exp}(1/6)$ , тобто середній час підготовки становить 6 годин:  $\mathbb{E}[T_i] = 6$ . Через скільки годин, у середньому, завершить підготовку останній із них?

У цьому випадку нас цікавить сподівання  $\mathbb{E}[T_{(3)}] \equiv \mathbb{E}[\max\{T_1, T_2, T_3\}]$ . Використовуючи (8.3.1), можна, звісно, знайти

$$f_{T_{(3)}}(t) = 3\lambda e^{-\lambda t}(1 - e^{-\lambda t})^2,$$

просте обчислення інтегралу від  $tf_{T_{(3)}}(t)$  є задачею трудомісткою.

Натомість можемо помітити, що

$$T_{(3)} = T_{(1)} + (T_{(2)} - T_{(1)}) + (T_{(3)} - T_{(2)}) ,$$

тобто сумаю часу, потрібного для завершення підготовки одним зі студентів, часу, потрібного додатково для завершення підготовки іншим студентом, і часу, потрібного додатково для завершення підготовки останнім студентом.

Згідно з (8.5.1),  $T_{(1)} \sim \text{Exp}(3\lambda)$ . А враховуючи властивість не мати пам'яти, можемо стверджувати, що по завершенні підготовки першим студентом процес підготовки інших двох студентів можна розглядати, ніби вони щойно сіли працювати. Таким чином,  $T_{(2)} - T_{(1)} \sim \text{Exp}(2\lambda)$  і, за такою ж логікою,  $T_{(3)} - T_{(2)} \sim \text{Exp}(\lambda)$ .

З урахуванням Твердження 8.4.3 та лінійності сподівання маємо:

$$\mathbb{E}[T_{(3)}] = \mathbb{E}[T_{(1)}] + \mathbb{E}[T_{(2)} - T_{(1)}] + \mathbb{E}[T_{(3)} - T_{(2)}] = \frac{1}{3\lambda} + \frac{1}{2\lambda} + \frac{1}{\lambda} = 2 + 3 + 6 = 11 .$$

Тобто незважаючи на те, що середній час підготовки кожного студента дорівнює 6 годин, сподіваний час підготовки *всіх* студентів становить 11 годин.  $\square$

## 9. Випадкові вектори та незалежні випадкові величини

У попередніх лекціях ми розглядали розподіли та сподівання випадкових величин. Проте на практиці дуже часто виникають ситуації, коли потрібно знати взаємозв'язок між різними випадковими величинами, визначеними для одного експерименту:

- для оцінювання ефективності нових ліків потрібно заміряти декілька різних показників — тиск, серцебиття, рівень холестерину тощо. Ці показники, очевидно, не є незалежні один від одного, а тому постає задача моделювати їх розподіл комплексно;
- під час аналізу взаємозв'язків між різними генетичними маркерами та певним захворюванням використання інформації про кожний маркер окремо може не дати змоги встановити справжню залежність, адже маркери можуть упливати на появу хвороби *в сукупності*;
- для аналізу деякого процесу, який розвивається в часі, роблять його заміри в конкретні моменти (кожний окремий замір є випадковою величиною), а потім розглядають розподіл такого *часового ряду* (time series) у сукупності. Часові ряди виринають у різних галузях, від економіки та фінансів до медицини.

Інші приклади, коли результат випадкового експерименту можна описати декількома величинами — число клієнтів у двох чергах, вага і зріст людини в деякій популяції, кількість добрих та врожайність, результати ЗНО та першої сесії деякого студента і безліч інших.

Отже, постає задача детального аналізу розподілів випадкових векторів, обчислень їхніх сподівань тощо. Для формалізації уведення поняття функції розподілу та щільності розподілу випадкового вектора спочатку потрібно розглянути простори, на яких визначено ці вектори, та міри, задані на цих просторах.

### 9.1. Елементи теорії міри для добутків просторів

#### 9.1.1. Добуток просторів і добуток мір

Розгляньмо декілька базових визначень.

**Визначення 9.1.1.** Нехай маємо два вимірні простори,  $(X, \mathcal{X})$  і  $(Y, \mathcal{Y})$ . *Добутком цих просторів* (product space) називають простір

$$X \times Y = \{\omega = (\omega_1, \omega_2) : \omega_1 \in X, \omega_2 \in Y\} . \quad (9.1.1)$$

□

**Визначення 9.1.2.** У добутку просторів  $X \times Y$  можна визначити *вимірний прямокутник* (measurable rectangle)  $A \times B$  такий, що  $A \in \mathcal{X}$ ,  $B \in \mathcal{Y}$ .  $\square$

**Зауваження 9.1.3.** Варто звернути увагу, що поняття «прямокутник» використано доволі умовно, оскільки множини  $A$  і  $B$  — це далеко не обов'язково інтервали. Наприклад, якщо  $X = Y = \mathbb{R}$ , а  $A = \mathbb{Q}$ ,  $B = (0; 1]$ , то  $A \times B$  також є вимірним прямокутником, хоча геометрична форма його зовсім нетривіальна.  $\square$

Розгляньмо  $\sigma$ -алгебру на добутку просторів, породжену вимірними прямокутниками, яку позначатимемо через  $\mathcal{X} \times \mathcal{Y}$ . Потрібно розуміти, що  $\mathcal{X} \times \mathcal{Y}$  є просто позначенням, така  $\sigma$ -алгебра не є Декартовим добутком двох  $\sigma$ -алгебр.

**Приклад 9.1.4.** Можемо показати, що  $\sigma$ -алгебра  $\mathcal{B} \times \mathcal{B}$  є нічим іншим, як Борелевою  $\sigma$ -алгеброю  $\mathcal{B}^2$ .

У Лекції 1 ми зазначали, що  $k$ -вимірну Борелеву  $\sigma$ -алгебру  $\mathcal{B}^k$  породжують  $k$ -вимірні прямокутники  $I_1 \times I_2 \times \dots \times I_k \in \mathbb{R}^k$ , де кожний  $I_j \subset \mathcal{I}$ , де  $\mathcal{I}$  — множина, яка містить порожню множину  $\emptyset$ , усі пів інтервали виду  $(a_1; a_2]$ ,  $a_1, a_2 \in \mathbb{R}$ ,  $a_1 < a_2$ , а також промені  $(-\infty; a]$ ,  $(b; \infty)$ ,  $a, b \in \mathbb{R}$ . Оскільки вимірні прямокутники в розумінні Визначення 9.1.2 є класом, ширшим від «звичайних» прямокутників, які породжують  $\mathcal{B}^2$ , то  $\sigma$ -алгебра  $\mathcal{B}^2 \subseteq \mathcal{B} \times \mathcal{B}$ .

З іншого боку, якщо  $A \in \mathcal{I}$ , то клас множин  $\mathcal{C}_A = \{B : A \times B \in \mathcal{B}^2\}$  містить саму числову вісь  $\mathbb{R}$ , оскільки

$$A \times \mathbb{R} = \bigcup_{n=1}^{\infty} A \times (-n; n] \in \mathcal{B}^2.$$

Окрім цього,  $\mathcal{C}_A$  замкнений відносно заперечень і зліченних об'єднань, а відтак сам є  $\sigma$ -алгеброю, що містить усі вимірні множини.

Аналогічно, нехай маємо  $\mathcal{C}_B = \{A : A \times B \in \mathcal{B}^2\}$ , де  $B$  — деяка вимірна множина (необов'язково  $B \in \mathcal{I}$ ). Тоді, як ми щойно показали,  $\mathcal{C}_B$  містить усі  $A \in \mathcal{I}$ , а відтак містить усі вимірні множини.

Отже всі вимірні прямокутники належать  $\mathcal{B}^2$ , тобто  $\mathcal{B} \times \mathcal{B} \subseteq \mathcal{B}^2$ .

Вкладення в обидва боки доводить рівність  $\mathcal{B}^2 = \mathcal{B} \times \mathcal{B}$ .  $\square$

У загальному випадку,  $\sigma$ -алгебра  $\mathcal{X} \times \mathcal{Y}$  є дуже потужною.

**Приклад 9.1.5.** Розгляньмо  $(X, \mathcal{X}) = (Y, \mathcal{Y}) = ((0; 1], \mathcal{B}(0; 1])$ . Тоді діагональ  $A = \{(x, y) : x = y\}$  належить  $\sigma$ -алгебрі  $\mathcal{B}(0; 1] \times \mathcal{B}(0; 1]$ . Справді, нехай  $x, y \in (0; 1]$ ,  $x \neq y$ . Тоді нехай  $X$  і  $Y$  — неперетинанні підінтервали з  $(0; 1]$  такі, що  $x \in X$ ,  $y \in Y$ , а їхні межі — раціональні числа. Очевидно,  $X \times Y$  не має спільних точок із діагоналлю  $A$ . Оскільки інтервали мають раціональні межі, то виходить, що *заперечення* діагонали,  $A^c$ , є зліченним об'єднанням прямокутників на кшталт  $X \times Y$ , а тому  $A^c \in \mathcal{B}(0; 1] \times \mathcal{B}(0; 1]$ . За властивостями  $\sigma$ -алгебри,  $A \in \mathcal{B}(0; 1]$ .  $\square$

Розгляньмо тепер питання побудови міри на деякому добутку просторів. Фундаментальним результатом у цьому контексті є така теорема.

**Теорема 9.1.6** (Теорема про добуток мір (Product measure theorem)). Нехай  $(X, \mathcal{X}, \mu)$  і  $(Y, \mathcal{Y}, \nu)$  — два мірні простори з  $\sigma$ -скінченними мірами. На  $\sigma$ -алгебрі  $\mathcal{X} \times \mathcal{Y}$  можна визначити міру  $\pi$  таку, що для будь-якої вимірної множини  $E \in \mathcal{X} \times \mathcal{Y}$

$$\pi(E) = \int \nu(\{y : (x, y) \in E\}) d\mu = \int \mu(\{x : (x, y) \in E\}) d\nu. \quad (9.1.2)$$

Тоді така міра є єдиною  $\sigma$ -скінченною мірою такою, що  $\pi(A \times B) = \mu(A)\nu(B)$  для  $A \in \mathcal{X}, B \in \mathcal{Y}$ .

Доведення цієї теореми можна знайти в Розд. 9.6.1.

**Визначення 9.1.7.** Міру  $\pi$  з Теореми 9.1.6 називають *добутком мір* (product measure) і позначають через  $\pi = \mu \times \nu$ .  $\square$

### 9.1.2. Теорема Фубіні

У курсі «Математичного аналізу» розглядають *Теорему Фубіні*<sup>1</sup> (Fubini Theorem) для дійсних функцій багатьох змінних. Для дальнього викладу матеріалу нам потрібно розглянути цю теорему в загальному випадку в застосуванні до випадкових величин і добутків мір.

**Теорема 9.1.8** (Теорема Фубіні (Fubini Theorem)). Нехай на добутку просторів  $(X \times Y, \mathcal{X} \times \mathcal{Y}, \mu \times \nu \equiv \pi)$  з  $\sigma$ -скінченним добутком мір визначено випадкову величину  $f : X \times Y \rightarrow \mathbb{R}$ . Тоді:

- якщо  $f \geq 0$  майже напевно<sup>2</sup>, то функції

$$\int_Y f(x, y) d\nu, \quad \int_X f(x, y) d\mu$$

вимірні відносно  $\mathcal{X}$  та  $\mathcal{Y}$  відповідно, і справедливо твердження

$$\int_{X \times Y} f(x, y) d\pi = \int_X \left( \int_Y f(x, y) d\nu \right) d\mu = \int_Y \left( \int_X f(x, y) d\mu \right) d\nu. \quad (9.1.3)$$

Інтеграл зліва називають *подвійним* (double), а два інші інтеграли — *повторними* (iterated);

- якщо  $f$  інтегровна відносно  $\pi$  (і до того ж необов'язково невід'ємна), то функції

$$\int_Y f(x, y) d\nu, \quad \int_X f(x, y) d\mu$$

скінченні і вимірні відносно  $\mathcal{X}$  та  $\mathcal{Y}$  відповідно на множинах  $A_0$  і  $B_0$  відповідно таких, що  $\mu(X \setminus A_0) = 0$  і  $\nu(Y \setminus B_0) = 0$ , і (9.1.3) також виконується.

Доведення цієї теореми здійснюють у курсі «Математичного аналізу». Альтернативне доведення можна знайти в Розд. 9.6.2.

<sup>1</sup>Названу на честь італійського математика Гвідо Фубіні (Guido Fubini, 1879–1943), який довів її в 1907 р.

<sup>2</sup>Цю частину, для невід'ємних функцій, називають також *теоремою Тонеллі* на честь італійського математика Леоніда Тонеллі (Leonida Tonelli, 1885–1946), який довів її у 1909 р.

<sup>3</sup>Тобто відповідні інтеграли можуть не існувати або бути нескінченними на деякій множині міри нуль.

**Приклад 9.1.9.** Розгляньмо простий приклад застосування Теореми 9.1.8.

Нехай  $D_r$  — замкнений круг на площині з центром у точці  $(0; 0)$  і радіусом  $r$ . Його площа є мірою Лебега відповідної множини, яку можна обчислити за Теоремою 9.1.8, оскільки підінтегральна функція невід'ємна. Здійснюючи перехід у полярні координати, маємо

$$\lambda_2(D_r) = \iint_{D_r} dx dy = \iint_{\substack{0 < \rho \leq r \\ 0 < \theta < 2\pi}} \rho d\rho d\theta = \int_0^r \left( \int_0^{2\pi} d\theta \right) \rho d\rho = \pi r^2.$$

□

**Зауваження 9.1.10.** Два випадки в Теоремі 9.1.8 розрізняють через те, що у випадку невід'ємних функцій допускають нескінченно велике значення інтегралу. Якщо функція не є невід'ємною, то можливі ситуації, коли інтеграл не існує, оскільки інтеграли

$$\int_{X \times Y} f^+ d\pi, \quad \int_{X \times Y} f^- d\pi$$

одночасно нескінчені. У цьому випадку повторні інтеграли можуть мати різні значення.

Для простого прикладу розгляньмо простір  $(\mathbb{N} \times \mathbb{N}, 2^\mathbb{N} \times 2^\mathbb{N}, \# \times \#)$ , на якому визначено функцію

$$f(x, y) = \begin{cases} 1, & x = y \\ -1, & x = y + 1 \\ 0, & \text{інакше} \end{cases}.$$

У матричному виді, де рядки відповідають  $x \in \mathbb{N}$ , а стовпці —  $y \in \mathbb{N}$ , функцію можна подати так:

$$f = \begin{bmatrix} 1 & 0 & 0 & 0 & \dots \\ -1 & 1 & 0 & 0 & \dots \\ 0 & -1 & 1 & 0 & \dots \\ 0 & 0 & -1 & 1 & \dots \\ \vdots & \vdots & 0 & -1 & \dots \\ \vdots & \vdots & \vdots & \vdots & \ddots \\ 0 & 0 & 0 & 0 & \dots \end{bmatrix}$$

Оскільки функція не є невід'ємною, для застосування Теореми 9.1.8 повинна виконуватися умова інтегровності. Проте функція  $|f|$  не є інтегровною:

$$\int_{\mathbb{N} \times \mathbb{N}} |f| d(\# \times \#) = \int_{x=y} 1 d(\# \times \#) + \int_{x=y+1} 1 d(\# \times \#) = \infty.$$

У той же час повторні інтеграли скінчені, але мають різні значення.

Справді, якщо розглянути повторний інтеграл

$$\int_{\mathbb{N}} \left( \int_{\mathbb{N}} f(x, y) \#(dy) \right) \#(dx),$$

то ми повинні спочатку розглянути функцію  $f(x, y)$  для деякого фіксованого  $x$ . Якщо  $x = 1$ , то

$$f(1, y) = \begin{cases} 1, & y = 1 \\ 0, & y > 1 \end{cases},$$

оскільки  $y$  не може бути нулем, а тому значення  $-1$  функція не набуває. Тоді

$$\int_{\mathbb{N}} f(1, y) \#(dy) = 1.$$

Для всіх інших  $x > 1$  маємо загальний вираз, наведений вище, і тоді

$$\int_{\mathbb{N}} f(1, y) \#(dy) = 1 - 1 = 0.$$

Отже

$$\int_{\mathbb{N}} \left( \int_{\mathbb{N}} f(x, y) \#(dy) \right) \#(dx) = 1 + 0 + \dots = 1.$$

Якщо ж розглянути функцію  $f(x, y)$  для деякого фіксованого  $y$ , то ми побачимо, що для всіх значень  $y \in \mathbb{N}$  маємо

$$\int_{\mathbb{N}} f(x, y) \#dx = 1 - 1 = 0.$$

Отже

$$\int_{\mathbb{N}} \left( \int_{\mathbb{N}} f(x, y) \#(dx) \right) \#(dy) = 0 + 0 + \dots = 0.$$

Як бачимо, значення повторних інтегралів не дорівнюють одне одному.  $\square$

### 9.1.3. Добутки вищих порядків

Вищеведені міркування можна поширити на випадок  $n$  просторів. Нехай  $(X, \mathcal{X}, \mu)$ ,  $(Y, \mathcal{Y}, \nu)$ ,  $(Z, \mathcal{Z}, \eta)$  — три мірні простори з  $\sigma$ -скінченними мірами. Тоді добутком просторів  $X \times Y \times Z$  вважатимемо добуток  $(X \times Y) \times Z$ <sup>4</sup>. Нехай  $\mathcal{X} \times \mathcal{Y} \times \mathcal{Z}$  —  $\sigma$ -алгебра на просторі  $X \times Y \times Z$ , породжена вимірними прямокутниками  $A \times B \times C$ ,  $A \in \mathcal{X}$ ,  $B \in \mathcal{Y}$ ,  $Z \in \mathcal{Z}$ .

На визначеній у такий спосіб  $\sigma$ -алгебрі  $\mathcal{X} \times \mathcal{Y} \times \mathcal{Z}$  можна визначити добуток мір  $\mu \times \nu \times \eta$  як  $(\mu \times \nu) \times \eta$  такий, що

$$\mu \times \nu \times \eta(A \times B \times C) = (\mu \times \nu)(A \times B) \cdot \eta(C) = \mu(A)\nu(B)\eta(C).$$

Доведення відповідних формул є абсолютно аналогічним викладеним раніше, оскільки вже дійсно відомо, що  $\mu \times \nu$  сама по собі є мірою.

Нарешті, очевидно, що Теорема 9.1.8 залишається справедливою і в цьому випадку, оскільки добуток трьох мір визначено як добуток двох мір, одна з яких сама по собі є добутком.

Відповідні міркування легко можна поширити на добутки просторів вищих порядків. Серед іншого, можна бачити, що міру Лебега в  $\mathbb{R}^k$ ,  $\lambda_k$ , також можна розглядати як добуток одновимірних мір Лебега:  $\lambda_k = \lambda \times \dots \times \lambda$ .

<sup>4</sup>За асоціативністю декартового добутку тривимірний добуток можна увести і в іншому порядку.

**Приклад 9.1.11.** Обчислімо об'єм кулі  $A$  одиничного радіуса в просторі  $\mathbb{R}^k$ :

$$V_k = \int_{\sum_{j=1}^k x_j^2 \leq 1} dx_1 \dots dx_k .$$

Позначмо  $B = \{(x_1, x_2) : x_1^2 + x_2^2 \leq 1\}$ ,  $C(x_1, x_2) = \left\{ (x_3, \dots, x_n) : \sum_{j=3}^k x_j^2 \leq 1 - x_1^2 - x_2^2 \right\}$ . Тоді за Теоремою 9.1.8 маємо

$$V_k = \int_A dx_1 \dots dx_k = \int_B \left( \int_{C(x_1, x_2)} dx_3 \dots dx_n \right) dx_1 dx_2 .$$

Цілком очевидно, що в інтегралі

$$\int_{C(x_1, x_2)} dx_3 \dots dx_n$$

можна перейти від координат  $x_j$ ,  $j = 3, \dots, k$ , до координат  $x_j \cdot \sqrt{1 - x_1^2 - x_2^2}$ , оскільки в цьому випадку матимемо кулю одиничного радіуса:

$$\sum_{j=3}^k \left( \frac{x_j}{\sqrt{1 - x_1^2 - x_2^2}} \right) \leq 1 ,$$

об'єм якої дорівнює  $V_{k-2}$ .

З урахуванням якобіану, який при цьому вискочить, маємо

$$\begin{aligned} V_k &= \int_B \left( V_{k-2} \cdot \left( \sqrt{1 - x_1^2 - x_2^2} \right)^{k-2} \right) dx_1 dx_2 \\ &= V_{k-2} \iint_{\substack{0 < \rho < 1 \\ 0 < \theta < 2\pi}} \left( \sqrt{1 - \rho^2} \right)^{k-2} \rho d\rho d\theta \\ &= \pi V_{k-2} \int_0^1 t^{\frac{k-2}{2}} dt \\ &= \frac{2\pi V_{k-2}}{k} . \end{aligned}$$

Якщо покласти, що  $V_0 = 1$ , і помітити, що  $V_1 = 2$  (довжина інтервалу  $(-1; 1)$ ), то виходить, що площа круга одиничного радіуса дорівнює

$$V_2 = \frac{2\pi \cdot 1}{2} = \pi ,$$

а об'єм кулі одиничного радіуса становить

$$V_3 = \frac{2\pi \cdot 2}{3} = \frac{4}{3}\pi .$$

Це відповідає загальновідомим геометричним фактам. У загальному випадку маємо

$$V_{2i-1} = \frac{2(2\pi)^{i-1}}{1 \cdot 3 \cdot 5 \cdot \dots \cdot (2i-1)} , \quad V_{2i} = \frac{(2\pi)^i}{2 \cdot 4 \cdot 6 \cdot \dots \cdot (2i)} , \quad i = 1, 2, \dots .$$

□

## 9.2. Розподіли випадкових векторів

### 9.2.1. Спільні функції та щільності розподілу

До цього моменту ми жодним чином не використали того факту, що працюємо з об'єктами з теорії ймовірностей — усі результати, у тому числі Теорема 9.1.8, справедливі для теорії міри взагалі. Розгляньмо тепер розподіли, які утворюють випадкові вектори, а також відповідні функції та щільності.

**Визначення 9.2.1.** Нехай маємо випадковий вектор  $\mathbf{X} = (X_1, \dots, X_k)^\top : (\Omega, \mathcal{A}, \mathbb{P}) \rightarrow (\mathbb{R}^k, \mathcal{B}^k)$ . Він утворює розподіл  $\mathbb{P}_{\mathbf{X}}$  такий, що

$$\mathbb{P}_{\mathbf{X}}(A) \equiv \mathbb{P}_{\mathbf{X}}\left((X_1, \dots, X_k)^\top \in A\right) = \mathbb{P}\left(\left\{\omega : (X_1(\omega), \dots, X_k(\omega))^\top \in A\right\}\right), \quad A \in \mathcal{B}^k. \quad (9.2.1)$$

Такий розподіл називають *спільним розподілом* (joint distribution) випадкового вектора.  $\square$

Як можна бачити, Визначення 9.2.1 принципово нічим не відрізняється від Визначення 4.1.6. Так само мало відрізняється і визначення функції розподілу.

**Визначення 9.2.2.** Нехай маємо випадковий вектор  $\mathbf{X} = (X_1, \dots, X_k)^\top : (\Omega, \mathcal{A}, \mathbb{P}) \rightarrow (\mathbb{R}^k, \mathcal{B}^k)$  зі спільним розподілом  $\mathbb{P}_{\mathbf{X}}$ . Функція відповідного розподілу має вид

$$\begin{aligned} F_{\mathbf{X}}(x_1, \dots, x_k) &= \mathbb{P}_{\mathbf{X}}(X_1 \leq x_1, \dots, X_k \leq x_k) \\ &= \mathbb{P}_{\mathbf{X}}(\omega : (X_1(\omega) \leq x_1) \cap \dots \cap (X_k(\omega) \leq x_k)) = \mathbb{P}_{\mathbf{X}}(S_{\mathbf{x}}), \end{aligned} \quad (9.2.2)$$

де  $S_{\mathbf{x}} = \{(y_1, \dots, y_k)^\top : y_i \leq x_i, i = 1, \dots, k\}$  — множина точок, які лежать «знизу і зліва» від  $\mathbf{x} \in \mathbb{R}^k$ .

Таку функцію розподілу називають *спільною функцією розподілу* (joint distribution function).  $\square$

Як і функція розподілу деякої випадкової величини (6.1.1), спільна функція розподілу має три важливі властивості, відповідниками яких є властивості з Твердження 6.1.2. Зокрема, спільна функція розподілу також є нормованою, її властивий аналог неперервності справа, і для неї справедливе узагальнення неспадності. Формальний розгляд цих властивостей можна знайти в Розд. 9.6.3.

У Розділі 6.1 ми зазначали, що не тільки розподіл  $\mathbb{P}_{\mathbf{X}}$  визначає собою деяку функцію розподілу  $F_{\mathbf{X}}$ , а й будь-яка функція розподілу визначає *едину* міру  $\nu$  в Борелевому просторі  $(\mathbb{R}, \mathcal{B})$  таку, що  $\nu((a; b]) = F(b) - F(a)$ ,  $a, b \in \mathbb{R}$  (Теорема 6.1.6). Схожий факт справедливий і для спільних функцій розподілу випадкових векторів, тобто будь-яка спільна функція розподілу задає *едину* відповідну їй міру на Борелевому просторі  $(\mathbb{R}^k, \mathcal{B}^k)$ . Про це докладніше написано в Розд. 9.6.3.

В одновимірному випадку ми доводили, що функція розподілу може мати не більш ніж зліченну кількість стрибків (Твердження 6.1.8), до того ж із Твердження 6.1.5 випливало, що якщо функція  $F_{\mathbf{X}}(x)$  має стрибок у точці  $c$ , то ця точка має ненульову ймовірність:  $\mathbb{P}_{\mathbf{X}}(X = c) = F_{\mathbf{X}}(c) - F_{\mathbf{X}}(c-) > 0$ . На відміну від одновимірного випадку, спільна функція розподілу може мати незліченну кількість точок розриву I роду, більше того, вони можуть становити собою множину міри 0.

**Визначення 9.2.3.** Якщо розподіл деякого випадкового вектора  $\mathbf{X} : (\Omega, \mathcal{A}, \mathbb{P}) \rightarrow (\mathbb{R}^k, \mathcal{B}^k)$  абсолютно неперервний відносно міри Лебега  $\lambda_k$ , то *спільною щільністю розподілу* (joint probability density function) є функція  $f_{\mathbf{X}}$  така, що

$$\mathbb{P}_{\mathbf{X}}(A) = \int_A f_{\mathbf{X}} d\lambda_k, \quad \forall A \in \mathcal{B}^k, \quad (9.2.3)$$

□

За аналогією з (6.3.2) маємо властивість

$$F_{\mathbf{X}}(x_1, \dots, x_k) = \int_{-\infty}^{x_k} \dots \int_{-\infty}^{x_1} f_{\mathbf{X}}(t_1, \dots, t_k) dt_1 \dots dt_k. \quad (9.2.4)$$

Відповідно, маємо узагальнення властивості (6.3.3)<sup>5</sup>:

$$f_{\mathbf{X}}(x_1, \dots, x_k) = \frac{\partial^k}{\partial x_1 \dots \partial x_k} F_{\mathbf{X}}(x_1, \dots, x_k). \quad (9.2.5)$$

**Визначення 9.2.4.** Якщо розподіл деякого випадкового вектора  $\mathbf{X} : (\Omega, \mathcal{A}, \mathbb{P}) \rightarrow (\mathbb{R}^k, \mathcal{B}^k)$  абсолютно неперервний відносно лічної міри  $\#_k$ , то *спільною функцією ймовірності* (joint probability mass function) є

$$p_{\mathbf{X}}(x_1, \dots, x_k) = \begin{cases} \mathbb{P}_{\mathbf{X}}(X_1 = x_1, \dots, X_k = x_k), & x \in \text{supp}(\mathbf{X}) \\ 0, & x \notin \text{supp}(\mathbf{X}) \end{cases}. \quad (9.2.6)$$

□

Як і в одновимірному випадку, спільна щільність<sup>6</sup> має такі властивості.

**Твердження 9.2.5.** Нехай випадковий вектор  $\mathbf{X}$  має спільний розподіл  $\mathbb{P}_{\mathbf{X}}$  зі спільною щільністю  $f_{\mathbf{X}}$ . Тоді:

- (i)  $f_{\mathbf{X}} \geq 0$ ;
- (ii)  $\int_{\mathbb{R}^k} f_{\mathbf{X}} d\lambda_k = 1$ .

Як і в одновимірному випадку, спільну щільність визначено з точністю до множини міри 0.

Як ілюстрацію спільної щільноти на Рис. 9.2.1 наведено щільність випадкового вектора з двовимірним стандартним нормальним розподілом.

**Приклад 9.2.6.** Загальновідомо, що тиск у шинах автомобіля  $X$  поліпшує (до певної межі) продуктивність пробігу  $Y$ . Нехай спільна щільність розподілу відповідних випадкових величин має вид

$$f_{X,Y}(x, y) = cx^2y \cdot \mathbb{1}\{0 < x^2 \leq y < 1\} \cdot \mathbb{1}\{x > 0\}.$$

<sup>5</sup>Для абсолютно неперервних випадкових величин функція розподілу абсолютно неперервна, тому порядок диференціювання не має значення.

<sup>6</sup>Аналогічні властивості справедливі і для функції ймовірності, оскільки функція ймовірності також є щільністю — відносно лічної міри.

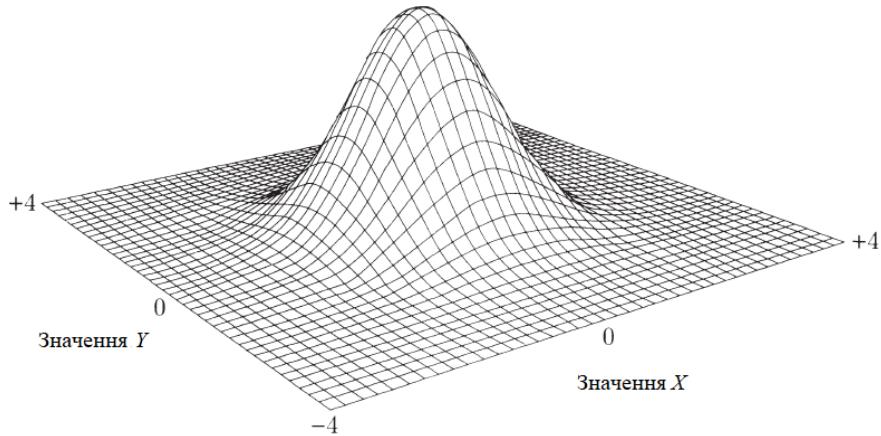


Рис. 9.2.1.: Спільна щільність  $\mathbf{X} = (X, Y)^\top$ , де  $X, Y \sim N(0, 1)$ ,  $X \perp\!\!\!\perp Y$  [9, с. 357]

З'ясуймо значення константи  $c$ :

$$1 = \iint_{\mathbb{R}^2} f_{X,Y}(x,y) = \iint_{0 < x^2 \leq y < 1, x > 0} cx^2y \, dx \, dy.$$

Відповідну область інтегрування зображенено на Рис. 9.2.2a.

Підраховуючи цей інтеграл, маємо

$$\iint_{x^2 \leq y < 1} cx^2y \, dx \, dy = c \int_0^1 \left( x^2 \int_{x^2}^1 y \, dy \right) \, dx = \frac{c}{2} \int_0^1 x^2(1 - x^4) \, dx = \frac{c}{2} \left( \frac{1}{3} - \frac{1}{7} \right) = \frac{2c}{21} = 1,$$

звідки  $c = \frac{21}{2}$ .

Використовуючи цю щільність, підрахуймо, наприклад,  $\mathbb{P}_{X,Y}(0 < X < \frac{3}{4}, \frac{1}{4} \leq Y < 1)$  на основі (9.2.3) (Рис. 9.2.2б):

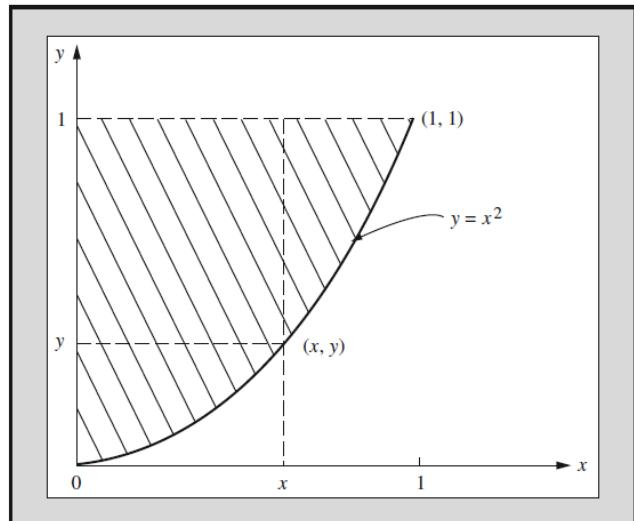
$$\begin{aligned} \mathbb{P}_{X,Y} \left( 0 < X < \frac{3}{4}, \frac{1}{4} \leq Y < 1 \right) &= c \int_0^{\frac{1}{2}} \int_{\frac{1}{4}}^1 x^2y \, dy \, dx + c \int_{\frac{1}{2}}^{\frac{3}{4}} \int_{x^2}^1 x^2y \, dy \, dx \\ &= c \int_0^{\frac{1}{2}} \left( x^2 \int_{\frac{1}{4}}^1 y \, dy \right) \, dx + c \int_{\frac{1}{2}}^{\frac{3}{4}} \left( x^2 \int_{x^2}^1 y \, dy \right) \, dx \\ &= \frac{c}{2} \int_0^{\frac{1}{2}} x^2 \left( 1 - \frac{1}{16} \right) \, dx + \frac{c}{2} \int_{\frac{1}{2}}^{\frac{3}{4}} x^2(1 - x^4) \, dx = \frac{41311}{65536} \approx 0.63. \end{aligned}$$

□

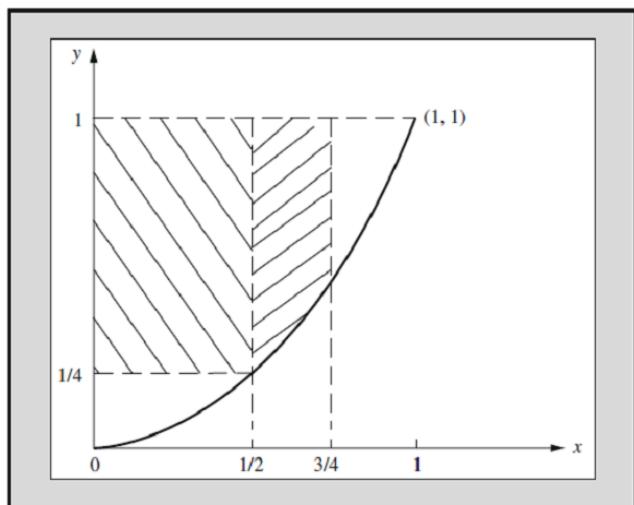
## 9.2.2. Маржинальні функції та щільності розподілу

Як ми бачили в Твердженні 7.3.6, якщо  $\mathbf{X} = (X_1, \dots, X_k)^\top$  є випадковим вектором, то  $X_j$  є випадковою величиною,  $j = 1, \dots, k$ . Відтак, вона має свій власний розподіл  $\mathbb{P}_j$ .

**Визначення 9.2.7.** Розподіл  $\mathbb{P}_j$  випадкової величини  $X_j$  —  $j$ -ої координати деякого випадкового вектора  $\mathbf{X} = (X_1, \dots, X_k)^\top$  — називають *маржинальним розподілом* (marginal distribution).



(a)



(б)

Рис. 9.2.2.: Ілюстрації до Прикладу 9.2.6: (а) носій спільної щільності розподілу; (б) множина  $A = \{(x, y)^\top : 0 < x < \frac{3}{4}, \frac{1}{4} \leq y < 1\}$  ([2], Рис. 7.4-5)

Розподіл  $\mathbb{P}_{j \in I}$  випадкового вектора  $\mathbf{X}_{j \in I} \equiv (X_j)$ ,  $j \in I$ , складеного з координат деякого випадкового вектора  $\mathbf{X} = (X_1, \dots, X_k)^\top$ , індекси яких належать множині  $I$ , також називають маржинальним розподілом, який одночасно є спільним розподілом для  $\mathbf{X}_{j \in I}$ .  $\square$

Розгляньмо функцію проекції  $g_I : \mathbb{R}^k \rightarrow \mathbb{R}^m$  таку, що  $g_I(x_1, \dots, x_k) = (x_j, j \in I)$ . Тоді  $g_I(\mathbf{X}) = \mathbf{X}_{j \in I}$ , і відповідний розподіл має вид

$$\begin{aligned} \mathbb{P}_{\mathbf{X}_{j \in I}}(\mathbf{X}_{j \in I} \in A) &= \mathbb{P}_{\mathbf{X}}(g_I(X_1, \dots, X_k) \in A) \\ &= \mathbb{P}_{\mathbf{X}}((\mathbf{X}_{j \in I} \in A) \cap (X_t \in \mathbb{R}, t \in \{1, \dots, k\} \setminus I)) . \end{aligned}$$

Зокрема, коли  $I = \{j\}$ , маємо

$$\mathbb{P}_{X_j}(X_j \in A) = \mathbb{P}_{\mathbf{X}}((X_j \in A) \cap (X_t \in \mathbb{R}, t \neq j)) . \quad (9.2.7)$$

Маржинальним розподілам відповідають *маржинальні функції розподілу* (marginal distribution function):

$$\begin{aligned} F_{X_j}(x) &= \mathbb{P}_{X_j}(X_j \leq x) = \mathbb{P}_{\mathbf{X}}(g_j(X_1, \dots, X_k) \leq x) \\ &= \mathbb{P}_{\mathbf{X}}(X_1 \in \mathbb{R}, \dots, X_{j-1} \in \mathbb{R}, X_j \leq x, X_{j+1} \in \mathbb{R}, \dots, X_k \in \mathbb{R}) \\ &= \mathbb{P}_{\mathbf{X}}\left(\lim_{\min\{x_1, \dots, x_{j-1}, x_{j+1}, \dots, x_k\} \rightarrow \infty} S_{(x_1, \dots, x_{j-1}, x_j, x_{j+1}, \dots, x_k)^\top}\right) \\ &= \lim_{\substack{x_1 \rightarrow \infty \\ \dots \\ x_{j-1} \rightarrow \infty \\ x_{j+1} \rightarrow \infty \\ \dots \\ x_k \rightarrow \infty}} F_{\mathbf{X}}(x_1, \dots, x_{j-1}, x_j, x_{j+1}, \dots, x_k) . \end{aligned} \quad (9.2.8)$$

Також можна розглянути поняття маржинальної щільності.

**Твердження 9.2.8.** Якщо  $\mathbb{P}_{\mathbf{X}}$  має щільність  $f_{\mathbf{X}}$ , то розподіл  $\mathbb{P}_{X_j}$  має *маржинальну щільність розподілу* (marginal probability density function)

$$f_j(x) = \int_{\mathbb{R}^{k-1}} f_{\mathbf{X}}(x_1, \dots, x_{j-1}, x, x_{j+1}, \dots, x_k) d\lambda_{k-1} , \quad (9.2.9)$$

де інтегрування відбувається за змінними  $x_i$ ,  $i \neq j$ .

*Доведення.* Справді, за Теоремою 6.3.2,

$$\mathbb{P}_{X_j}(A) = \int_A f_j(x) d\lambda ,$$

тому потрібно показати, що інтеграл правої частини рівності (9.2.9) за множиною  $A$  дорівнює  $\mathbb{P}_{X_j}(A)$ .

З урахуванням того, що багатовимірна міра Лебега є добутком одновимірних мір Лебега і зокрема що  $\lambda_k = \lambda_{k-1} \times \lambda$ , можна застосувати Теорему 9.1.8 і подати (повторний) інтеграл правої

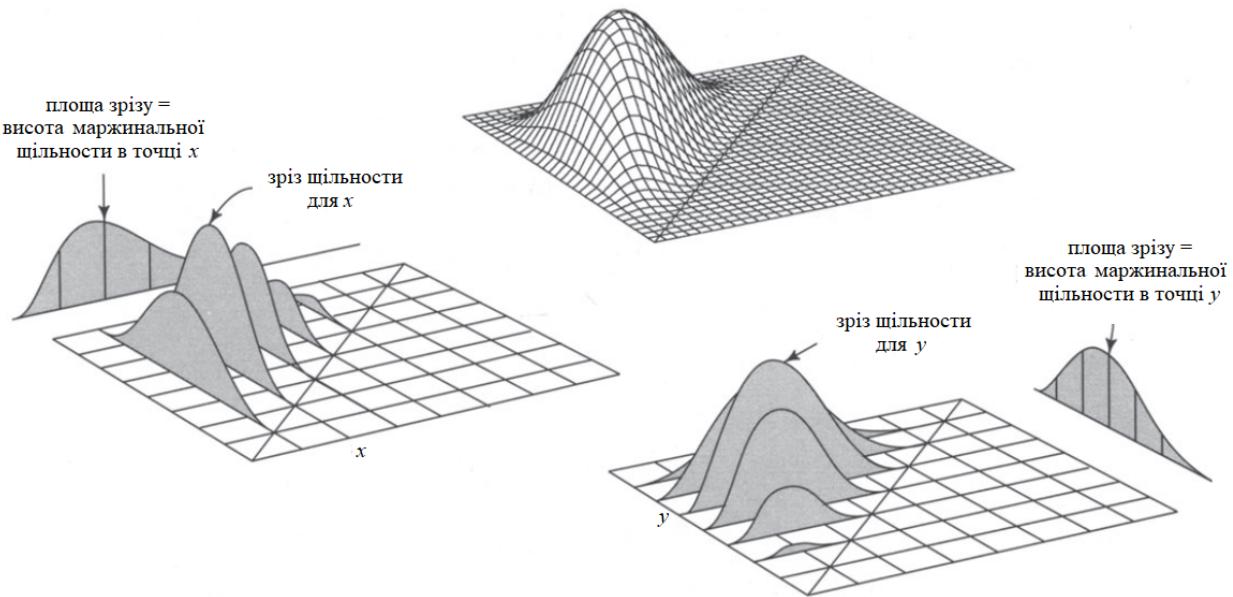


Рис. 9.2.3.: Ілюстративне зображення поняття спільної та маржинальних щільностей ([9], Рис. 6.3.1)

частини рівності (9.2.9) за множиною  $A$  як (кратний) інтеграл:

$$\begin{aligned} \int_A f_j(x) d\lambda &= \int_A \left( \int_{\mathbb{R}^{k-1}} f_{\mathbf{X}}(x_1, \dots, x_{j-1}, x, x_{j+1}, \dots, x_k) d\lambda_{k-1} \right) d\lambda \\ &= \int_{\mathbb{R} \times \dots \times \mathbb{R} \times A \times \mathbb{R} \times \dots \times \mathbb{R}} f_{\mathbf{X}}(x_1, \dots, x_{j-1}, x, x_{j+1}, \dots, x_k) d\lambda_k \\ &= \mathbb{P}_{\mathbf{X}}((X_j \in A) \cap (X_t \in \mathbb{R}, t \neq j)) , \end{aligned}$$

що дорівнює  $\mathbb{P}_{X_j}(X_j \in A)$  відповідно до (9.2.7).  $\square$

Ілюстрацію маржинальної щільності наведено на Рис. 9.2.3.

Як наслідок із цього твердження, можна розглянути спільну щільність  $f_{j \in I}$  для випадкового вектора  $\mathbf{X}_{j \in I} \equiv (X_j)$ ,  $j \in I$ , складеного з координат деякого випадкового вектора  $\mathbf{X} = (X_1, \dots, X_k)^\top$ , індекси яких належать множині  $I$ :

$$f_{j \in I}(x_j, j \in I) = \int_{\mathbb{R}^{k-|I|}} f_{\mathbf{X}}(x_1, \dots, x_k) d\lambda_{k-|I|} ,$$

де інтегрування відбувається за змінними  $x_i$ ,  $i \in \{1, \dots, k\} \setminus I$ .

**Приклад 9.2.9** (Рівномірний розподіл). На основі міри  $\lambda_k$  можна визначити рівномірний розподіл на обмеженому прямокутнику  $A = (a_1; b_1] \times \dots \times (a_k; b_k]$ . Маржинальні функції рівномірного розподілу дорівнюють

$$F_i(x) = \begin{cases} 0 , & x \leq a_i \\ \frac{x - a_i}{b_i - a_i} , & x \in (a_i; b_i] \\ 1 , & x > b_i \end{cases} ,$$

і за аналогією з уведенням міри  $\lambda_k$  покладімо

$$F(x_1, \dots, x_k) = \prod_{i=1}^k F_i(x_i) .$$

Тоді

$$\begin{aligned} F(x_1, \dots, x_k) &= \begin{cases} 0, & \prod_{i=1}^k \mathbb{1}\{x_i \leq a_i\} = 0^7 \\ \frac{\prod_{i=1}^k (x_i - a_i)}{\prod_{i=1}^k (b_i - a_i)}, & \prod_{i=1}^k \mathbb{1}\{x_i \in (a_i; b_i]\} = 1 \\ 1, & \text{інакше} \end{cases} \\ &= \begin{cases} 0, & \prod_{i=1}^k \mathbb{1}\{x_i \leq a_i\} = 0 \\ \frac{\lambda_k \left( \prod_{i=1}^k (a_i; x_i] \right)}{\lambda_k(A)}, & \prod_{i=1}^k \mathbb{1}\{x_i \in (a_i; b_i]\} = 1 \\ 1, & \text{інакше} \end{cases} . \end{aligned}$$

У цьому випадку позначаємо:  $\mathbf{X} \sim \text{U}(A)$ .

Для такого рівномірного розподілу щільністю, за (9.2.5), є

$$f(x_1, \dots, x_k) = \frac{1}{\prod_{i=1}^k (b_i - a_i)} \cdot \prod_{i=1}^k \mathbb{1}\{x_i \in (a_i; b_i]\} .$$

У загальнивши, можна дістати рівномірний розподіл на деякій довільній множині  $B \in \mathcal{B}^k$ :

$$\mathbb{P}_{\mathbf{X}}(C) = \frac{\lambda_k(C \cap B)}{\lambda_k(B)} , \quad C \subseteq B , C \in \mathcal{B}^k ,$$

де щільність розподілу має вид

$$f_{\mathbf{X}}(x_1, \dots, x_k) = \frac{1}{\lambda_k(C)} \cdot \mathbb{1}\{(x_1, \dots, x_k)^\top \in C\} , \quad (9.2.10)$$

так що

$$\int_C f_{\mathbf{X}}(x_1, \dots, x_k) d\lambda_k = 1 .$$

□

**Приклад 9.2.10** (Таблиця спряженості). Розгляньмо приклад дискретного випадкового вектора  $\mathbf{X} = (X, Y)^\top$ , для якого маржинальними розподілами є розподіли Бернуллі. У цьому випадку спільну функцію ймовірності можна однозначно задати перерахуванням чотирьох значень —  $\mathbb{P}_{\mathbf{X}}(X = 0, Y = 0)$ ,  $\mathbb{P}_{\mathbf{X}}(X = 0, Y = 1)$ ,  $\mathbb{P}_{\mathbf{X}}(X = 1, Y = 0)$ ,  $\mathbb{P}_{\mathbf{X}}(X = 1, Y = 1)$ , які прийнято подавати у виді *таблиці спряженості* (contingency table).

Такі таблиці часто виринають у контексті дослідження впливу деякого фактора (наприклад, нових ліків чи деякої економічної політики) на деяку величину (наприклад, одужання від хвороби

<sup>7</sup>Тобто що принаймні один  $x_i$  не перевищує  $a_i$ .

Табл. 9.2.1.: Таблиця спряженості для Прикладу 9.2.10

		$Y = 1$	$Y = 0$	Разом
$X = 1$	0.05	0.20	0.25	
$X = 0$	0.03	0.72	0.75	
Разом	0.08	0.92	1.00	

чи певний економічний результат).

Для прикладу покладімо  $X = 1$ , якщо особа є курцем, а  $Y = 1$ , якщо особа захворіла на рак легень протягом життя. Нехай спільна функція розподілу відповідає Табл. 9.2.1<sup>8</sup>.

Маржинальні функції ймовірності можна зчитати з рядка та стовпця «Разом»:

$$\begin{aligned}\mathbb{P}_Y(Y = 1) &= \int_{\mathbb{R}} \mathbb{P}_{\mathbf{X}}(X = x, Y = 1) \#(dx) \\ &= \mathbb{P}_{\mathbf{X}}(X = 1, Y = 1) + \mathbb{P}_{\mathbf{X}}(X = 0, Y = 1) = 0.05 + 0.03 = 0.08 ,\end{aligned}$$

$$\begin{aligned}\mathbb{P}_Y(Y = 0) &= \int_{\mathbb{R}} \mathbb{P}_{\mathbf{X}}(X = x, Y = 0) \#(dx) \\ &= \mathbb{P}_{\mathbf{X}}(X = 1, Y = 0) + \mathbb{P}_{\mathbf{X}}(X = 0, Y = 0) = 0.20 + 0.72 = 0.92 .\end{aligned}$$

Аналогічно  $\mathbb{P}_X(X = 1) = 0.05 + 0.20 = 0.25$ ,  $\mathbb{P}_X(X = 0) = 0.03 + 0.72 = 0.75$ .

Отже з Табл. 9.2.1 випливає:

$$X \sim \text{Bern}(0.25) , \quad Y \sim \text{Bern}(0.08) . \quad \square$$

**Приклад 9.2.11.** Повернімося до Прикладу 9.2.6 та підрахуймо маржинальні щільності розподілів:

$$\begin{aligned}f_X(x) &= \int_{\mathbb{R}} cx^2 y \cdot \mathbb{1}\{0 < x^2 \leq y < 1\} \cdot \mathbb{1}\{x > 0\} dy = cx^2 \cdot \mathbb{1}\{0 < x < 1\} \int_{x^2}^1 y dy \\ &= \frac{21}{4} x^2 (1 - x^4) \cdot \mathbb{1}\{0 < x < 1\} , \\ f_Y(y) &= \int_{\mathbb{R}} cx^2 y \cdot \mathbb{1}\{0 < x^2 \leq y < 1\} \cdot \mathbb{1}\{x > 0\} dx = cy \cdot \mathbb{1}\{0 < y < 1\} \int_0^{\sqrt{y}} x^2 dx \\ &= \frac{7}{2} y^2 \sqrt{y} \cdot \mathbb{1}\{0 < y < 1\} .\end{aligned}$$

□

<sup>8</sup>Відповідні числа взято з [1, с. 308], вони є суттєво ілюстративними.

### 9.3. Сподівання випадкового вектора та коваріація

#### 9.3.1. Сподівання випадкового вектора

Аналогічно поняттю сподівання випадкової величини, уведеного в Розд. 6.4, можна розглянути поняття сподівання випадкового вектора.

**Визначення 9.3.1.** Сподіванням (expectation) випадкового вектора  $\mathbf{X} = (X_1, \dots, X_k)^\top : (\Omega, \mathcal{A}, \mathbb{P}) \rightarrow \mathbb{R}^k$  є вектор, складений зі сподівань його координат:

$$\mathbb{E}[\mathbf{X}] = \begin{pmatrix} \mathbb{E}[X_1] \\ \vdots \\ \mathbb{E}[X_k] \end{pmatrix}. \quad (9.3.1)$$

□

За властивістю лінійності (одновимірного) сподівання маємо, для деякої матриці  $\mathbf{A} \in \mathbb{R}^n \times \mathbb{R}^k$  та деякого вектора  $\mathbf{b} \in \mathbb{R}^n$ :

$$\mathbb{E}[\mathbf{A}\mathbf{X} + \mathbf{b}] = \mathbf{A}\mathbb{E}[\mathbf{X}] + \mathbf{b}. \quad (9.3.2)$$

Справедливим є аналог Теореми 6.4.2.

**Теорема 9.3.2.** Нехай  $\mathbf{X} = (X_1, \dots, X_k)^\top : (\Omega, \mathcal{A}, \mathbb{P}) \rightarrow \mathbb{R}^k$  — деякий випадковий вектор, який породжує розподіл  $\mathbb{P}_{\mathbf{X}}$ , який абсолютно неперервний відносно деякої міри  $\mu$  і має (спільну) щільність  $f_{\mathbf{X}}$ . Нехай  $g : \mathbb{R}^k \rightarrow \mathbb{R}$  — деяка вимірна функція, так що  $g(X_1, \dots, X_k) : \Omega \rightarrow \mathbb{R}$  є випадковою величиною. Тоді

$$\int g(X_1(\omega), \dots, X_k(\omega)) d\mathbb{P} = \int g(x_1, \dots, x_k) d\mathbb{P}_{\mathbf{X}} = \int g(x_1, \dots, x_k) \cdot f_{\mathbf{X}}(x_1, \dots, x_k) d\mu,$$

у тому розумінні, що якщо існує один інтеграл, то існують усі інші, і вони дорівнюють один одному.

**Приклад 9.3.3.** Нехай на ділянці дороги довжиною  $L$  стається аварія в точці  $X$ , а машина швидкої допомоги стоїть у точці  $Y$ . Нехай  $(X, Y)^\top$  має рівномірний розподіл:  $(X, Y)^\top \sim U((0; L] \times (0; L])$ .

Підрахуймо середню відстань між місцем аварії та швидкою:

$$\begin{aligned}
 \mathbb{E} [|X - Y|] &= \iint_{\mathbb{R}^2} |x - y| \cdot f_{X,Y}(x, y) dx dy \\
 &= \frac{1}{L^2} \iint_{\mathbb{R}^2} |x - y| \cdot \mathbb{1}\{x \in (0; L]\} \cdot \mathbb{1}\{y \in (0; L]\} dx dy \\
 &= \frac{1}{L^2} \int_0^L \int_0^L |x - y| dx dy \\
 &= \frac{1}{L^2} \left( \int_0^L \int_0^x (x - y) dy dx + \int_0^L \int_x^L (y - x) dy dx \right) \\
 &= \frac{1}{L^2} \int_0^L \left( \frac{L^2}{2} + x^2 - xL \right) dx \\
 &= \frac{L}{3}.
 \end{aligned}$$

На основі цього результату можна відразу обчислити сподівання  $M_{\max} = \max\{X, Y\}$ ,  $M_{\min} = \min\{X, Y\}$ , оскільки  $M_{\max} + M_{\min} = X + Y$ , а  $M_{\max} - M_{\min} = |X - Y|$ :

$$\begin{aligned}
 \mathbb{E}[M_{\max} + M_{\min}] &= \mathbb{E}[X + Y] = L, \\
 \mathbb{E}[M_{\max} - M_{\min}] &= \mathbb{E}[|X - Y|] = \frac{L}{3}.
 \end{aligned}$$

Із цієї системи двох рівнянь бачимо, що

$$\mathbb{E}[M_{\max}] = \frac{2L}{3}, \quad \mathbb{E}[M_{\min}] = \frac{L}{3}.$$

Зокрема, можна бачити, що  $\mathbb{E}[M_{\max}]$  та  $\mathbb{E}[M_{\min}]$  розташовано на однаковій відстані від  $\frac{L}{2}$ , що пояснюється симетрією.  $\square$

### 9.3.2. Коваріація та кореляція двох випадкових величин

Аналогом дисперсії — розкиду значень випадкової величини відносно її сподівання — у багатовимірному випадку є матриця коваріацій. Щоб мати можливість розглянути це поняття, спочатку торкнімося поняття коваріації двох випадкових величин.

**Визначення 9.3.4.** *Коваріацією* (covariance) двох випадкових величин  $X_i$  і  $X_j$  є

$$\text{Cov}(X_i, X_j) = \mathbb{E}[(X_i - \mathbb{E}[X_i])(X_j - \mathbb{E}[X_j])]. \quad (9.3.3)$$

$\square$

За аналогією з позначенням дисперсії, коваріацію часто позначають як  $\sigma_{XY}$ .

За аналогією з одновимірним випадком можна довести деякі властивості коваріації.

**Твердження 9.3.5.** Коваріація  $\text{Cov}(X, Y)$  двох випадкових величин  $X$  та  $Y$  має такі властивості:

- $\text{Cov}(X, Y) = \text{Cov}(Y, X);$
- $\text{Cov}(X, X) = \text{Var}(X);$
- $\text{Cov}(X, Y) = \mathbb{E}[XY] - \mathbb{E}[X]\mathbb{E}[Y];$
- $\text{Cov}(X, a) = 0, a \in \mathbb{R};$
- $\text{Cov}(aX + bY, cV + dW) = ac\text{Cov}(X, V) + ad\text{Cov}(X, W) + bc\text{Cov}(Y, V) + bd\text{Cov}(Y, W);$
- $\text{Var}(X + Y) = \text{Var}(X) + \text{Var}(Y) + 2\text{Cov}(X, Y).$

*Доведення.* (i) Це безпосередньо випливає з визначення;

(ii) це також безпосередньо випливає з визначення;

(iii)

$$\begin{aligned}\text{Cov}(X, Y) &= \mathbb{E}[(X - \mathbb{E}[X])(Y - \mathbb{E}[Y])] \\ &= \mathbb{E}[XY - X\mathbb{E}[Y] - Y\mathbb{E}[X] + \mathbb{E}[X]\mathbb{E}[Y]] \\ &= \mathbb{E}[XY] - 2\mathbb{E}[X]\mathbb{E}[Y] + \mathbb{E}[X]\mathbb{E}[Y] = \mathbb{E}[XY] - \mathbb{E}[X]\mathbb{E}[Y] ;\end{aligned}$$

(iv) це безпосередньо випливає з того факту, що  $\mathbb{E}[a] = a;$

(v) це безпосередньо випливає з властивості лінійності інтегралу;

(vi)

$$\begin{aligned}\text{Var}(X + Y) &= \mathbb{E}[(X + Y)^2] - (\mathbb{E}[(X + Y)])^2 \\ &= \mathbb{E}[X^2] + 2\mathbb{E}[XY] + \mathbb{E}[Y]^2 - ((\mathbb{E}[X])^2 + 2\mathbb{E}[X]\mathbb{E}[Y] + (\mathbb{E}[Y])^2) \\ &= \text{Var}(X) + \text{Var}(Y) + 2\text{Cov}(X, Y) .\end{aligned}$$

□

З останньої властивості Твердження 9.3.5 випливає

$$\text{Var}(X - Y) = \text{Var}(X) + \text{Var}(Y) - 2\text{Cov}(X, Y) .$$

Частою помилкою є використання знаку «мінус» перед  $\text{Var}(Y)$ , адже  $\text{Var}(-Y) = (-1)^2\text{Var}(Y) = \text{Var}(Y)$ .

Відповідну властивість можна узагальнити на випадок  $k$  випадкових величин:

$$\text{Var}\left(\sum_{i=1}^k X_i\right) = \sum_{i=1}^k \text{Var}(X_i) + 2 \sum_{i < j} \text{Cov}(X_i, X_j) . \quad (9.3.4)$$

Пригадаймо, що середньоквадратичне відхилення (4.3.5) було запропоновано, оскільки випадкова величина та її дисперсія мають різні одиниці виміру. Коваріація  $\text{Cov}(X, Y)$  також має інші одиниці вимір, аніж  $X$  чи  $Y$ . Понад те, величини  $X$  та  $Y$  можуть самі мати різні одиниці виміру. Тому на практиці доцільніше використовувати нормовану величину, яка взагалі не має одиниць виміру, і до того ж має зручну інтерпретацію.

**Визначення 9.3.6.** *Коефіцієнтом кореляції* (correlation) двох випадкових величини  $X$  та  $Y$  є

$$\text{Corr}(X_i, X_j) = \frac{\text{Cov}(X_i, X_j)}{\sqrt{\text{Var}(X)} \sqrt{\text{Var}(Y)}}. \quad (9.3.5)$$

□

Коефіцієнт кореляції часто позначають як  $\rho_{XY}$ .

**Твердження 9.3.7.** Коефіцієнт кореляції (9.3.5) між довільними випадковими величинами  $X$  та  $Y$  лежить на проміжку  $[-1; 1]$ .

*Доведення.* Розгляньмо стандартизовані відповідно до (7.5.7) випадкові величини

$$X^* = \frac{X - \mathbb{E}[X]}{\sigma_X}, \quad Y^* = \frac{Y - \mathbb{E}[Y]}{\sigma_Y},$$

для яких  $\mathbb{E}[X^*] = \mathbb{E}[Y^*] = 0$ ,  $\text{Var}(X^*) = \text{Var}(Y^*) = 1$ . Тоді  $\mathbb{E}[X^*Y^*] = \text{Corr}(X, Y)$ .

Оскільки  $\mathbb{E}[(X^*)^2] = \mathbb{E}[(Y^*)^2] = 1$ ,

$$\begin{aligned} 0 &\leq \mathbb{E}[(X^* - Y^*)^2] = \mathbb{E}[(X^*)^2] - 2\mathbb{E}[X^*Y^*] + \mathbb{E}[(Y^*)^2] = 2 - 2\mathbb{E}[X^*Y^*], \\ 0 &\leq \mathbb{E}[(X^* + Y^*)^2] = \mathbb{E}[(X^*)^2] + 2\mathbb{E}[X^*Y^*] + \mathbb{E}[(Y^*)^2] = 2 + 2\mathbb{E}[X^*Y^*], \end{aligned}$$

звідки випливає  $-1 \leq \mathbb{E}[X^*Y^*] = \text{Corr}(X, Y) \leq 1$ .

□

Коваріація двох випадкових величин  $\text{Cov}(X, Y)$  показує ступінь лінійного зв'язку між ними:

- якщо  $\text{Cov}(X, Y) > 0$ , то між величинами існує додатний лінійний зв'язок, тобто більші значення однієї величини свідчать про більші значення іншої;
- якщо  $\text{Cov}(X, Y) < 0$ , то між величинами існує від'ємний лінійний зв'язок, тобто більші значення однієї величини свідчать про менші значення іншої;
- якщо  $\text{Cov}(X, Y) = 0$ , то лінійного зв'язку між величинами немає.

Те саме можна сказати про  $\text{Corr}(X, Y)$ , але оскільки коефіцієнт кореляції обмежений у своїх значеннях, можна казати, що значення  $\text{Corr}(X, Y)$ , близькі за модулем до 1, свідчать про сильнішу лінійну залежність.

Це випливає з доведення Твердження 9.3.7. Справді,  $\text{Corr}(X, Y) = 1$  тоді й тільки тоді, коли  $\mathbb{E}[(X^* - Y^*)^2] = 0$ , тобто тоді й тільки тоді, коли  $X^* = Y^*$  майже напевно. Іншими словами, це станеться тоді й тільки тоді, коли

$$Y = \mathbb{E}[Y] + \frac{\sigma_Y}{\sigma_X}(X - \mathbb{E}[X]),$$

тобто тоді й тільки тоді, коли  $Y = aX + b$  для деяких констант  $a, b \in \mathbb{R}$ ,  $a > 0$ .

Аналогічно можна показати, що  $\text{Corr}(X, Y) = -1$ , коли

$$Y = \mathbb{E}[Y] - \frac{\sigma_Y}{\sigma_X}(X - \mathbb{E}[X]),$$

тобто тоді й тільки тоді, коли  $Y = aX + b$  для деяких констант  $a, b \in \mathbb{R}$ ,  $a < 0$ .

На Рис. 9.3.1 наведено деякий набір реалізацій випадкових величин  $X$  та  $Y$  із різними значеннями коваріації. Варто звернути увагу на випадок (г), оскільки це є типовий приклад, коли коваріація дорівнює нулю (лінійного зв'язку немає), хоча між випадковими величинами існує очевидний зв'язок (у даному випадку — квадратичний).

**Приклад 9.3.8.** Повернімося до Прикладу 9.2.6 та підрахуймо кореляцію між  $X$  та  $Y$ :

$$\mathbb{E}[X] = \int_{\mathbb{R}} x \cdot \frac{21}{4} x^2 (1 - x^4) \cdot \mathbb{1}\{0 < x < 1\} dx = \frac{21}{4} \left( \frac{x^4}{4} \Big|_0^1 - \frac{x^8}{8} \Big|_0^1 \right) = \frac{21}{32},$$

$$\mathbb{E}[X^2] = \int_{\mathbb{R}} x^2 \cdot \frac{21}{4} x^2 (1 - x^4) \cdot \mathbb{1}\{0 < x < 1\} dx = \frac{21}{4} \left( \frac{x^5}{5} \Big|_0^1 - \frac{x^9}{9} \Big|_0^1 \right) = \frac{7}{15},$$

$$\text{Var}(X) = \mathbb{E}[X^2] - (\mathbb{E}[X])^2 \approx 0.036,$$

$$\mathbb{E}[Y] = \int_0^1 y \cdot \frac{7}{2} y^2 \sqrt{y} \cdot \mathbb{1}\{0 < y < 1\} dy = \frac{7}{2} \cdot \frac{2}{9} y^{9/2} \Big|_0^1 = \frac{7}{9},$$

$$\mathbb{E}[Y^2] = \int_0^1 y^2 \cdot \frac{7}{2} y^2 \sqrt{y} \cdot \mathbb{1}\{0 < y < 1\} dy = \frac{7}{2} \cdot \frac{2}{11} y^{11/2} \Big|_0^1 = \frac{7}{11},$$

$$\text{Var}(Y) = \mathbb{E}[Y^2] - (\mathbb{E}[Y])^2 \approx 0.031,$$

$$\mathbb{E}[XY] = \int_{\mathbb{R}^2} xy \cdot \frac{21}{2} x^2 y \cdot \mathbb{1}\{0 < x^2 \leq y < 1\} \cdot \mathbb{1}\{x > 0\} dy dx = \frac{21}{2} \int_0^1 x^3 \left( \int_{x^2}^1 y^2 dy \right) dx = \frac{21}{40},$$

$$\text{Cov}(X, Y) = \mathbb{E}[XY] - \mathbb{E}[X]\mathbb{E}[Y] = \frac{7}{480},$$

$$\text{Corr}(X, Y) = \frac{7/480}{\sqrt{0.036}\sqrt{0.031}} \approx 0.434 > 0.$$

Те, що кореляція додатна, інтуїтивно очікувано: вищі значення тиску в шинах відповідають вищим показникам продуктивності пробігу.  $\square$

### 9.3.3. Матриця коваріацій випадкового вектора

**Визначення 9.3.9.** Нехай  $\mathbf{X} = (X_1, \dots, X_k)^\top : (\Omega, \mathcal{A}, \mathbb{P}) \rightarrow \mathbb{R}^k$  — деякий випадковий вектор. Тоді його *матрицею коваріацій* (covariance matrix) є матриця

$$\begin{aligned} \text{Cov}(\mathbf{X}) &= \mathbb{E}[(\mathbf{X} - \mathbb{E}[\mathbf{X}])(\mathbf{X} - \mathbb{E}[\mathbf{X}])^\top] \\ &= \mathbb{E} \left[ \begin{pmatrix} X_1 - \mathbb{E}[X_1] \\ \vdots \\ X_k - \mathbb{E}[X_k] \end{pmatrix} \cdot \left( (X_1 - \mathbb{E}[X_1], \dots, X_k - \mathbb{E}[X_k]) \right)^\top \right] \\ &= \begin{pmatrix} \text{Var}(X_1) & \text{Cov}(X_1, X_2) & \dots & \text{Cov}(X_1, X_k) \\ \text{Cov}(X_2, X_1) & \text{Var}(X_2) & \dots & \text{Cov}(X_2, X_k) \\ \vdots & \vdots & \ddots & \vdots \\ \text{Cov}(X_k, X_1) & \text{Cov}(X_k, X_2) & \dots & \text{Var}(X_k) \end{pmatrix}. \end{aligned} \tag{9.3.6}$$

$\square$

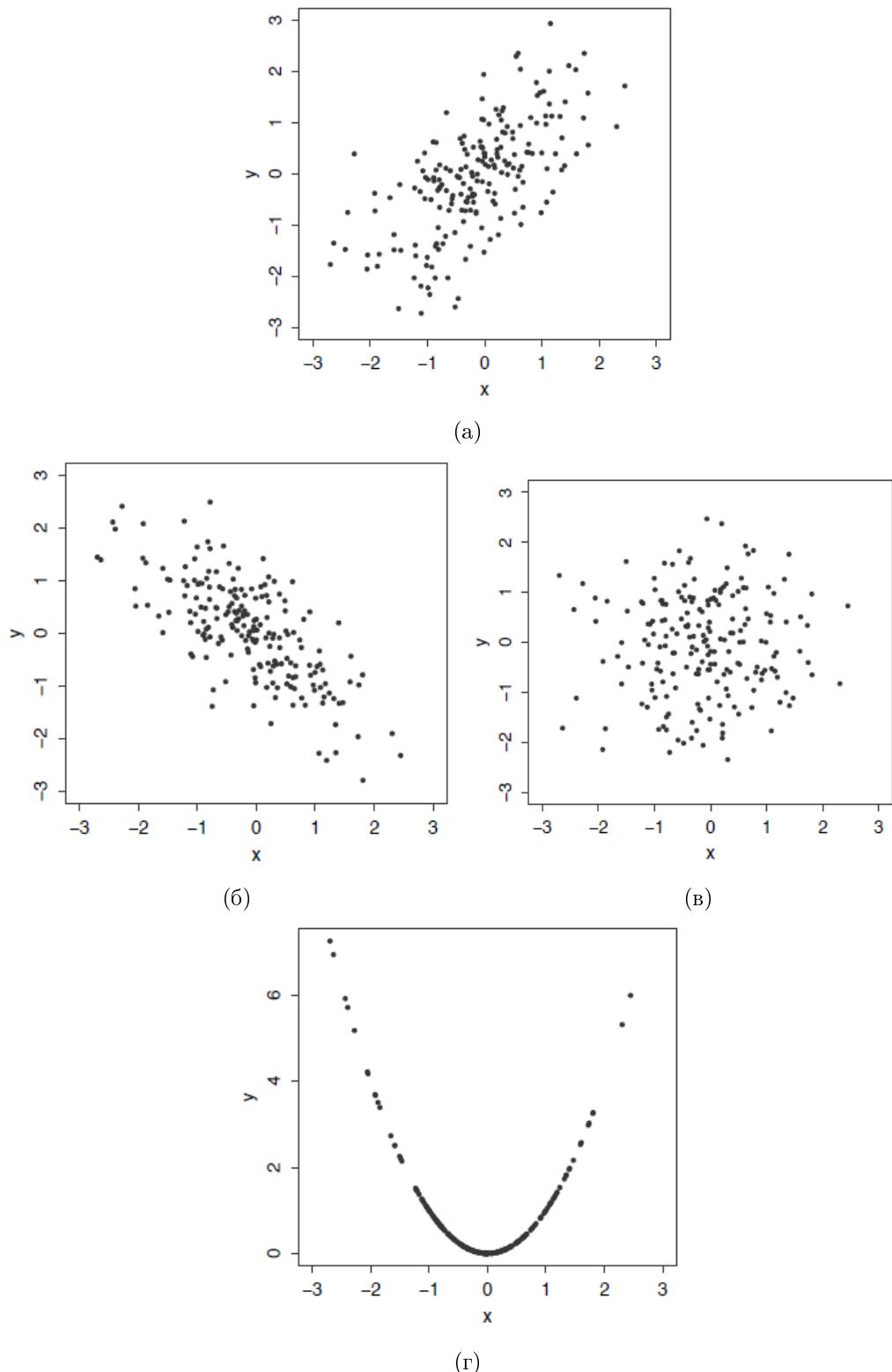


Рис. 9.3.1.: Реалізації випадкових величин  $X$  та  $Y$ : (а)  $\text{Cov}(X, Y) > 0$ ; (б)  $\text{Cov}(X, Y) < 0$ ; (в)  $\text{Cov}(X, Y) = 0$ , між величинами немає зв'язку; (г)  $\text{Cov}(X, Y) = 0$ , між величинами існує квадратичний зв'язок ([1], Рис. 7.9)

Для матриці коваріацій справедливі такі властивості.

**Твердження 9.3.10.** Матриця коваріацій  $\text{Cov}(\mathbf{X}) \equiv \Sigma$  деякого випадкового вектора  $\mathbf{X}$  має такі властивості:

- (i) вона симетрична:  $\Sigma^\top = \Sigma$ ;
- (ii)  $\text{Cov}(\mathbf{AX} + \mathbf{b}) = \mathbf{A}\Sigma\mathbf{A}^\top$ <sup>9</sup>;
- (iii) вона невід'ємно визначена<sup>10</sup>:  

$$\mathbf{a}^\top \Sigma \mathbf{a} \geq 0, \quad \mathbf{a} \in \mathbb{R}^k.$$

*Доведення.* (i) Це випливає з симетричності коваріації;

- (ii) це випливає з лінійності сподівання (9.3.2);
- (iii) розгляньмо деякий вектор  $\mathbf{a} \in \mathbb{R}^k$ . Тоді  $\mathbf{a}^\top \mathbf{X} = \sum_{i=1}^k a_i X_i$  є деякою (одновимірною) випадковою величиною. Відтак, дисперсія такої випадкової величини  $\text{Var}(\mathbf{a}^\top \mathbf{X}) \geq 0$ . З іншого боку, оскільки  $\mathbb{E}[\mathbf{a}^\top \mathbf{X}] = \mathbb{E}\left[\sum_{i=1}^k a_i X_i\right] = \sum_{i=1}^k a_i \mathbb{E}[X_i] = \mathbf{a}^\top \mathbb{E}[\mathbf{X}]$ , маємо:

$$\begin{aligned} \text{Var}(\mathbf{a}^\top \mathbf{X}) &= \mathbb{E}\left[\left(\mathbf{a}^\top \mathbf{X} - \mathbf{a}^\top \mathbb{E}[\mathbf{X}]\right)^2\right] = \\ &= \mathbb{E}\left[\left(\mathbf{a}^\top \mathbf{X} - \mathbf{a}^\top \mathbb{E}[\mathbf{X}]\right)\left(\mathbf{a}^\top \mathbf{X} - \mathbf{a}^\top \mathbb{E}[\mathbf{X}]\right)^\top\right] = \\ &= \mathbb{E}\left[\mathbf{a}^\top (\mathbf{X} - \mathbb{E}[\mathbf{X}]) (\mathbf{X} - \mathbb{E}[\mathbf{X}])^\top \mathbf{a}\right] = \mathbf{a}^\top \text{Cov}(\mathbf{X}) \mathbf{a} \geq 0. \end{aligned}$$

□

## 9.4. Незалежність випадкових величин

### 9.4.1. Визначення і критерії незалежності

Розгляньмо одне з найважливіших визначень теорії ймовірностей — визначення незалежних випадкових величин.

**Визначення 9.4.1.** Випадкові величини  $X_1, \dots, X_k$  *незалежні* (independent), якщо

$$\mathbb{P}_{\mathbf{X}}(X_1 \in H_1, \dots, X_k \in H_k) = \mathbb{P}_{X_1}(X_1 \in H_1) \dots \mathbb{P}_{X_k}(X_k \in H_k), \quad H_i \in \mathcal{B}, \quad i = 1, \dots, k. \quad (9.4.1)$$

У загальному випадку випадкові вектори  $\mathbf{X}_1 : (\Omega, \mathcal{A}, \mathbb{P}) \rightarrow \mathbb{R}^{d_1}, \dots, \mathbf{X}_k : (\Omega, \mathcal{A}, \mathbb{P}) \rightarrow \mathbb{R}^{d_k}$  незалежні, якщо

$$\mathbb{P}_{\mathbf{X}}(\mathbf{X}_1 \in H_1, \dots, \mathbf{X}_k \in H_k) = \mathbb{P}_{\mathbf{X}_1}(\mathbf{X}_1 \in H_1) \dots \mathbb{P}_{\mathbf{X}_k}(\mathbf{X}_k \in H_k), \quad H_i \in \mathcal{B}^{d_i}, \quad i = 1, \dots, k. \quad (9.4.2)$$

Нескінчена множина випадкових величин чи векторів незалежна, якщо незалежною є будь-яка скінчена підмножина цих величин чи векторів. □

<sup>9</sup>Порівняйте з аналогічною властивістю для дисперсій:  $\text{Var}(aX + b) = a^2 \text{Var}(X)$ .

<sup>10</sup>Порівняйте з аналогічною властивістю для дисперсій:  $\text{Var}(X) \geq 0$ .

**Зауваження 9.4.2.** Визначення 3.4.1 каже про те, що повна незалежність набору подій передбачає незалежність будь-якої підмножини подій (усіх пар, усіх трійок тощо), що автоматично враховано в (9.4.1), оскільки завжди можна покласти  $H_i = \mathbb{R}$  для деяких  $i$ .  $\square$

Визначення 9.4.1 незалежності випадкових величин не має значної користі, оскільки перевірка (9.4.1) на практиці нетривіальна. Натомість із цієї характеристизації випливають декілька ключових результатів.

**Теорема 9.4.3.** Випадкові величини  $X_1, \dots, X_k$  незалежні тоді й тільки тоді, коли

$$\mathbb{P}_{\mathbf{X}}(X_1 \leq x_1, \dots, X_k \leq x_k) = \prod_{i=1}^k \mathbb{P}_{X_i}(X_i \leq x_i), \quad x_i \in (-\infty; \infty) \cup \{\infty\}, \quad i = 1, \dots, k. \quad (9.4.3)$$

Доведення цієї теореми можна знайти в Розд. 9.6.4.

Очевидно, що якщо випадковий вектор  $\mathbf{X}$  має функцію розподілу  $F_{\mathbf{X}}$ , а випадкові величини  $X_1, \dots, X_k$  — функції розподілу  $F_{X_1}, \dots, F_{X_k}$  відповідно, то (9.4.3) набуває форми

$$F_{\mathbf{X}}(x_1, \dots, x_k) = F_{X_1}(x_1) \cdot \dots \cdot F_{X_k}(x_k), \quad x_i \in (-\infty; \infty) \cup \{\infty\}, \quad i = 1, \dots, k. \quad (9.4.4)$$

Схожий результат справедливий для розподілів.

**Теорема 9.4.4.** Нехай величини  $X_1, \dots, X_k$  мають розподіли  $\mathbb{P}_{X_1}, \dots, \mathbb{P}_{X_k}$  відповідно. Вони є незалежні тоді й тільки тоді, коли розподіл  $\mathbb{P}_{\mathbf{X}}$  випадкового вектора  $\mathbf{X} = (X_1, \dots, X_k)$  є добутком відповідних маржинальних розподілів:

$$\mathbb{P}_{\mathbf{X}} = \mathbb{P}_{X_1} \times \dots \times \mathbb{P}_{X_k}. \quad (9.4.5)$$

Нарешті, якщо кожна випадкова змінна  $X_i$  має щільність  $f_i$ , то справедливий такий критерій.

**Теорема 9.4.5.** Випадкові величини  $X_1, \dots, X_k$  зі щільностями  $f_1, \dots, f_k$  відповідно, незалежні тоді й тільки тоді, коли випадковий вектор  $\mathbf{X} = (X_1, \dots, X_k)$  має щільність  $f_{\mathbf{X}}$  таку, що

$$f_{\mathbf{X}}(x_1, \dots, x_k) = f_1(x_1) \cdot \dots \cdot f_k(x_k), \quad x_i \in \mathbb{R}, \quad i = 1, \dots, k. \quad (9.4.6)$$

*Доведення.* Доведення випливає з (9.4.4) з урахуванням (9.2.4):

$$\begin{aligned} \int_{-\infty}^{x_k} \dots \int_{-\infty}^{x_1} f_{\mathbf{X}}(t_1, \dots, t_k) dt_1 \dots dt_k &= F_{\mathbf{X}}(x_1, \dots, x_k) \\ &= F_{X_1}(x_1) \cdot \dots \cdot F_{X_k}(x_k) \\ &= \left( \int_{-\infty}^{x_1} f_1(t) dt \right) \cdot \dots \cdot \left( \int_{-\infty}^{x_k} f_k(t) dt \right) \\ &= \int_{-\infty}^{x_k} \dots \int_{-\infty}^{x_1} f_1(t_1) \cdot \dots \cdot f_k(t_k) dt_1 \dots dt_k. \end{aligned} \quad \square$$

**Приклад 9.4.6.** Розгляньмо розподіл, рівномірний на одиничному квадраті (Рис. 9.4.1a):

$$f_{X,Y}(x, y) = \mathbb{1}\{x \in (0; 1]\} \cdot \mathbb{1}\{y \in (0; 1]\}.$$

Випадкові величини  $X$  і  $Y$  у цьому випадку незалежні за Теоремою 9.4.5, оскільки

$$f_{X,Y}(x, y) = f_X(x)f_Y(y) ,$$

де

$$f_X(x) = \mathbb{1}\{x \in (0; 1]\} , \quad f_Y(y) = \mathbb{1}\{y \in (0; 1]\} .$$

При цьому маржинальні розподіли також рівномірні.

Можна пересвідчитися в тому, що це справді маржинальні розподіли, використавши (9.2.9):

$$f_X(x) = \int_{\mathbb{R}} f_{X,Y}(x, y) dy = \int_0^1 \mathbb{1}\{x \in (0; 1]\} dy = \mathbb{1}\{x \in (0; 1]\} ,$$

і аналогічно для  $f_Y(y)$ .

Розгляньмо тепер розподіл, рівномірний на одиничному крузі (Рис. 9.4.16):

$$f_{X,Y}(x, y) = \frac{1}{\pi} \cdot \mathbb{1}\{x^2 + y^2 \leq 1\} .$$

Величини  $X$  та  $Y$  у цьому випадку *не є* незалежними. Справді, для  $x^2 \leq 1$  маємо

$$f_X(x) = \int_{\mathbb{R}} f_{X,Y}(x, y) dy = \frac{1}{\pi} \int_{\mathbb{R}} \mathbb{1}\{x^2 + y^2 \leq 1\} dy = \frac{1}{\pi} \int_{-\sqrt{1-x^2}}^{\sqrt{1-x^2}} dy = \frac{2}{\pi} \sqrt{1-x^2} ,$$

а, відповідно, для всіх  $x^2 > 1$  маємо 0.

Аналогічно для  $Y$ :

$$f_Y(y) = \frac{2}{\pi} \sqrt{1-y^2} \cdot \mathbb{1}\{y^2 \leq 1\} .$$

Очевидно, добуток двох маржинальних щільностей не дає спільної щільності (інакше спільна щільність охоплювала б увесь зафарбований квадрат на Рис. 9.4.16), а самі маржинальні щільності не відповідають рівномірному розподілу (оскільки величини схильні набувати значень біля 0 з більшою ймовірністю).

Варто зазначити, що ключовою умовою для застосування (9.4.6) є її виконання *для всіх* точок простору, а не тільки точок із носія спільної щільності. Справді, функцію  $f_{X,Y}$  можна розглядати як добуток двох щільностей рівномірного розподілу ( $g(x) = 1/\pi$  і  $h(y) = 1$ ), проте це справедливо тільки для точок  $x^2 + y^2 \leq 1$ . Наприклад,

$$f_{X,Y}(1, 1) = 0 \neq \frac{1}{\pi} \cdot 1 = g(1)h(1) . \quad \square$$

**Приклад 9.4.7.** Нехай  $T_1 \sim \text{Exp}(\lambda_1)$ ,  $T_2 \sim \text{Exp}(\lambda_2)$ ,  $T_1 \perp\!\!\!\perp T_2$ . Наприклад,  $T_1$  і  $T_2$  можуть відповідати часу очікування настання різних подій (очікування у двох різних чергах, тривалість безперебійної роботи двох компонент тощо). Щікавим є ймовірність того, що подія першого типу настане швидше від події другого,  $\mathbb{P}_{T_1, T_2}(T_1 < T_2)$ . Для обчислення такої ймовірності можна використати (9.2.3), помітивши, що через незалежність двох величин спільна щільність є добутком

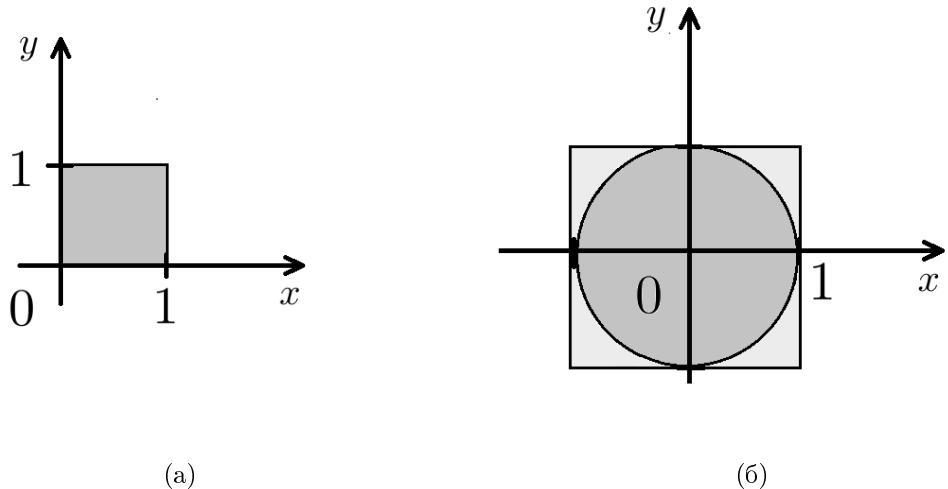


Рис. 9.4.1.: Спільні щільності рівномірних розподілів: (а) на одиничному квадраті; (б) на одиничному кружі

щільностей маржинальних:

$$\begin{aligned}
 \mathbb{P}_{T_1, T_2} (T_1 < T_2) &= \iint_{t_1 < t_2} f_{T_1, T_2}(t_1, t_2) dt_1 dt_2 \\
 &= \iint_{t_1 < t_2} f_{T_1}(t_1) f_{T_2}(t_2) dt_1 dt_2 \\
 &= \int_0^\infty \int_0^{t_2} \lambda_1 e^{-\lambda_1 t_1} \lambda_2 e^{-\lambda_2 t_2} dt_1 dt_2 \\
 &= \int_0^\infty \left( \int_0^{t_2} \lambda_1 e^{-\lambda_1 t_1} dt_1 \right) \lambda_2 e^{-\lambda_2 t_2} dt_2 \\
 &= \int_0^\infty \left( 1 - e^{-\lambda_1 t_2} \right) \lambda_2 e^{-\lambda_2 t_2} dt_2 \\
 &= \int_0^\infty \lambda_2 e^{-\lambda_2 t_2} dt_2 - \int_0^\infty \lambda_2 e^{-(\lambda_1 + \lambda_2)t_2} dt_2 \\
 &= 1 - \frac{\lambda_2}{\lambda_1 + \lambda_2} = \frac{\lambda_1}{\lambda_1 + \lambda_2}.
 \end{aligned}$$

□

Також безпосереднім наслідком наведених вище міркувань є критерій незалежності дискретних випадкових величин:

$$\mathbb{P}_{\mathbf{X}} (X_1 = x_1, \dots, X_k = x_k) = \mathbb{P}_{X_1} (X_1 = x_1) \cdot \dots \cdot \mathbb{P}_{X_k} (X_k = x_k) . \quad (9.4.7)$$

**Приклад 9.4.8.** За таблицею спряженості Табл. 9.2.1 з Прикладу 9.2.10 можна бачити, що величини  $X$  та  $Y$  не є незалежними. Справді, для виконання незалежності повинно бути, наприклад,  $\mathbb{P}_{X, Y}(X = 0, Y = 0) = \mathbb{P}_X(X = 0) \mathbb{P}_Y(Y = 0)$ , хоча  $\mathbb{P}_{X, Y}(X = 0, Y = 0) = 0.72$ , але  $\mathbb{P}_X(X = 0) \cdot \mathbb{P}_Y(Y = 0) = 0.75 \cdot 0.92 = 0.69 \neq 0.72$ . □

#### 9.4.2. Функції та сподівання від незалежних випадкових величин

Якщо випадкові величини незалежні, то незалежні також є функції від цих величин.

**Теорема 9.4.9.** Якщо величини  $X_{i,j}$  незалежні,  $i = 1, \dots, n$ ,  $j = 1, \dots, m(i)$ , а функції  $f_i : \mathbb{R}^{m(i)} \rightarrow \mathbb{R}$  вимірні, то незалежні є випадкові величини  $f_i(X_{i,1}, \dots, X_{i,m(i)})$ .

Доведення цієї теореми можна знайти в Розд. 9.6.

**Приклад 9.4.10.** Якщо  $X_1, \dots, X_k$  незалежні, то незалежними є, наприклад,  $X = X_1$  і  $Y = X_2 \cdot \dots \cdot X_k$ .  $\square$

Для незалежних випадкових величин також суттєво спрощується підрахунок сподівань.

**Теорема 9.4.11.** Нехай  $X$  і  $Y$  — дві незалежні випадкові величини з розподілами  $\mathbb{P}_X$  і  $\mathbb{P}_Y$  відповідно. Якщо  $h : \mathbb{R}^2 \rightarrow \mathbb{R}$  — деяка вимірна функція така, що вона невід'ємна ( $h \geq 0$ ) чи інтегровна ( $\mathbb{E}[|h(X, Y)|] < \infty$ ), то

$$\mathbb{E}[h(X, Y)] = \int_{\mathbb{R}} \int_{\mathbb{R}} h(x, y) d\mathbb{P}_X d\mathbb{P}_Y. \quad (9.4.8)$$

Зокрема, якщо  $h(x, y) = f(x)g(y)$ , де  $f, g : \mathbb{R} \rightarrow \mathbb{R}$  вимірні функції такі, що вони невід'ємні  $f, g \geq 0$  чи інтегровні  $\mathbb{E}[|f(X)|] < \infty$ ,  $\mathbb{E}[|g(Y)|] < \infty$ , то

$$\mathbb{E}[f(X)g(Y)] = \mathbb{E}[f(X)] \mathbb{E}[g(Y)]. \quad (9.4.9)$$

*Доведення.* З урахуванням Теореми 9.4.4 вираз (9.4.8) випливає з Теорем 6.4.2 та 9.1.8:

$$\mathbb{E}[h(X, Y)] = \iint_{\mathbb{R}^2} h d(\mathbb{P}_X \times \mathbb{P}_Y) = \int_{\mathbb{R}} \int_{\mathbb{R}} h(x, y) d\mathbb{P}_X d\mathbb{P}_Y.$$

Для доведення (9.4.9) спочатку розгляньмо випадок невід'ємних функцій. Тоді можна застосувати (9.4.8) і Теорему 9.1.8:

$$\begin{aligned} \mathbb{E}[f(X)g(Y)] &= \int_{\mathbb{R}} \int_{\mathbb{R}} f(x)g(y) d\mathbb{P}_X d\mathbb{P}_Y \\ &= \int_{\mathbb{R}} g(y) \left( \int_{\mathbb{R}} f(x) d\mathbb{P}_X \right) d\mathbb{P}_Y \\ &= \int_{\mathbb{R}} \mathbb{E}[f(X)] g(y) d\mathbb{P}_Y = \mathbb{E}[f(X)] \mathbb{E}[g(Y)]. \end{aligned}$$

Маючи цей результат, можна відразу зробити висновок, що

$$\mathbb{E}[|f(X)g(Y)|] = \mathbb{E}[|f(X)|] \mathbb{E}[|g(Y)|] < \infty,$$

оскільки  $|f| \geq 0$ ,  $|g| \geq 0$ , а скінченність сподівань від цих функцій надано за умовою. Відтак випадкова величина  $f \cdot g$  інтегровна, і можна знову застосувати Теорему 9.1.8 і показати (9.4.9).  $\square$

Із цієї доведеної теореми безпосередньо випливає такий результат.

**Теорема 9.4.12.** Нехай  $X_1, \dots, X_k$  — незалежні випадкові величини такі, що вони невід'ємні ( $X_i \geq 0$  майже напевно,  $i = 1, \dots, k$ ) чи інтегровні ( $\mathbb{E}[|X_i|] < \infty$ ,  $i = 1, \dots, k$ ). Тоді

$$\mathbb{E} \left[ \prod_{i=1}^k X_i \right] = \prod_{i=1}^k \mathbb{E}[X_i] , \quad (9.4.10)$$

у тому розумінні, що сподівання зліва існує і дорівнює добутку сподівань справа.

**Доведення.** Покладімо  $X = X_1, Y = X_2 \cdot \dots \cdot X_k$ . За Теоремою 9.4.9  $X \perp\!\!\!\perp Y$ . Тоді можна застосувати Теорему 9.4.11, поклавши  $f(x) = |x|$ ,  $g(y) = |y|$ :

$$\mathbb{E}[|X_1 \cdot \dots \cdot X_k|] = \mathbb{E}[|X_1|] \mathbb{E}[|X_2 \cdot \dots \cdot X_k|] .$$

За індукцією випливає такий результат, що для  $1 \leq m \leq n$

$$\mathbb{E}[|X_m \cdot \dots \cdot X_k|] = \prod_{i=m}^k \mathbb{E}[|X_i|] .$$

Якщо  $X_i \geq 0$  майже напевно, то  $|X_i| = X_i$  (майже напевно) і ми відразу маємо шуканий результат, поклавши  $m = 1$ .

У загальному випадку для інтегровних випадкових величин можна покласти  $m = 2$  і зробити висновок, що

$$\mathbb{E}[|Y|] = \mathbb{E}[|X_2 \cdot \dots \cdot X_k|] = \prod_{i=2}^k \mathbb{E}[|X_i|] < \infty ,$$

оскільки кожне окрім сподівання в добутку скінченне. Відтак випадкова величина  $Y$  інтегровна, і тоді застосування Теореми 9.4.11 з  $f(x) = x$  і  $g(y) = y$  дає

$$\mathbb{E}[X_1 \cdot \dots \cdot X_k] = \mathbb{E}[X_1] \mathbb{E}[X_2 \cdot \dots \cdot X_k] ,$$

і шуканий результат випливає за індукцією.  $\square$

Із того факту, що для незалежних випадкових величин сподівання добутку є добутком сподівань негайно можна встановити таку властивість.

**Твердження 9.4.13.** Якщо випадкові величини  $X$  і  $Y$  незалежні, то  $\text{Cov}(X, Y) = 0$ .

**Доведення.**

$$\text{Cov}(X, Y) = \mathbb{E}[XY] - \mathbb{E}[X]\mathbb{E}[Y] = \mathbb{E}[X]\mathbb{E}[Y] - \mathbb{E}[X]\mathbb{E}[Y] = 0 .$$

$\square$

Такі випадкові величини називають *некорелюваними* (uncorrelated). При цьому дуже важливо пам'ятати, що *зворотне твердження в загальному випадку нesправедливе*.

**Приклад 9.4.14.** Нехай  $\mathbb{P}_X(X = 0) = \mathbb{P}_X(X = 1) = \mathbb{P}_X(X = -1) = \frac{1}{3}$ , а  $Y = \mathbb{1}\{X = 0\}$ . Тоді

$$\mathbb{E}[X] = 0 \cdot \frac{1}{3} + 1 \cdot \frac{1}{3} - 1 \cdot \frac{1}{3} = 0 .$$

З іншого боку,

$$\begin{aligned}\mathbb{E}[XY] &= (0 \cdot 0) \cdot \mathbb{P}_{X,Y}(X = 0, Y = 0) + (1 \cdot 0) \cdot \mathbb{P}_{X,Y}(X = 1, Y = 0) + (-1 \cdot 0) \cdot \mathbb{P}_{X,Y}(X = -1, Y = 0) \\ &\quad + (0 \cdot 1) \cdot \mathbb{P}_{X,Y}(X = 0, Y = 1) + (1 \cdot 1) \cdot \mathbb{P}_{X,Y}(X = 1, Y = 1) + (-1 \cdot 1) \cdot \mathbb{P}_{X,Y}(X = -1, Y = 1) \\ &= \mathbb{P}_{X,Y}(X = 1, Y = 1) - \mathbb{P}_{X,Y}(X = -1, Y = 1) = 0,\end{aligned}$$

оскільки відповідні події неймовірні.

Отже  $\text{Cov}(X, Y) = 0$ , хоча ці величини явно не є незалежними.  $\square$

**Приклад 9.4.15.** Нехай  $X$  та  $Y$  — дві випадкові величини, що не є незалежними, зі скінченними сподіваннями та дисперсіями, до того ж  $\text{Var}(X) = \text{Var}(Y)$ . Нехай  $U = X + Y$ ,  $V = X - Y$ . Тоді

$$\begin{aligned}\mathbb{E}[UV] &= \mathbb{E}[(X + Y)(X - Y)] = \mathbb{E}[X^2] - \mathbb{E}[Y^2], \\ \mathbb{E}[U]\mathbb{E}[V] &= \mathbb{E}[X + Y]\mathbb{E}[X - Y] = (\mathbb{E}[X])^2 - (\mathbb{E}[Y])^2, \\ \text{Cov}(U, V) &= \mathbb{E}[X^2] - \mathbb{E}[Y^2] - ((\mathbb{E}[X])^2 - (\mathbb{E}[Y])^2) = \text{Var}(X) - \text{Var}(Y) = 0,\end{aligned}$$

тобто змінні  $U$  і  $V$  некорельовані, хоча вони, очевидно, є залежні.  $\square$

З урахуванням Твердження 9.4.13 можна спростити підрахунок дисперсії суми випадкових величин за формулою (9.3.4) у випадку незалежних величин:

$$\text{Var}(X_1 + \dots + X_k) = \sum_{i=1}^k \text{Var}(X_i). \quad (9.4.11)$$

**Приклад 9.4.16.** Нехай  $X_1, \dots, X_n$  — деякі незалежні випадкові величини з однаковим розподілом, який невідомий досліднику. Нехай  $\mathbb{E}[X_1] = \dots = \mathbb{E}[X_n] = \mu_X$ ,  $\text{Var}(X_1) = \dots = \text{Var}(X_n) = \sigma_X^2$ , проте справжні значення цих параметрів також невідомі досліднику. У статистиці ми маємо справу з певним набором даних, або ж вибіркою  $(x_1, \dots, x_n)$ , яку можна вважати набором певних реалізацій відповідних випадкових величин — конкретних значень, яких вони набули по завершенню деякого випадкового експерименту, наслідки якого ми аналізуємо. Задачею статистиків є визначення параметрів розподілів, згідно з якими розподілені дані відповідної вибірки.

Чи не найпростішою числововою характеристикою вибірки є її *середнє вибіркове значення* (sample mean)  $\bar{X}$ . Оскільки всі величини незалежні, можемо підрахувати сподівання та дисперсію середнього вибіркового значення:

$$\begin{aligned}\mathbb{E}[\bar{X}] &= \mathbb{E}\left[\frac{1}{n} \sum_{i=1}^n X_i\right] = \frac{1}{n} \sum_{i=1}^n \mathbb{E}[X_i] = \frac{1}{n} \sum_{i=1}^n \mu_X = \mu_X, \\ \text{Var}(\bar{X}) &= \text{Var}\left(\frac{1}{n} \sum_{i=1}^n X_i\right) = \frac{1}{n^2} \sum_{i=1}^n \text{Var}(X_i) = \frac{1}{n^2} \sum_{i=1}^n \sigma_X^2 = \frac{\sigma_X^2}{n}.\end{aligned}$$

$\square$

## 9.5. Суми незалежних випадкових величин

### 9.5.1. Згортка щільностей розподілів

На практиці дуже часто виникає потреба знайти розподіл суми випадкових величин. Якщо ці випадкові величини незалежні, існує простий спосіб обчислення відповідних розподілів.

**Твердження 9.5.1.** Нехай  $X$  та  $Y$  — дві незалежні випадкові величини зі щільностями розподілу  $f_X$  і  $f_Y$  відповідно. Тоді щільність суми  $S = X + Y$  дорівнює

$$f_S(s) = \int_{-\infty}^{\infty} f_X(s-y) f_Y(y) dy. \quad (9.5.1)$$

*Доведення.* Функція розподілу суми має вираз

$$\begin{aligned} F_S(s) &= \mathbb{P}_S(S \leq s) = \iint_{x+y \leq s} f_{X,Y}(x,y) dx dy \\ &= \iint_{x+y \leq s} f_X(x) f_Y(y) dx dy \\ &= \int_{-\infty}^{\infty} \left( \int_{-\infty}^{s-y} f_X(x) dx \right) f_Y(y) dy \\ &= \int_{-\infty}^{\infty} F_X(s-y) f_Y(y) dy. \end{aligned}$$

Диференціюючи цей вираз, дістаємо:

$$\begin{aligned} f_S(s) &= \frac{d}{ds} F_S(s) = \frac{d}{ds} \left( \int_{-\infty}^{\infty} F_X(s-y) f_Y(y) dy \right) \\ &= \int_{-\infty}^{\infty} \frac{d}{ds} F_X(s-y) f_Y(y) dy \\ &= \int_{-\infty}^{\infty} f_X(s-y) f_Y(y) dy, \end{aligned}$$

де ми застосували формулу Ляйбніца для інтегрування (Leibniz integral rule):

$$\frac{d}{dx} \left( \int_{a(x)}^{b(x)} f(x,t) dt \right) = \int_{a(x)}^{b(x)} \frac{\partial}{\partial x} f(x,t) dt + f(x, b(x)) \frac{db}{dx} - f(x, a(x)) \frac{da}{dx}.$$

□

Інтеграл у (9.5.1) має назву *згортки* (convolution).

У застосуванні до дискретних величин формула (9.5.1) набуває виду

$$\mathbb{P}_S(S = s) = \sum_{j=-\infty}^{\infty} \mathbb{P}_X(X = s-j) \mathbb{P}_Y(Y = j). \quad (9.5.2)$$

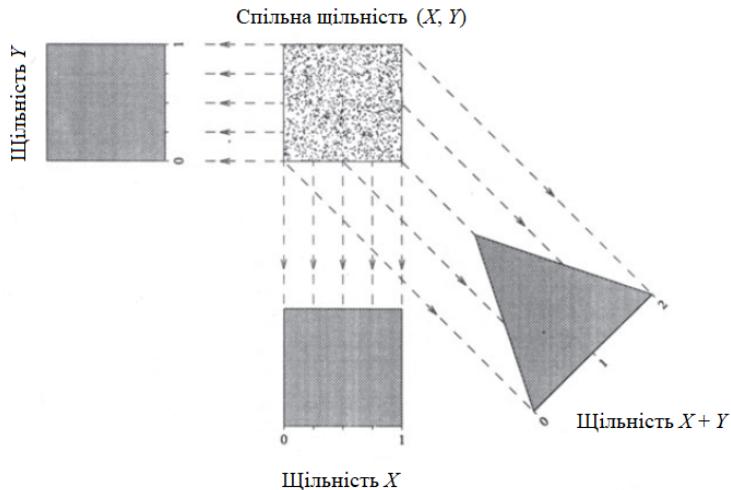


Рис. 9.5.1.: Набір випадкових точок, згенерованих відповідно до спільної щільності розподілу  $X, Y \sim U((0; 1])$ ,  $X \perp\!\!\!\perp Y$ , а також маржинальні щільності та щільність суми  $X + Y$  [9, с. 380]

**Приклад 9.5.2.** Розгляньмо обчислення щільності розподілу суми  $S = X + Y$  двох рівномірно розподілених випадкових величин  $X, Y \sim U((\alpha; \beta])$ ,  $X \perp\!\!\!\perp Y$ . Згідно з (9.5.1),

$$\begin{aligned}
 f_S(s) &= \int_{-\infty}^{\infty} f_X(s-y) f_Y(y) dy = \int_{-\infty}^{\infty} \frac{1}{\beta - \alpha} \mathbb{1}\{s-y \in (\alpha; \beta]\} \cdot \frac{1}{\beta - \alpha} \mathbb{1}\{y \in (\alpha; \beta]\} dy \\
 &= \frac{1}{(\beta - \alpha)^2} \int_{-\infty}^{\infty} \mathbb{1}\{s - \beta \leq y < s - \alpha\} \cdot \mathbb{1}\{\alpha < y \leq \beta\} dy \\
 &= \frac{1}{(\beta - \alpha)^2} \int_{\max\{s-\beta, \alpha\}}^{\min\{s-\alpha, \beta\}} dy \\
 &= \begin{cases} \frac{s - 2\alpha}{(\beta - \alpha)^2}, & s \in (2\alpha; \alpha + \beta] \\ \frac{2\beta - s}{(\beta - \alpha)^2}, & s \in (\alpha + \beta; 2\beta] \\ 0, & s \notin (2\alpha; 2\beta] \end{cases}.
 \end{aligned}$$

Процес підрахунку щільностей сум незалежних рівномірно розподілених випадкових величин можна продовжити. Для простоти розгляду зосередьмося на стандартних рівномірних величинах. У цьому випадку щільність суми  $S_2 = X + Y$  дорівнює

$$f_{S_2}(s) = \begin{cases} s, & s \in (0; 1] \\ 2 - s, & s \in (1; 2] \\ 0, & s \notin (0; 2] \end{cases}.$$

Це є щільністю так званого *трикутного* (triangular) розподілу (Рис. 9.5.1).

Для суми трьох стандартних рівномірних величин  $S_3 = X + Y + W = S_2 + W$  щільність матиме

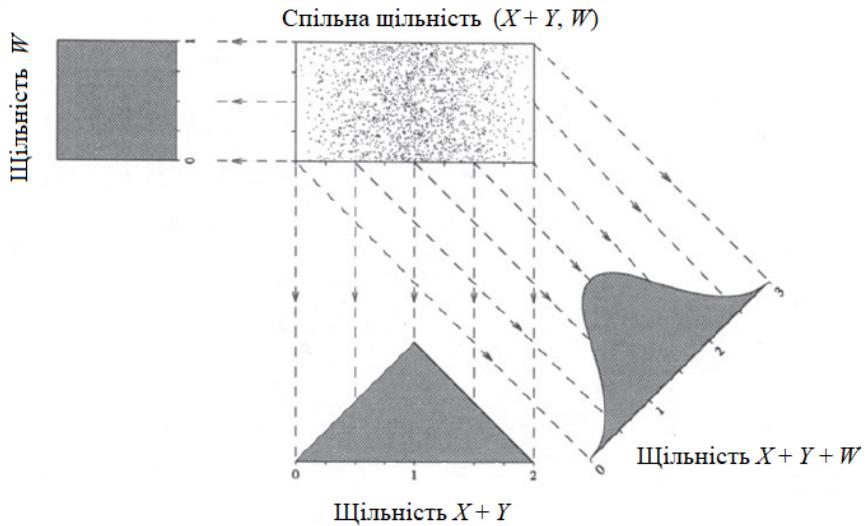


Рис. 9.5.2.: Набір випадкових точок, згенерованих відповідно до спільної щільності розподілу  $S_2 = X + Y$  та  $W$ , де величини  $X, Y, W \sim U((0; 1])$  незалежні, а також маржинальні щільності та щільність суми  $X + Y + W$  [9, с. 380]

вираз

$$\begin{aligned}
 f_{S_3}(s) &= \int_{-\infty}^{\infty} f_W(s-y) f_{S_2}(y) dy \\
 &= \int_{-\infty}^{\infty} \mathbb{1}\{0 < s-y \leq 1\} f_{S_2}(y) dy \\
 &= \int_{s-1}^s f_{S_2}(y) dy \\
 &= \mathbb{P}_{S_2}(s-1 < S_2 < s) .
 \end{aligned}$$

Наприклад, якщо  $0 < s < 1$ , то  $s-1 < 0$ , і матимемо

$$\mathbb{P}_{S_2}(s-1 < S_2 < s) = \int_0^s f_{S_2}(s) ds = \int_0^s s ds = \frac{s^2}{2} .$$

Аналогічно можна порахувати вирази для  $s$ , що лежать в інших проміжках:

$$f_{S_3}(s) = \begin{cases} \frac{s^2}{2} , & s \in (0; 1] \\ -s^2 + 3s - \frac{3}{2} , & s \in (1; 2] \\ \frac{1}{2}(3-s)^2 , & s \in (2; 3] \\ 0 , & s \notin (0; 3] \end{cases} .$$

Відповідний результат зображеного на Рис. 9.5.2.

Аналітичний вираз суми  $n$  випадкових рівномірних величин дуже складний, але графіки щільностей, наведені на Рис. 9.5.3, свідчать, що щільність відповідної суми дуже швидко збігається

до щільності нормального розподілу. Це є ілюстрація фундаментального результата теорії ймовірностей, який ми вивчатимемо далі в цьому курсу — *центральної граничної теореми*.

□

**Приклад 9.5.3.** Нехай три числа рахують з похибкою округлення, що менша від  $10^{-6}$  за абсолютною значенням. Покладімо, що похибки округлення незалежні і мають рівномірний розподіл. Підрахуймо ймовірність, що сума округлених чисел відрізняється від справжньої суми більше, ніж на  $2 \cdot 10^{-6}$ .

Позначмо  $X_i$  = «похибка в  $i$ -ому числі в одиницях  $10^{-6}$ ». Тоді  $X_i \sim U((-1; 1])$ ,  $U_i = (X_i + 1)/2 \sim U((0; 1])$ ,  $S_3 = U_1 + U_2 + U_3$ , і

$$\begin{aligned} \mathbb{P}_{X_1, X_2, X_3} (|X_1 + X_2 + X_3| > 2) &= 2\mathbb{P}_{U_1, U_2, U_3} (2U_1 - 1 + 2U_2 - 1 + 2U_3 - 1 > 2) \\ &= 2\mathbb{P}_{S_3} \left( S_3 > \frac{5}{2} \right) \\ &= 2 \int_{\frac{5}{2}}^{\infty} f_{S_3}(s) ds \\ &= 2 \int_{\frac{5}{2}}^2 \frac{1}{2} (3 - s)^2 ds \\ &= \frac{1}{24}. \end{aligned}$$

□

### 9.5.2. Застосування згортки до відомих розподілів

Застосуймо (9.5.1) до обчислення щільностей сум інших розподілів, які ми вивчали раніше в цьому курсі.

**Твердження 9.5.4.** Розподіл суми  $S = \sum_{i=1}^k X_i$  незалежних випадкових величин  $X_i, i = 1, \dots, k$  такий:

- (i) якщо  $X_i \sim \text{Bern}(p)$ ,  $i = 1, \dots, k$ , то  $S \sim \text{Binom}(k, p)$ ;
- (ii) якщо  $X_i \sim \text{Binom}(n_i, p)$ ,  $i = 1, \dots, k$ , то  $S \sim \text{Binom}\left(\sum_{i=1}^k n_i, p\right)$ ;
- (iii) якщо  $X_i \sim \text{Pois}(\lambda_i)$ ,  $i = 1, \dots, k$ , то  $S \sim \text{Pois}\left(\sum_{i=1}^k \lambda_i\right)$ ;
- (iv) якщо  $X_i \sim N(\mu_i, \sigma_i^2)$ ,  $i = 1, \dots, k$ , то  $S \sim N\left(\sum_{i=1}^k \mu_i, \sum_{i=1}^k \sigma_i^2\right)$ .

Доведення цього твердження можна знайти в Розд. 9.6.5.

**Зауваження 9.5.5.** Те, що сума величин Бернуллі є біномною величиною, цілком очевидно за визначеннями біномної величини як загального числа успіхів у деякій серії випробувань з однаковою ймовірністю успіху  $p$ .

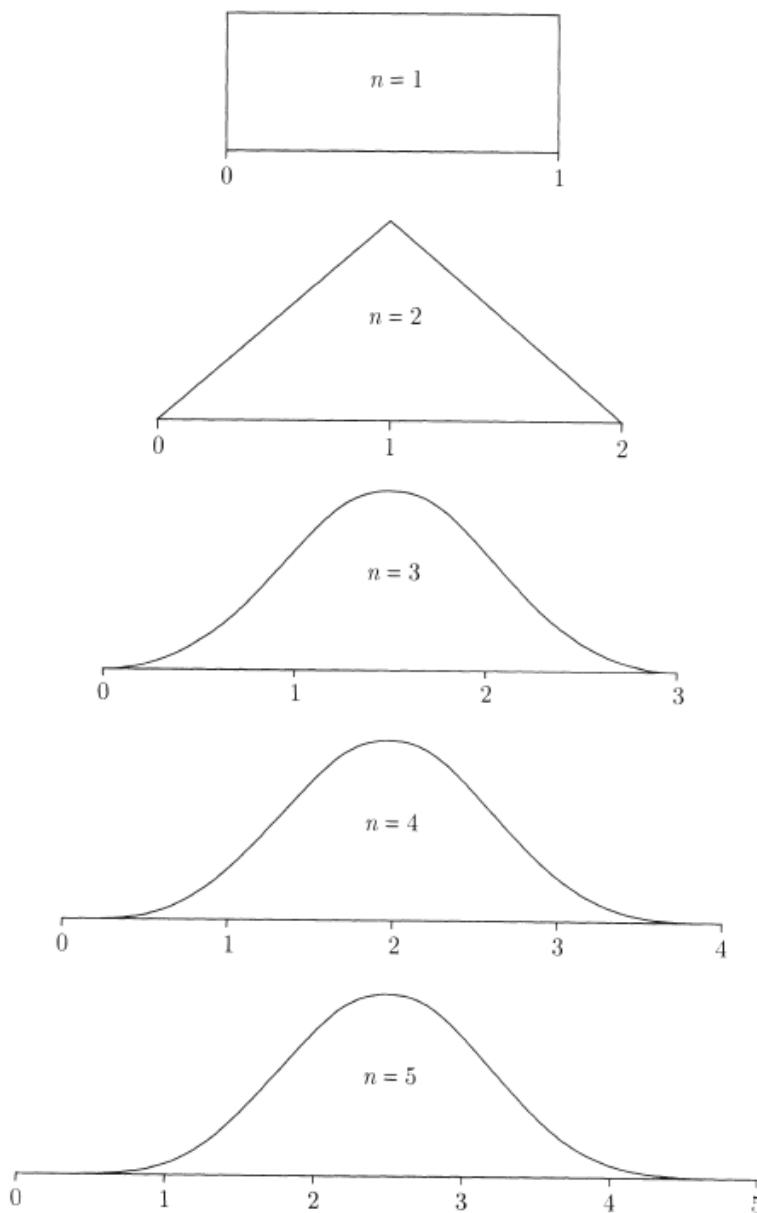


Рис. 9.5.3.: Графіки щільностей сум  $n$  незалежних величин зі стандартним рівномірним розподiлом [9, с. 381]

Те, що сума біномних величин є біномною величиною, цілком очевидно, адже кожна з них відповідає числу успіхів у деякій серії випробувань з однаковою ймовірністю успіху  $p$ . Відповідно, їх сума — це просто сума числа успіхів, тобто сама по собі є числом успіхів, а отже біномною величиною.  $\square$

**Приклад 9.5.6.** У двох партіях продукції обсягом  $n_1 = 10$  і  $n_2 = 15$  товари можуть, незалежно один від одного, бути браковані з імовірністю 6.25%. Підрахуймо ймовірність того, що загальне число бракованих товарів більше 5.

Нехай  $X_1$  та  $X_2$  — кількості бракованих товарів у кожній партії,  $X_1 \sim \text{Binom}(10, 0.0625)$ ,  $X_2 \sim \text{Binom}(15, 0.0625)$ , а тому відповідно до Твердження 9.5.4,  $X_1 + X_2 \sim \text{Binom}(25, 0.0625)$ , а отже

$$\begin{aligned}\mathbb{P}_{X_1+X_2}(X_1 + X_2 > 5) &= 1 - \mathbb{P}_{X_1+X_2}(X_1 + X_2 \leq 5) \\ &= 1 - \sum_{i=0}^5 \mathbb{P}_{X_1+X_2}(X_1 + X_2 = i) \\ &= 1 - \sum_{i=0}^5 \binom{25}{i} (0.0625)^i (0.9375)^{25-i} \\ &= 0.0038.\end{aligned}$$

 $\square$ 

**Приклад 9.5.7.** Нехай 5 радіоактивних джерел незалежно від одного випромінюють  $X_i$ ,  $i = 1, \dots, 5$ , частинок відповідно зі швидкістю 0.08 частинок на певний проміжок часу. Підрахуймо ймовірність того, що загальне число частинок за певний проміжок часу не перевищить 3.

За умовами,  $X_i \sim \text{Pois}(0.08)$ , а тому відповідно до Твердження 9.5.4,  $S \equiv \sum_{i=1}^5 X_i \sim \text{Pois}(0.4)$ , а отже

$$\mathbb{P}_S(S \leq 3) = \sum_{j=0}^3 \frac{e^{-0.4}(0.4)^j}{j!} \approx 0.9992.$$

 $\square$ 

## 9.6. Доведення окремих тверджень

### 9.6.1. Теорема про добуток мір

Доведімо Теорему 9.1.6.

**Визначення 9.6.1.** Для деякої множини  $E \in \mathcal{X} \times \mathcal{Y}$  та точок  $x \in X$  і  $y \in Y$  можна визначити *перетини множини  $E$*  (sections of  $E$ )  $E_x = \{y : (x, y) \in E\}$  та  $E_y = \{x : (x, y) \in E\}$ .  $\square$

Ілюстрацію цих понять подано на Рис. 9.6.1, де замість  $x \in X$  та  $y \in Y$  використано позначення  $\omega_1 \in \Omega_1$  та  $\omega_2 \in \Omega_2$ , відповідно.

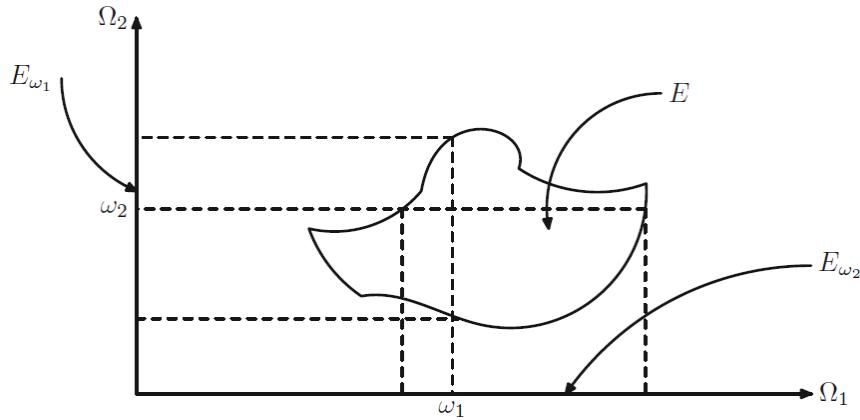


Рис. 9.6.1.: Ілюстрація до поняття перетинів ([4, с. 81])

**Зауваження 9.6.2.** Зокрема, нескладно бачити, що

$$\begin{aligned}
 (X \times Y)_x &= \{y : (x, y) \in X \times Y\} = Y, \\
 (X \times Y)_y &= \{x : (x, y) \in X \times Y\} = X, \\
 (A \times B)_x &= \{y : (x, y) \in A \times B\} = \begin{cases} B, & x \in A \\ \emptyset, & x \notin A \end{cases}, \\
 (A \times B)_y &= \{x : (x, y) \in A \times B\} = \begin{cases} A, & y \in B \\ \emptyset, & y \notin B \end{cases}.
 \end{aligned}$$

□

**Лема 9.6.3.** Якщо  $E \in \mathcal{X} \times \mathcal{Y}$ , то для кожного  $x \in X$  перетин  $E_x \in \mathcal{Y}$ , а для кожного  $y \in Y$  перетин  $E_y \in \mathcal{X}$ .

*Доведення.* Для кожного  $x \in X$  можна визначити відображення  $T_x : Y \rightarrow X \times Y$  таке, що  $T_x(y) = (x, y)$ . Якщо  $E = A \times B$  — вимірний прямокутник, то

$$T_x^{-1}(E) = \begin{cases} B, & x \in A \\ \emptyset, & x \notin A \end{cases}.$$

В обох випадках  $T_x^{-1}(E) \in \mathcal{Y}$ .

За Теоремою 7.3.1,  $T_x$  є вимірною функцією, а відтак  $E_x = T_x^{-1}(E) \in \mathcal{Y}$  для будь-якої множини  $E \in \mathcal{X} \times \mathcal{Y}$ .

Аналогічно доводять випадок для  $E_y$ .

□

Аналогічно перетинам множини можна увести поняття перетинів функцій.

**Визначення 9.6.4.** Для деякого  $x \in X$  функцію  $f_x : E_x \rightarrow \mathbb{R}$  таку, що  $f_x(y) = f(x, y)$ ,  $(x, y) \in E$  називають *перетином функції*  $f$  у точці  $x$  (section of  $f$  at  $x$ ). Аналогічно уводять поняття перетину функції  $f$  у точці  $y \in Y$ .

□

**Лема 9.6.5.** Якщо функція  $f : (X \times Y, \mathcal{X} \times \mathcal{Y}) \rightarrow (\mathbb{R}, \mathcal{B})$  вимірна відносно  $\mathcal{X} \times \mathcal{Y}$ , то для кожного  $x \in X$  функція  $f_x(y) = f(x, y)$  є вимірною відносно  $\mathcal{Y}$ , а для кожного  $y \in Y$  функція  $f_y(x) = f(y, x)$  є вимірною відносно  $\mathcal{X}$ .

*Доведення.* Нехай  $B \in \mathcal{B}$ . Тоді

$$f_x^{-1}(B) = \{y \in Y : f_x(y) \in B\} = \{y \in Y : f(x, y) \in B\} = \{y \in Y : (x, y) \in f^{-1}(B)\} = (f^{-1}(B))_x ,$$

тобто  $f_x^{-1}(B)$  є перетином множини  $f^{-1}(B)$  у точці  $x$ . Але оскільки  $f$  є вимірною, то  $f^{-1}(B) \in \mathcal{X} \times \mathcal{Y}$ , тобто  $f_x^{-1}(B) \in \mathcal{Y}$  за Лемою 9.6.3. Із цього випливає, що  $f_x$  є вимірною відносно  $\mathcal{Y}$ .

Аналогічно доводять випадок для  $f_y$ .  $\square$

Розгляньмо два мірні простори  $(X, \mathcal{X}, \mu)$  і  $(Y, \mathcal{Y}, \nu)$ , де обидві міри  $\sigma$ -скінченні. Для деякої множини  $E \in \mathcal{X} \times \mathcal{Y}$  можна визначити дві функції:

$$f(x) = \nu(E_x) , \quad g(y) = \mu(E_y) . \quad (9.6.1)$$

Ці функції коректні, оскільки множини  $E_x$  і  $E_y$  вимірні за Лемою 9.6.3, тому до них можна застосувати відповідні міри.

Згідно з Лемою 9.6.5, функція  $f(x) = \nu(\{y : (x, y) \in E\})$  є коректно визначеною вимірною функцією від  $x$ . Без доведення використаймо таку лему.

**Лема 9.6.6.** Для будь-якого  $E \in \mathcal{X} \times \mathcal{Y}$  функції (9.6.1) невід'ємні, вимірні, і

$$\int f d\mu = \int g d\nu .$$

Використовуючи уведені поняття й леми, можемо довести Теорему 9.1.6.

*Доведення.* Спочатку покажімо, що міра  $\pi$  справді є мірою в розумінні Визначення 2.1.2. По-перше, із Леми 9.6.6 випливає, що функції  $f(x) = \nu(E_x)$  і  $g(y) = \mu(E_y)$  є невід'ємні і вимірні, а також що  $\int f d\mu = \int g d\nu$ , тобто міру  $\pi$  визначено коректно. Цілком очевидно, що  $\pi$  є невід'ємною, оскільки міри  $\mu$  і  $\nu$  самі по собі невід'ємні. Потрібно показати  $\sigma$ -адитивність  $\pi$ .

Нехай  $E_n \in \mathcal{X} \times \mathcal{Y}$ ,  $n = 1, 2, \dots$ ,  $E_i \cap E_j = \emptyset$ ,  $i \neq j$ ,  $E = \bigcup_n E_n$ , і  $f_n(x) = \nu(E_{nx})$ . Оскільки, за визначенням,

$$E_x = \left( \bigcup_n E_n \right)_x = \bigcup_n E_{nx} ,$$

ми маємо, за  $\sigma$ -адитивністю міри:

$$f(x) = \nu(E_x) = \nu \left( \bigcup_n E_{nx} \right) = \sum_n \nu(E_{nx}) = \sum_n f_n(x) , \quad x \in X ,$$

де  $f_n(x) = \nu(E_{nx})$ .

Кожна  $f_n$  є невід'ємною і вимірною за Лемою 9.6.6. Оскільки  $\sum_{k=1}^n f_k$  є вимірною як сума вимірних функцій, а  $\lim_{n \rightarrow \infty} \sum_{k=1}^n f_k(x) = f(x)$ , то  $f$  також є невід'ємною і вимірною за Твердженням 7.3.12. Тоді за Теоремою 6.2.6,

$$\int f d\mu = \int \sum_n f_n d\mu = \sum_n \int f_n d\mu ,$$

тобто  $\pi(E) = \sum_n \pi(E_n)$ .

Покажімо тепер, що для деякої  $E = A \times B$ ,  $A \in \mathcal{X}$ ,  $B \in \mathcal{Y}$ ,  $\pi(E) = \mu(A)\nu(B)$ . Справді, використовуючи Завдання 9.6.2, маємо:

$$\begin{aligned} f(x) &= \nu(E_x) = \nu(B) \mathbb{1}\{x \in A\} , \\ g(y) &= \mu(E_y) = \mu(A) \mathbb{1}\{y \in B\} . \end{aligned}$$

Таким чином,  $f$  і  $g$  є невід'ємні і вимірні (оскільки вони є добутками деякої константи та індикаторної функції). Тоді

$$\pi(E) = \int f \, d\mu = \nu(B) \cdot \int \mathbb{1}\{x \in A\} \, d\mu = \mu(A)\nu(B) .$$

Покажімо тепер, що  $\pi$  є  $\sigma$ -скінченною мірою. Нехай  $\{A_i\}$ ,  $i = 1, 2, \dots$ , та  $\{B_j\}$ ,  $j = 1, 2, \dots$  — розбиття просторів  $X$  та  $Y$ , відповідно, такі, що  $\mu(A_i) < \infty$ ,  $\nu(B_j) < \infty \forall i, j$ . Такі розбиття існують за  $\sigma$ -скінченістю мір  $\mu$  і  $\nu$ . Тоді  $\{A_i \times B_j\}$ ,  $i, j = 1, 2, \dots$  — розбиття добутку просторів  $X \times Y$ , а  $\pi(A_i \times B_j) = \mu(A_i)\nu(B_j) < \infty \forall i, j$ . Це і є визначення  $\sigma$ -скінченності.

Нарешті, залишилося показати, що міра  $\pi$  — єдина міра така, що  $\pi(A \times B) = \mu(A)\nu(B)$  для деякого вимірного прямокутника. Розгляньмо деяку іншу функцію  $\tau$  таку, що для всіх вимірних прямокутників  $A \times B$ ,  $A \in \mathcal{X}$ ,  $B \in \mathcal{Y}$ , справедливо  $\tau(A \times B) = \mu(A)\nu(B)$ . Якщо  $E = \bigcup_{i=1}^r A_i \times B_i$ ,  $A_i \in \mathcal{X}$ ,  $B_i \in \mathcal{Y}$ ,  $i = 1, \dots, r$ , то

$$\sum_{i=1}^r \tau(A_i \times B_i) = \sum_{i=1}^r \mu(A_i)\nu(B_i) = \sum_{i=1}^r \pi(A_i \times B_i) = \pi \left( \sum_{i=1}^r A_i \times B_i \right) = \pi(E) .$$

Якщо розглянути інше подання множини  $E$  як об'єднання  $E = \bigcup_{j=1}^s A'_j \times B'_j$ ,  $A'_j \in \mathcal{X}$ ,  $B'_j \in \mathcal{Y}$ ,  $j = 1, \dots, s$ , то

$$\sum_{j=1}^s \tau(A'_j \times B'_j) = \sum_{j=1}^s \mu(A'_j)\nu(B'_j) = \sum_{j=1}^s \pi(A'_j \times B'_j) = \pi \left( \sum_{j=1}^s A'_j \times B'_j \right) = \pi(E) ,$$

тобто функція  $\tau$  коректна. Мало того, це не просто функція, а міра, оскільки  $\pi$  — це міра, і не просто міра, а міра на алгебрі, утвореній зі скінченних об'єднань (і заперечень) вимірних прямокутників. А відтак за Теоремою 2.4.1  $\tau$  має єдине подовження на  $\sigma$ -алгебру, породжену вимірними прямокутниками, і таким подовженням не може бути ніщо інше, як  $\pi$ .  $\square$

### 9.6.2. Теорема Фубіні

Доведімо Теорему 9.1.8.

*Доведення.* Розгляньмо спочатку випадок невід'ємної випадкової величини.

За Лемою 9.6.5, функція  $f(x, y)$  для деякого фіксованого  $x \in X$  є функцією від  $y$ , вимірною відносно  $\mathcal{Y}$ . Потрібно показати, що інтеграл  $\int_Y f(x, y) \, d\nu$  існує, що він сам по собі є функцією від  $x$ , вимірною відносно  $\mathcal{X}$ , і, нарешті, що для нього виконується (9.1.3).

Доведення, як завжди, будуватимемо покроково.

**Індикаторні функції** Якщо  $f(x, y) = \mathbb{1}\{(x, y) \in E\}$  для деякої множини  $E \in X \times Y$ , то можна застосувати Теорему 9.1.6, оскільки в цьому випадку для фіксованого  $x$  маємо  $f(x, y) = \mathbb{1}\{y \in E_x\}$ ,

і тоді

$$\int_Y f(x, y) d\nu = \int_Y \mathbb{1}_{\{y \in E_x\}} d\nu = \nu(E_x) ,$$

а (9.1.3) перетворюється в  $\pi(E) = \int \nu(E_x) d\mu$ , що вже доведено.

**Прості функції** Нехай  $f$  — деяка невід'ємна проста функція:

$$f(x, y) = \sum_{i=1}^n a_i \mathbb{1}_{\{(x, y) \in E_i\}} .$$

Тоді інтеграл  $\int_Y f(x, y) d\nu$ , за властивістю лінійності, є сумаю інтегралів

$$\int_Y f(x, y) d\nu = \sum_{i=1}^n \int_Y a_i \mathbb{1}_{\{(x, y) \in E_i\}} d\nu = \sum_{i=1}^n \int_Y a_i \mathbb{1}_{\{y \in E_{ix}\}} d\nu = \sum_{i=1}^n a_i \nu(E_{ix}) ,$$

тобто є сумаю вимірних відносно  $\mathcal{X}$  функцій, а отже сам є вимірний відносно  $\mathcal{X}$ . Більше того, властивість лінійності дає змогу розписати (9.1.3) так:

$$\int_{X \times Y} f(x, y) d\pi = \sum_{i=1}^n a_i \pi(E_i) = \sum_{i=1}^n a_i \int_X \nu(E_{ix}) d\mu = \int_X \left( \int_Y f(x, y) d\nu \right) d\mu .$$

**Невід'ємні функції** Нарешті деяка невід'ємна функція  $f(x, y)$  є границею послідовності простих функцій, і просте застосування Теореми 6.2.6 відразу дає шуканий результат.

Аналогічне доведення справедливе і для інтегралу  $\int_X f(x, y) d\mu$ .

Розгляньмо тепер загальний випадок деякої інтегровної відносно  $\pi$  функції  $f(x, y)$ :

$$\int_{X \times Y} |f(x, y)| d\pi < \infty .$$

Оскільки функція  $|f(x, y)|$  невід'ємна, то для неї виконується (9.1.3), зокрема

$$\int_X \left( \int_Y |f(x, y)| d\nu \right) d\mu < \infty .$$

Розгляньмо множину  $A_0 \subseteq X$  таку, що для кожного  $x \in A_0$

$$\int_Y |f(x, y)| d\nu < \infty .$$

Тоді

$$\int_Y f(x, y) d\nu = \int_Y f^+(x, y) d\nu - \int_Y f^-(x, y) d\nu .$$

Оскільки  $f^+$  і  $f^-$  — невід'ємні функції, то за аналогією з доведенням першої частини теореми обидва інтеграли справа є вимірними відносно  $\mathcal{X}$  функціями, а оскільки  $f^+ \leq |f|$ ,  $f^- \leq |f|$ , то й інтегровними. Відповідно, і інтеграл зліва є вимірним та інтегровним.

Якщо це так, то  $\int_X (\int_Y f(x, y) d\nu) d\mu < \infty$ , а отже

$$\begin{aligned} \int_X \left( \int_Y f(x, y) d\nu \right) d\mu &= \int_X \left( \int_Y f^+(x, y) d\nu \right) d\mu - \int_X \left( \int_Y f^-(x, y) d\nu \right) d\mu \\ &= \int_{X \times Y} f^+(x, y) d\pi - \int_{X \times Y} f^-(x, y) d\pi = \int_{X \times Y} f(x, y) d\pi < \infty. \end{aligned}$$

При цьому  $\mu(X \setminus A_0) = 0$  за властивостями інтегралу, зокрема, якщо  $\int f d\mu < \infty$ , то  $f < \infty$  майже скрізь.

Аналогічне доведення справедливе і для інтегралу  $\int_X f(x, y) d\mu$  та деякої множини  $B_0 \subseteq Y$ , на якій

$$\int_X |f(x, y)| d\mu < \infty.$$

□

### 9.6.3. Властивості спільних функцій розподілу

Розглянемо на  $\sigma$ -алгебрі  $\mathcal{B}^k$  у  $k$ -вимірному просторі  $\mathbb{R}^k$  клас обмежених прямокутників виду

$$A = \{\mathbf{x} : a_i < x_i \leq b_i, i = 1, \dots, k\} = (a_1; b_1] \times \dots \times (a_k; b_k] \equiv I_1 \times \dots \times I_k.$$

Кожний такий прямокутник має  $2^k$  вершин — точок  $\mathbf{x} = (x_1, \dots, x_k)$ , де кожна координата  $i$  дорівнює або  $a_i$ , або  $b_i$ ,  $i = 1, \dots, k$ . Нехай  $\text{sign}_A \mathbf{x}$  — *знак вершини*, який дорівнює  $+1$ , якщо кількість координат  $x_i$ , що дорівнюють деяким  $a_i$ , парна, і  $-1$  — якщо вона непарна.

Можна визначити *приріст* функції  $F_{\mathbf{X}}$  таким чином:

$$\Delta_A F_{\mathbf{X}} = \sum_{\{\mathbf{x} = (x_1, \dots, x_k) : x_i = a_i \text{ або } x_i = b_i, i = 1, \dots, k\}} \text{sign}_A \mathbf{x} \cdot F_{\mathbf{X}}(\mathbf{x}). \quad (9.6.2)$$

Визначений таким чином приріст є нічим іншим, як мірою відповідного прямокутника:  $\mathbb{P}_{\mathbf{X}}(A) = \Delta_A F_{\mathbf{X}}$ . У цьому нескладно пересвідчитися, розглянувши випадки одно- і двомірного просторів, а загальний випадок доведемо трішки нижче.

**Приклад 9.6.7.** Якщо  $F_{\mathbf{X}} : \mathbb{R} \rightarrow \mathbb{R}$ , то «прямокутник» має вид  $A = (a; b]$ , він має  $2^1 = 2$  «вершини»,  $a$  і  $b$ , і знаками вершин є  $\text{sign}_A a = -1$ ,  $\text{sign}_A b = +1$ . Відповідно,

$$\mathbb{P}_{\mathbf{X}}((a; b]) = F_{\mathbf{X}}(b) - F_{\mathbf{X}}(a) = \Delta_{(a; b]} F_{\mathbf{X}},$$

що є одночасно застосуванням (9.6.2) і властивості (ii) з Твердження 6.1.5.

Якщо  $F_{\mathbf{X}} : \mathbb{R}^2 \rightarrow \mathbb{R}$ , то прямокутник має вид  $A = (a_1; b_1] \times (a_2; b_2]$ , він має  $2^2 = 4$  вершини,  $(a_1, a_2)$ ,  $(a_1, b_2)$ ,  $(b_1, a_2)$  і  $(b_1, b_2)$ , і знаками вершин є  $\text{sign}_A(a_1, a_2) = \text{sign}_A(b_1, b_2) = +1$ ,  $\text{sign}_A(a_1, b_1) = \text{sign}_A(b_1, a_2) = -1$  (Рис. 9.6.2). Відповідно, за (9.6.2),

$$\Delta_A F_{\mathbf{X}} = F_{\mathbf{X}}(b_1, b_2) + F_{\mathbf{X}}(a_1, a_2) - F_{\mathbf{X}}(a_1, b_2) - F_{\mathbf{X}}(b_1, a_2).$$

Але цей же вираз є мірою відповідного прямокутника  $\mathbb{P}_{\mathbf{X}}(A) = \mathbb{P}_{\mathbf{X}}(a_1 < X_1 \leq b_1, a_2 < X_2 \leq b_2)$ . Справді,

$$\mathbb{P}_{\mathbf{X}}(X_1 \leq b_1, X_2 \leq b_2) = \mathbb{P}_{\mathbf{X}}(X_1 \leq a_1, X_2 \leq b_2) + \mathbb{P}_{\mathbf{X}}(a_1 < X_1 \leq b_1, X_2 \leq b_2),$$

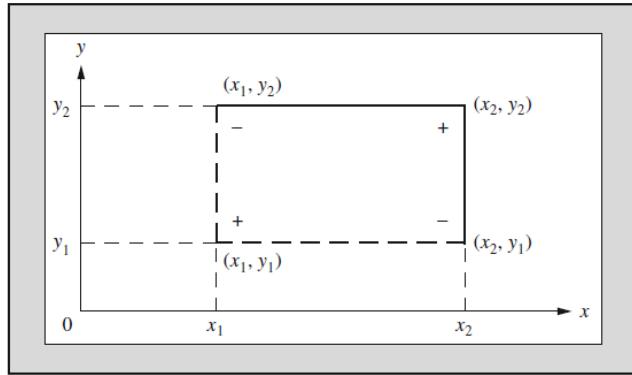


Рис. 9.6.2.: Ілюстрація до знаків вершин прямокутника на площині ([2], Рис. 7.3)

звідки

$$\mathbb{P}_{\mathbf{X}}(a_1 < X_1 \leq b_1, X_2 \leq b_2) = F_{\mathbf{X}}(b_1, b_2) - F_{\mathbf{X}}(a_1, b_2).$$

Аналогічно, оскільки

$$\mathbb{P}_{\mathbf{X}}(a_1 < X_1 \leq b_1, X_2 \leq b_2) = \mathbb{P}_{\mathbf{X}}(a_1 < X_1 \leq b_1, X_2 \leq a_2) + \mathbb{P}_{\mathbf{X}}(a_1 < X_1 \leq b_1, a_2 < X_2 \leq b_2),$$

маємо:

$$\begin{aligned} \mathbb{P}_{\mathbf{X}}(a_1 < X_1 \leq b_1, a_2 < X_2 \leq b_2) &= \mathbb{P}_{\mathbf{X}}(a_1 < X_1 \leq b_1, X_2 \leq b_2) - \mathbb{P}_{\mathbf{X}}(a_1 < X_1 \leq b_1, X_2 \leq a_2) \\ &= (F_{\mathbf{X}}(b_1, b_2) - F_{\mathbf{X}}(a_1, b_2)) - (F_{\mathbf{X}}(b_1, a_2) - F_{\mathbf{X}}(a_1, a_2)) \\ &= F_{\mathbf{X}}(b_1, b_2) + F_{\mathbf{X}}(a_1, a_2) - F_{\mathbf{X}}(a_1, b_2) - F_{\mathbf{X}}(b_1, a_2). \end{aligned}$$

Таким чином,  $\mathbb{P}_{\mathbf{X}}(A) = \Delta_A F_{\mathbf{X}}$ . □

Тепер можемо розглянути декілька властивостей спільних функцій розподілу.

**Твердження 9.6.8.** Спільна функція розподілу  $F_{\mathbf{X}}(x_1, \dots, x_k)$  має такі властивості:

(i) є неспадною за кожною змінною  $j = 1, \dots, k$ :

$$F_{\mathbf{X}}(x_1, \dots, x_{j-1}, a, x_{j+1}, \dots, x_k) \leq F_{\mathbf{X}}(x_1, \dots, x_{j-1}, b, x_{j+1}, \dots, x_k), \quad a \leq b.$$

Окрім цього, для обмежених прямокутників виду  $A = \{\mathbf{x} : a_i < x_i \leq b_i, i = 1, \dots, k\}$  приrostки  $\Delta_A F_{\mathbf{X}} \geq 0$ ;

(ii) є неперервною зверху у тому сенсі, що якщо  $x_i^{(n)} \downarrow x_i, i = 1, \dots, k$ , то  $\mathbf{x}^{(n)} \downarrow \mathbf{x}$ ,  $S_{\mathbf{x}^{(n)}} \downarrow S_{\mathbf{x}}$ , а відтак  $F_{\mathbf{X}}(\mathbf{x}^{(n)}) \rightarrow F_{\mathbf{X}}(\mathbf{x})$ ;

(iii) є нормованою:

$$\begin{aligned} \lim_{\min\{x_1, \dots, x_k\} \rightarrow -\infty} F_{\mathbf{X}}(x_1, \dots, x_k) &= 0, \\ \lim_{\min\{x_1, \dots, x_k\} \rightarrow \infty} F_{\mathbf{X}}(x_1, \dots, x_k) &= 1. \end{aligned}$$

*Доведення.* (i) Неспадність за кожною змінною безпосередньо випливає з неспадності міри з Теореми 2.2.1,

$$\mathbb{P}_{\mathbf{X}} \left( S_{(x_1, \dots, x_{j-1}, a, x_{j+1}, \dots, x_k)^\top} \right) \leq \mathbb{P}_{\mathbf{X}} \left( S_{(x_1, \dots, x_{j-1}, b, x_{j+1}, \dots, x_k)^\top} \right) , \quad a \leq b .$$

Що стосується приростів, то кожний обмежений прямокутник можна подати як

$$A = S_{(b_1, \dots, b_k)^\top} \setminus \left( S_{(a_1, b_2, \dots, b_k)^\top} \cup S_{(b_1, a_2, \dots, b_k)^\top} \cup \dots \cup S_{(b_1, b_2, \dots, a_k)^\top} \right) .$$

Застосовуючи до цього виразу формулу (2.2.3), мамо:

$$\begin{aligned} \mathbb{P}_{\mathbf{X}} (A) &= \mathbb{P}_{\mathbf{X}} \left( S_{(b_1, \dots, b_k)^\top} \right) - \mathbb{P}_{\mathbf{X}} \left( S_{(a_1, b_2, \dots, b_k)^\top} \cup S_{(b_1, a_2, \dots, b_k)^\top} \cup \dots \cup S_{(b_1, b_2, \dots, a_k)^\top} \right) \\ &= \mathbb{P}_{\mathbf{X}} \left( S_{(b_1, \dots, b_k)^\top} \right) - \left( \sum_{i=1}^k \mathbb{P}_{\mathbf{X}} \left( S_{(b_1, \dots, b_{i-1}, a_i, b_{i+1}, \dots, b_k)^\top} \right) \right. \\ &\quad - \sum_{i < j} \mathbb{P}_{\mathbf{X}} \left( S_{(b_1, \dots, b_{i-1}, a_i, b_{i+1}, \dots, b_k)^\top} \cap S_{(b_1, \dots, b_{j-1}, a_j, b_{j+1}, \dots, b_k)^\top} \right) \\ &\quad + \dots \\ &\quad \left. + (-1)^{k+1} \mathbb{P}_{\mathbf{X}} \left( S_{(a_1, b_2, \dots, b_k)^\top} \cap \dots \cap S_{(b_1, b_2, \dots, a_k)^\top} \right) \right) . \end{aligned}$$

Нескладно бачити, що кожний доданок відповідає значенню функції розподілу в деякій точці. Наприклад,

$$\mathbb{P}_{\mathbf{X}} \left( S_{(b_1, \dots, b_{i-1}, a_i, b_{i+1}, \dots, b_k)^\top} \right) = F_{\mathbf{X}}(b_1, \dots, b_{i-1}, a_i, b_{i+1}, \dots, b_k) , \quad i = 1, \dots, k ,$$

а, скажімо

$$\begin{aligned} &\mathbb{P}_{\mathbf{X}} \left( S_{(b_1, \dots, b_{i-1}, a_i, b_{i+1}, \dots, b_k)^\top} \cap S_{(b_1, \dots, b_{j-1}, a_j, b_{j+1}, \dots, b_k)^\top} \right) \\ &= \mathbb{P}_{\mathbf{X}} \left( S_{(b_1, \dots, b_{i-1}, a_i, b_{i+1}, \dots, b_{j-1}, a_j, b_{j+1}, \dots, b_k)^\top} \right) \\ &= F_{\mathbf{X}}(b_1, \dots, b_{i-1}, a_i, b_{i+1}, \dots, b_{j-1}, a_j, b_{j+1}, \dots, b_k) , \quad i < j , \end{aligned}$$

Загальне число доданків із різними знаками у виразі для  $\mathbb{P}_{\mathbf{X}} (A)$  дорівнює

$$\binom{k}{0} + \binom{k}{1} + \binom{k}{2} + \dots + \binom{k}{k} = 2^k ,$$

а відтак можна помітити, що

$$\mathbb{P}_{\mathbf{X}} (A) = \Delta_A F_{\mathbf{X}} ,$$

що є невід'ємною величиною в силу невід'ємності міри  $\mathbb{P}_{\mathbf{X}}$ ;

(ii) оскільки за Теоремою 2.2.8 будь-яка міра є неперервною, у тому числі неперервною зверху, можемо розглянути спадну послідовність векторів  $\mathbf{x}^{(n)} = (x_1^{(n)}, \dots, x_k^{(n)})$  таких, що  $x_i^{(n)} =$

$x_i + h_n^i \downarrow x_i$  (наприклад, якщо  $h_n^i = 1/n$ ),  $i = 1, \dots, k$ . Тоді, вочевидь,

$$\lim_{n \rightarrow \infty} S_{\mathbf{x}^{(n)}} = \lim_{n \rightarrow \infty} \{ \mathbf{y} : y_i \leq x_i + h_n^i, i = 1, \dots, k \} = \{ \mathbf{y} : y_i \leq x_i, i = 1, \dots, k \} = S_{\mathbf{x}},$$

а тоді

$$\lim_{n \rightarrow \infty} \mathbb{P}_{\mathbf{X}}(S_{\mathbf{x}^{(n)}}) = \mathbb{P}_{\mathbf{X}}\left(\lim_{n \rightarrow \infty} S_{\mathbf{x}^{(n)}}\right) = \mathbb{P}_{\mathbf{X}}(S_{\mathbf{x}}),$$

тобто

$$\lim_{n \rightarrow \infty} F_{\mathbf{X}}(\mathbf{x}^{(n)}) = \lim_{\mathbf{x}^{(n)} \downarrow \mathbf{x}} F_{\mathbf{X}}(\mathbf{x}^{(n)}) = F_{\mathbf{X}}(\mathbf{x});$$

(iii) для доведення

$$\lim_{\min\{x_1, \dots, x_k\} \rightarrow -\infty} F_{\mathbf{X}}(x_1, \dots, x_k) = 0$$

потрібно помітити, що умова переходу до границі вимагає, щоб *найменша* координата прямувала до  $-\infty$ , інші можуть залишатися скінченими. Тоді достатньо розглянути деяку послідовність  $\mathbf{x}^{(n)} = (x_1^{(n)}, \dots, x_k^{(n)})$  таких, що  $x_i^{(n)} = x_i - h_n^i \downarrow -\infty$  (наприклад, якщо  $h_n^i = n$ ), для деякого  $i \in \{1, \dots, k\}$ .

У цьому випадку

$$\lim_{n \rightarrow \infty} S_{\mathbf{x}^{(n)}} = \lim_{n \rightarrow \infty} \{ \mathbf{y} : y_1 \leq x_1, \dots, y_i \leq x_i - h_n^i, \dots, y_k \leq x_k, i = 1, \dots, k \} = \emptyset,$$

а тоді

$$\lim_{\min\{x_1, \dots, x_k\} \rightarrow -\infty} F_{\mathbf{X}}(x_1, \dots, x_k) = \lim_{n \rightarrow \infty} \mathbb{P}_{\mathbf{X}}(S_{\mathbf{x}^{(n)}}) = \mathbb{P}_{\mathbf{X}}(\emptyset) = 0.$$

Схожим чином можна показати  $\lim_{\min\{x_1, \dots, x_k\} \rightarrow \infty} F_{\mathbf{X}}(x_1, \dots, x_k) = 1$ , помітивши, що

$$\lim_{\min\{x_1, \dots, x_k\} \rightarrow \infty} F_{\mathbf{X}}(x_1, \dots, x_k) = \lim_{\substack{x_1 \rightarrow \infty \\ \dots \\ x_k \rightarrow \infty}} F_{\mathbf{X}}(x_1, \dots, x_k),$$

тобто що *всі координати* повинні прямувати до нескінченності. Тоді можна розглянути зростаючу послідовність  $\mathbf{x}^{(n)} = (x_1^{(n)}, \dots, x_k^{(n)})$  таких, що  $x_i^{(n)} = x_i + h_n^i \uparrow \infty$ , для *кожного*  $i \in \{1, \dots, k\}$ . Для такої послідовності  $\lim_{n \rightarrow \infty} S_{\mathbf{x}^{(n)}} = \Omega$ , а  $\mathbb{P}_{\mathbf{X}}(\Omega) = 1$ .

□

**Теорема 9.6.9.** Нехай  $F : \mathbb{R}^k \rightarrow \mathbb{R}$  — деяка спільна функція розподілу, тобто функція, яка задовольняє умови (i)–(ii) з Твердження 9.6.8. Тоді в Борелевому просторі  $(\mathbb{R}^k, \mathcal{B}^k)$  існує єдина міра  $\nu$  така, що  $\nu(A) = \Delta_A F$ ,  $A \in \mathcal{B}^k$ .

Якщо ця функція додатково задовольняє умову (iii) з Твердження 9.6.8, то міра  $\nu$  є ймовірнісною мірою, тобто існує випадковий вектор  $\mathbf{X} : (\Omega, \mathcal{A}, \mathbb{P}) \rightarrow \mathbb{R}^k$  на деякому ймовірнісному просторі такий, що  $F(x_1, \dots, x_k) = \mathbb{P}_{\mathbf{X}}(X_1 \leq x_1, \dots, X_k \leq x_k)$ .

Наприклад, для рівномірного розподілу з Прикладу 9.2.9, як нескладно переконатися,  $\Delta_A F = (F(b_1) - F(a_1)) \cdot \dots \cdot (F(b_k) - F(a_k))$  і тому умови (i)–(ii) з Твердження 9.6.8 виконуються.

#### 9.6.4. Критерій незалежності випадкових величин

**Визначення 9.6.10.** Клас множин  $\mathcal{P}$  називають  $\pi$ -системою ( $\pi$ -system), якщо вона замкнена відносно перетинів: з  $A, B \in \mathcal{P}$  випливає  $A \cap B \in \mathcal{P}$ .  $\square$

**Визначення 9.6.11.** Клас множин  $\mathcal{L}$  називають  $\lambda$ -системою ( $\lambda$ -system), якщо:

- (i)  $\Omega \in \mathcal{L}$ ;
- (ii) якщо  $A, B \in \mathcal{L}$  і  $A \subset B$ , то  $B \setminus A \in \mathcal{L}$ ;
- (iii) якщо неспадна послідовність  $A_n \in \mathcal{L}$  збігається до множини  $A$ , то  $A \in \mathcal{L}$ .

 $\square$ 

**Теорема 9.6.12** (Теорема Дінкіна про  $\pi$ - $\lambda$  (Dynkin's  $\pi$ - $\lambda$  Theorem)<sup>11</sup>). Якщо  $\mathcal{P}$  — деяка  $\pi$ -система, а  $\mathcal{L}$  — деяка  $\lambda$ -система така, що вона містить  $\mathcal{P}$ ,  $\mathcal{P} \subseteq \mathcal{L}$ , то вона ж містить і  $\sigma$ -алгебру, яку породжує  $\mathcal{P}$ :  $\mathcal{A}(\mathcal{P}) \subseteq \mathcal{L}$ .

Цю теорему ми доводити не будемо, оскільки це виходить за межі власне теорії ймовірностей. Ми використаємо її результат для доведення іншого потрібного для нас твердження.

**Твердження 9.6.13.** Нехай  $\mathcal{P}_1, \dots, \mathcal{P}_k$  — незалежні  $\pi$ -системи. Тоді незалежними є  $\mathcal{A}(\mathcal{P}_1), \dots, \mathcal{A}(\mathcal{P}_k)$ .

**Доведення.** Нехай  $A_2 \in \mathcal{P}_2, \dots, A_k \in \mathcal{P}_k$  і  $F = A_2 \cap \dots \cap A_k$ , а  $\mathcal{L} = \{A : \mathbb{P}(A \cap F) = \mathbb{P}(A) \mathbb{P}(F)\}$ . Перевірмо за Визначенням 9.6.11, що  $\mathcal{L}$  є  $\lambda$ -системою:

- (i) оскільки  $\mathbb{P}(\Omega \cap F) = \mathbb{P}(\Omega) \mathbb{P}(F)$ , простір  $\Omega \in \mathcal{L}$ ;
- (ii) нехай  $B, C \in \mathcal{L}$ ,  $B \subset C$ . Тоді

$$(C \setminus B) \cap F = C \cap B^c \cap F = C \cap F \cap (B^c \cup F^c) = (C \cap F) \setminus (B \cap F) .$$

За властивістю (iii) міри з Теореми 2.2.1, а також того факту, що  $\mathcal{L}$  містить події, незалежні з  $F$ , випливає:

$$\mathbb{P}((C \setminus B) \cap F) = \mathbb{P}(B \cap F) - \mathbb{P}(A \cap F) = \mathbb{P}(B) \mathbb{P}(F) - \mathbb{P}(A) \mathbb{P}(F) = \mathbb{P}(B \setminus A) \mathbb{P}(F) ,$$

тобто  $B \setminus A \in \mathcal{L}$ ;

- (iii) розгляньмо неспадну послідовність подій  $B_n \in \mathcal{L}$ , яка збігається до  $B$ . Тоді послідовність подій  $B_n \cap F$  також є неспадною, і вона збігається до  $B \cap F$ . За властивістю неперервності міри з Теореми 2.2.8, а також того факту, що  $\mathcal{L}$  містить події, незалежні з  $F$ , маємо:

$$\mathbb{P}(B \cap F) = \lim_{n \rightarrow \infty} \mathbb{P}(B_n \cap F) = \lim_{n \rightarrow \infty} \mathbb{P}(B_n) \mathbb{P}(F) = \mathbb{P}(B) \mathbb{P}(F) ,$$

тобто  $B \in \mathcal{L}$ .

<sup>11</sup>Названо так на честь радянсько-американського математика Євгена Дінкіна (1924–2014)

Оскільки  $\mathcal{L}$  справді є  $\lambda$ -системою, а  $\mathcal{P}_1 \subset \mathcal{L}$ , то за Теоремою 9.6.12  $\mathcal{A}(\mathcal{P}_1) \subset \mathcal{L}$ . Таким чином, якщо  $A_1 \in \mathcal{A}(\mathcal{P}_1)$ ,  $A_j \in \mathcal{P}_j$ ,  $j = 2, \dots, k$ , то

$$\mathbb{P}\left(\bigcap_{i=1}^k A_i\right) = \mathbb{P}(A_1) \mathbb{P}\left(\bigcap_{i=2}^k A_i\right) = \prod_{i=1}^n \mathbb{P}(A_i) ,$$

а отже незалежними є класи множин  $\mathcal{A}(\mathcal{P}_1), \mathcal{P}_2, \dots, \mathcal{P}_k$ .

Повторне застосування цих міркувань дає незалежність усіх  $\mathcal{A}(\mathcal{P}_j)$ ,  $j = 1, \dots, k$ .  $\square$

Із Визначення 9.4.1 випливає незалежність відповідних  $\sigma$ -алгебр.

**Твердження 9.6.14.** Якщо випадкові величини  $X_1, \dots, X_k$  незалежні, то незалежними є відповідні породжені  $\sigma$ -алгебри  $\mathcal{A}(X_1), \dots, \mathcal{A}(X_k)$ .

Якщо випадкові величини  $X_1, \dots, X_k$  вимірні відносно деяких незалежних  $\sigma$ -алгебр  $\mathcal{A}_1, \dots, \mathcal{A}_k$ , то ці величини є незалежними.

*Доведення.* Нехай випадкові величини  $X_1, \dots, X_k$  незалежні. Якщо  $A_1 \in \mathcal{A}(X_1), \dots, A_k \in \mathcal{A}(X_k)$ , то з Визначення 12.2.8 випливає

$$A_1 = \{\omega : X_1(\omega) \in B_1\}, \dots, A_k = \{\omega : X_k(\omega) \in B_k\}$$

для деяких  $B_1, \dots, B_k \in \mathbb{R}$ . Згідно з (9.4.1) маємо

$$\begin{aligned} \mathbb{P}(A_1 \cap \dots \cap A_k) &= \mathbb{P}_{\mathbf{X}}(X_1 \in B_1, \dots, X_k \in B_k) \\ &= \mathbb{P}_{X_1}(X_1 \in B_1) \cdot \dots \cdot \mathbb{P}_{X_k}(X_k \in B_k) = \mathbb{P}(A_1) \cdot \dots \cdot \mathbb{P}(A_k) , \end{aligned}$$

і аналогічно для будь-яких комбінацій цих подій (пар, трійок тощо). Це повністю відповідає визначенню незалежності подій (3.4.1) та Визначення 3.4.4 для  $k$  подій. А оскільки множини  $A_1, \dots, A_k$  є довільні, маємо незалежність відповідних  $\sigma$ -алгебр.

Розгляньмо доведення в інший бік. Якщо незалежними є  $\sigma$ -алгебри  $\mathcal{A}_1, \dots, \mathcal{A}_k$ , то можна взяти деякі множини  $B_1, \dots, B_k \in \mathbb{R}$  такі, що  $A_j = \{\omega : X_j(\omega) \in B_j\} \in \mathcal{A}_j$ ,  $j = 1, \dots, k$ . Множини  $A_j$  незалежні за незалежністю  $\sigma$ -алгебр, і тому

$$\begin{aligned} \mathbb{P}_{\mathbf{X}}(X_1 \in B_1, \dots, X_k \in B_k) &= \mathbb{P}(A_1 \cap \dots \cap A_k) \\ &= \mathbb{P}(A_1) \cdot \dots \cdot \mathbb{P}(A_k) = \mathbb{P}_{X_1}(X_1 \in B_1) \cdot \dots \cdot \mathbb{P}_{X_k}(X_k \in B_k) , \end{aligned}$$

і аналогічно для будь-яких комбінацій цих подій (пар, трійок тощо).  $\square$

Доведімо Теорему 9.4.3.

*Доведення.* Неодмінність очевидно випливає з (9.4.1), оскільки в події  $X_i \in (-\infty; x_i]$ ,  $i = 1, \dots, k$ , промінь очевидно належить Борелевій  $\sigma$ -алгебрі.

Для доведення достатності спочатку варто помітити, що множина

$$\mathcal{P}_i = \{\omega : X_i(\omega) \leq x_i, x_i \in (-\infty, \infty]\}$$

є  $\pi$ -системою, адже

$$\{\omega : X_i(\omega) \leq x\} \cap \{\omega : X_i(\omega) \leq y\} = \{\omega : X_i(\omega) \leq \min\{x, y\}\} \in \mathcal{P}_i .$$

Як відомо з Розділу 1.5.3, клас променів  $(-\infty; x]$  породжує Борелеву  $\sigma$ -алгебру  $\mathcal{B}$ . За Лемою 7.6.4, клас множин  $\mathcal{P}_i$  породжує  $\sigma$ -алгебру  $\mathcal{A}(X_i)$ . Відтак, за Твердженням 9.6.13, оскільки  $\mathcal{P}_i$  є незалежні, то і  $\mathcal{A}(\mathcal{P}_i) = \mathcal{A}(X_i)$  є незалежні, а значить, за Твердженням 9.6.14, і самі  $X_i$  є незалежні.  $\square$

Доведімо Теорему 9.4.4.

*Доведення.* Нехай величини  $X_1, \dots, X_k$  незалежні. Розгляньмо вимірні прямокутники  $A_1 \times \dots \times A_k$ ,  $A_j \in \mathcal{B}$ ,  $j = 1, \dots, k$ . За визначенням розподілу (4.1.1) та з урахуванням незалежності випадкових величин маємо:

$$\begin{aligned}\mathbb{P}_{\mathbf{X}}(A_1 \times \dots \times A_k) &= \mathbb{P}_{\mathbf{X}}(\mathbf{X} \in A_1 \times \dots \times A_k) \\ &= \mathbb{P}_{\mathbf{X}}(X_1 \in A_1, \dots, X_k \in A_k) \\ &= \prod_{i=1}^k \mathbb{P}_{X_i}(X_i \in A_i) = \prod_{i=1}^k \mathbb{P}_{X_i}(A_i),\end{aligned}$$

де  $A_i \in \mathcal{B}$ ,  $i = 1, \dots, k$ . Іншими словами,

$$\mathbb{P}_{\mathbf{X}}(A_1 \times \dots \times A_k) = (\mathbb{P}_{X_1} \times \dots \times \mathbb{P}_{X_k})(A_1 \times \dots \times A_k),$$

тобто міри  $\mathbb{P}_{\mathbf{X}}$  і  $\mathbb{P}_{X_1} \times \dots \times \mathbb{P}_{X_k}$  дорівнюють на множині вимірних прямокутників, яка є  $\pi$ -системою, що породжує  $\mathcal{B}^k$ . У доведенні Теореми 2.4.1, яке ми опустили, фігурує твердження, що якщо дві міри дорівнюють одна одній для  $\pi$ -системи, то вони також дорівнюють одна одній і для всієї  $\sigma$ -алгебри, яку ця  $\pi$ -система породжує. Таким чином, виконується (9.4.5).

Якщо ж покласти, що виконується (9.4.5), тобто що

$$\mathbb{P}_{\mathbf{X}}(A_1 \times \dots \times A_k) = (\mathbb{P}_{X_1} \times \dots \times \mathbb{P}_{X_k})(A_1 \times \dots \times A_k)$$

для всіх  $A_i \in \mathcal{B}$ ,  $i = 1, \dots, k$ , то з властивості добутку мір (Теорема 9.1.6) випливає

$$(\mathbb{P}_{X_1} \times \dots \times \mathbb{P}_{X_k})(A_1 \times \dots \times A_k) = \prod_{i=1}^k \mathbb{P}_{X_i}(A_i),$$

що в поєднанні з  $\mathbb{P}_{\mathbf{X}}(A_1 \times \dots \times A_k)$  дає визначення незалежності випадкових величин.  $\square$

Доведімо Теорему 9.4.9.

*Доведення.* Для доведення цього твердження спочатку доведімо такий факт: якщо  $\sigma$ -алгебри  $\mathcal{C}_{i,j}$ ,  $i = 1, \dots, n$ ,  $j = 1, \dots, m(i)$ , незалежні, то незалежні є  $\sigma$ -алгебри  $\mathcal{M}_i = \mathcal{A}\left(\bigcup_{j=1}^{m(i)} \mathcal{C}_{i,j}\right)$ .

Справді, нехай  $\mathcal{P}_i$  — деякий клас, який містить множини виду  $\bigcap_{j=1}^{m(i)} A_{i,j}$ , де  $A_{i,j} \in \mathcal{C}_{i,j}$ . Тоді за побудовою  $\mathcal{P}_i$  є  $\pi$ -системою і з Твердженням 9.6.13 випливає незалежність  $\mathcal{A}(\mathcal{P}_i)$ ,  $i = 1, \dots, k$ .

При цьому, оскільки  $\bigcup_{j=1}^{m(i)} \mathcal{C}_{i,j} \subseteq \mathcal{P}_i$ <sup>12</sup>, то  $\mathcal{M}_i \subseteq \mathcal{A}(\mathcal{P}_i)$ . А оскільки  $\mathcal{P}_i \subseteq \mathcal{M}_i$ , то  $\mathcal{A}(\mathcal{P}_i) \subseteq \mathcal{M}_i$ . Відповідно,  $\mathcal{M}_i = \mathcal{A}(\mathcal{P}_i)$ , і тому  $\mathcal{M}_i$  є незалежні.

Для доведення самого твердження достатньо покласти  $\mathcal{C}_{i,j} = \mathcal{A}(X_{i,j})$  і помітити, що випадкова величина  $f_i(X_{i,1}, \dots, X_{i,m(i)})$  вимірна відносно  $\mathcal{M}_i$ . Конкретніше, оскільки  $f_i$  вимірна відносно  $\mathcal{B}^k$ , то

$$f_i^{-1}(A) \in \mathcal{B}^k, \quad A \in \mathcal{B}.$$

<sup>12</sup> Адже в силу властивостей  $\sigma$ -алгебри  $A_{i,j}^c \in \mathcal{C}_{i,j}$  і за правилом де Моргана якщо перетини належать  $\mathcal{P}_i$ , то й об'єднання також належать  $\mathcal{P}_i$ .

Випадковий вектор  $(X_{i,1}, \dots, X_{i,m(i)})$  за визначенням вимірний відносно  $\sigma$ -алгебри  $\mathcal{A}(X_{i,1}, \dots, X_{i,m(i)})$ :

$$(X_{i,1}, \dots, X_{i,m(i)})^{-1}(f_i^{-1}(A)) \in \mathcal{A}((X_{i,1}, \dots, X_{i,m(i)})) , \quad f_i^{-1}(A) \in \mathcal{B}^k .$$

Але за властивостями оберненої функції,

$$(X_{i,1}, \dots, X_{i,m(i)})^{-1}(f_i^{-1}(A)) = (f_i(X_{i,1}, \dots, X_{i,m(i)}))^{-1}(A) \in \mathcal{A}((X_{i,1}, \dots, X_{i,m(i)})) , \quad A \in \mathcal{B} .$$

Також можна показати, що

$$\mathcal{A}((Y_1, \dots, Y_m)) = \mathcal{A}(\mathcal{A}(Y_1) \cup \dots \cup \mathcal{A}(Y_m)) .$$

Ми цього факту детально доводити не будемо<sup>13</sup>.

Звідси

$$\mathcal{A}((X_{i,1}, \dots, X_{i,m(i)})) = \mathcal{A}(\mathcal{A}(X_{i,1}) \cup \dots \cup \mathcal{A}(X_{i,m(i)})) = \mathcal{M}_i ,$$

тобто  $f_i(X_{i,1}, \dots, X_{i,m(i)})$  є вимірною відносно  $\mathcal{M}_i$ .

Нарешті, за Твердженням 9.6.14, незалежність  $\mathcal{M}_i$  має наслідком незалежність випадкових величин  $f_i(X_{i,1}, \dots, X_{i,m(i)})$ .  $\square$

### 9.6.5. Згортки відомих розподілів

Доведімо Твердження 9.5.4.

*Доведення.* (i) Здійснімо доведення за індукцією.

Нехай  $X_1, X_2 \sim \text{Bern}(p)$ . За формулою (9.5.2) для  $S_2 = X_1 + X_2$  маємо:

$$\begin{aligned} \mathbb{P}_{S_2}(S_2 = s) &= \sum_{j=-\infty}^{\infty} \mathbb{P}_{X_1}(X_1 = s-j) \mathbb{P}_{X_2}(X_2 = j) \\ &= \mathbb{P}_{X_1}(X_1 = s-0) \mathbb{P}_{X_2}(X_2 = 0) + \mathbb{P}_{X_1}(X_1 = s-1) \mathbb{P}_{X_2}(X_2 = 1) \\ &= (1-p)\mathbb{P}_{X_1}(X_1 = s) + p\mathbb{P}_{X_1}(X_1 = s-1) . \end{aligned}$$

Оскільки носій  $S_2$  дорівнює  $\text{supp}(S_2) = \{0, 1, 2\}$ , маємо

$$\begin{aligned} \mathbb{P}_{S_2}(S_2 = 0) &= (1-p) \cdot (1-p) + p \cdot 0 = (1-p)^2 , \\ \mathbb{P}_{S_2}(S_2 = 1) &= (1-p) \cdot p + p \cdot (1-p) = 2p(1-p) , \\ \mathbb{P}_{S_2}(S_2 = 2) &= (1-p) \cdot 0 + p \cdot p = p^2 , \end{aligned}$$

тобто

$$\mathbb{P}_{S_2}(S_2 = s) = \binom{2}{s} p^s (1-p)^{2-s} , \quad s = 0, 1, 2 .$$

<sup>13</sup> Ескізно доведення цього твердження таке. Множина  $\mathcal{A}(Y_j)$  складається з множин виду  $C = \{Y_j^{-1}(A)\}$  для деякої  $A \in \mathcal{B}$ . Однак тоді  $C = (Y_1, \dots, Y_m)^{-1}(A \times \mathbb{R} \times \dots \times \mathbb{R}) \in \mathcal{A}((Y_1, \dots, Y_m))$ . Таким чином,  $\mathcal{A}(Y_j) \subseteq \mathcal{A}((Y_1, \dots, Y_m))$ , а отже  $\mathcal{A}(\mathcal{A}(Y_1) \cup \dots \cup \mathcal{A}(Y_m)) \subseteq \mathcal{A}((Y_1, \dots, Y_m))$ . З іншого боку, якщо  $A_1, \dots, A_m \in \mathcal{B}$ , то

$$(Y_1, \dots, Y_m)^{-1}(A_1 \times \dots \times A_m) = Y_1^{-1}(A_1) \cap \dots \cap Y_m^{-1}(A_m) \in \mathcal{A}(\mathcal{A}(Y_1) \cup \dots \cup \mathcal{A}(Y_m)) ,$$

тобто  $(Y_1, \dots, Y_m)^{-1}(C) \in \mathcal{A}((Y_1, \dots, Y_m))$  для всіх вимірних прямокутників  $C$ . Оскільки вимірні прямокутники породжують  $\mathcal{B}^k$ , то вищеприведена формула справедлива для всіх  $C \in \mathcal{B}^k$ , а тому  $\mathcal{A}(\mathcal{A}(Y_1) \cup \dots \cup \mathcal{A}(Y_m)) \subseteq \mathcal{A}(Y_1 \cup \dots \cup Y_m)$ .

Відтак  $S_2 \sim \text{Binom}(2, p)$ .

Тепер нехай  $S_{k-1} = \sum_{i=1}^{k-1} \sim \text{Binom}(k-1, p)$ . Нехай  $X_k \sim \text{Bern}(p)$ . Тоді для  $S_k = S_{k-1} + X_k$  за формулою згортки маємо

$$\begin{aligned}
 \mathbb{P}_{S_k}(S_k = s) &= \sum_{j=-\infty}^{\infty} \mathbb{P}_{S_{k-1}}(S_{k-1} = s-j) \mathbb{P}_{X_k}(X_k = j) \\
 &= \mathbb{P}_{S_{k-1}}(S_{k-1} = s-0) \mathbb{P}_{X_k}(X_k = 0) + \mathbb{P}_{S_{k-1}}(S_{k-1} = s-1) \mathbb{P}_{X_k}(X_k = 1) \\
 &= (1-p) \mathbb{P}_{S_{k-1}}(S_{k-1} = s) + p \mathbb{P}_{S_{k-1}}(S_{k-1} = s-1) \\
 &= (1-p) \cdot \binom{k-1}{s} p^s (1-p)^{k-1-s} + p \cdot \binom{k-1}{s-1} p^{s-1} (1-p)^{k-1-(s-1)} \\
 &= p^s (1-p)^{k-s} \left( \binom{k-1}{s} + \binom{k-1}{s-1} \right) \\
 &= \binom{k}{s} p^s (1-p)^{k-s},
 \end{aligned}$$

де ми використали так зване *правило Паскаля* (Pascal's rule):

$$\binom{k-1}{s} + \binom{k-1}{s-1} = \binom{k}{s},$$

доведення якого очевидне після розкриття відповідних біномних коефіцієнтів та приведення суми до спільного знаменника:

$$\begin{aligned}
 \binom{k-1}{s} + \binom{k-1}{s-1} &= \frac{(k-1)!}{s!(k-1-s)!} + \frac{(k-1)!}{(s-1)!(k-1-(s-1))!} \\
 &= \frac{(k-1)! \cdot (k-s) + (k-1)! \cdot s}{s!(k-s)!} \\
 &= \frac{k!}{s!(k-s)!} = \binom{k}{s}.
 \end{aligned}$$

Отже  $S_k \sim \text{Binom}(k, p)$ ;

(ii) нехай  $X_1 \sim \text{Binom}(n_1, p)$ ,  $X_2 \sim \text{Binom}(n_2, p)$ ,  $n_1 < n_2$ . За формулою (9.5.2) маємо:

$$\begin{aligned}
 \mathbb{P}_S(S = s) &= \sum_{j=-\infty}^{\infty} \mathbb{P}_{X_1}(X_1 = s-j) \mathbb{P}_{X_2}(X_2 = j) \\
 &= \sum_{j=0}^s \binom{n_1}{s-j} p^{s-j} (1-p)^{n_1-s+j} \cdot \binom{n_2}{j} p^j (1-p)^{n_2-j} \\
 &= p^s (1-p)^{n_1+n_2-s} \sum_{j=0}^s \binom{n_1}{s-j} \cdot \binom{n_2}{j} \\
 &= \binom{n_1+n_2}{s} p^s (1-p)^{n_1+n_2-s},
 \end{aligned}$$

оскільки

$$\sum_{j=0}^s \binom{n_1}{s-j} \cdot \binom{n_2}{j} = \binom{n_1 + n_2}{s}$$

за тотожністю Вандермонда (5.4.2). Іншими словами,  $X_1 + X_2 \sim \text{Binom}(n_1 + n_2, p)$ . За індукцію випливає, що

$$\sum_{i=1}^k X_i \sim \text{Binom}\left(\sum_{i=1}^k n_i, p\right) ;$$

(iii) нехай  $X_1 \sim \text{Pois}(\lambda_1)$ ,  $X_2 \sim \text{Pois}(\lambda_2)$ . За формулою (9.5.2) маємо:

$$\begin{aligned} \mathbb{P}_S(S=s) &= \sum_{j=-\infty}^{\infty} \mathbb{P}_{X_1}(X_1=s-j) \mathbb{P}_{X_2}(X_2=j) \\ &= \sum_{j=0}^s \frac{e^{-\lambda_1} \lambda_1^{s-j}}{(s-j)!} \cdot \frac{e^{-\lambda_2} \lambda_2^j}{j!} \\ &= \frac{e^{-(\lambda_1+\lambda_2)}}{s!} \sum_{j=0}^s \frac{s!}{j!(s-j)!} \lambda_1^{s-j} \lambda_2^j \\ &= \frac{e^{-(\lambda_1+\lambda_2)}}{s!} (\lambda_1 + \lambda_2)^s, \end{aligned}$$

тобто  $X_1 + X_2 \sim \text{Pois}(\lambda_1 + \lambda_2)$ . За індукцію випливає, що

$$\sum_{i=1}^k X_i \sim \text{Pois}\left(\sum_{i=1}^k \lambda_i\right) ;$$

(iv) нехай маємо дві незалежні величини  $X_1 \sim N(0, \sigma^2)$ ,  $X_2 \sim N(0, 1)$ . Застосування (9.5.1) разом із прийомом із Прикладу 10.1.6 (виділення щільності деякого нормального розподілу,

інтеграл від якої дорівнює 1) дає:

$$\begin{aligned}
 f_{X_1+X_2}(s) &= \int_{-\infty}^{\infty} f_{X_1}(s-y) f_{X_2}(y) dy \\
 &= \int_{-\infty}^{\infty} \frac{1}{2\pi\sigma} e^{-\frac{(s-y)^2}{2\sigma^2} - \frac{y^2}{2}} dy \\
 &= \int_{-\infty}^{\infty} \frac{1}{2\pi\sigma} e^{-\frac{s^2 - 2sy + y^2 - y^2\sigma^2}{2\sigma^2}} dy \\
 &= \int_{-\infty}^{\infty} \frac{1}{2\pi\sigma} e^{-\left(\frac{y^2(1+\sigma^2)}{2\sigma^2} - y\frac{s}{\sigma^2} + \frac{s^2}{2\sigma^2}\right)} dy \\
 &= \int_{-\infty}^{\infty} \frac{1}{2\pi\sigma} e^{-\left(\left(\sqrt{\frac{1+\sigma^2}{2\sigma^2}}y - \frac{s}{2\sigma^2} \cdot \sqrt{\frac{2\sigma^2}{1+\sigma^2}}\right)^2 + \frac{s^2}{2\sigma^2} - \left(\frac{s}{2\sigma^2}\right)^2 \cdot \frac{2\sigma^2}{(1+\sigma^2)}\right)} dy \\
 &= \frac{1}{\sqrt{2\pi\sigma}} e^{-\left(\frac{s^2}{2\sigma^2} - \left(\frac{s}{2\sigma^2}\right)^2 \cdot \frac{2\sigma^2}{(1+\sigma^2)}\right)} \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{\left(y - \frac{s}{2\sigma^2} \cdot \frac{2\sigma^2}{1+\sigma^2}\right)^2}{2 \cdot \frac{\sigma^2}{1+\sigma^2}}} dy \\
 &= \frac{\sigma}{\sqrt{1+\sigma^2}} \cdot \frac{1}{\sqrt{2\pi\sigma}} e^{-\left(\frac{s^2}{2\sigma^2} - \left(\frac{s}{\sigma^2}\right)^2 \cdot \frac{2\sigma^2}{4(1+\sigma^2)}\right)} \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{\left(y - \frac{s}{1+\sigma^2}\right)^2}{2 \cdot \frac{\sigma^2}{1+\sigma^2}}} dy \\
 &= \frac{1}{\sqrt{2\pi(1+\sigma^2)}} e^{-\frac{s^2}{2(1+\sigma^2)}},
 \end{aligned}$$

оскільки під інтегралом стоїть щільність випадкової величини з розподілом  $N\left(\frac{s}{1+\sigma^2}, \frac{\sigma^2}{1+\sigma^2}\right)$ .

Отже  $X_1 + X_2 \sim N(0, 1 + \sigma^2)$ .

Розгляньмо тепер  $X_1 \sim N(\mu_1, \sigma_1^2)$ ,  $X_2 \sim N(\mu_2, \sigma_2^2)$ ,  $X_1 \perp\!\!\!\perp X_2$ :

$$X_1 + X_2 = \sigma_2 \left( \frac{X_1 - \mu_1}{\sigma_2} + \frac{X_2 - \mu_2}{\sigma_2} \right) + \mu_1 + \mu_2.$$

Оскільки, відповідно до Завдання 7.5.9,  $\frac{X_1 - \mu_1}{\sigma_2} \sim N\left(0, \frac{\sigma_1^2}{\sigma_2^2}\right)$ , а  $\frac{X_2 - \mu_2}{\sigma_2} \sim N(0, 1)$ , можемо використати попередній результат і помітити, що

$$\frac{X_1 - \mu_1}{\sigma_2} + \frac{X_2 - \mu_2}{\sigma_2} \sim N\left(0, 1 + \frac{\sigma_1^2}{\sigma_2^2}\right).$$

Тоді, відповідно до Завдання 7.5.9,

$$X_1 + X_2 \sim N\left(\mu_1 + \mu_2, \sigma_2^2 \left(1 + \frac{\sigma_1^2}{\sigma_2^2}\right)\right) = N\left(\mu_1 + \mu_2, \sigma_1^2 + \sigma_2^2\right).$$

За індукцією випливає, що

$$\sum_{i=1}^k X_i \sim N\left(\sum_{i=1}^k \mu_i, \sum_{i=1}^k \sigma_i^2\right).$$

□

## 10. Функції від випадкових векторів та їхні розподіли

У минулій лекції ми детально розглянули випадкові вектори, їхні числові характеристики та поняття незалежності випадкових величин. У поточній лекції ми продовжимо аналіз випадкових векторів, зокрема, з'ясуємо, як обчислювати розподіли функції від випадкових векторів. Також ми розглянемо декілька важливих на практиці розподілів випадкових векторів.

### 10.1. Формула заміни змінних для випадкових векторів

Раніше ми розглядали формули заміни змінних для дискретних і неперервних випадкових величин. Вони мають свої аналоги і для випадкових векторів.

Так, зокрема, якщо маємо випадковий вектор  $\mathbf{X} = (X_1, \dots, X_k)^\top : (\Omega, \mathcal{A}, \mathbb{P}) \rightarrow \mathbb{R}^k$ , а  $g : \mathbb{R}^k \rightarrow \mathbb{R}^n$  — деяка Борелева функція, то  $\mathbf{Y} = (Y_1, \dots, Y_n) = g(\mathbf{X})$  має розподіл

$$\mathbb{P}_{\mathbf{Y}}(\mathbf{Y} \in B) = \mathbb{P}_{\mathbf{X}}(g(\mathbf{X}) \in B) = \mathbb{P}_{\mathbf{X}}(X \in g^{-1}(B)) , \quad B \in \mathcal{B}^k , \quad (10.1.1)$$

де  $g^{-1}(B) = \{(x_1, \dots, x_k)^\top \in \mathbb{R}^k : g(x_1, \dots, x_k) \in B\}$ .

Нехай  $\mathbf{X}$  — дискретний випадковий вектор. Тоді (10.1.1) набуває форми

$$\begin{aligned} \mathbb{P}_{\mathbf{Y}}(Y_1 = y_1, \dots, Y_n = y_n) &= \mathbb{P}_{\mathbf{X}}\left(g(X_1, \dots, X_k) = (y_1, \dots, y_n)^\top\right) \\ &= \sum_{\mathbf{x}: g(\mathbf{x}) = \mathbf{y}} \mathbb{P}_{\mathbf{X}}(X_1 = x_1, \dots, X_k = x_k) . \end{aligned} \quad (10.1.2)$$

Для неперервних випадкових векторів існує результат, аналогічний Теоремі 7.4.4.

**Теорема 10.1.1** (Формула заміни змінних для випадкових векторів (Change of variables formula for random vectors)). Нехай  $\mathbf{X} = (X_1, \dots, X_k)^\top$  — випадковий вектор зі (спільною) щільністю  $f_{\mathbf{X}}$ , неперервною на  $\text{supp}(\mathbf{X})$ . Нехай  $h_1, \dots, h_k : \text{supp}(\mathbf{X}) \rightarrow \mathbb{R}$  такі, що  $(h_1, \dots, h_k)$  задає взаємно однозначне відображення між  $\text{supp}(\mathbf{X})$  та його образом, який позначатимемо через  $T$ . Тоді якщо  $y_i = h_i(\mathbf{x})$ ,  $\mathbf{x} = (x_1, \dots, x_k)^\top$ , можна подати  $x_i = g_i(\mathbf{y})$ ,  $\mathbf{y} = (y_1, \dots, y_k)^\top$ ,  $i = 1, \dots, k$ .

Нехай частинні похідні  $g_{ij}(\mathbf{y}) = \frac{\partial}{\partial y_j} g_i(\mathbf{y})$ ,  $i, j = 1, \dots, k$ , існують і є неперервними в точці  $\mathbf{y} \in T$ . Також нехай якобіан

$$J = \begin{vmatrix} g_{11}(\mathbf{y}) & \dots & g_{1k}(\mathbf{y}) \\ \vdots & \ddots & \vdots \\ g_{k1}(\mathbf{y}) & \dots & g_{kk}(\mathbf{y}) \end{vmatrix}$$

не дорівнює нулю на  $T$ .

У цьому випадку

$$f_{\mathbf{Y}}(\mathbf{y}) = f_{\mathbf{X}}(g_1(\mathbf{y}), \dots, g_k(\mathbf{y})) \cdot |J| \cdot \mathbb{1}\{\mathbf{y} \in T\} . \quad (10.1.3)$$

*Доведення.* Доведення цієї теореми безпосередньо випливає з застосування формул заміни

них у повторному інтегралі. Так, мірою деякої множини  $B \in T$ , за (9.2.3), є

$$\mathbb{P}_{\mathbf{Y}}(\mathbf{Y} \in B) = \int_B f_{\mathbf{Y}} d\lambda_k .$$

За Теоремою 9.1.8,

$$\mathbb{P}_{\mathbf{Y}}(\mathbf{Y} \in B) = \int_{\mathbf{y} \in B} \dots \int f_{\mathbf{Y}}(y_1, \dots, y_k) dy_1 \dots dy_k ,$$

і якщо здійснити заміну змінних  $y_i = g_i(\mathbf{x})$ ,  $i = 1, \dots, k$ , то відразу дістанемо (10.1.3).  $\square$

Якобіан у твердженні Теореми 10.1.1 виконує роль «похідної оберненої функції».

Для застосування теореми обов'язково потрібно перевіряти виконання зазначених умов, а також правильно визначати область  $T$ , яка є носієм відповідного вектора  $\mathbf{Y}$ .

**Приклад 10.1.2.** Нехай випадкові величини  $X_1, X_2 \sim U((\alpha; \beta])$ ,  $X_1 \perp\!\!\!\perp X_2$ . Підрахуймо щільність розподілу величини  $S = X_1 + X_2$ .

Для використання Теореми 10.1.1 потрібно, щоб результатом перетворення цього двовимірного вектора також був деякий інший двовимірний вектор. Іншими словами, нам потрібно формально розглянути вектор  $\mathbf{Y} = (Y_1 = S, Y_2)^\top$ , у якому першою координатою буде шукана величина  $S$ , а другою координатою буде величина  $Y_2$ , перетворення якої дуже просте.

Наприклад, застосуймо такі функції:

$$y_1 = h_1(x_1, x_2) = x_1 + x_2 , \quad y_2 = h_2(x_1, x_2) = x_2 .$$

Відображення  $(h_1, h_2)$  взаємно однозначне:

$$x_1 = g_1(y_1, y_2) = y_1 - y_2 , \quad x_2 = g_2(y_1, y_2) = y_2 ,$$

а відповідний якобіан дорівнює

$$J = \begin{vmatrix} g_{11}(y_1, y_2) & g_{12}(y_1, y_2) \\ g_{21}(y_1, y_2) & g_{22}(y_1, y_2) \end{vmatrix} = \begin{vmatrix} \frac{\partial}{\partial y_1} g_1(y_1, y_2) & \frac{\partial}{\partial y_2} g_1(y_1, y_2) \\ \frac{\partial}{\partial y_1} g_2(y_1, y_2) & \frac{\partial}{\partial y_2} g_2(y_1, y_2) \end{vmatrix} = \begin{vmatrix} 1 & -1 \\ 0 & 1 \end{vmatrix} = 1 .$$

Носієм вектора  $(X_1, X_2)^\top$  є множина  $\text{supp}(\mathbf{X}) = (\alpha; \beta] \times (\alpha; \beta]$ . Образом цієї множини буде такий носій вектора  $(Y_1, Y_2)^\top = (X_1 + X_2, X_1)^\top$  (Рис. 10.1.1):

$$T = \left\{ (y_1, y_2)^\top : 2\alpha < y_1 \leq 2\beta, \alpha < y_2 \leq \beta, \alpha < y_1 - y_2 \leq \beta \right\} .$$

Через незалежність двох величин  $X_1$  та  $X_2$  бачимо, що їхня спільна щільність розподілу є добутком відповідних маржинальних щільностей:

$$f_{X_1, X_2}(x_1, x_2) = \frac{1}{\beta - \alpha} \cdot \mathbb{1}\{\alpha < x_1 < \beta\} \cdot \frac{1}{\beta - \alpha} \cdot \mathbb{1}\{\alpha < x_2 < \beta\} .$$

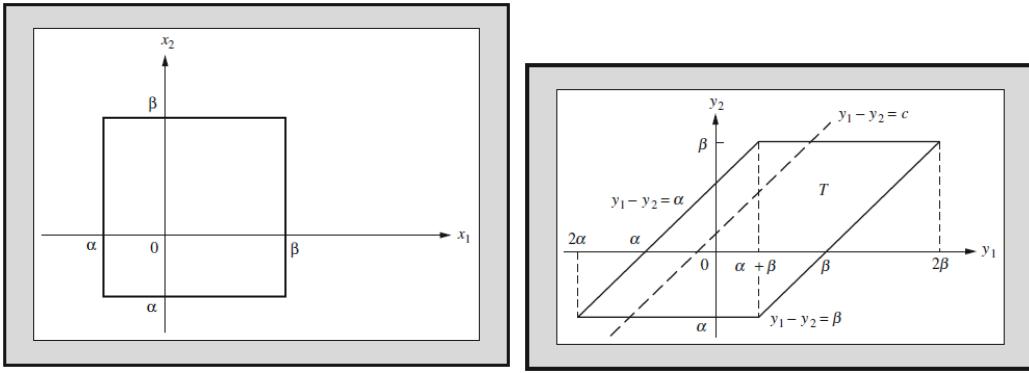


Рис. 10.1.1.: Ілюстрація до Прикладу 10.1.2: (а) носій  $\text{supp}(\mathbf{X})$ ; (б) носій  $T = \text{supp}(\mathbf{Y})$  ([2], Рис. 11.1-2)

Отже за (10.1.3) маємо:

$$\begin{aligned}
 f_{\mathbf{Y}}(y_1, y_2) &= f_{\mathbf{X}}(g_1(y_1, y_2), g_2(y_1, y_2)) \cdot |J| \cdot \mathbb{1} \left\{ (y_1, y_2)^\top \in T \right\} \\
 &= f_{\mathbf{X}}(y_1 - y_2, y_2) \cdot |J| \cdot \mathbb{1} \left\{ (y_1, y_2)^\top \in T \right\} \\
 &= \frac{1}{\beta - \alpha} \cdot \mathbb{1} \{ \alpha < y_1 - y_2 < \beta \} \cdot \frac{1}{\beta - \alpha} \cdot \mathbb{1} \{ \alpha < y_2 < \beta \} \cdot 1 \cdot \mathbb{1} \left\{ (y_1, y_2)^\top \in T \right\} \\
 &= \frac{1}{(\beta - \alpha)^2} \cdot \mathbb{1} \left\{ (y_1, y_2)^\top \in T \right\} .
 \end{aligned}$$

Тоді за (9.2.9) маємо:

$$\begin{aligned}
 f_S(y_1) &= \int_{-\infty}^{\infty} \frac{1}{(\beta - \alpha)^2} \cdot \mathbb{1} \left\{ (y_1, y_2)^\top \in T \right\} dy_2 \\
 &= \begin{cases} \int_{\alpha}^{y_1 - \alpha} \frac{1}{(\beta - \alpha)^2} dy_2, & y_1 \in (2\alpha; \alpha + \beta] \\ \int_{y_1 - \beta}^{\beta} \frac{1}{(\beta - \alpha)^2} dy_2, & y_1 \in (\alpha + \beta; 2\beta] \\ 0, & y_1 \notin (2\alpha; 2\beta] \end{cases} \\
 &= \begin{cases} \frac{y_1 - 2\alpha}{(\beta - \alpha)^2}, & y_1 \in (2\alpha; \alpha + \beta] \\ \frac{2\beta - y_1}{(\beta - \alpha)^2}, & y_1 \in (\alpha + \beta; 2\beta] \\ 0, & y_1 \notin (2\alpha; 2\beta] \end{cases} .
 \end{aligned}$$

Такий метод підрахунку щільностей сум випадкових змінних є доволі загальним і працює для будь-яких змінних, у тому числі залежних, на відміну від методу, розглянутого в Розд. 9.5.  $\square$

**Приклад 10.1.3.** Нехай  $\mathbf{X} = (X_1, X_2)^\top$  — випадковий вектор зі спільною щільністю  $f_{\mathbf{X}}$ . Підрахуймо спільну щільність для вектора  $\mathbf{Y} = (Y_1, Y_2)^\top = (X_1 + X_2, X_1 - X_2)^\top$ .

Використаймо Теорему 10.1.1 з такими функціями:

$$y_1 = h_1(x_1, x_2) = x_1 + x_2, \quad y_2 = h_2(x_1, x_2) = x_1 - x_2.$$

Відображення  $(h_1, h_2)$  взаємно однозначне:

$$x_1 = g_1(y_1, y_2) = \frac{y_1 + y_2}{2}, \quad x_2 = g_2(y_1, y_2) = \frac{y_1 - y_2}{2},$$

а відповідний якобіан дорівнює

$$J = \begin{vmatrix} g_{11}(y_1, y_2) & g_{12}(y_1, y_2) \\ g_{21}(y_1, y_2) & g_{22}(y_1, y_2) \end{vmatrix} = \begin{vmatrix} \frac{\partial}{\partial y_1} g_1(y_1, y_2) & \frac{\partial}{\partial y_2} g_1(y_1, y_2) \\ \frac{\partial}{\partial y_1} g_2(y_1, y_2) & \frac{\partial}{\partial y_2} g_2(y_1, y_2) \end{vmatrix} = \begin{vmatrix} \frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & -\frac{1}{2} \end{vmatrix} = -\frac{1}{2}.$$

Отже за (10.1.3) маємо:

$$\begin{aligned} f_{\mathbf{Y}}(y_1, y_2) &= f_{\mathbf{X}}(g_1(y_1, y_2), g_2(y_1, y_2)) \cdot |J| \cdot \mathbb{1} \left\{ (y_1, y_2)^\top \in T \right\} \\ &= f_{\mathbf{X}} \left( \frac{y_1 + y_2}{2}, \frac{y_1 - y_2}{2} \right) \cdot \left| -\frac{1}{2} \right| \cdot \mathbb{1} \left\{ (y_1, y_2)^\top \in T \right\} \\ &= \frac{1}{2} f_{\mathbf{X}} \left( \frac{y_1 + y_2}{2}, \frac{y_1 - y_2}{2} \right) \cdot \mathbb{1} \left\{ (y_1, y_2)^\top \in T \right\}. \end{aligned}$$

Наприклад, нехай  $X_1, X_2 \sim N(0, 1)$ ,  $X_1 \perp\!\!\!\perp X_2$ . Тоді носієм вектора  $(X_1, X_2)^\top \in \text{supp}((X_1, X_2)^\top) = \mathbb{R}^2$ . Його образом під дією відображення також є  $T = \mathbb{R}^2$ . Отже з урахуванням незалежності маємо

$$\begin{aligned} f_{\mathbf{Y}}(y_1, y_2) &= \frac{1}{2} f_{\mathbf{X}} \left( \frac{y_1 + y_2}{2}, \frac{y_1 - y_2}{2} \right) \cdot \mathbb{1} \left\{ (y_1, y_2)^\top \in \mathbb{R}^2 \right\} \\ &= \frac{1}{2} f_{X_1} \left( \frac{y_1 + y_2}{2} \right) \cdot f_{X_2} \left( \frac{y_1 - y_2}{2} \right) \\ &= \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2} \cdot \left( \frac{y_1 + y_2}{2} \right)^2} \cdot \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2} \cdot \left( \frac{y_1 - y_2}{2} \right)^2} \\ &= \frac{1}{4\pi} e^{-\frac{(y_1 + y_2)^2}{8} - \frac{(y_1 - y_2)^2}{8}} \\ &= \frac{1}{4\pi} e^{-\frac{(y_1^2 + y_2^2)}{4}} \\ &= \frac{1}{\sqrt{2\pi} \cdot 2} e^{-\frac{(y_1 - 0)^2}{2 \cdot 2}} \cdot \frac{1}{\sqrt{2\pi} \cdot 2} e^{-\frac{(y_2 - 0)^2}{2 \cdot 2}}, \end{aligned}$$

тобто відповідні величини є нормальними:  $X_1 + X_2 \sim N(0, 2)$ ,  $X_1 - X_2 \sim N(0, 2)$ , і понад те  $X_1 + X_2 \perp\!\!\!\perp X_1 - X_2$ . Більше того, можна показати, що ця властивість (сума й різниця двох незалежних величин незалежна) справедлива *виключно* для нормальних випадкових величин.  $\square$

**Зауваження 10.1.4.** Часто Теорему 10.1.1 використовують для пошуку щільності розподілу деякої функції від випадкових величин, а щільності інших координат мало цікаві. Так, у Прикладі 10.1.3 нас цікавив розподіл як суми, так і різниці двох величин, а в Прикладі 10.1.2 нас

цікавив розподіл *тільки* суми, тому другу координату ми визначили з міркувань спрощення підрахунку оберненого відображення та якобіану.  $\square$

У випадках, коли функції  $h_i, i = 1, \dots, k$ , не є взаємно однозначними, інколи можливо розбити носій вектора  $\mathbf{X}$  на неперетинанні підмножини так, що на кожній із них ці функції будуть взаємно однозначними.

**Твердження 10.1.5.** Нехай  $\mathbf{X} = (X_1, \dots, X_k)^\top$  — випадковий вектор зі (спільною) щільністю  $f_{\mathbf{X}}$ , неперервною на  $\text{supp}(\mathbf{X})$ . Нехай  $\mathbf{Y} = (Y_1, \dots, Y_k)^\top = (h_1(\mathbf{X}), \dots, h_k(\mathbf{X}))^\top$ . Нехай існує розбиття  $A_0, A_1, \dots, A_n$  носія  $\text{supp}(\mathbf{X})$  таке, що  $\mathbb{P}_{\mathbf{X}}(\mathbf{X} \in A_0) = 0$ .

Нехай на  $A_1, \dots, A_n$  визначено функції  $h_{11}, \dots, h_{1n}, \dots, h_{k1}, \dots, h_{kn}$  відповідно, такі що:

- $h_{ij}(\mathbf{x}) = h_i(\mathbf{x})$  для  $\mathbf{x} \in A_j$ ,  $i = 1, \dots, k$ ,  $j = 1, \dots, n$ ;
- множина  $B = \{\mathbf{y} : (h_{1j}(x_1), \dots, h_{kj}(x_k))^\top = \mathbf{y}, \mathbf{x} \in A_j\}$  однаакова для всіх  $i = 1, \dots, k$ ;
- що  $(h_{1j}, \dots, h_{kj})$  задає взаємно однозначно відображення між  $A_j$  та  $B$ ,  $j = 1, \dots, n$ ;
- частинні похідні  $g_{imj}(\mathbf{y}) = \frac{\partial}{\partial y_m} h_{ij}^{-1}(\mathbf{y}) \equiv \frac{\partial}{\partial y_m} g_{ij}(\mathbf{y})$ ,  $i, m = 1, \dots, k$ ,  $j = 1, \dots, n$ , існують і є неперервними на  $B$ ;
- якобіан

$$J_j = \begin{vmatrix} g_{11j}(\mathbf{y}) & \dots & g_{1kj}(\mathbf{y}) \\ \vdots & \ddots & \vdots \\ g_{k1j}(\mathbf{y}) & \dots & g_{kkj}(\mathbf{y}) \end{vmatrix}$$

не дорівнює нулю на  $B$ ,  $j = 1, \dots, n$ .

Тоді

$$f_{\mathbf{Y}}(\mathbf{y}) = \mathbb{1}\{\mathbf{y} \in B\} \cdot \sum_{j=1}^n f_{\mathbf{X}}(g_{1j}(\mathbf{y}), \dots, g_{kj}(\mathbf{y})) \cdot |J_j| . \quad (10.1.4)$$

**Приклад 10.1.6.** Нехай  $X_1, X_2 \sim N(0, 1)$ ,  $X_1 \perp\!\!\!\perp X_2$ . Нас цікавить розподіл їх відношення,  $X_1/X_2$ . Можемо покласти другою координатою одну з двох величин, проте скористаймося слідчим моментом і одночасно знайдімо розподіл модуля стандартної нормальні величини:

$$h_1(x_1, x_2) = \frac{x_1}{x_2}, \quad h_2(x_1, x_2) = |x_2| .$$

Розподіл модуля стандартної нормальної величини має назву *напівнормального* (half-normal).

Як можна побачити, таке відображення не є взаємно однозначним на  $\mathbb{R}^2$ , а в точці  $x_2 = 0$  маемо ділення на нуль. Тому утворімо розбиття  $\mathbb{R}^2$  на три множини:

$$A_0 = \{(x_1, 0)^\top \in \mathbb{R}^2\} , \quad A_1 = \{(x_1, x_2)^\top \in \mathbb{R}^2 : x_2 < 0\} , \quad A_2 = \{(x_1, x_2)^\top \in \mathbb{R}^2 : x_2 > 0\} ,$$

де  $\mathbb{P}_{\mathbf{X}}(A_0) = 0$ , оскільки це лінія у двовимірному просторі.

Тоді на множині  $A_1$  відображення  $(h_1, h_2)$  є взаємно однозначним, і

$$g_{11}(y_1, y_2) = -y_1 y_2 , \quad g_{21}(y_1, y_2) = -y_2 .$$

Відповідний якобіан дорівнює

$$J_1 = \begin{vmatrix} g_{111}(y_1, y_2) & g_{121}(y_1, y_2) \\ g_{211}(y_1, y_2) & g_{221}(y_1, y_2) \end{vmatrix} = \begin{vmatrix} \frac{\partial}{\partial y_1} g_{11}(y_1, y_2) & \frac{\partial}{\partial y_2} g_{11}(y_1, y_2) \\ \frac{\partial}{\partial y_1} g_{21}(y_1, y_2) & \frac{\partial}{\partial y_2} g_{21}(y_1, y_2) \end{vmatrix} = \begin{vmatrix} -y_2 & -y_1 \\ 0 & -1 \end{vmatrix} = y_2 < 0.$$

Образом  $\mathbb{R}^2$  під дією відображення  $(h_1, h_2)$  на множині  $A_1$  буде  $B = \{(y_1, y_2)^\top \in \mathbb{R}^2 : y_2 > 0\}$ .

Для множини  $A_2$  маємо  $g_{12}(y_1, y_2) = y_1 y_2$ ,  $g_{22}(y_1, y_2) = y_2$ ,  $J_2 = y_2 > 0$ . Образом  $\mathbb{R}^2$  також буде та сама множина  $B$ .

Отже, використовуючи (10.1.4) та враховуючи незалежність  $X_1 \perp\!\!\!\perp X_2$ , маємо:

$$\begin{aligned} f_Y(y_1, y_2) &= \mathbb{1}\{(y_1, y_2)^\top \in B\} \cdot \left( f_{\mathbf{X}}(-y_1 y_2, -y_2) \cdot |y_2| + f_{\mathbf{X}}(y_1 y_2, y_2) \cdot |y_2| \right) \\ &= \mathbb{1}\{y_2 > 0\} \cdot \left( f_{X_1}(-y_1 y_2) \cdot f_{X_2}(-y_2) \cdot |y_2| + f_{X_1}(y_1 y_2) \cdot f_{X_2}(y_2) \cdot |y_2| \right) \\ &= \mathbb{1}\{y_2 > 0\} \left( \frac{1}{\sqrt{2\pi}} e^{-\frac{(-y_1 y_2)^2}{2}} \cdot \frac{1}{\sqrt{2\pi}} e^{-\frac{(-y_2)^2}{2}} \cdot |y_2| + \frac{1}{\sqrt{2\pi}} e^{-\frac{(y_1 y_2)^2}{2}} \cdot \frac{1}{\sqrt{2\pi}} e^{-\frac{y_2^2}{2}} \cdot |y_2| \right) \\ &= \mathbb{1}\{y_2 > 0\} \cdot \frac{y_2}{\pi} e^{-\frac{(y_1^2+1)y_2^2}{2}}. \end{aligned}$$

Знайдімо обидві маржинальні щільності. Для  $Y_1 = X_1/X_2$  маємо:

$$\begin{aligned} f_{Y_1}(y_1) &= \int_0^\infty \frac{y_2}{\pi} e^{-\frac{(y_1^2+1)y_2^2}{2}} dy_2 \\ &= \int_0^\infty \frac{1}{2\pi} e^{-\frac{(y_1^2+1)}{2}t} dt \\ &= \frac{1}{2\pi} \cdot \frac{2}{y_1^2 + 1} \\ &= \frac{1}{\pi(y_1^2 + 1)}, \end{aligned}$$

що є нічим іншим, як щільністю стандартного розподілу Коші, який ми зустрічали в Прикладі 6.4.8.

Для  $Y_2 = |X_2|$  маємо:

$$f_{Y_2}(y_2) = \int_{-\infty}^\infty \frac{y_2}{\pi} e^{-\frac{(y_1^2+1)y_2^2}{2}} \cdot \mathbb{1}\{y_2 > 0\} dy_1.$$

Для підрахунку цього інтегралу використаймо стандартний підхід із виділенням щільності відомого розподілу, інтеграл якої заздалегідь дорівнює 1. У нашому випадку, очевидно, потрібно виділити щільність деякого нормальногорозподілу.

Для цього потрібно, щоб показник експоненти мав вид  $-\frac{(x-a)^2}{2b^2}$  для деяких  $a, b$ . Ми маємо:

$$-\frac{(y_1^2+1)y_2^2}{2} = -\frac{y_2^2}{2} + \left( -\frac{(y_1 - 0)^2}{2 \cdot \frac{1}{y_2^2}} \right).$$

Отже ми бачимо, що можемо утворити щільність нормальногорозподілу з параметрами  $\mu = 0$  і

$$\sigma^2 = \frac{1}{y_2^2}.$$

Щоб додовнити щільність, потрібно помножити її подiлити на вiдповiдний нормуючий коефiцiєнт:

$$e^{-\frac{y_2^2}{2}} \cdot e^{-\frac{y_1^2}{2 \cdot \frac{1}{y_2^2}}} = e^{-\frac{y_2^2}{2}} \cdot \frac{\sqrt{2\pi \cdot \frac{1}{y_2^2}} e^{-\frac{(y_1-0)^2}{2 \cdot \frac{1}{y_2^2}}}}{\sqrt{2\pi \cdot \frac{1}{y_2^2}}} = \sqrt{2\pi \cdot \frac{1}{y_2^2}} \cdot e^{-\frac{y_2^2}{2}} f_W(y_1),$$

де  $W \sim N\left(0, \frac{1}{y_2^2}\right)$ .

Нарештi,

$$\begin{aligned} \int_{-\infty}^{\infty} \frac{y_2}{\pi} e^{-\frac{(y_1+1)y_2^2}{2}} \cdot \mathbb{1}\{y_2 > 0\} dy_1 &= \int_{-\infty}^{\infty} \frac{y_2}{\pi} \sqrt{2\pi \cdot \frac{1}{y_2^2}} e^{-\frac{y_2^2}{2}} \cdot \mathbb{1}\{y_2 > 0\} \cdot f_W(y_1) dy_1 \\ &= \frac{y_2}{\pi} \sqrt{2\pi \cdot \frac{1}{y_2^2}} e^{-\frac{y_2^2}{2}} \cdot \mathbb{1}\{y_2 > 0\} \\ &= \sqrt{\frac{2}{\pi}} e^{-\frac{y_2^2}{2}} \cdot \mathbb{1}\{y_2 > 0\}, \end{aligned}$$

тобто

$$f_{Y_2}(y_2) = 2\phi(y_2) \cdot \mathbb{1}\{y_2 > 0\}.$$

□

## 10.2. Гамма-розподiл

Розгляньмо важливий розподiл, який є узагальненням експоненцiйного. У Розд. 8.4 ми з'ясували, що випадкова величина з експоненцiйним розподiлом  $X \sim \text{Exp}(\lambda)$  вiдповiдає часу очiкування до настання певної подiї в процесi Пуассона з iнтенсивнiстю  $\lambda$ . Розгляньмо розподiл, який характеризує час очiкування до настання *декiлькох подiй* (ми про це вiрше згадували в Зауваженнi 8.5.1).

**Визначення 10.2.1.** Випадкова величина  $X$  має *Гамма-розподiл* (Gamma distribution) iз параметрами  $k > 0$ ,  $\lambda > 0$ ,  $X \sim \text{Gamma}(k, \lambda)$ , якщо її щільнiсть дорiвнює

$$f(x) = \frac{1}{\Gamma(k)} x^{k-1} e^{-\lambda x} \lambda^k \cdot \mathbb{1}\{x > 0\}, \quad (10.2.1)$$

де  $\Gamma(k)$  — гамма-функцiя:

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

□

Нескладно бачити, що щільнiсть (10.2.1) iнтегрується до 1, оскiльки, зробивши замiну  $\lambda x = y$ , маємо:

$$\int_0^{\infty} \frac{1}{\Gamma(k)} x^{k-1} e^{-\lambda x} \lambda^k dx = \frac{1}{\Gamma(k)} \int_0^{\infty} y^{k-1} \lambda^{1-k} e^{-y} \lambda^k \lambda dy = \frac{1}{\Gamma(k)} \int_0^{\infty} y^{k-1} e^{-y} dy = \frac{1}{\Gamma(k)} \cdot \Gamma(k) = 1.$$

Функція розподілу для Гамма-розподілу в загальному випадку не має простого аналітичного запису.

Графіки щільностей Гамма-розподілу із різними значеннями параметрів наведено на Рис. 10.2.1. У вічі впадають дві загальні тенденції:

- що більше значення параметра  $k$ , то симетричнішим є розподіл;
- що більше значення параметра  $\lambda$ , то концентрованішим є розподіл.

Нехай  $X \sim \text{Gamma}(k, 1)$ . Згідно з (7.5.2), для величини  $Y = \frac{X}{\lambda}$ ,  $\lambda > 0$ , маємо:

$$f_Y(y) = f_X(\lambda y) \cdot \lambda = \frac{1}{\Gamma(k)} (\lambda y)^{k-1} e^{-\lambda y} \cdot \mathbb{1}\{\lambda y > 0\} \cdot \lambda = \frac{1}{\Gamma(k)} y^{k-1} e^{-\lambda y} \lambda^k \cdot \mathbb{1}\{y > 0\} ,$$

тобто  $Y \sim \text{Gamma}(k, \lambda)$ . Іншими словами, Гамма-розподіли утворюють сім'ю відносно масштабування (але не відносно зсуву).

Це спостереження спрощує виведення формул для сподівання та дисперсії Гамма-розподілу.

**Твердження 10.2.2.** Якщо  $X \sim \text{Gamma}(k, \lambda)$ , то  $\mathbb{E}[X] = \frac{k}{\lambda}$ ,  $\text{Var}(X) = \frac{k}{\lambda^2}$ .

*Доведення.* Розглянемо спочатку  $Y \sim \text{Gamma}(k, 1)$ .

Як відомо з курсу «Математичний аналіз»,  $\Gamma(k+1) = k\Gamma(k)$ . Тоді за визначенням сподівання маємо:

$$\begin{aligned} \mathbb{E}[Y] &= \int_0^\infty y \cdot \frac{1}{\Gamma(k)} y^{k-1} e^{-y} dy \\ &= k \int_0^\infty \frac{1}{\Gamma(k+1)} y^k e^{-y} dy \\ &= k , \end{aligned}$$

оскільки підінтегральний вираз є щільністю розподілу  $\text{Gamma}(k+1, 1)$ , а тому інтеграл є 1.

Аналогічно, оскільки  $\Gamma(k+2) = (k+1)\Gamma(k+1) = k(k+1)\Gamma(k)$ , відповідні міркування дають  $\mathbb{E}[Y^2] = k(k+1)$ .

Отже

$$\text{Var}(Y) = \mathbb{E}[Y^2] - (\mathbb{E}[Y])^2 = k(k+1) - k^2 = k .$$

Тепер розглянемо  $\frac{Y}{\lambda} = X \sim \text{Gamma}(k, \lambda)$ . За властивостями сподівань та дисперсії відразу маємо шуканий результат:

$$\begin{aligned} \mathbb{E}[X] &= \mathbb{E}\left[\frac{Y}{\lambda}\right] = \frac{k}{\lambda} , \\ \text{Var}(X) &= \text{Var}\left(\frac{Y}{\lambda}\right) = \frac{k}{\lambda^2} . \end{aligned}$$

□

Як можна помітити, розподіл  $\text{Gamma}(1, \lambda)$  тотожний експоненціальному розподілу з параметром  $\lambda$ . Так, якщо  $X \sim \text{Gamma}(1, \lambda)$ , а  $Y \sim \text{Exp}(\lambda)$ , то

$$f_X(x) = \frac{1}{\Gamma(1)} x^{1-1} e^{-\lambda x} \lambda^1 \cdot \mathbb{1}\{x > 0\} = \lambda e^{-\lambda x} \cdot \mathbb{1}\{x > 0\} = f_Y(x) .$$

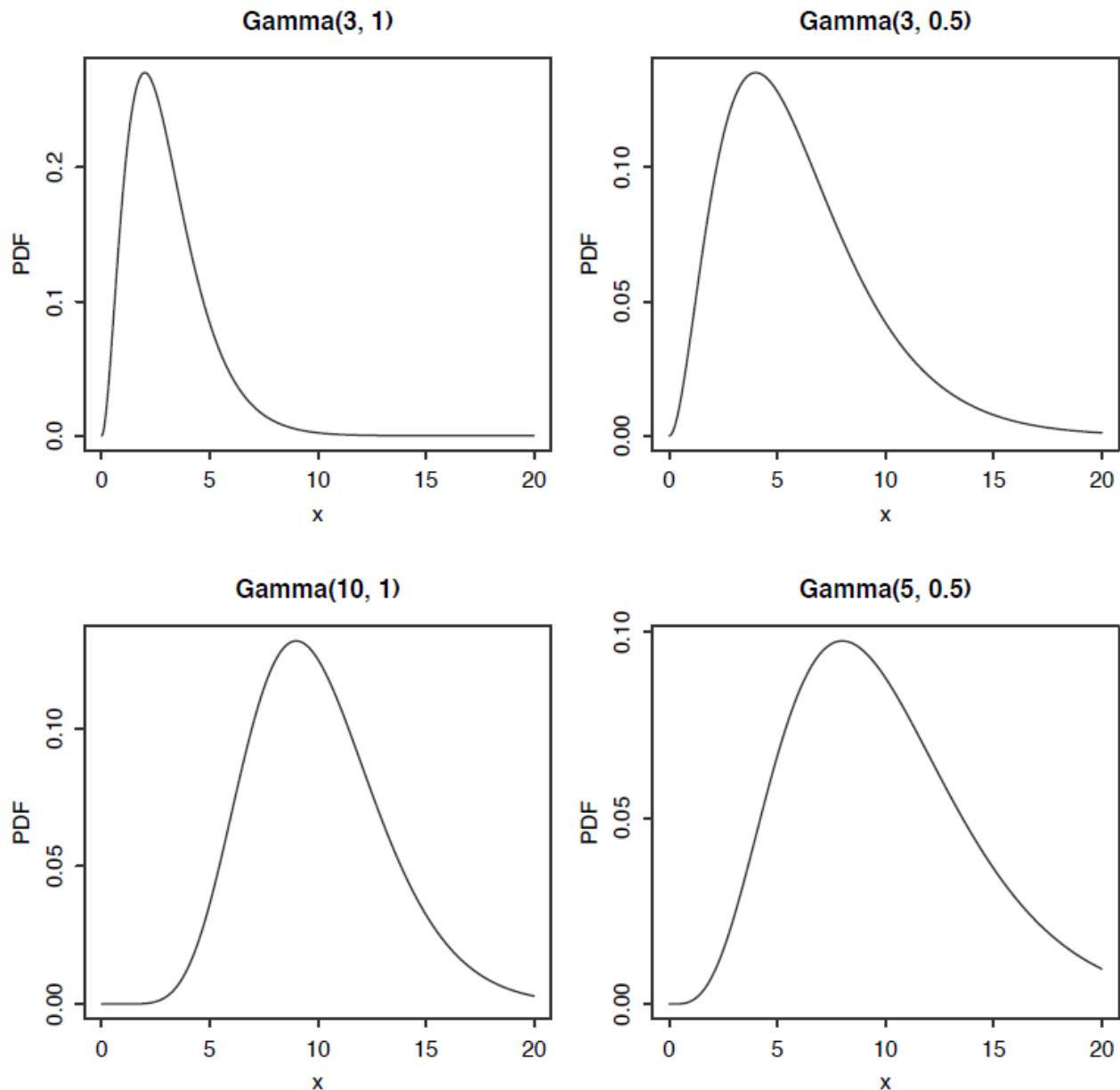


Рис. 10.2.1.: Щільності Гамма-розподiлу з рiзними значеннями параметрiв  $k$  i  $\lambda$  ([1], Рис. 8.7)

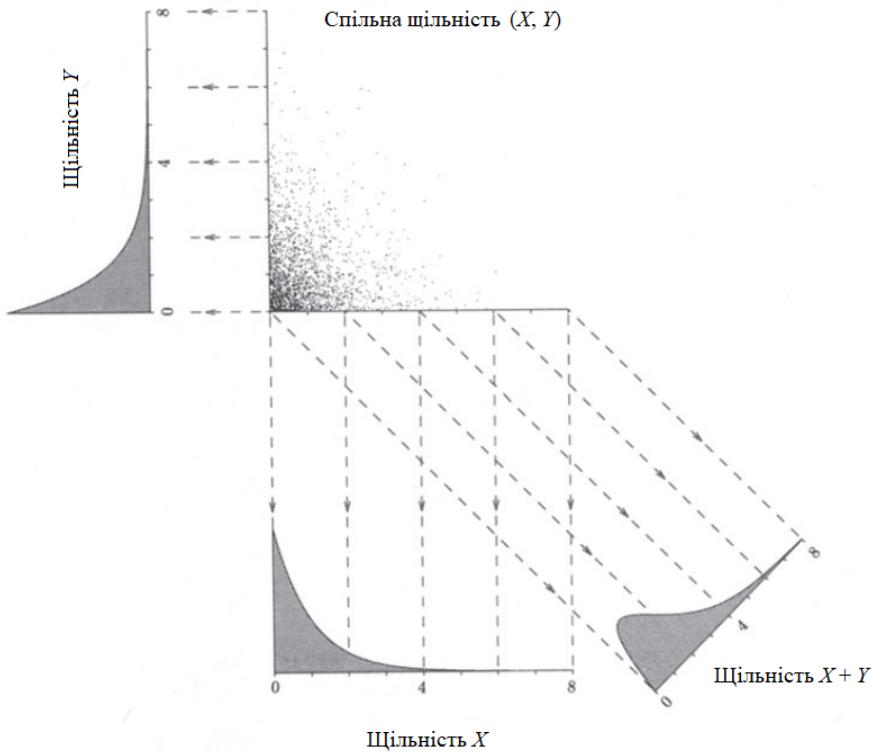


Рис. 10.2.2.: Набір випадкових точок, згенерованих відповідно до спільної щільності розподілу  $X, Y \sim \text{Exp}(1) = \text{Gamma}(1, 1)$ ,  $X \perp\!\!\!\perp Y$ , а також маржинальні щільності та щільність суми  $X + Y \sim \text{Gamma}(2, 1)$  [9, с. 375]

Додатковий параметр  $k$  Гамма-розподілу надає йому більше функціональної гнучкості. Це дає змогу застосовувати його для моделювання різного роду явищ на кшталт дощових опадів, витрат страхової компанії чи генерації вітрової енергетики. Гамма-розподіл відповідає тільки невід'ємним значенням, а функціональна гнучкість дає змогу моделювати як сильно сконцентровані розподіли на кшталт експоненційного, так і симетричні та схожі на нормальні.

Більше того, сума незалежних експоненційних розподілів має Гамма-розподіл. Це випливає з такого загального твердження.

**Твердження 10.2.3.** Нехай  $X_i \sim \text{Gamma}(k_i, \lambda)$ ,  $i = 1, \dots, n$ . Якщо вони всі незалежні, то розподіл суми  $S = \sum_{i=1}^n X_i$  такий:  $S \sim \text{Gamma}(\sum_{i=1}^n k_i, \lambda)$ .

Доведення цього твердження можна знайти в Розд. 10.5.

Ілюстрацію відповідних міркувань для випадку  $X_1, X_2 \sim \text{Exp}(1)$ ,  $X_1 \perp\!\!\!\perp X_2$  наведено на Рис. 10.2.2.

Відповідний результат дає змогу інтерпретувати випадкову величину  $X \sim \text{Gamma}(k, \lambda)$  як час очікування настання  $k$  подій у деякому процесі Пуассона з інтенсивністю  $\lambda$ . На Рис. 10.2.3 схематично зображено такий процес. Як було з'ясовано в Розд. 8.5,  $X_1, X_2, \dots \sim \text{Exp}(\lambda)$ . А відповідно до вищепередбачених міркувань,  $T_k \sim \text{Gamma}(k, \lambda)$ ,  $k \geq 1$ . При цьому варто зазначити, що на відміну від  $X_k$ ,  $k \geq 1$ , величини  $T_k$  не є незалежними, оскільки за побудовою  $T_1 < T_2 < \dots$

Тепер стає цілком очевидним, чому сума Гамма-величин з параметрами  $k_1, \dots, k_n$  є Гамма-величиною з параметром  $k_1 + \dots + k_n$ . Кожна з них відповідає часу очікування настання певного

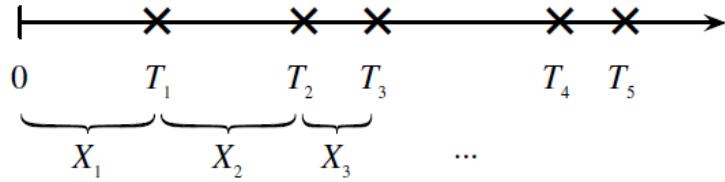


Рис. 10.2.3.: Схематичне зображення процесу Пуассона [1], Рис. 8.8

числа подій у процесі Пуассона. Відповідна, їх сума — це просто загальний час очікування настання  $k_1 + \dots + k_n$  подій.

До речі, можна також стверджувати, що як експоненційний розподіл є неперервним аналогом розподілу геометричного, так і Гамма-розподіл є неперервним аналогом розподілу від'ємного біномного.

Понад те, позначивши через  $N_t$  число подій, які сталися за певний період  $t$ ,  $N_t \sim \text{Pois}(\lambda t)$ , маємо залежність:

$$\mathbb{P}_{T_r}(T_r > t) = \mathbb{P}_{N_t}(N_t < r) , \quad r \in \mathbb{N} , \quad t \geq 0 .$$

Це цікавий результат, адже дає змогу обчислити ймовірності за участю  $T_r$  без потреби проводити чисельного інтегрування. Оскільки функція Гамма-розподілу в загальному випадку не має аналітичного виразу, це суттєво спрощує обчислення.

**Приклад 10.2.4.** Нехай деяка компонента має очікувану тривалість роботи з експоненційним розподілом:  $X_1 \sim \text{Exp}\left(\frac{1}{24}\right)$  (інтенсивність виходів із ладу — один на добу). Нехай після виходу з ладу цієї компоненти її міняють на іншу аналогічну. Підрахуймо ймовірність, що така система пропрацює щонайменше 3 доби.

Позначивши тривалість роботи другої компоненти через  $X_2 \sim \text{Exp}\left(\frac{1}{24}\right)$ , дістанемо, що

$$X_1 + X_2 \sim \text{Gamma}\left(2, \frac{1}{24}\right) ,$$

а відтак

$$\mathbb{P}_{X_1+X_2}(X_1 + X_2 > 3 \cdot 24) = \mathbb{P}_{N_{72}}(N_{72} < 2) = e^{-\frac{72}{24}} \left( \left(\frac{72}{24}\right)^0 + \left(\frac{72}{24}\right)^1 \right) \approx 0.199 .$$

□

**Приклад 10.2.5** (Розподіл  $\chi_r^2$ ). Із Прикладу 7.4.1 відомо, що квадрат стандартної нормальної величини  $Z^2$  має щільність розподілу

$$f_{Z^2}(x) = \frac{1}{\sqrt{2\pi x}} e^{-x/2} \cdot \mathbb{1}\{x > 0\} , \quad (10.2.2)$$

яка, як було зазначено, відповідає розподілу  $\chi^2$  з 1 ступенем свободи:  $Z^2 \sim \chi_1^2$ .

У загальному випадку, *сума*  $X = \sum_{i=1}^r Z_i^2$ ,  $Z_i \sim N(0, 1)$ ,  $i = 1, \dots, r$ , де всі величини незалежні, має *розподіл*  $\chi^2$  з  $r$  *ступенями свободи*,  $X \sim \chi_r^2$ .

Щоб знайти його щільність розподілу, потрібно помітити, що щільність (10.2.2) є нічим іншим, як щільністю Гамма-розподілу з параметрами  $k = \frac{1}{2}$  і  $\lambda = \frac{1}{2}$ :

$$\frac{1}{\sqrt{2\pi x}} e^{-x/2} \cdot \mathbb{1}\{x > 0\} = \frac{1}{\Gamma(\frac{1}{2})} x^{\frac{1}{2}-1} e^{-x/2} \left(\frac{1}{2}\right)^{\frac{1}{2}} \cdot \mathbb{1}\{x > 0\} ,$$

оскільки, як відомо з курсу «Математичний аналіз»,  $\Gamma(\frac{1}{2}) = \sqrt{\pi}$ .

Отже, відповідно до Твердження 9.5.4,

$$\sum_{i=1}^r Z_i^2 \sim \text{Gamma}\left(\frac{r}{2}, \frac{1}{2}\right) ,$$

а тому щільність розподілу  $X \sim \chi_r^2$  має вираз

$$f_X(x) = \frac{1}{2^{\frac{r}{2}} \Gamma\left(\frac{r}{2}\right)} x^{\frac{r}{2}-1} e^{-\frac{x}{2}} \cdot \mathbb{1}\{x > 0\} . \quad (10.2.3)$$

Тобто розподіл  $\chi_r^2$  є нічим іншим, як Гамма-розподілом із параметрами  $k = \frac{r}{2}$  і  $\lambda = \frac{1}{2}$ .

Використовуючи результати Твердження 10.2.2, відразу дістаємо:

$$\mathbb{E}[\chi_r^2] = \frac{r/2}{1/2} = r , \quad \text{Var}(\chi_r^2) = \frac{r/2}{1/2^2} = 2r .$$

Розподіл  $\chi_r^2$  грає визначну роль у статистиці. Зокрема, він тісно пов'язаний із розподілом *випадкової дисперсії*. Також на його основі визначають два ключові розподіли для перевірки *статистичних гіпотез*:

- $t$ -розподіл з  $r$  ступенями свободи (він же розподіл Ст'юдента, Student's  $t$ -distribution). Випадкова величина  $T \sim t_r$ , якщо  $T = \frac{X}{\sqrt{Y/r}}$ , де  $X \sim N(0, 1)$ ,  $Y \sim \chi_r^2$ . До речі, нескладно бачити, що якщо  $r = 1$ , то ми маємо  $T = \frac{X}{\sqrt{Y}} = \frac{X}{|Z|}$ , де  $X, Z \sim N(0, 1)$ . За симетрією можна бачити, що розподіл  $T$  такий самий, як і в  $\frac{X}{Y}$ , тобто розподіл  $t_1$  є нічим іншим, як розподілом Коші;
- $F$ -розподіл з  $r_1$  і  $r_2$  ступенями свободи (він же розподіл Фішера-Снедекора, Fisher-Snedecor distribution). Випадкова величина  $F \sim F_{r_1, r_2}$ , якщо  $F = \frac{X/r_1}{Y/r_2}$ , де  $X \sim \chi_{r_1}^2$ ,  $Y \sim \chi_{r_2}^2$ .

Щільності цих розподілів можна дістати прямим застосуванням Теореми 10.1.1, проте це виходить за межі нашого курсу і є предметом курсу «Математична статистика».  $\square$

## 10.3. Мультиномійний розподіл

Розгляньмо тепер декілька важливих на практиці розподілів випадкових векторів. Першим будемо розглядати розподіл, який є узагальненням біномного. Як ми вивчали раніше, величина з біномним розподілом відповідає загальному числу успіхів у серії незалежних випробувань, де успіх може статися з певною ймовірністю. Узагальненням цієї моделі є серія незалежних випробувань, кожне з яких може завершитися *одним із декількох можливих* варіантів. Такими,

наприклад, є такі «випробування», як підкидання гральних кісточок (6 можливих варіантів), вибір  $r$  куль із набору куль  $m$  різних кольорів ( $m$  можливих варіантів), класифікація осіб за групою крові (4 можливі варіанти) тощо.

**Визначення 10.3.1.** Нехай  $n$  об'єктів можна незалежно одне від одного віднести до однієї з  $k$  категорій. У категорію  $j$  об'єкт можна віднести з імовірністю  $p_j$ ,  $\sum_{j=1}^k p_j = 1$ .

Нехай  $X_j$  — число об'єктів у категорії  $j$ ,  $\sum_{j=1}^k X_j = n$ . Тоді  $\mathbf{X} = (X_1, \dots, X_k)^\top$  має *мультиномій розподіл* (multinomial distribution) із параметрами  $n$  і  $\mathbf{p} = (p_1, \dots, p_k)^\top$ :  $\mathbf{X} \sim \text{Mult}_k(n, \mathbf{p})$ .  $\square$

Знайдімо функцію ймовірності такої випадкової величини.

**Твердження 10.3.2.** Якщо випадковий вектор  $\mathbf{X} \sim \text{Mult}_k(n, \mathbf{p})$ , то його функція ймовірності дорівнює

$$\begin{aligned} \mathbb{P}_{\mathbf{X}}(X_1 = n_1, \dots, X_k = n_k) &= \frac{n!}{n_1! \dots n_k!} p_1^{n_1} \dots p_k^{n_k} \\ &\times \prod_{j=1}^k \mathbb{1}\{n_j \in \mathbb{Z}^+\} \cdot \mathbb{1}\{n_1 + \dots + n_k = n\}, \end{aligned} \quad (10.3.1)$$

де  $\mathbb{Z}^+ = \{0, 1, 2, \dots\}$ .

*Доведення.* Очевидно, що якщо  $\sum_{j=1}^k n_j \neq n$ , то така подія неймовірна, звідси потреба в індикаторній функції.

Якщо ж  $\sum_{j=1}^k n_j = n$  і всі числа — невід'ємні цілі, то за визначенням (4.1.1) маємо:

$$\mathbb{P}_{\mathbf{X}}(X_1 = n_1, \dots, X_k = n_k) = \mathbb{P}(\{\omega : X_1(\omega) = n_1, \dots, X_k(\omega) = n_k\}) = \sum_{\omega: X_1(\omega) = n_1, \dots, X_k(\omega) = n_k} \mathbb{P}(\{\omega\}),$$

де  $\Omega = \{\omega \in \{1, \dots, k\}^n\}$ , тобто кожна елементарна подія — вектор із індексів категорій. Кількість усіх елементарних подій, які відповідають конкретному набору  $n_1, \dots, n_k$ , еквівалентна кількості варіантів розміщення  $n$  елементів по  $k$  категоріях. Загальне число можливих перестановок —  $n!$ , до цього треба врахувати, що перестановки в рамках окремих категорій не впливають на загальне розміщення, а отже загальне число таких варіантів становить  $\frac{n!}{n_1! \dots n_k!}$ .

Позначмо деякий такий варіант через  $\omega_{n_1, \dots, n_k}^i$ ,  $i = 1, \dots, \frac{n!}{n_1! \dots n_k!}$ . Оскільки всі випробування незалежні, маємо  $\mathbb{P}(\{\omega_{n_1, \dots, n_k}^i\}) = p_1^{n_1} \dots p_k^{n_k}$ .

Отже

$$\begin{aligned} \mathbb{P}_{\mathbf{X}}(X_1 = n_1, \dots, X_k = n_k) &= \sum_{\omega: X_1(\omega) = n_1, \dots, X_k(\omega) = n_k} \mathbb{P}(\{\omega\}) = \sum_{i=1}^{\frac{n!}{n_1! \dots n_k!}} \mathbb{P}(\{\omega_{n_1, \dots, n_k}^i\}) \\ &= \frac{n!}{n_1! \dots n_k!} p_1^{n_1} \dots p_k^{n_k}. \end{aligned}$$

Залишилось тільки показати, що сума всіх імовірностей дорівнює 1. Цей результат безпосередньо випливає з так званої мультиноміальної теореми:

$$(a_1 + \dots + a_k)^n = \sum_{\substack{(n_1, \dots, n_k): \\ n_1 + \dots + n_k = n}} \frac{n!}{n_1! \dots n_k!} a_1^{n_1} \dots a_k^{n_k}, \quad (10.3.2)$$

яку ми доводити не будемо, хоча це не складно зробити на основі біномної теореми за індукцією.  $\square$

**Зауваження 10.3.3.** Потрібно завжди пам'ятати, що носієм мультиномної величини є множина чисел  $n_1, \dots, n_k$  така, що їх сума дорівнює  $n$ . Зокрема, це означає, що запис (10.3.1) є до певної міри **надлишковим**, адже його можна переписати так:

$$\begin{aligned} & \mathbb{P}_{\mathbf{X}} (X_1 = n_1, \dots, X_k = n - (n_1 + \dots + n_{k-1})) \\ &= \frac{n!}{n_1! \dots (n - (n_1 + \dots + n_{k-1}))!} p_1^{n_1} \dots (1 - (p_1 + \dots + p_{k-1}))^{n - (n_1 + \dots + n_{k-1})} \\ & \times \prod_{j=1}^{k-1} \mathbb{1}\{n_j \in \mathbb{Z}^+\} \cdot \mathbb{1}\{n_1 + \dots + n_{k-1} \leq n\} . \end{aligned}$$

Іншими словами, для опису мультиномного розподілу достатньо  $k-1$  імовірностей  $p_1, \dots, p_{k-1}$ , оскільки  $k$ -а імовірність однозначно залежить від усіх інших. У схожий спосіб кажемо, що величина  $X$  має біномний розподіл,  $X \sim \text{Binom}(n, p)$ :

$$\mathbb{P}_X (X = k) = \binom{n}{k} p^k (1 - p)^{n-k} ,$$

хоча також могли б стверджувати, що вектор  $\mathbf{X} = (X, Y = n - X)^\top$  має мультиномний розподіл,  $\mathbf{X} \sim \text{Mult}_2(n, (p, q)^\top)$ , де  $q = 1 - p$ :

$$\mathbb{P}_{X,Y} (X = k, Y = m) = \frac{n!}{k!m!} p^k q^m = \frac{n!}{k!(n-k)!} p^k (1 - p)^{n-k} = \binom{n}{k} p^k (1 - p)^{n-k} .$$

$\square$

**Приклад 10.3.4.** Нехай правильну гральну кісточку підкидають 10 разів. Імовірність того, що одиниця випала 2 рази, двійка — 1 раз, трійка — 3 рази, четвірка — 1 раз, п'ятірка — 2 рази, шістка — 1 раз, становить

$$\mathbb{P}_{\mathbf{X}} (X_1 = 2, X_2 = 1, X_3 = 3, X_4 = 1, X_5 = 2, X_6 = 1) = \frac{10!}{2!1!3!1!2!1!} \cdot \frac{1}{6^{10}} \approx 0.003 ,$$

оскільки відповідний вектор

$$\mathbf{X} = (X_1, \dots, X_6)^\top \sim \text{Mult}_6 \left( 10, \left( \frac{1}{6}, \dots, \frac{1}{6} \right)^\top \right) .$$

$\square$

Розгляньмо маржинальні розподіли на основі мультиномного. Інтуїтивно повинно бути зрозуміло, що кожна  $X_j$ ,  $j = 1, \dots, k$ , являє собою кількість об'єктів, що потрапили в категорію  $j$ . Кожне таке потрапляння можна інтерпретувати як «успіх», тоді «невдача» — це потрапляння в будь-яку іншу категорію. Відтак  $X_j$  повинна мати біномний розподіл.

З іншого боку, якщо об'єднати декілька категорій в одну, то ми знову дістанемо мультиномний розподіл, оскільки об'єднання категорій можна сприймати як нову категорію.

Відповідні міркування формалізує таке твердження.

**Твердження 10.3.5.** Нехай  $\mathbf{X} \sim \text{Mult}_k(n, \mathbf{p})$ . Розгляньмо величини  $X_1, \dots, X_s$  і  $Y = n - (X_1 + \dots + X_s)$ <sup>1</sup>. Нехай  $q = 1 - (p_1 + \dots + p_s)$ . Тоді  $\mathbf{X}' = (X_1, \dots, X_s, Y)^\top \sim \text{Mult}_{s+1}(n, (p_1, \dots, p_s, q)^\top)$ .

*Доведення.* Відповідний результат випливає безпосередньо з функції ймовірності (10.3.1) та (9.2.9). Нехай  $n_{1:s} = n_1 + \dots + n_s$  і

$$A_s = \{(n_{s+1}, \dots, n_k) : n_{s+1}, \dots, n_k \in \mathbb{Z}^+, n_{s+1} + \dots + n_k = n - n_{1:s}\}.$$

Тоді

$$\begin{aligned} \mathbb{P}_{\mathbf{X}'}(X_1 = n_1, \dots, X_s = n_s, Y = n - n_{1:s}) &= \sum_{A_s} \mathbb{P}_{\mathbf{X}}(X_1 = n_1, \dots, X_k = n_k) \\ &= \sum_{A_s} \frac{n!}{n_1! \dots n_k!} p_1^{n_1} \dots p_k^{n_k} \\ &= \frac{n!}{n_1! \dots n_s! (n - n_{1:s})!} p_1^{n_1} \dots p_s^{n_s} \sum_{A_s} \frac{(n - n_{1:s})!}{n_{s+1}! \dots n_k!} p_{s+1}^{n_{s+1}} \dots p_k^{n_k} \\ &= \frac{n!}{n_1! \dots n_s! (n - n_{1:s})!} p_1^{n_1} \dots p_s^{n_s} \cdot q^{n - n_{1:s}}, \end{aligned}$$

де ми використали мультиномійну теорему (10.3.2) для  $q = 1 - (p_1 + \dots + p_s) = p_{s+1} + \dots + p_k$ .  $\square$

**Приклад 10.3.6.** Повернімося до Прикладу 10.3.4. Підрахуймо ймовірність, що двійка випала на другій, четвертій та шостій кісточках.

У цьому випадку  $Y = 10 - (X_2 + X_4 + X_6)$ ,  $q = 1 - \frac{3}{6} = \frac{1}{2}$ , а відтак  $(X_2, X_4, X_6, Y)^\top \sim \text{Mult}_4(10, (\frac{1}{6}, \frac{1}{6}, \frac{1}{6}, \frac{1}{2})^\top)$ , і тому

$$\mathbb{P}_{X_2, X_4, X_6}(X_2 = X_4 = X_6 = 2) = \frac{10!}{2!2!2!4!} \cdot \frac{1}{6^2} \cdot \frac{1}{6^2} \cdot \frac{1}{6^2} \cdot \frac{1}{2^4} \approx 0.025.$$

$\square$

**Зауваження 10.3.7.** Із Твердження 10.3.5 безпосередньо випливає той факт, що  $X_j \sim \text{Binom}(n, p_j)$ ,  $j = 1, \dots, k$ . Відповідно, випадкова величина  $Y = n - (X_1 + \dots + X_s) = X_{s+1} + \dots + X_k \sim \text{Binom}(n, q)$ . Іншими словами, із Твердження 4.5.9 випливає, що сума маржинальних розподілів також є біномним розподілом:  $\sum_{i=1}^s X_i \sim \text{Binom}(n, \sum_{i=1}^s p_i)$ .

Це дуже цікавий результат, адже в загальному випадку, якщо  $X_1 \sim \text{Binom}(n, p_1)$ ,  $X_2 \sim \text{Binom}(n, p_2)$ , то їх сума *не обов'язково є біномною величиною*. Наприклад, нехай  $\mathbf{X} = (X_1, X_2)^\top$  має розподіл, поданий у Табл. 10.3.1.

Нескладно бачити, що *маржинально*  $X_1, X_2 \sim \text{Binom}(2, \frac{1}{2})$ . Проте розподіл їх суми має такий

<sup>1</sup>Нумерація величин, звісно, не грає жодної ролі через повну симетрію.

Табл. 10.3.1.: Спільний розподіл вектора  $\mathbf{X} = (X_1, X_2)^\top$  для Прикладу 10.3.1

	$X_2 = 0$	$X_2 = 1$	$X_2 = 2$
$X_1 = 0$	$\frac{1}{4}$	0	0
$X_1 = 1$	0	$\frac{1}{2}$	0
$X_1 = 2$	0	0	$\frac{1}{4}$

вид:

$$\mathbb{P}_{X_1+X_2}(X_1 + X_2 = s) = \begin{cases} \frac{1}{4}, & s = 0 \\ 0, & s = 1 \\ \frac{1}{2}, & s = 2 \\ 0, & s = 3 \\ \frac{1}{4}, & s = 4 \\ 0, & \text{інакше} \end{cases},$$

що немає нічого спільного з біномним розподілом. Таким чином, властивість суми біномних величин бути біномною величиною є особливістю мультиноміального розподілу.  $\square$

Відповідне спостереження можна використати для обчислення сподівання та коваріацій мультиноміального розподілу.

**Твердження 10.3.8.** Якщо  $\mathbf{X} \sim \text{Mult}_k(n, \mathbf{p})$ , то

$$\mathbb{E}[\mathbf{X}] = \begin{pmatrix} np_1 \\ \vdots \\ np_k \end{pmatrix}, \quad \text{Cov}(\mathbf{X}) = \begin{pmatrix} np_1(1-p_1) & -np_1p_2 & \dots & -np_1p_k \\ -np_1p_2 & np_2(1-p_2) & \dots & -np_2p_k \\ \vdots & \vdots & \ddots & \vdots \\ -np_kp_1 & -np_kp_2 & \dots & np_k(1-p_k) \end{pmatrix}.$$

**Доведення.** Вираз для сподівання є прямим застосуванням (9.3.1) до того факту, що маржинальні розподіли є біномними.

Коваріацію  $\text{Cov}(X_i, X_j)$ ,  $1 \leq i, j \leq k$ ,  $i \neq j$ , можна обчислити з властивості (vi) Твердження 9.3.5, пригадавши вираз дисперсії біномної величини:

$$\begin{aligned} \text{Cov}(X_i, X_j) &= \frac{1}{2}(\text{Var}(X_i + X_j) - \text{Var}(X_i) - \text{Var}(X_j)) \\ &= \frac{1}{2}(n(p_i + p_j)(1 - (p_i + p_j)) - np_i(1 - p_i) - np_j(1 - p_j)) \\ &= -np_i p_j. \end{aligned}$$

Нарешті, елементи на головній діагоналі матриці  $\text{Cov}(\mathbf{X})$  є дисперсіями відповідних маржинальних біномних розподілів.  $\square$

Варто звернути уваги, що величини  $X_i$  та  $X_j$  мають від'ємну кореляцію, що інтуїтивно зрозуміло: якщо в категорію  $i$  потрапило багато об'єктів, то в категорію  $j$  може потрапити менша кількість об'єктів.

**Приклад 10.3.9.** Групи крові в людській популяції розподілено з (приблизно) такими частотами:  $p_A = 0.40$ ,  $p_B = 0.10$ ,  $p_{AB} = 0.05$ ,  $p_O = 0.45$ . Нехай у донорській операції беруть участь 20 донорів. Позначмо через  $X_A$ ,  $X_B$ ,  $X_{AB}$  і  $X_O$  число осіб із відповідними групами крові. Тоді  $\mathbf{X} = (X_A, X_B, X_{AB}, X_O)^\top \sim \text{Mult}_4(20, (p_A, p_B, p_{AB}, p_O)^\top)$ .

Сподіване число представників кожної групи серед донорів буде

$$\mathbb{E}[\mathbf{X}] = \begin{pmatrix} \mathbb{E}[X_A] \\ \mathbb{E}[X_B] \\ \mathbb{E}[X_{AB}] \\ \mathbb{E}[X_O] \end{pmatrix} = \begin{pmatrix} 20 \cdot 0.40 \\ 20 \cdot 0.10 \\ 20 \cdot 0.05 \\ 20 \cdot 0.45 \end{pmatrix} = \begin{pmatrix} 8 \\ 2 \\ 1 \\ 9 \end{pmatrix},$$

із середньоквадратичними відхиленнями

$$\begin{aligned} \sigma_A &= \sqrt{\text{Var}(X_A)} = \sqrt{20 \cdot 0.40 \cdot 0.60} \approx 2.19, \\ \sigma_B &= \sqrt{\text{Var}(X_B)} = \sqrt{20 \cdot 0.10 \cdot 0.90} \approx 1.34, \\ \sigma_{AB} &= \sqrt{\text{Var}(X_{AB})} = \sqrt{20 \cdot 0.05 \cdot 0.95} \approx 0.97, \\ \sigma_O &= \sqrt{\text{Var}(X_O)} = \sqrt{20 \cdot 0.45 \cdot 0.55} \approx 2.22. \end{aligned}$$

Коефіцієнти кореляції між величиною  $X_A$  та рештою величин такі:

$$\begin{aligned} \rho_{A,B} &= \frac{\text{Cov}(X_A, X_B)}{\sigma_A \sigma_B} \approx \frac{-20 \cdot 0.40 \cdot 0.10}{2.19 \cdot 1.34} \approx -0.272, \\ \rho_{A,AB} &= \frac{\text{Cov}(X_A, X_{AB})}{\sigma_A \sigma_{AB}} \approx \frac{-20 \cdot 0.40 \cdot 0.05}{2.19 \cdot 0.97} \approx -0.187, \\ \rho_{A,O} &= \frac{\text{Cov}(X_A, X_O)}{\sigma_A \sigma_O} \approx \frac{-20 \cdot 0.40 \cdot 0.45}{2.19 \cdot 2.22} \approx -0.739, \end{aligned}$$

тобто кореляція найбільша (за абсолютною значенням) між величинами  $X_A$  та  $X_O$ , а найменша — між  $X_A$  і  $X_{AB}$ , і всі вони від'ємні.  $\square$

## 10.4. Багатовимірний нормальний розподіл

Іншим надзвичайно важливим на практиці розподілом є багатовимірний нормальний розподіл, який узагальнює нормальній розподіл на  $k$ -вимірні простори.

**Визначення 10.4.1.** Випадковий вектор  $\mathbf{X} = (X_1, \dots, X_k)^\top$  має *багатовимірний нормальний розподіл* (multivariate normal distribution),  $\mathbf{X} \sim N(\boldsymbol{\mu}, \Sigma)$ , якщо

$$\mathbf{X} = \mathbf{A} \cdot \mathbf{Z} + \boldsymbol{\mu},$$

де  $\Sigma = \mathbf{A} \mathbf{A}^\top$ ,  $\mathbf{A} \in \mathbb{R}^k \times \mathbb{R}^n$  — матриця така, що  $\Sigma$  невироджена<sup>2</sup>,  $\mathbf{Z} = (Z_1, \dots, Z_n)^\top$ ,  $Z_i \sim N(0, 1)$ , усі  $Z_i$  незалежні,  $i = 1, \dots, n$ ,  $\boldsymbol{\mu} = (\mu_1, \dots, \mu_n)^\top \in \mathbb{R}^n$ .  $\square$

Багатовимірний нормальний розподіл має альтернативне визначення.

<sup>2</sup> Якщо матриця вироджена, то розподіл не матиме щільності. Ми такі випадки не розглядаємо.

**Визначення 10.4.2.** Випадковий вектор  $\mathbf{X} = (X_1, \dots, X_k)^\top \sim N(\boldsymbol{\mu}, \Sigma)$  якщо будь-яка лінійна комбінація

$$Y = \mathbf{c}^\top \mathbf{X} = \sum_{i=1}^k c_i X_i$$

його координат,  $\mathbf{c} \in \mathbb{R}^k$ , є випадковою величиною з нормальним розподілом.  $\square$

**Твердження 10.4.3.** Визначення 10.4.1 та 10.4.2 еквівалентні.

*Доведення.* Цілком очевидно, що з Визначення 10.4.1 випливає Визначення 10.4.2, оскільки для будь-якого  $\mathbf{c} \in \mathbb{R}^k$

$$Y = \mathbf{c}^\top \mathbf{X} = \mathbf{c}^\top \mathbf{A} \mathbf{Z} + \mathbf{c}^\top \boldsymbol{\mu} \equiv b_0 + \sum_{i=1}^k b_i Z_i ,$$

де  $b_0, b_1, \dots, b_k \in \mathbb{R}$ . Відповідно до Завдання 7.5.9,  $b_i Z_i \sim N(0, b_i^2)$ , а за рахунок Твердження 9.5.4 маємо

$$Y \sim N \left( b_0 + \sum_{i=1}^k b_i, \sum_{i=1}^k b_i^2 \right) .$$

Доведення взаємозв'язку визначень в інший бік виходить за межі нашого курсу.  $\square$

**Завдання 10.4.4.** Із Визначення 10.4.2 фактично випливає, що маржинальними розподілами для багатовимірного нормального розподілу є також (одновимірні) нормальні розподіли. Зворотне твердження виконується далеко не завжди.

Наприклад, нехай  $X \sim N(0, 1)$ , а  $S$  – дискретна величина така, що  $\mathbb{P}_S(S = 1) = \mathbb{P}_S(S = -1) = 0.5$ . Така величина має так званий *розподіл Радемахера*<sup>3</sup> (Rademacher distribution). Нехай  $X \perp\!\!\!\perp S$ . Покажімо, що випадкова величина  $Y = XS$  матиме стандартний нормальній розподіл.

Можемо застосувати Теорему 10.1.1, розглянувши  $h_1(x, s) = xs$ ,  $h_2(x, s) = s$ . Тоді обернені функції будуть дорівнювати  $g_1(u, v) = \frac{u}{v}$ ,  $g_2(u, v) = v$ . Відповідний якобіан дорівнює

$$J = \begin{vmatrix} g_{11}(u, v) & g_{12}(u, v) \\ g_{21}(u, v) & g_{22}(u, v) \end{vmatrix} = \begin{vmatrix} \frac{\partial}{\partial u} g_1(u, v) & \frac{\partial}{\partial v} g_1(u, v) \\ \frac{\partial}{\partial u} g_2(u, v) & \frac{\partial}{\partial v} g_2(u, v) \end{vmatrix} = \begin{vmatrix} \frac{1}{v} & -\frac{u}{v^2} \\ 0 & 1 \end{vmatrix} = \frac{1}{v} \neq 0 .$$

Образом носія  $\text{supp}((X, S)^\top) = \mathbb{R} \times \{-1, 1\}$  під дією відповідного відображення буде ця ж множина:  $T = \text{supp}((X, S)^\top)$ . Усі інші умови теореми також виконуються.

За рахунок незалежності відповідних величин їхня спільна щільність дорівнює

$$f_{X,S}(x, s) = \phi(x) \cdot \mathbb{P}_S(S = s) .$$

<sup>3</sup>Названа так на честь американського математика німецького походження Ганса Адольфа Радемахера (Hans Adolph Rademacher, 1892–1969).

Остаточно маємо:

$$\begin{aligned} f_{Y,S}(u, v) &= f_{X,S}\left(\frac{u}{v}, v\right) \cdot \frac{1}{|v|} \cdot \mathbb{1}\left\{(u, v)^\top \in \mathbb{R} \times \{1, -1\}\right\} \\ &= \phi\left(\frac{u}{v}\right) \cdot \mathbb{P}_S(S = v) \cdot \frac{1}{|v|} \cdot \mathbb{1}\left\{(u, v)^\top \in \mathbb{R} \times \{1, -1\}\right\}, \end{aligned}$$

звідки маржинальний розподіл можна дістати згідно з (9.2.9),

$$\begin{aligned} f_Y(u) &= \sum_{v=-\infty}^{\infty} \phi\left(\frac{u}{v}\right) \cdot \mathbb{P}_S(S = v) \cdot \frac{1}{|v|} \cdot \mathbb{1}\left\{(u, v)^\top \in \mathbb{R} \times \{1, -1\}\right\} \\ &= \phi\left(\frac{u}{-1}\right) \cdot \mathbb{P}_S(S = -1) \cdot \frac{1}{|-1|} + \phi\left(\frac{u}{1}\right) \cdot \mathbb{P}_S(S = 1) \cdot \frac{1}{|1|} \\ &= \frac{1}{2}(\phi(-u) + \phi(u)) \\ &= \phi(u), \end{aligned}$$

за рахунок парності функції  $\phi$ .

Отже  $X$  та  $Y$  мають нормальний розподіл, проте  $\mathbb{P}_{X,Y}(X + Y = 0) = \mathbb{P}_S(S = -1) = \frac{1}{2}$ , тобто розподіл суми (лінійної комбінації з одиничними коефіцієнтами) явно не є нормальним. Відповідно, спільний розподіл вектора  $(X, Y)^\top$  не є двовимірним нормальним розподілом.  $\square$

Знайдімо спільну щільність багатовимірного нормального розподілу.

**Лема 10.4.5.** Якщо  $Z_1, \dots, Z_k \sim N(0, 1)$  — незалежні випадкові величини, то спільна щільність розподілу випадкового вектора  $(Z_1, \dots, Z_k)^\top$  має вираз

$$f_{\mathbf{Z}}(z_1, \dots, z_k) = \frac{1}{\sqrt{(2\pi)^k}} e^{-\frac{1}{2}(z_1^2 + \dots + z_k^2)} = \frac{1}{\sqrt{(2\pi)^k}} e^{-\frac{1}{2}\mathbf{z}^\top \mathbf{z}}, \quad (10.4.1)$$

де  $\mathbf{z} = (z_1, \dots, z_k)^\top$ .

*Доведення.* Це безпосередньо випливає з (7.2.1) та (9.4.6).  $\square$

Тепер можемо дістати спільну щільність довільного нормального вектора.

**Твердження 10.4.6.** Нехай  $\mathbf{X} = (X_1, \dots, X_k)^\top$  можна подати як

$$\mathbf{X} = \mathbf{A} \cdot \mathbf{Z} + \boldsymbol{\mu},$$

де  $\mathbf{A} \in \mathbb{R}^{k \times \mathbb{R}^n}$  — деяка матриця,  $\mathbf{Z} = (Z_1, \dots, Z_n)^\top$  — вектор незалежних стандартних нормальних величин,  $\boldsymbol{\mu} = (\mu_1, \dots, \mu_k)^\top \in \mathbb{R}^k$ . Тоді спільна щільність  $f_{\mathbf{X}}$  має вираз

$$f_{\mathbf{X}}(x_1, \dots, x_k) = \frac{1}{\sqrt{(2\pi)^k |\Sigma|}} e^{-\frac{1}{2}(\mathbf{x} - \boldsymbol{\mu})^\top \Sigma^{-1} (\mathbf{x} - \boldsymbol{\mu})}, \quad (10.4.2)$$

де  $\mathbf{x} = (x_1, \dots, x_k)^\top$ ,  $\Sigma = \mathbf{A} \mathbf{A}^\top$ , до того ж  $\Sigma$  невироджена.

*Доведення.* Розгляньмо випадок, коли  $n = k$ . Як потім стане зрозуміло, це необов'язково, але так простіше довести відповідне твердження.

Вираз (10.4.1) дає спільну щільність вектора  $\mathbf{Z}$ . До нього можна застосувати Теорему 10.1.1. Позначмо кожний  $i$ -ий рядок матриці  $\mathbf{A}$  через  $\mathbf{a}_i = (a_{i1}, \dots, a_{ik})$ . Тоді

$$x_i = h_i(z_1, \dots, z_k) = \mathbf{a}_i \cdot \mathbf{z} + \mu_i, \quad i = 1, \dots, k.$$

Відображення  $(h_1, \dots, h_k)$  взаємно однозначне, оскільки матриця  $\mathbf{A}$  невироджена, а тому існує обернена матриця, яку позначмо через  $\mathbf{B} \equiv \mathbf{A}^{-1}$ :

$$\mathbf{Z} = \mathbf{B}(\mathbf{X} - \boldsymbol{\mu}).$$

Позначивши кожний  $i$ -ий рядок оберненої матриці  $\mathbf{B}$  через  $\mathbf{b}_i = (b_{i1}, \dots, b_{ik})$ , дістаємо:

$$z_i = g_i(x_1, \dots, x_k) = \mathbf{b}_i(\mathbf{x} - \boldsymbol{\mu}) = \sum_{j=1}^k b_{ij}(x_j - \mu_j).$$

Відповідний якобіан дорівнює

$$J = \begin{vmatrix} g_{11}(\mathbf{x}) & \dots & g_{1k}(\mathbf{x}) \\ \vdots & \ddots & \vdots \\ g_{k1}(\mathbf{x}) & \dots & g_{kk}(\mathbf{x}) \end{vmatrix} = \begin{vmatrix} b_{11} & b_{12} & \dots & b_{1k} \\ b_{21} & b_{22} & \dots & b_{2k} \\ \vdots & \vdots & \ddots & \vdots \\ b_{k1} & b_{k2} & \dots & b_{kk} \end{vmatrix} = |\mathbf{B}| = \frac{1}{|\mathbf{A}|}.$$

Носієм вектора  $\mathbf{Z}$  є простір  $\text{supp}(\mathbf{Z}) = \mathbb{R}^k$ , а відтак і його образом також буде простір  $T = \mathbb{R}^k$ . Оскільки матриця  $\mathbf{A}$  невироджена, то якобіан не обертається в нуль на множині  $T$ .

Потрібно помітити, що

$$\mathbf{z} = \begin{pmatrix} z_1 \\ \vdots \\ z_k \end{pmatrix} = \begin{pmatrix} \mathbf{b}_1(\mathbf{x} - \boldsymbol{\mu}) \\ \vdots \\ \mathbf{b}_k(\mathbf{x} - \boldsymbol{\mu}) \end{pmatrix} = \mathbf{B}(\mathbf{x} - \boldsymbol{\mu}).$$

Отже, за (10.1.3) та Лемою 10.4.5 маємо:

$$\begin{aligned} f_{\mathbf{X}}(\mathbf{x}) &= f_{\mathbf{Z}}(g_1(\mathbf{x}), \dots, g_k(\mathbf{x})) \cdot |J| \cdot \mathbb{1} \left\{ \mathbf{x} \in \mathbb{R}^k \right\} \\ &= \frac{1}{|\mathbf{A}|} \cdot \prod_{i=1}^k \frac{1}{\sqrt{2\pi}} e^{-\frac{z_i^2}{2}} \\ &= \frac{1}{|\mathbf{A}|} \cdot \frac{1}{\sqrt{(2\pi)^k}} e^{-\frac{1}{2} \sum_{i=1}^k z_i^2} \\ &= \frac{1}{|\mathbf{A}|} \cdot \frac{1}{\sqrt{(2\pi)^k}} e^{-\frac{1}{2} \mathbf{z}^\top \mathbf{z}} \\ &= \frac{1}{|\mathbf{A}|} \cdot \frac{1}{\sqrt{(2\pi)^k}} e^{-\frac{1}{2} (\mathbf{B}(\mathbf{x} - \boldsymbol{\mu}))^\top (\mathbf{B}(\mathbf{x} - \boldsymbol{\mu}))} \\ &= \frac{1}{|\mathbf{A}|} \cdot \frac{1}{\sqrt{(2\pi)^k}} e^{-\frac{1}{2} (\mathbf{x} - \boldsymbol{\mu})^\top (\mathbf{B}^\top \mathbf{B})(\mathbf{x} - \boldsymbol{\mu})}. \end{aligned} \tag{10.4.3}$$

Оскільки  $\Sigma = \mathbf{A}\mathbf{A}^\top$ ,  $|\Sigma| = |\mathbf{A}|^2$ , а  $\Sigma^{-1} = (\mathbf{A}^\top)^{-1} \mathbf{A}^{-1} = \mathbf{B}^\top \mathbf{B}$ , звідки випливає

$$f_{\mathbf{X}}(\mathbf{x}) = \frac{1}{\sqrt{(2\pi)^k |\Sigma|}} e^{-\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu})^\top \Sigma^{-1}(\mathbf{x}-\boldsymbol{\mu})}.$$

У Визначенні 10.4.1 принципово, щоб  $\Sigma = \mathbf{A}\mathbf{A}^\top$ . Насправді сама собою матриця  $\mathbf{A}$  жодної ролі не грає. Тому якщо  $\mathbf{A}\mathbf{A}^\top = \mathbf{A}'(\mathbf{A}')^\top$  для деякої  $\mathbf{A}' \neq \mathbf{A}$ , то вектори  $\mathbf{A}\mathbf{Z} + \boldsymbol{\mu}$  і  $\mathbf{A}'\mathbf{Z} + \boldsymbol{\mu}$  матимуть однакові розподіли. Тому доведення твердження справедливе також для будь-якої  $\mathbf{A}'$  такої, що  $\Sigma = \mathbf{A}'(\mathbf{A}')^\top$ .  $\square$

**Зауваження 10.4.7.** Варто зазначити, що матриця  $\Sigma$  є симетричною і додатно визначеною. Симетричність випливає безпосередньо з визначення  $\Sigma = \mathbf{A}\mathbf{A}^\top$ . Додатну визначеність можна показати у спосіб, схожий із доведенням властивості (iii) Твердження 9.3.10. Нехай  $\mathbf{x} \in \mathbb{R}^k$ . Тоді  $\mathbf{A}^\top(\mathbf{x} - \boldsymbol{\mu}) \equiv \mathbf{y} \in \mathbb{R}^k$ . Відтак,

$$\begin{aligned} (\mathbf{x} - \boldsymbol{\mu})^\top \Sigma (\mathbf{x} - \boldsymbol{\mu}) &= (\mathbf{x} - \boldsymbol{\mu})^\top \mathbf{A}\mathbf{A}^\top (\mathbf{x} - \boldsymbol{\mu}) \\ &= \left( \mathbf{A}^\top(\mathbf{x} - \boldsymbol{\mu}) \right)^\top \left( \mathbf{A}^\top(\mathbf{x} - \boldsymbol{\mu}) \right) \\ &= \sum_{i=1}^k y_i^2 \geq 0, \end{aligned}$$

до того ж рівність нулю досягається тільки тоді, коли  $\mathbf{y} = 0$ , а отже коли  $\mathbf{x} = \boldsymbol{\mu}$ . Отже, матриця  $\Sigma$  (строго) додатно визначена.

Цей факт гарантує, зокрема, те, що детермінант матриці додатний, тому з нього коректно брати корінь.  $\square$

Для повноти викладу матеріалу потрібно показати, що щільність (10.4.2) відповідає тільки багатовимірному нормальному розподілу.

**Лема 10.4.8.** Якщо випадковий вектор  $\mathbf{X}$  має щільність (10.4.2), де  $\boldsymbol{\mu} \in \mathbb{R}^k$  — довільний вектор, а  $\Sigma$  — симетрична додатно визначена матриця, то він має багатовимірний нормальній розподіл.

*Доведення.* Відповідно до Зауваження 10.4.7, матриця  $\Sigma$  з виразу (10.4.2) повинна бути додатно визначена. Як відомо з курсу «Лінійна алгебра», будь-яку квадратну невироджену матрицю можна подати у формі спектрального розвинення:

$$\Sigma = \mathbf{Q}\Lambda\mathbf{Q}^\top,$$

де  $\Lambda$  — діагональна матриця з власними числами матриці  $\Sigma$  на головній діагоналі,  $\mathbf{Q}$  — квадратова матриця, де кожний  $i$ -ий стовпець відповідає  $i$ -ому власному числу. У силу додатної визначеності матриці  $\Sigma$ , її власні числа додатні, і тому

$$\Sigma = \mathbf{Q}\sqrt{\Lambda}\sqrt{\Lambda}\mathbf{Q}^\top = \mathbf{Q}\sqrt{\Lambda}\left(\sqrt{\Lambda}\mathbf{Q}\right)^\top \equiv \mathbf{A}\mathbf{A}^\top.$$

Отже для довільного вектора  $\mu \in \mathbb{R}^k$  та довільної додатно визначеної симетричної матриці  $\Sigma$

справедливо

$$\begin{aligned} f_{\mathbf{X}}(x_1, \dots, x_k) &= \frac{1}{\sqrt{(2\pi)^k |\Sigma|}} e^{-\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu})^\top \Sigma^{-1}(\mathbf{x}-\boldsymbol{\mu})} \\ &= \frac{1}{\sqrt{(2\pi)^k |\mathbf{A}\mathbf{A}^\top|}} e^{-\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu})^\top (\mathbf{A}\mathbf{A}^\top)^{-1}(\mathbf{x}-\boldsymbol{\mu})}. \end{aligned}$$

Цей вираз повністю відповідає аналогічному виразу (10.4.3) з доведення Твердження 10.4.6, тобто це відповідає щільності деякої лінійної функції від вектора стандартних нормальніх величин, а тому є щільністю багатовимірного нормального розподілу.  $\square$

На практиці дуже часто застосовують двовимірний нормальний розподіл, щільність якого випливає з (10.4.2):

$$\begin{aligned} f_{X_1, X_2}(x_1, x_2) &= \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}} e^{-\frac{1}{2\sigma_1^2\sigma_2^2(1-\rho^2)} \left( (x_1 - \mu_1, x_2 - \mu_2) \begin{pmatrix} \sigma_2^2 & -\rho\sigma_1\sigma_2 \\ -\rho\sigma_1\sigma_2 & \sigma_1^2 \end{pmatrix} \begin{pmatrix} x_1 - \mu_1 \\ x_2 - \mu_2 \end{pmatrix} \right)} \\ &= \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}} e^{-\frac{1}{2(1-\rho^2)} \left( \left( \frac{x_1 - \mu_1}{\sigma_1} \right)^2 - 2\rho \left( \frac{x_1 - \mu_1}{\sigma_1} \right) \left( \frac{x_2 - \mu_2}{\sigma_2} \right) + \left( \frac{x_2 - \mu_2}{\sigma_2} \right)^2 \right)}, \end{aligned}$$

де  $\rho = \text{Corr}(X_1, X_2)$ , оскільки  $\rho = \sigma_{12}\sigma_1\sigma_2$ ,  $|\Sigma| = \sigma_1^2\sigma_2^2(1-\rho^2)$ , а

$$\Sigma^{-1} = \frac{1}{\sigma_1^2\sigma_2^2(1-\rho^2)} \begin{pmatrix} \sigma_2^2 & -\rho\sigma_1\sigma_2 \\ -\rho\sigma_1\sigma_2 & \sigma_1^2 \end{pmatrix}.$$

Прикладами векторів із двовимірним нормальним розподілом є (зріст батька, зріст сина), (бали за ЗНО, бали за першу сесію), (зріст, вага), (напруга, спротив), (обсяг добрив, урожайність), (рівень холестерину, тиск) тощо.

Спільні щільності двовимірних нормальних випадкових векторів наведено на Рис. 10.4.1.

**Приклад 10.4.9.** Нехай

$$\begin{pmatrix} X_1 \\ X_2 \end{pmatrix} \sim N \left( \begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix}, \begin{pmatrix} \sigma_1^2 & \sigma_{12} \\ \sigma_{12} & \sigma_2^2 \end{pmatrix} \right).$$

Знайдімо ймовірність, що одна з його координат більші від іншої. Оскільки  $\mathbb{P}_{X_1, X_2}(X_1 < X_2) = \mathbb{P}_{X_1, X_2}(X_1 - X_2 < 0)$ , розгляньмо величину  $D = X_1 - X_2$ . Оскільки вектор має двовимірний нормальний розподіл, різниця його координат (як і будь-яка інша лінійна комбінація) також має нормальний розподіл. Відтак достатньо дістати її сподівання та дисперсію:

$$\begin{aligned} \mathbb{E}[X_1 - X_2] &= \mathbb{E}[X_1] - \mathbb{E}[X_2] = \mu_1 - \mu_2 \\ \text{Var}(X_1 - X_2) &= \text{Var}(X_1) + \text{Var}(X_2) - 2\text{Cov}(X_1, X_2) = \sigma_1^2 + \sigma_2^2 - 2\sigma_{12}. \end{aligned}$$

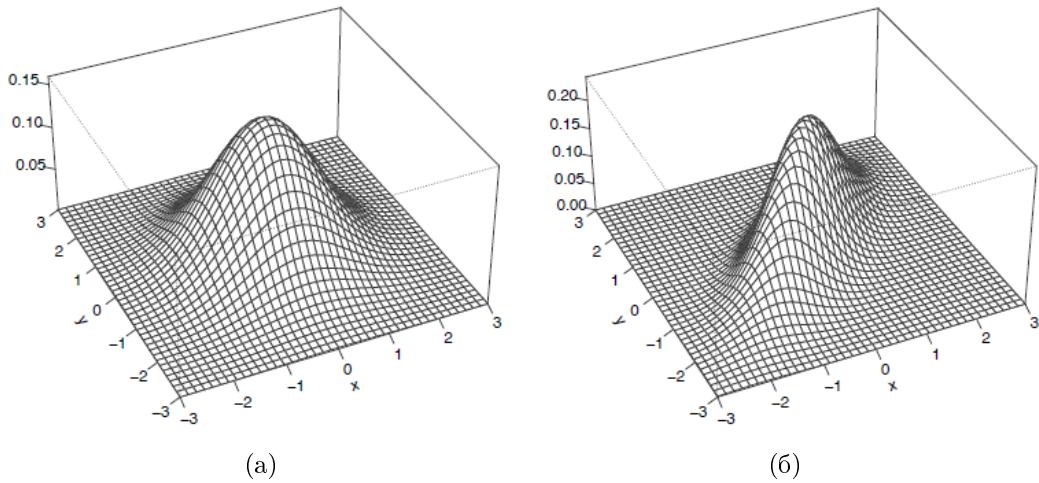


Рис. 10.4.1.: Спільні щільності розподілу для двох багатовимірних нормальних векторів: (а) маржинальні розподіли — стандартні нормальні і незалежні; (б) маржинальні розподіли — стандартні нормальні з коефіцієнтом кореляції 0.75 ([1], Рис. 7.10)

Тоді

$$\begin{aligned}
 \mathbb{P}_{X_1, X_2}(X_1 < X_2) &= \mathbb{P}_D(D < 0) \\
 &= \mathbb{P}_D\left(\frac{D - (\mu_1 - \mu_2)}{\sqrt{\sigma_1^2 + \sigma_2^2 - 2\sigma_{12}}} < \frac{-(\mu_1 - \mu_2)}{\sqrt{\sigma_1^2 + \sigma_2^2 - 2\sigma_{12}}}\right) \\
 &= \Phi\left(\frac{\mu_2 - \mu_1}{\sqrt{\sigma_1^2 + \sigma_2^2 - 2\sigma_{12}}}\right).
 \end{aligned}$$

□

Як і в одновимірному випадку, де параметри розподілу  $\mu$  і  $\sigma^2$  відповідали сподіванню та дисперсії нормальної величини, так параметри  $\boldsymbol{\mu}$  і  $\Sigma$  відіграють особливу роль.

**Твердження 10.4.10.** Якщо  $\mathbf{X} \sim N(\boldsymbol{\mu}, \Sigma)$ , то  $\mathbb{E}[\mathbf{X}] = \boldsymbol{\mu}$ ,  $\text{Cov}(\mathbf{X}) = \Sigma$ .

*Доведення.* Розглянемо багатовимірний стандартний нормальний вектор  $\mathbf{Z} = (Z_1, \dots, Z_k)^\top$ , де  $Z_i \sim N(0, 1)$ ,  $i = 1, \dots, k$  — незалежні величини. Тоді, за Визначенням 9.3.1,

$$\mathbb{E}[\mathbf{Z}] = \begin{pmatrix} \mathbb{E}[Z_1] \\ \vdots \\ \mathbb{E}[Z_k] \end{pmatrix} = \begin{pmatrix} 0 \\ \vdots \\ 0 \end{pmatrix}.$$

Також, оскільки  $\text{Cov}(Z_i, Z_i) = \text{Var}(Z_i) = 1$ , а через незалежність  $\text{Cov}(Z_i, Z_j) = 0$ ,  $i \neq j$ , маємо:

$$\text{Cov}(\mathbf{Z}) = \mathbf{I},$$

де  $\mathbf{I}$  — одинична матриця.

Згідно з Визначенням 10.4.1,  $\mathbf{X} = \mathbf{A}\mathbf{Z} + \boldsymbol{\mu}$  для деяких  $\mathbf{A}$  і  $\boldsymbol{\mu}$ ,  $\Sigma = \mathbf{A}\mathbf{A}^\top$ .

Використовуючи властивість (9.3.2), відразу маємо:

$$\mathbb{E}[\mathbf{X}] = \mathbb{E}[\mathbf{A}\mathbf{Z} + \boldsymbol{\mu}] = \mathbf{A}\mathbb{E}[\mathbf{Z}] + \boldsymbol{\mu} = \mathbf{A}\mathbf{0} + \boldsymbol{\mu} = \boldsymbol{\mu}.$$

Використовуючи властивість (ii) із Твердження 9.3.10, відразу маємо:

$$\text{Cov}(\mathbf{X}) = \text{Cov}(\mathbf{A}\mathbf{Z} + \boldsymbol{\mu}) = \mathbf{A}\mathbf{I}\mathbf{A}^\top = \mathbf{A}\mathbf{A}^\top = \Sigma.$$

□

**Зauważення 10.4.11.** Щільність одновимірного нормального розподілу можна переписати як

$$f_X(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2}(x-\mu)\cdot(\sigma^2)^{-1}\cdot(x-\mu)}.$$

Такий запис дає можливість ліпше зрозуміти, чому вираз (10.4.2) справді є узагальненням щільності в одновимірному випадку. □

Особливістю багатовимірного нормального розподілу є не тільки той факт, що всі маржинальні розподіли є нормальними, а й будь-яка підмножина його координат має багатовимірний нормальний розподіл. Доведення цього факту на основі Визначення 10.4.2 майже тривіальне, хоча виконання інтегрування щільності (10.4.2) — задача непроста.

**Твердження 10.4.12.** Якщо  $\mathbf{X} = (X_1, \dots, X_k)^\top \sim N(\boldsymbol{\mu}, \Sigma)$ , де

$$\boldsymbol{\mu} = \begin{pmatrix} \mu_1 \\ \vdots \\ \mu_k \end{pmatrix}, \quad \Sigma = \begin{pmatrix} \text{Var}(X_1) & \text{Cov}(X_1, X_2) & \dots & \text{Cov}(X_1, X_k) \\ \text{Cov}(X_2, X_1) & \text{Var}(X_2) & \dots & \text{Cov}(X_2, X_k) \\ \vdots & \vdots & \ddots & \vdots \\ \text{Cov}(X_k, X_1) & \text{Cov}(X_k, X_2) & \dots & \text{Var}(X_k) \end{pmatrix}.$$

Тоді вектор, складений із будь-якої множини координат  $\mathbf{X}$ , також має багатовимірний нормальний розподіл:  $\mathbf{X}' = (X_{i_1}, \dots, X_{i_m})^\top \sim N(\boldsymbol{\mu}', \Sigma')$ ,  $i_1, \dots, i_m \in \{1, 2, \dots, k\}$ ,  $i_s \neq i_t$  для будь-яких  $s, t$ , де

$$\boldsymbol{\mu}' = \begin{pmatrix} \mu_{i_1} \\ \mu_{i_2} \\ \vdots \\ \mu_{i_m} \end{pmatrix}, \quad \Sigma' = \begin{pmatrix} \text{Var}(X_{i_1}) & \text{Cov}(X_{i_1}, X_{i_2}) & \dots & \text{Cov}(X_{i_1}, X_{i_m}) \\ \text{Cov}(X_{i_2}, X_{i_1}) & \text{Var}(X_{i_2}) & \dots & \text{Cov}(X_{i_2}, X_{i_m}) \\ \vdots & \vdots & \ddots & \vdots \\ \text{Cov}(X_{i_m}, X_{i_1}) & \text{Cov}(X_{i_m}, X_{i_2}) & \dots & \text{Var}(X_{i_m}) \end{pmatrix}.$$

**Доведення.** Згідно з Визначенням 10.4.2, випадковий вектор  $\mathbf{X}$  матиме багатовимірний нормальний розподіл тоді й тільки тоді, коли будь-яка лінійна комбінація його координат  $\mathbf{c}^\top \mathbf{X}$ ,  $\mathbf{c} \in \mathbb{R}^k$ , є випадковою величиною з нормальним розподілом. Розгляньмо  $\mathbf{c}'_j = \mathbf{I}_j^\top$  — вектор-рядок, у якому всі координати дорівнюють нулю, окрім координати  $i_j$ , яка дорівнює одиниці,  $j = 1, \dots, m$ . Тоді  $X_{i_j} = \mathbf{c}'_j \mathbf{X}$ ,  $j = 1, \dots, m$ , буде нормальним величиною. А відтак і вектор  $\mathbf{X}'$ , складений із таких  $X_{i_j}$ ,  $j = 1, \dots, m$ , повинен мати багатовимірний нормальний розподіл.

Сподівання й матрицю коваріації вектора  $\mathbf{X}'$  знайдімо з Визначення 10.4.1. Вектор  $\mathbf{X}'$  можна

дістати шляхом лінійного перетворення вектора  $\mathbf{X}$ :  $\mathbf{X}' = \mathbf{M}\mathbf{X}$ , де

$$\mathbf{M} = \begin{pmatrix} \mathbf{I}_1^\top \\ \vdots \\ \mathbf{I}_m^\top \end{pmatrix}.$$

Тоді

$$\mathbf{X}' = \mathbf{M}(\mathbf{A}\mathbf{Z} + \boldsymbol{\mu}) = \mathbf{M}\mathbf{A}\mathbf{Z} + \mathbf{M}\boldsymbol{\mu}.$$

Нескладно бачити, що

$$\mathbf{M}\boldsymbol{\mu} = \boldsymbol{\mu}',$$

а

$$\mathbf{M}\mathbf{A}(\mathbf{M}\mathbf{A})^\top = \mathbf{M}\mathbf{A}\mathbf{A}^\top\mathbf{M}^\top = \mathbf{M}\boldsymbol{\Sigma}\mathbf{M}^\top = \boldsymbol{\Sigma}',$$

тобто  $\mathbf{X}' \sim N(\boldsymbol{\mu}', \boldsymbol{\Sigma}')$ .  $\square$

**Приклад 10.4.13.** Розгляньмо конкретний приклад. Нехай

$$\begin{pmatrix} X_1 \\ X_2 \\ X_3 \end{pmatrix} \sim N \left( \begin{pmatrix} \mu_1 \\ \mu_2 \\ \mu_3 \end{pmatrix}, \begin{pmatrix} \sigma_1^2 & \sigma_{12} & \sigma_{13} \\ \sigma_{12} & \sigma_2^2 & \sigma_{23} \\ \sigma_{13} & \sigma_{23} & \sigma_3^2 \end{pmatrix} \right).$$

Нас цікавить розподіл  $\mathbf{X}' = (X_3, X_1)^\top$ . У цьому випадку

$$\mathbf{M} = \begin{pmatrix} 0 & 0 & 1 \\ 1 & 0 & 0 \end{pmatrix},$$

адже

$$\begin{pmatrix} 0 & 0 & 1 \\ 1 & 0 & 0 \end{pmatrix} \begin{pmatrix} X_1 \\ X_2 \\ X_3 \end{pmatrix} = \begin{pmatrix} X_3 \\ X_1 \end{pmatrix}.$$

Тоді

$$\boldsymbol{\mu}' = \begin{pmatrix} 0 & 0 & 1 \\ 1 & 0 & 0 \end{pmatrix} \begin{pmatrix} \mu_1 \\ \mu_2 \\ \mu_3 \end{pmatrix} = \begin{pmatrix} \mu_3 \\ \mu_1 \end{pmatrix},$$

а

$$\boldsymbol{\Sigma}' = \begin{pmatrix} 0 & 0 & 1 \\ 1 & 0 & 0 \end{pmatrix} \begin{pmatrix} \sigma_1^2 & \sigma_{12} & \sigma_{13} \\ \sigma_{12} & \sigma_2^2 & \sigma_{23} \\ \sigma_{13} & \sigma_{23} & \sigma_3^2 \end{pmatrix} \begin{pmatrix} 0 & 1 \\ 0 & 0 \\ 1 & 0 \end{pmatrix} = \begin{pmatrix} \sigma_3^2 & \sigma_{13} \\ \sigma_{13} & \sigma_1^2 \end{pmatrix}.$$

$\square$

У Твердженні 9.4.13 ми зазначали, що з незалежності двох випадкових величин  $X$  та  $Y$  випливає, що  $\text{Cov}(X, Y) = 0$ , але зворотне в загальному випадку не є справедливим. Як виявляється, багатовимірний нормальний розподіл унікальний тим, що якраз для нього зворотне твердження виконується.

**Твердження 10.4.14.** Нехай  $\mathbf{X} \sim N(\boldsymbol{\mu}, \Sigma)$  можна подати як конкатенацію двох випадкових векторів меншої розмірності:  $\mathbf{X} = (\mathbf{X}_1^\top, \mathbf{X}_2^\top)^\top$ ,  $\mathbf{X} : (\Omega, \mathcal{A}, \mathbb{P}) \rightarrow \mathbb{R}^k$ ,  $\mathbf{X}_1 : (\Omega, \mathcal{A}, \mathbb{P}) \rightarrow \mathbb{R}^m$ ,  $\mathbf{X}_2 : (\Omega, \mathcal{A}, \mathbb{P}) \rightarrow \mathbb{R}^{k-m}$ ,  $1 \leq m < k$ .

Якщо кореляція між будь-якою координатою  $\mathbf{X}_1$  і будь-якою координатою  $\mathbf{X}_2$  нульова, то  $\mathbf{X}_1 \perp\!\!\!\perp \mathbf{X}_2$ .

**Доведення.** Згідно з Теоремою 9.4.5, потрібно показати, що спільну щільність вектора  $\mathbf{X}$  можна подати як добуток щільностей векторів  $\mathbf{X}_1$  і  $\mathbf{X}_2$ . Якщо ці вектори покоординатно некорельовані, то матриця коваріацій є блоково-діагональною:

$$\Sigma = \begin{pmatrix} \Sigma_1 & \mathbf{0} \\ \mathbf{0} & \Sigma_2 \end{pmatrix},$$

де  $\text{Cov}(\mathbf{X}_1) = \Sigma_1$ ,  $\text{Cov}(\mathbf{X}_2) = \Sigma_2$ . Тоді її детермінант дорівнює  $|\Sigma| = |\Sigma_1||\Sigma_2|$  й обернена матриця буде блоково-діагональною:

$$\Sigma^{-1} = \begin{pmatrix} \Sigma_1^{-1} & \mathbf{0} \\ \mathbf{0} & \Sigma_2^{-1} \end{pmatrix},$$

отже

$$\begin{aligned} f_{\mathbf{X}}(\mathbf{x}) &= \frac{1}{\sqrt{(2\pi)^k |\Sigma|}} e^{-\frac{1}{2}(\mathbf{x} - \boldsymbol{\mu})^\top \Sigma^{-1}(\mathbf{x} - \boldsymbol{\mu})} \\ &= \frac{1}{\sqrt{(2\pi)^k |\Sigma_1||\Sigma_2|}} e^{-\frac{1}{2}((\mathbf{x}_1 - \boldsymbol{\mu}_1)^\top, (\mathbf{x}_2 - \boldsymbol{\mu}_2)^\top) \begin{pmatrix} \Sigma_1^{-1} & \mathbf{0} \\ \mathbf{0} & \Sigma_2^{-1} \end{pmatrix} \begin{pmatrix} \mathbf{x}_1 - \boldsymbol{\mu}_1 \\ \mathbf{x}_2 - \boldsymbol{\mu}_2 \end{pmatrix}} \\ &= \frac{1}{\sqrt{(2\pi)^m |\Sigma_1|} \sqrt{(2\pi)^{k-m} |\Sigma_2|}} e^{-\frac{1}{2}((\mathbf{x}_1 - \boldsymbol{\mu}_1)^\top \Sigma_1^{-1}(\mathbf{x}_1 - \boldsymbol{\mu}_1) + (\mathbf{x}_2 - \boldsymbol{\mu}_2)^\top \Sigma_2^{-1}(\mathbf{x}_2 - \boldsymbol{\mu}_2))} \\ &= f_{\mathbf{X}_1}(\mathbf{x}_1) \cdot f_{\mathbf{X}_2}(\mathbf{x}_2). \end{aligned}$$

□

**Зауваження 10.4.15.** Застосування Твердження 10.4.14 потребує виконання умови, щоб вектори  $\mathbf{X}_1$  і  $\mathbf{X}_2$  мали спільний багатовимірний нормальній розподіл. Інакше з некорельованості незалежність випливатиме далеко необов'язково.

Повернімося до Зауваження 10.4.4. Змінні  $X$  та  $Y$  були нормальні, але як було показано, їхній спільний розподіл не був нормальній. Кореляція між ними нульова:

$$\text{Cov}(X, Y) = \mathbb{E}[XY] - \mathbb{E}[X]\mathbb{E}[Y] = \mathbb{E}[X^2S] - 0 = \mathbb{E}[X^2]\mathbb{E}[S] = 1 \cdot 0 = 0,$$

де ми використали незалежність  $S$  і  $X$  (а отже  $X^2$ ). Але незважаючи на це, змінні, як було зазначено, залежні. □

## 10.5. Доведення окремих тверджень

Доведімо Твердження 10.2.3.

**Доведення.** Нехай  $X_i \sim \text{Gamma}(n_i, \lambda)$ . Застосування (9.5.1) разом із прийомом із Прикладу 10.1.6 (виділення щільності деякого нормального розподілу, інтеграл від якої дорівнює 1) дає:

$$\begin{aligned}
 f_{X_1+X_2}(s) &= \int_{-\infty}^{\infty} f_{X_1}(s-y) f_{X_2}(y) dy \\
 &= \int_0^{\infty} \frac{1}{\Gamma(n_1)} (s-y)^{n_1-1} e^{-\lambda(s-y)} \lambda^{n_1} \cdot \mathbb{1}\{s-y > 0\} \cdot \frac{1}{\Gamma(n_2)} y^{n_2-1} e^{-\lambda y} \lambda^{n_2} \cdot \mathbb{1}\{y > 0\} dy \\
 &= \frac{1}{\Gamma(n_1+n_2)} e^{-\lambda s} \lambda^{n_1+n_2} \int_0^s \frac{\Gamma(n_1+n_2)}{\Gamma(n_1)\Gamma(n_2)} (s-y)^{n_1-1} y^{n_2-1} dy \\
 &= \frac{1}{\Gamma(n_1+n_2)} s^{n_1+n_2-1} e^{-\lambda s} \lambda^{n_1+n_2} \int_0^s \frac{\Gamma(n_1+n_2)}{\Gamma(n_1)\Gamma(n_2)} \left(1 - \frac{y}{s}\right)^{n_1-1} \left(\frac{y}{s}\right)^{n_2-1} \cdot \frac{1}{s} dy \\
 &= \frac{1}{\Gamma(n_1+n_2)} s^{n_1+n_2-1} e^{-\lambda s} \lambda^{n_1+n_2} \int_0^1 \frac{\Gamma(n_1+n_2)}{\Gamma(n_1)\Gamma(n_2)} (1-u)^{n_1-1} u^{n_2-1} du.
 \end{aligned}$$

Можна помітити, що інтеграл

$$\int_0^1 (1-u)^{n_1-1} u^{n_2-1} du = B(n_1, n_2)$$

є *бета-функцією*, яка, як відомо з курсу «Математичний аналіз», дорівнює  $\frac{\Gamma(n_1)\Gamma(n_2)}{\Gamma(n_1+n_2)}$ . тому

$$\frac{\Gamma(n_1+n_2)}{\Gamma(n_1)\Gamma(n_2)} \int_0^1 (1-u)^{n_1-1} u^{n_2-1} du = \frac{\Gamma(n_1+n_2)}{\Gamma(n_1)\Gamma(n_2)} \cdot \frac{\Gamma(n_1)\Gamma(n_2)}{\Gamma(n_1+n_2)} = 1,$$

а отже  $S \sim \text{Gamma}(n_1+n_2, \lambda)$ .

За індукцією випливає, що

$$\sum_{i=1}^k X_i \sim \text{Gamma}\left(\sum_{i=1}^k n_i, \lambda\right).$$

□

# 11. Сподівання та інші характеристики випадкових величин

У Лекції 6 ми увели поняття сподівання  $\mathbb{E}[X]$  випадкової величини  $X : (\Omega, \mathcal{A}, \mathbb{P}) \rightarrow (\mathbb{R}, \mathcal{B})$  як її інтегралу за мірою  $\mathbb{P}$ :  $\mathbb{E}[X] = \int X d\mathbb{P}$ . За допомогою Теореми 4.3.6 ми показали, як можна звести підрахунок такого інтегралу до звичних інтегралів Рімана (для абсолютно неперервних випадкових величин) та рядів (для дискретних випадкових величин). Також ми дізналися, чому дорівнюють сподівання випадкових величин із поширеними на практиці розподілами.

У цій лекції ми продовжимо розглядати сподівання та їхні властивості, а також розглянемо інші характеристики випадкових величин, які корисно використовувати для їх опису.

## 11.1. Властивості сподівань

### 11.1.1. Властивості, що випливають з поняття інтегралу

Оскільки сподівання — це інтеграл за мірою, низка його властивостей є безпосереднім наслідком властивостей інтегралів. У Твердженні 6.2.10 ми увели декілька таких властивостей, які доцільно повторити тут для повноти викладу, але вже в термінах сподівань.

**Твердження 11.1.1.** Нехай  $X$  та  $Y$  — інтегровні випадкові величини. Тоді:

(i) якщо  $X \geq 0$  майже напевно, то  $\mathbb{E}[X] \geq 0$ ;

(ii) лінійність сподівання: для всіх  $a, b \in \mathbb{R}$ ,

$$\mathbb{E}[aX + bY] = a\mathbb{E}[X] + b\mathbb{E}[Y] ;$$

(iii) монотонність: якщо  $X \leq Y$  майже напевно, то  $\mathbb{E}[X] \leq \mathbb{E}[Y]$ ;

(iv) якщо  $X = Y$  майже напевно, то  $\mathbb{E}[X] = \mathbb{E}[Y]$ ;

(v)  $|\mathbb{E}[X]| \leq \mathbb{E}[|X|]$ .

Ми не будемо доводити цих властивостей, оскільки це базові властивості інтегралу за мірою, і їх доводять у рамках інших курсів.

Раніше ми розглядали Теорему 6.2.6 про монотонну збіжність для інтегралів вимірних функцій. Оскільки сподівання — це інтеграл, доречно повторити тут цю теорему.

**Теорема 11.1.2** (Теорема про монотонну збіжність (Monotone convergence theorem)). Розгляньмо деяку неспадну послідовність невід'ємних випадкових величин  $X_1, X_2, \dots$  таких, що  $X_1(\omega) \leq X_2(\omega) \leq \dots$  для всіх  $\omega$ , і  $\lim_{n \rightarrow \infty} X_n = X$  майже напевно. Тоді

$$\lim_{n \rightarrow \infty} \mathbb{E}[X_n] = \mathbb{E}\left[\lim_{n \rightarrow \infty} X_n\right] = \mathbb{E}[X] .$$

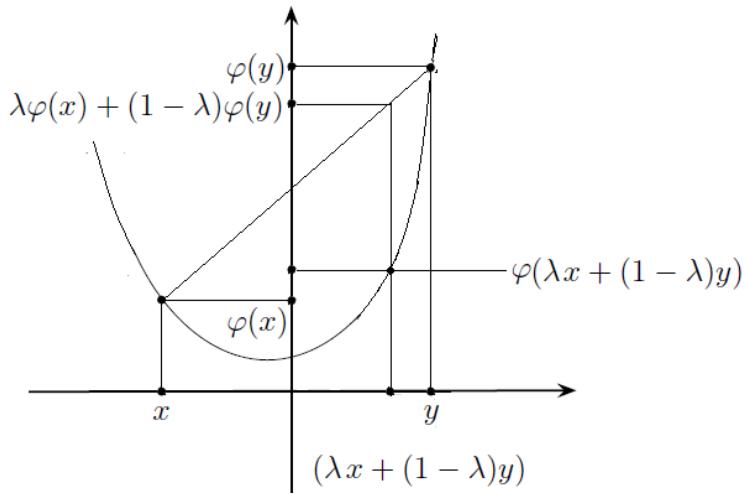


Рис. 11.1.1.: Ілюстрація визначення опуклої функції

Можна довести аналогічний результат для послідовностей, які необов'язково неспадні, проте які обмежені зверху.

**Теорема 11.1.3** (Теорема про мажоровану збіжність (Dominated convergence theorem)). Розгляньмо деяку послідовність випадкових величин  $X_1, X_2, \dots$  таких, що  $\lim_{n \rightarrow \infty} X_n = X$  майже напевно. Нехай  $|X_n| \leq Y$  майже напевно для всіх  $n$ , а  $Y$  інтегровна. Тоді

$$\lim_{n \rightarrow \infty} \mathbb{E}[X_n] = \mathbb{E}\left[\lim_{n \rightarrow \infty} X_n\right] = \mathbb{E}[X].$$

### 11.1.2. Нерівність Єнсена

Для формулювання надзвичайно важливої нерівності Єнсена потрібно згадати, які функції називають опуклими.

**Визначення 11.1.4.** Функція  $\varphi : \mathbb{R} \rightarrow \mathbb{R}$  опукла (convex)<sup>1</sup> якщо

$$\lambda\varphi(x) + (1 - \lambda)\varphi(y) \geq \varphi(\lambda x + (1 - \lambda)y), \quad \lambda \in (0; 1), \quad x, y \in \mathbb{R}. \quad (11.1.1)$$

□

Геометрична інтерпретація цього визначення полягає в тому, що якщо з'єднати дві точки на графіку функції  $\varphi$  прямую, то відповідний сегмент лежатиме вище від графіку  $\varphi$ . Ілюстрацію цього визначення наведено на Рис. 11.1.1.

**Теорема 11.1.5** (Нерівність Єнсена (Jensen's inequality)<sup>2</sup>). Якщо функція  $\varphi : \mathbb{R} \rightarrow \mathbb{R}$  опукла, а випадкові величини  $X$  і  $\varphi(X)$  інтегровні, тобто  $\mathbb{E}[|X|] < \infty$  і  $\mathbb{E}[|\varphi(X)|] < \infty$ , то

$$\mathbb{E}[\varphi(X)] \geq \varphi(\mathbb{E}[X]). \quad (11.1.2)$$

<sup>1</sup>Наведене формулювання опуклості, власне кажучи, і є його визначенням, оскільки працює для будь-яких функцій. Відомий із курсу «Математичний аналіз» спосіб перевірити опуклість функції взяття її другої похідної працює тільки для функцій, які таку похідну мають.

<sup>2</sup>Названа так на честь данського математика Йогана Єнсена (Johan Ludwig William Valdemar Jensen, 1859–1925).

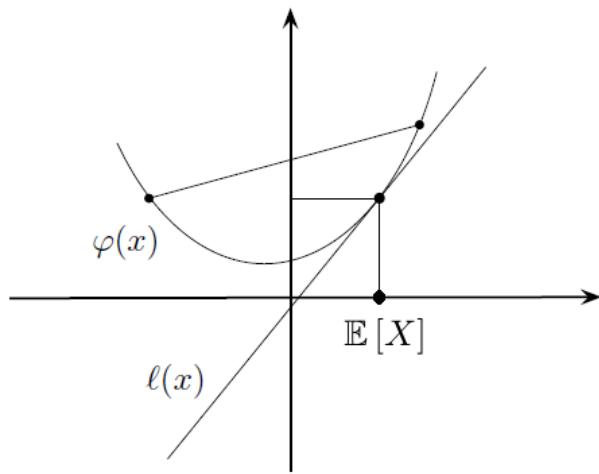


Рис. 11.1.2.: Ілюстрація доведення Теореми 11.1.5

*Доведення.* Нехай  $\ell(x) = ax + b$  — деяка лінійна функція така, що  $\ell(\mathbb{E}[X]) = \varphi(\mathbb{E}[X])$  і  $\varphi(x) \geq \ell(x)$  (Рис. 11.1.2)<sup>3</sup>.

Згідно з властивостями (ii)–(iii) Твердження 6.4.6,

$$\mathbb{E}[\varphi(X)] \geq \mathbb{E}[aX + b] = a\mathbb{E}[X] + b = \ell(\mathbb{E}[X]) = \varphi(\mathbb{E}[X]).$$

□

**Зауваження 11.1.6.** Декілька безпосередніх наслідків із нерівності Енсена:

- якщо функція вгнута (concave), тобто опукла вгору, то  $\mathbb{E}[\varphi(X)] \leq \varphi(\mathbb{E}[X])$ ;
- $\mathbb{E}[\varphi(X)] = \varphi(\mathbb{E}[X])$  тоді й тільки тоді, коли  $\varphi(X) = a + bX$  майже напевно;
- $\mathbb{E}[|X|] \geq |\mathbb{E}[X]|$ , оскільки модуль — це опукла функція;
- $\mathbb{E}[X^r] \geq (\mathbb{E}[X])^r$  для  $r > 1$  для додатних випадкових величин  $X$ , оскільки піднесення до такого ступеня — це також опукла функція;
- $\mathbb{E}\left[\frac{1}{X}\right] \geq \frac{1}{\mathbb{E}[X]}$  для додатних випадкових величин  $X$ , оскільки ділення на  $X$  — це також опукла функція;
- $\mathbb{E}[\ln X] \leq \ln(\mathbb{E}[X])$  для додатних випадкових величин  $X$ , оскільки логарифм — це вгнута функція;
- $\mathbb{E}[X^r] \leq (\mathbb{E}[X])^r$  для  $0 < r < 1$  для додатних випадкових величин  $X$ , оскільки піднесення до такого ступеня — це також вгнута функція.

□

<sup>3</sup>Можна формально показати, що така  $\ell$  завжди існує.

**Приклад 11.1.7.** На основі нерівності Єнсена (11.1.2) можна довести декілька відомих нерівностей. В усіх випадках нехай  $X$  — деяка дискретна випадкова величина з носієм  $\text{supp}(X) = \{x_1, \dots, x_n\}$  та дискретним рівномірним розподілом. Тоді:

- якщо  $f(x) = x^2$ , то маємо нерівність

$$\left( \frac{1}{n} \sum_{i=1}^n x_i \right)^2 \leq \frac{1}{n} \sum_{i=1}^n x_i^2 \quad \Rightarrow \quad \frac{1}{n} \sum_{i=1}^n x_i \leq \sqrt{\frac{1}{n} \sum_{i=1}^n x_i^2}.$$

тобто середнє аритметичне не перевищує кореня з середнього відповідних квадратів;

- якщо  $f(x) = 1/x$ , а всі  $x_i > 0$ ,  $i = 1, \dots, n$  (тільки в цьому випадку функція буде опуклою), то маємо нерівність

$$\left( \frac{1}{n} \sum_{i=1}^n x_i \right)^{-1} \leq \frac{1}{n} \cdot \sum_{i=1}^n \frac{1}{x_i} \quad \Rightarrow \quad \frac{n}{\frac{1}{x_1} + \dots + \frac{1}{x_n}} \leq \frac{1}{n} \sum_{i=1}^n x_i,$$

тобто гармонійне середнє додатних чисел не перевищує середнього аритметичного цих же чисел;

- нарешті, якщо  $f(x) = \ln x$ , а всі  $x_i > 0$ ,  $i = 1, \dots, n$ , то маємо нерівність

$$\frac{1}{n} \sum_{i=1}^n \ln x_i \leq \ln \left( \frac{1}{n} \sum_{i=1}^n x_i \right) \quad \Rightarrow \quad e^{\frac{1}{n} \sum_{i=1}^n \ln x_i} \leq \frac{1}{n} \sum_{i=1}^n x_i \quad \Rightarrow \quad \sqrt[n]{x_1 \dots x_n} \leq \frac{1}{n} \sum_{i=1}^n x_i,$$

тобто середнє геометричне додатних чисел не перевищує середнього аритметичного цих же чисел.

□

**Приклад 11.1.8** (Максимізація сподіваної корисності). Розгляньмо відомий парадокс з економічної теорії — Санкт-Петербурзький парадокс (St. Petersburg paradox)<sup>4</sup>.

Багатий незнайомець пропонує зіграти у таку гру. Підкидатимуть правильну монетку доти, доки не випаде герб. Якщо герб випаде з першого разу, гравець дістане 2 грн, якщо з другого — 4 грн, якщо з третього — 8 грн, і т.д. Яку суму гравець готовий заплатити, щоб узяти участь у такій грі?

Згідно з так званою *теорією сподіваної корисності* (expected utility theory), гравець між різними варіантами обирає той, який несе для нього максимальну корисність. У нашій задачі гравець обирає між грою та утриманням від гри. Якщо гравець має  $w$  грошей і вирішить утриматися від гри, він не дістане нічого і залишиться з сумою  $w$  грошей. Якщо ж він вирішить узяти участь у грі, то дістане  $\mathbb{E}[w - c + X]$ , де  $c$  — ціна, яку він готовий заплатити за участь у грі, а  $X$  — випадковий виграні.

<sup>4</sup>Названий так на честь міста, де працював швейцарський математик Даніель Бернуллі (Daniel Bernoulli, 1700–1782), який у 1738 р. провів його детальний аналіз.

Чому дорівнює сподіваний виграш запропонованої гри,  $\mathbb{E}[X]$ ? Імовірність випаду герба на  $n$ -ому кроці дорівнює  $\frac{1}{2^n}$ , і виграш буде  $2^n$  грошей. Отже

$$\mathbb{E}[X] = \sum_{i=1}^{\infty} \frac{1}{2^n} \cdot 2^n = \infty,$$

тобто сподіваний виграш нескінченно великий. Порівнюючи нескінченність із гарантованим виграшем у  $w$  грошей, гравець завжди буде готовий узяти участь у грі, *незалежно від ціни грі*. Іншими словами, гравець повинен бути готовий заплатити будь-яку суму за право взяти участь у такій грі.

Такий результат суперечить загальній інтуїції, адже в реальному житті мало хто готовий був би покласти все своє майно заради права зіграти в такого роду грі, навіть незважаючи на те, що формально сподіваний виграш необмежено великий.

Одним із можливих розв'язань цієї ситуації в економіці є наділення гравців деякою вгнутою функцією корисності (utility function)  $u$ , яка кожному можливому виграшу  $x$  зіставляє його корисність  $u(x)$ . Для вгнутої диференційованої функції відомо, що  $u''(x) < 0$ , тобто похідна  $u'(x)$  є спадною функцією. Відтак що більше значення  $x$ , то *повільніше* зростає корисність для гравця. Це має називу *спадна гранична корисність* (diminishing marginal utility) і пояснює, зокрема, чому зайва тисяча гривень для пенсіонера має більшу корисність, ніж для мільйонера.

Якщо покласти  $u(x) = \sqrt{x}$ , то

$$\mathbb{E}[\sqrt{X}] = \sum_{i=1}^{\infty} \frac{\sqrt{w - c + 2^n}}{2^n} < \infty,$$

що можна показати за допомогою аналізу збіжності відповідного числового ряду, наприклад, за допомогою ознаки д'Аламбера:

$$\lim_{n \rightarrow \infty} \frac{\sqrt{w - c + 2^{n+1}}}{2^{n+1}} \cdot \frac{2^n}{\sqrt{w - c + 2^n}} = \frac{1}{2} \sqrt{\lim_{n \rightarrow \infty} \frac{\sqrt{w - c + 2^{n+1}}}{\sqrt{w - c + 2^n}}} = \frac{1}{2} < 1.$$

Гравець погодиться взяти участь у грі, якщо

$$\sum_{i=1}^{\infty} \frac{\sqrt{w - c + 2^n}}{2^n} \leq \sqrt{w},$$

тобто якщо

$$\sum_{i=1}^{\infty} \frac{\sqrt{w - c + 2^n} - \sqrt{w}}{2^n} \geq 0.$$

Наприклад, чисельними методами можна підрахувати, що мільйонер ( $w = 1\,000\,000$ ) готовий буде заплатити до 22.87 за право взяти участь у грі (більша ціна вже вестиме до від'ємного сподіваного приросту в корисності), а власник  $w = 1\,000$  — до 12.93.

Оскільки функція корисності вгнута, згідно з (11.1.2) маємо

$$\mathbb{E}[u(X)] \leq u(\mathbb{E}[X]),$$

тобто гравець завжди надаватиме перевагу гарантованому виграшу в  $\mathbb{E}[X]$ , який несе корисність  $u(\mathbb{E}[X])$ , перед ризикованою ситуацією зі сподіваною корисністю  $\mathbb{E}[u(X)]$ . На цьому побудовано

принцип страхування. Наприклад, нехай особа має статки розміром  $w = 10\,000$ , які вона може втратити з імовірністю 0.05 (наприклад, у випадку землетрусу чи іншого стихійного лиха). Її функція корисності дорівнює  $u(x) = \sqrt{x}$ . Тоді її сподівана корисність дорівнюватиме

$$0.95 \cdot \sqrt{10\,000} + 0.05 \cdot \sqrt{0} = 95.$$

Звісно, особа воліла б убезпечити себе від такої ризикованої ситуації. Страхова фірма пропонує придбати поліс, який повністю покріє втрачені статки. Ціна полісу —  $c$ . У цьому випадку особа має вибір між сплатою полісу і дістанням  $\sqrt{10\,000 - c}$  корисності за будь-якого розкладу подій та ігноруванням страхової компанії і дістанням корисності 95. Відповідно, особа буде готова заплатити ціну  $c$ , для якої виконується

$$\sqrt{10\,000 - c} \geq 95 \Rightarrow c \leq 975.$$

□

### 11.1.3. Нерівність Гольдера

**Теорема 11.1.9** (Нерівність Гольдера (Hölder's inequality)<sup>5</sup>). Якщо  $p, q \in [1; \infty]$ , і  $\frac{1}{p} + \frac{1}{q} = 1$ , то

$$\mathbb{E}[|XY|] \leq \|X\|_p \|Y\|_q, \quad (11.1.3)$$

де

$$\|X\|_r = \begin{cases} (\mathbb{E}[|X|^r])^{\frac{1}{r}}, & r \in [1; \infty) \\ \inf \{M : \mathbb{P}(|X| > M) = 0\}, & r = \infty \end{cases}.$$

*Доведення.* Доведімо спочатку, що для  $1 \leq p, q < \infty$

$$|xy| \leq \frac{|x|^p}{p} + \frac{|y|^q}{q}.$$

Якщо  $x = 0$  чи  $y = 0$ , це твердження тривіально виконується, тому достатньо довести для випадку  $x > 0$  і  $y > 0$ , тобто що

$$xy \leq \frac{x^p}{p} + \frac{y^q}{q}.$$

Нехай  $y > 0$ . Тоді можемо визначити функцію  $\varphi : \mathbb{R}^+ \rightarrow \mathbb{R}$ :

$$\varphi(x) = \left( \frac{x^p}{p} + \frac{y^q}{q} - xy \right) \cdot \mathbb{1}\{x > 0\}.$$

Якщо продиференціювати цю функцію, дістанемо

$$\begin{aligned} \varphi'(x) &= x^{p-1} - y, \\ \varphi''(x) &= (p-1)x^{p-2} > 0. \end{aligned}$$

<sup>5</sup>Названа так на честь німецького математика Отто Гольдера (Ludwig Otto Hölder, 1859–1937), який довів її у 1889 р.

<sup>6</sup>Тобто мається на увазі, що  $p$  і  $q$  можуть бути в тому числі нескінченно великими.

Отже функція досягає свого мінімуму в точці

$$x^* = y^{\frac{1}{p-1}} .$$

Оскільки за умовою теореми  $\frac{1}{p} + \frac{1}{q} = 1$ , маємо

$$x^* = y^{\frac{q}{p}} ,$$

а отже

$$\varphi(x^*) = y^q \left( \frac{1}{p} + \frac{1}{q} \right) - y^{\frac{q}{p}+1} = y^q - y^q = 0 .$$

Оскільки  $x^*$  — точка мінімуму, для всіх інших точок маємо

$$\varphi(x^*) = 0 \leq \varphi(x) \Rightarrow xy \leq \frac{x^p}{p} + \frac{y^q}{q} .$$

Якщо  $\mathbb{E}[|X|^p] = 0$ , то  $|X|^p = 0$  майже напевно<sup>7</sup>, тобто  $|XY| = 0$  майже напевно, а отже  $\mathbb{E}[|XY|] = 0$ . Тому достатньо довести твердження для випадку  $\mathbb{E}[|X|^p] > 0$  і  $\mathbb{E}[|Y|^q] > 0$ . Застосуємо вищенаведену нерівність для  $x = \frac{X}{(\mathbb{E}[|X|^p])^{\frac{1}{p}}}$ ,  $y = \frac{Y}{(\mathbb{E}[|Y|^q])^{\frac{1}{q}}}$ :

$$\frac{|XY|}{(\mathbb{E}[|X|^p])^{\frac{1}{p}}(\mathbb{E}[|Y|^q])^{\frac{1}{q}}} \leq \frac{|X|^p}{p\mathbb{E}[|X|^p]} + \frac{|Y|^q}{q\mathbb{E}[|Y|^q]} ,$$

звідки за властивістю монотонності сподівання випливає

$$\frac{\mathbb{E}[|XY|]}{(\mathbb{E}[|X|^p])^{\frac{1}{p}}(\mathbb{E}[|Y|^q])^{\frac{1}{q}}} \leq \frac{\mathbb{E}[|X|^p]}{p\mathbb{E}[|X|^p]} + \frac{\mathbb{E}[|Y|^q]}{q\mathbb{E}[|Y|^q]} = \frac{1}{p} + \frac{1}{q} = 1 ,$$

або ж

$$\mathbb{E}[|XY|] \leq (\mathbb{E}[|X|^p])^{\frac{1}{p}}(\mathbb{E}[|Y|^q])^{\frac{1}{q}} .$$

Випадок із  $p = 1$ ,  $q = \infty$ <sup>8</sup> можна просто довести, адже оскільки фактично  $\|Y\|_q = \sup|Y|$ , то маємо

$$|XY| \leq |X| \cdot \sup|Y| ,$$

звідки за монотонністю сподівання маємо

$$\mathbb{E}[|XY|] \leq \mathbb{E}[|X|] \cdot \sup|Y| = \|Y\|_1 \cdot \|Y\|_\infty .$$

□

**Зауваження 11.1.10.** Частковим випадком нерівності Гольдера є нерівність Коши-Буняковського

<sup>7</sup>За великим рахунком, такої властивості явно ми не доводили, але це інтуїтивно зрозуміло і читач може перевірити себе в справедливості цього твердження.

<sup>8</sup>Аналогічно для  $p = \infty$ ,  $q = 1$ .

съкого<sup>9</sup> (Cauchy-Schwarz inequality). Так, поклавши  $p = q = 2$ , маємо

$$\mathbb{E} [|XY|] \leq \sqrt{\mathbb{E} [X^2]} \sqrt{\mathbb{E} [Y^2]} , \quad (11.1.4)$$

а з урахуванням Теореми 11.1.5 також

$$|\mathbb{E} [XY]| \leq \sqrt{\mathbb{E} [X^2]} \sqrt{\mathbb{E} [Y^2]} .$$

Альтернативну форму запису цієї нерівності можна дістати, якщо розглянути центровані випадкові величини

$$\tilde{X} = X - \mathbb{E} [X] , \quad \tilde{Y} = Y - \mathbb{E} [Y] ,$$

для яких справедливі такі властивості:

$$\begin{aligned} \mathbb{E} \left[ \left( \tilde{X} \right)^2 \right] &= \mathbb{E} \left[ (X - \mathbb{E} [X])^2 \right] = \text{Var} (X) , \\ \mathbb{E} \left[ \left( \tilde{Y} \right)^2 \right] &= \mathbb{E} \left[ (Y - \mathbb{E} [Y])^2 \right] = \text{Var} (Y) , \\ \mathbb{E} [\tilde{X} \tilde{Y}] &= \mathbb{E} [(X - \mathbb{E} [X])(Y - \mathbb{E} [Y])] = \text{Cov} (X, Y) . \end{aligned}$$

Отже

$$|\text{Cov} (X, Y)| = \left| \mathbb{E} [\tilde{X} \tilde{Y}] \right| \leq \sqrt{\mathbb{E} \left[ \left( \tilde{X} \right)^2 \right] \mathbb{E} \left[ \left( \tilde{Y} \right)^2 \right]} = \sqrt{\text{Var} (X) \text{Var} (Y)} .$$

□

**Приклад 11.1.11.** Розгляньмо застосування нерівності Коші-Буняковського до оцінки зверху ймовірності, що невід'ємна випадкова величина  $X$  дорівнює 0. Наприклад,  $X$  може позначати кількість неправильних відповідей під час тесту ( $\mathbb{P}_X (X = 0)$  буде ймовірністю написати тест ідеально) або число пар осіб, у яких спільний день народження ( $\mathbb{P}_X (X = 0)$  буде ймовірністю того, що всі дні народження різні).

Невід'ємну величину можна подати як добуток  $X = X \cdot \mathbb{1} \{X > 0\}$ , до якого можна застосувати (11.1.4):

$$\mathbb{E} [X] = \mathbb{E} [X \cdot \mathbb{1} \{X > 0\}] \leq \sqrt{\mathbb{E} [X^2] \mathbb{E} \left[ (\mathbb{1} \{X > 0\})^2 \right]} = \sqrt{\mathbb{E} [X^2] \mathbb{P}_X (X > 0)} ,$$

звідки випливає

$$\mathbb{P}_X (X > 0) \geq \frac{(\mathbb{E} [X])^2}{\mathbb{E} [X^2]} ,$$

а відтак

$$\mathbb{P}_X (X = 0) = 1 - \mathbb{P}_X (X > 0) \leq 1 - \frac{(\mathbb{E} [X])^2}{\mathbb{E} [X^2]} = \frac{\text{Var} (X)}{\mathbb{E} [X^2]} .$$

<sup>9</sup>У 1821 р. цю нерівність довів Коші для сум. У 1859 р. її аналог для інтегралів довів український математик Віктор Буняковський (1804–1889). Проте в англомовній літературі нерівність названо на честь німецького математика Германна Шварца (Karl Hermann Amandus Schwarz, 1843–1921), який довів її аж у 1888 р.

Наприклад, якщо  $X = \sum_{j=1}^n X_j$ , де  $X_j \sim \text{Bern}(p_j)$  незалежні,  $j = 1, \dots, n$ , маємо

$$\begin{aligned}\text{Var}(X) &= \sum_{j=1}^n \text{Var}(X_j) = \sum_{j=1}^n p_j(1-p_j) = \sum_{j=1}^n p_j - \sum_{j=1}^n p_j^2, \\ \mathbb{E}[X^2] &= \text{Var}(X) + (\mathbb{E}[X])^2 = \sum_{j=1}^n p_j - \sum_{j=1}^n p_j^2 + \left(\sum_{j=1}^n p_j\right)^2,\end{aligned}$$

а отже

$$\mathbb{P}_X(X = 0) \leq \frac{\sum_{j=1}^n p_j - \sum_{j=1}^n p_j^2}{\sum_{j=1}^n p_j - \sum_{j=1}^n p_j^2 + \left(\sum_{j=1}^n p_j\right)^2} \leq \frac{1}{1 + \sum_{j=1}^n p_j} = \frac{1}{1 + \mathbb{E}[X]}.$$

Іншими словами, у цьому конкретному випадку маємо, що що більше значення сподівання, то менш імовірним є нульове значення величини  $X$ .

Наприклад, нехай у кімнаті присутні  $n$  осіб. На практичному занятті ми з'ясували, як підрахувати ймовірність того, що принаймні в одній парі осіб буде спільний день народження. Розгляньмо модифіковану задачу підрахунку ймовірності того, що принаймні в одній парі осіб день народження буде *або спільний, або не відрізняється більше, ніж на один день*. Напряму цю задачу розв'язати доволі складно, тому ми спробуємо з'ясувати, яке повинно бути  $n$ , щоб гарантувати, що така ймовірність буде перевищувати 0.5.

Нехай  $X$  = «загальне число пар, у яких відстань між днями народження не перевищує 1 дня». Можна записати

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

де  $X_i = \mathbb{1}\{\text{у } i\text{-ої парі відстань між днями народження не перевищує 1 день}\}$ . Усього таких пар  $\binom{n}{2}$ . А ймовірність, що у деякої окремо взятої парі відстань між днями народження не перевищує 1 дня, дорівнює  $\frac{3}{365}$ , оскільки це сума трьох подій: або день народження одинаковий (імовірність  $\frac{1}{365}$ ), або день народження першого передує дню народження другого (імовірність  $\frac{1}{365}$ ), або день народження першого слідує за днем народження другого (імовірність  $\frac{1}{365}$ ).

Отже,

$$\mathbb{P}_X(X = 0) \leq \frac{1}{1 + \mathbb{E}[X]} = \frac{1}{1 + \binom{n}{2} \cdot \frac{3}{365}}.$$

Імовірність того, що принаймні в одній парі осіб відстань між днями народження не перевищує 1 дня, буде більша за 0.5, якщо  $\mathbb{P}_X(X = 0) \leq 0.5$ . Методом перебору значень  $n$  можна з'ясувати, що

$$\begin{aligned}\mathbb{P}_X(X = 0) &\leq 0.503, & n &= 16, \\ \mathbb{P}_X(X = 0) &\leq 0.472, & n &= 17.\end{aligned}$$

Отже, уже для  $n = 17$  імовірність того, що принаймні в одній парі осіб відстань між днями народження не перевищує 1 дня, гарантовано буде більша за 0.5. Варто відзначити, що ця ймовірність може перевищувати 0.5 і для менших значень  $n$ , але ми цього не можемо знати без повного розв'язання задачі. Нерівність дає нам гарантію саме для  $n = 17$ .  $\square$

### 11.1.4. Нерівності Маркова і Чебишова

У Прикладі 9.4.16 ми розглянули поняття середнього вибіркового значення  $\bar{X}$  для вибірки  $X_1, \dots, X_n$  незалежних величин з однаковим розподілом. Серед іншого, ми обчислили, що

$$\mathbb{E}[\bar{X}] = \mathbb{E}[X] , \quad \text{Var}(\bar{X}) = \frac{\text{Var}(X)}{n} .$$

Іншими словами, дисперсія середнього багатьох випадкових величин суттєво менша від дисперсії окремої величини, тобто що більше даних ми маємо, то точніше ми можемо оцінити сподівання випадкової величини. Проте на практиці нас часто цікавить, *наскільки* точно ми можемо оцінити сподівання, тобто як оцінити ймовірність похибки.

Як мотиваційний приклад розгляньмо підкидання правильної монетки. Нехай  $X_i = 1$ , якщо на  $i$ -ому випаді був герб, і  $X_i = -1$ , якщо число. Вочевидь,  $\mathbb{E}[X_i] = 0$ ,  $\text{Var}(X_i) = 1$ . Розгляньмо найгірший сценарій, коли монетка всі рази випала гербом, тоді  $\bar{X} = 1$ , що доволі далеко від справжнього сподівання. Розгляньмо тепер найліпший сценарій, коли монетка половину разів випала гербом, а половину — числом. Тоді, вочевидь,  $\bar{X} = 0$ , тобто середнє вибіркове дорівнює сподіванню.

На практиці нас цікавить поведінка явища у деякому *типовому сценарії*. Для цього потрібно розглянути деякі нерівності, які дають змогу оцінити ймовірність того, що випадкова величина відхиляється від свого сподівання.

**Теорема 11.1.12** (Нерівність Маркова (Markov's inequality)<sup>10</sup>). Нехай  $g : \mathbb{R} \rightarrow [0; \infty)$  — неспадна вимірна функція. Тоді для будь-якої випадкової величини  $X$  має місце

$$\mathbb{P}_X(X \geq c) \leq \frac{\mathbb{E}[g(X)]}{g(c)} , \quad c > 0 . \quad (11.1.5)$$

*Доведення.* Оскільки  $g$  неспадна і невід'ємна, для будь-якого  $c > 0$  справедливо

$$g(c) \cdot \mathbb{1}\{c \leq X\} \leq g(X) .$$

За властивістю (iii) з Твердження 6.4.6, маємо

$$\mathbb{E}[g(c) \cdot \mathbb{1}\{c \leq X\}] = g(c)\mathbb{P}_X(X \geq c) \leq \mathbb{E}[g(X)] ,$$

тобто

$$\mathbb{P}_X(X \geq c) \leq \frac{\mathbb{E}[g(X)]}{g(c)} .$$

□

**Зауваження 11.1.13.** Класичне формулювання нерівності Маркова можна дістати, якщо по-класті  $X \geq 0$  майже напевно,  $g(x) = x$ :

$$\mathbb{P}_X(X \geq c) \leq \frac{\mathbb{E}[X]}{c} , \quad c > 0 . \quad (11.1.6)$$

<sup>10</sup>Названа так на честь російського математика Андрея Маркова (1856–1922).

Можна прибрати вимогу про невід'ємність величини  $X$ , розглянувши величину  $|X|$ :

$$\mathbb{P}_X (|X| \geq c) \leq \frac{\mathbb{E}[|X|]}{c}, \quad c > 0.$$

Або, що те ж саме:

$$\mathbb{P}_X (|X| \geq c \cdot \mathbb{E}[|X|]) \leq \frac{1}{c}, \quad c > 0.$$

Суть нерівності Маркова проста: (невід'ємна) випадкова величина не може дуже сильно перевищувати своє сподівання.  $\square$

Із нерівності Маркова безпосередньо випливають деякі часткові випадки. Чи не найважливішим наслідком є *нерівність Чебишова* (Chebyshev's inequality)<sup>11</sup>, яку можна дістати, поклавши  $g(x) = x^2$  та розглянувши величину  $|X - \mathbb{E}[X]|$ :

$$\mathbb{P}_X (|X - \mathbb{E}[X]| \geq c) \leq \frac{\text{Var}(X)}{c^2}, \quad c > 0. \quad (11.1.7)$$

Альтернативна форма запису цієї нерівності:

$$\mathbb{P}_X (|X - \mathbb{E}[X]| \geq c \cdot \sigma_X) \leq \frac{1}{c^2}, \quad c > 0.$$

Іншими словами, випадкова величина не може суттєво відхилятися від свого сподівання на деяку відстань, що вимірюється в середньоквадратичних відхиляннях.

Якщо порівняти нерівності Маркова й Чебишова, то нерівність Чебишова пропонує точнішу оцінку, проте натомість вона вимагає існування скінченної дисперсії. Для нерівності Маркова достатньо було існування скінченного сподівання.

**Приклад 11.1.14.** У статистиці, як правило, мають справу з деяким набором даних, який відповідає реалізаціям деяких випадкових величин, справжній розподіл яких досліднику невідомий. Нерівність Чебишова (11.1.7) можна застосувати для визначення, якого розміру повинна бути вибірка, щоб середнє значення по цій вибірці гарантовано, з певною ймовірністю  $p$  лежало на певній відстані від справжнього сподівання.

Конкретніше, нехай маємо  $n$  випадкових величин  $X_1, \dots, X_n$ , які мають одинаковий розподіл і є незалежні. Ми не знаємо, який саме розподіл вони мають. Нехай  $\mathbb{E}[X_i] = \mu_X$ ,  $\text{Var}(X_i) = \sigma_X^2$ ,  $i = 1, \dots, n$ . Ми, знову ж таки, не знаємо, чому саме дорівнюють ці параметри, проте ми знаємо сподівання та дисперсію середнього вибіркового значення (Приклад 9.4.16), а саме  $\mathbb{E}[\bar{X}] = \mu_X$ ,  $\text{Var}(\bar{X}) = \frac{\sigma_X^2}{n}$ . До середнього вибіркового значення можна застосувати (11.1.7):

$$\mathbb{P}_{\bar{X}} \left( |\bar{X} - \mu_X| \geq c \cdot \frac{\sigma_X}{\sqrt{n}} \right) \leq \frac{1}{c^2} \Rightarrow \mathbb{P}_{\bar{X}} (|\bar{X} - \mu_X| \leq c \cdot \sigma_X) \geq 1 - \frac{1}{nc^2}.$$

Якщо ми будемо вимагати, щоб

$$1 - \frac{1}{nc^2} \geq p,$$

тобто що ймовірність потрапляння середнього вибіркового у проміжок відстанню  $c$  середньоквадратичних від справжнього сподівання  $\mu_X$  гарантовано перевищує деяку наперед задану ймовірність  $p$ .

<sup>11</sup>Названа так на честь російського математика Пафнютія Чебишова (1821–1894).

ність  $p$ , то ми повинні покласти

$$n \geq \frac{1}{c^2(1-p)},$$

незалежно від конкретних параметрів  $\mu_X$  та  $\sigma_X^2$ .

Наприклад, якщо ми хочемо, щоб середнє вибіркове значення відхилилося від справжнього сподівання не більше ніж на  $0.5\sigma_X$  (хоча ми не знаємо ні справжнього сподівання, ні середньо-квадратичного відхилення) з імовірністю  $p = 0.95$ , нам потрібно взяти вибірку розміром

$$n \geq \frac{1}{(0.5)^2(1-0.95)} = 80,$$

що насправді є дуже малим значенням, враховуючи, що в сучасну еру великих даних вибірки можуть мати декілька мільйонів значень.  $\square$

Варто зазначити, що нерівність Маркова (Чебишова) працює для всіх розподілів, а тому може бути не дуже точною. Для кожного конкретного розподілу, використовуючи його функцію розподілу, можна дістати точніші оцінки знизу.

**Приклад 11.1.15.** Нехай  $X \sim \text{Exp}(\lambda)$ . Оскільки  $\mathbb{E}[X] = \frac{1}{\lambda}$ ,  $\text{Var}(X) = \frac{1}{\lambda^2}$ , нерівність (11.1.7) дає для  $c > 1$  таку оцінку:

$$\mathbb{P}_X \left( \left| X - \frac{1}{\lambda} \right| \geq \frac{c}{\lambda} \right) \leq \frac{1}{c^2}.$$

Використовуючи функцію розподілу  $X$ , маємо

$$\begin{aligned} \mathbb{P}_X \left( \left| X - \frac{1}{\lambda} \right| \geq \frac{c}{\lambda} \right) &= \mathbb{P}_X \left( X - \frac{1}{\lambda} \leq -\frac{c}{\lambda} \right) + \mathbb{P}_X \left( X - \frac{1}{\lambda} \geq \frac{c}{\lambda} \right) \\ &= 1 - e^{-\lambda \max\{0, (\frac{1}{\lambda} - \frac{c}{\lambda})\}} + e^{-\lambda \max\{0, (\frac{1}{\lambda} + \frac{c}{\lambda})\}} \\ &= 1 - e^{-\max\{0, 1-c\}} + e^{-\max\{0, 1+c\}}. \end{aligned}$$

Ми використали позначення  $\max$  для того, щоб явно не розписувати окремо випадок, коли події  $X - \frac{1}{\lambda} \leq -\frac{c}{\lambda}$  і  $X - \frac{1}{\lambda} \geq \frac{c}{\lambda}$  є неймовірними, тобто коли  $1 - c \geq 0$  і  $1 + c \geq 0$  відповідно.

На Рис. 11.1.3 зображено графіки відповідних оцінок як функції від  $c > 1$  (Листинг 11.1.1). Як можна бачити, оцінка, яка випливає з формули функції експоненційного розподілу, значно точніша (для  $c = 1$  оцінка з нерівності Чебишова взагалі безглузда: вона говорить, що ймовірність не перевищує 1, тобто може бути будь-яка!), що особливо кидається у вічі для малих  $c$ . Що більше  $c$ , то менша ймовірність, що значення  $X$  буде дуже сильно відхилятися від свого сподівання, тому обидві криві прямують до 0.

Листинг 11.1.1: Код для Прикладу 11.1.15

```
import numpy as np
import matplotlib.pyplot as plt
from matplotlib import rc

5
def f(c):
    return np.minimum(1,
                      1 - np.exp(-np.maximum(0, 1 - c)) + np.exp(-np.maximum(0, c + 1)))
10
```

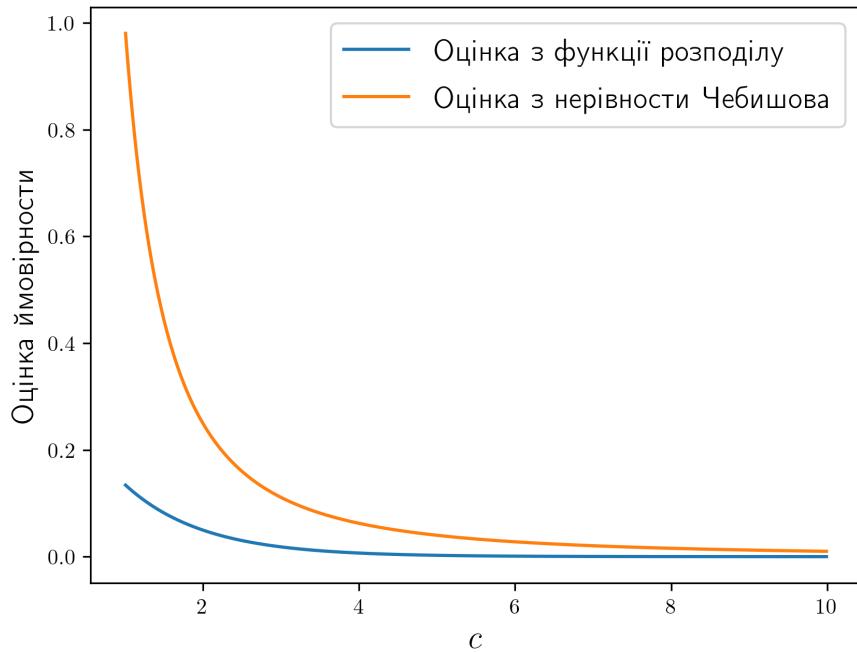


Рис. 11.1.3.: Графік верхніх оцінок імовірностей відхилень випадкової величини з експоненційним розподілом від свого сподівання на  $c$  середньоквадратичних відхилень як функція від  $c$  для Прикладу 11.1.15

```

def chebyshev(c):
    return np.minimum(1, 1 / (c**2))

15
rc('text', usetex=True)
rc('text.latex', unicode=True)
rc('text.latex', preamble=r'\usepackage[utf8]{inputenc}')
rc('text.latex', preamble=r'\usepackage[ukrainian]{babel}')

20
xx = np.arange(1.01, 10, step=0.01)

plt.plot(xx, f(xx), label="Оцінка з функції розподілу")
plt.plot(xx, chebyshev(xx), label="Оцінка з нерівності Чебишова")
25
plt.xlabel(r"$c$ ", fontsize=15)
plt.ylabel("Оцінка ймовірності", fontsize=15)
plt.xticks(fontsize=10)
plt.yticks(fontsize=10)
plt.legend(fontsize=15)
30
plt.savefig("../images/Lecture 11/expomarkov.png", dpi=300)

```

□

## 11.2. Міри центральної тенденції

Як ми зазначали раніше, розподіл  $\mathbb{P}_X$  деякої випадкової величини  $X$  повністю описує її функція розподілу  $F_X$ . Проте в низці як теоретичних, так і прикладних ситуацій достатньо дістати декілька ключових характеристик, які дають змогу описати випадкову величину.

Основною такою характеристикою, безсумнівно, є сподівання, яке має інтуїтивну інтерпретацію середньозваженого значення випадкової величини. Разом зі сподіванням також використовують деякі альтернативні характеристики зі схожою інтерпретацією, які сукупно мають назву *міри центральної тенденції* (measures of central tendency)<sup>12</sup>. До таких характеристик, окрім сподівання, також належать медіана і мода.

**Визначення 11.2.1.** Число  $M$  називають *медіаною* (median) випадкової величини  $X$  із розподілом  $\mathbb{P}_X$ , якщо

$$\mathbb{P}_X(X \leq M) \geq \frac{1}{2}, \quad \mathbb{P}_X(X \geq M) \geq \frac{1}{2}. \quad (11.2.1)$$

□

Найпростіша ситуація, коли одночасно виконуються ці дві нерівності — це коли  $\mathbb{P}(X \leq M) = \frac{1}{2}$ , тобто коли  $M \in F_X^{-1}(0.5)$ . Проте це справедливо тільки для неперервних функцій розподілу, які в принципі набувають значення 0.5, а не, скажімо, мають стрибок.

Інтуїтивно медіану можна інтерпретувати як таке значення  $M$ , що як справа, так і зліва від  $M$  зосереджена половина маси розподілу випадкової величини.

Варто зазначити, що медіана існує для будь-якого розподілу, чого не можна сказати про сподівання.

**Визначення 11.2.2.** Число  $m$  називають *модою* (mode) випадкової величини  $X$  із розподілом  $\mathbb{P}_X$ , якщо

- для дискретних величин —  $\mathbb{P}_X(X = m) \geq \mathbb{P}_X(X = x)$  для всіх  $x$ ;
- для (абсолютно) неперервних величин —  $f_X(m) \geq f_X(x)$  для всіх  $x$ .

□

Інтуїтивно моду можна інтерпретувати як таке значення  $m$ , у якому зосереджено найбільшу масу розподілу випадкової величини.

Медіана і мода, на відміну від сподівання, можуть бути не єдині. Наприклад, на Рис. 11.2.1 зображене щільність деякого розподілу, який має незліченно багато медіан  $M \in [-1; 1]$  та дві моди —  $m = \pm 3$ .

**Приклад 11.2.3.** Класичним прикладом, який ілюструє принципову відмінність сподівання від медіани, є приклад із заробітними платами працівників у певній компанії чи галузі промисловості.

Розгляньмо випадкову величину  $X$  = «заробітна плата випадково вибраного працівника». Це може бути як неперервна випадкова величина з деяким розподілом, якщо можливі значення

<sup>12</sup>У цьому контексті слово «міра» використано в загальнозважаному значенні, жодної мови про міру як функцію від множини тут немає.

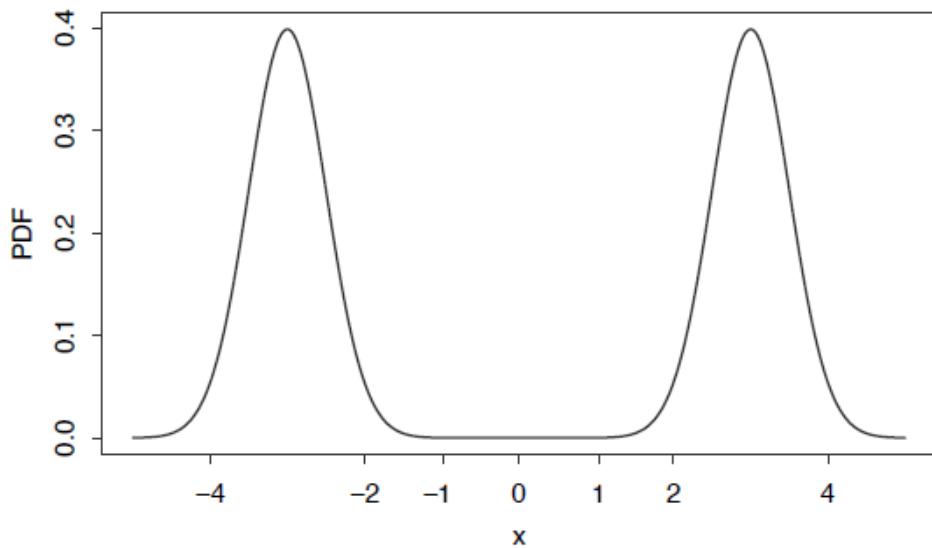


Рис. 11.2.1.: Приклад розподілу з двома модами та незліченою кількістю медіан ([1], Рис. 6.1)

заробітних плат можуть бути довільні, або дискретна, якщо існує деякий фіксований перелік посадових окладів, на основі яких можна обчислити скінчений набір можливих заробітних плат.

Нас цікавить відповідь на питання, яка зарплата є «типовою». Відповідь на це питання залежить від того, із якою метою ми цим цікавимося. На практиці різні числові характеристики розподілу, як і його графік (гістограма або стовпчикова діаграма), можуть пролити різне світло на природу досліджуваного явища, і, як правило, однією характеристикою ніхто не обмежується. Розгляньмо деякі відмінності в застосуванні трьох мір центральної тенденції.

Якщо  $X$  дискретна, і всі зарплати різні, мода не має особливої користі, оскільки кожна зарплата буде модою. Якщо ж зарплати потрапляють у певні категорії (наприклад, на кафедрі працюють старші викладачі, доценти і професори, у кожного з яких фіксований посадовий оклад), то мода може бути й єдина, проте що більш рівномірний розподіл зарплат, то менше інформації вона несе. Скажімо, якщо на кафедрі працюють 10 старших викладачів, 9 доцентів і 8 професорів, то модою буде зарплата старшого викладача, хоча вона зовсім не описує розподілу зарплат у жодному корисному сенсі. У таких випадках доцільно подавати графік (гістограму) розподілу.

Розгляньмо тепер медіану. Для величини  $X$  медіаною є значення, яке перебуває «посередині», тобто це такий рівень зарплати, що половина працівників отримує менше, а половина — більше цього рівня. Порівняно зі сподіванням медіана менш чутлива до викидів. Наприклад, якщо зарплата генерального директора значно перевищує зарплату працівників середньої ланки, то сподівання буде сильно завищено і не розкриватиме справжньої ситуації з зарплатами в компанії.

Розподіл заробітних плат, чи доходів у загальнішому випадку, у рамках деякої популяції (компанія, галузь промисловості, країна), як правило, є сильно асиметричним, тому застосування медіані для відображення «типової» зарплати є стандартною практикою.  $\square$

Для медіан справедлива цікава властивість, якої не мають сподівання.

**Твердження 11.2.4.** Нехай  $X$  — випадкова величина з носієм  $\text{supp}(X) = (a; b)$ . Нехай  $r$  — деяка взаємно однозначна функція  $r : (a; b) \rightarrow \mathbb{R}$ . Тоді якщо  $M$  є медіаною  $X$ , то  $r(M)$  є медіаною  $r(X)$ .

**Доведення.** Нехай  $Y = r(X)$ . Оскільки  $r$  взаємно однозначна, вона повинна бути або строго зростаюча, або строго спадна. Якщо  $r$  зростаюча, то  $Y \geq r(M)$  тоді й тільки тоді, коли  $X \geq M$ , а відтак

$$\mathbb{P}_Y(Y \geq r(M)) = \mathbb{P}_X(X \geq M) \geq \frac{1}{2}$$

за визначенням медіани  $M$ . Аналогічно

$$\mathbb{P}_Y(Y \leq r(M)) = \mathbb{P}_X(X \leq M) \geq \frac{1}{2}.$$

Якщо ж  $r$  спадна, то матимемо  $Y \geq r(M)$  тоді й тільки тоді, коли  $X \leq M$ , і доведення аналогічне.  $\square$

Варто звернути увагу, що для сподівань такої властивості не існує. Натомість ми маємо хіба що нерівність Єнсена.

Розгляньмо конкретний приклад обчислень мір центральної тенденції для дискретного розподілу.

**Приклад 11.2.5.** Розгляньмо послідовність незалежних підкидань *неправильної* монетки з імовірністю випаду герба, що дорівнює  $p$ . Підкидання тривають до випаду першого герба. Як ми знаємо, випадкова величина, яка описує час до появи першого герба, має геометричний розподіл зі щільністю (5.2.1):

$$\mathbb{P}_X(X = k) = (1 - p)^k p \cdot \mathbb{1}\{k \in \mathbb{Z}^+\}.$$

Одним із можливих питань, на які ми можемо дати відповідь — чому дорівнює «середнє» значення випадкової величини. Сподіванням цієї випадкової величини є  $\mathbb{E}[X] = \frac{1-p}{p}$ . Зокрема, якщо монетка правильна —  $p = 0.5$  — ми можемо сподіватися, у середньому, дістати  $\mathbb{E}[X] = 1$  випад числа перед першим випадом герба.

Альтернативно ми можемо поцікавитися, чому дорівнює значення випадкової величини, яке «лежить посередині», тобто чому дорівнює медіана  $M$ , яка задовольняє дві нерівності:

$$\begin{aligned} \mathbb{P}_X(X \leq M) &= \sum_{k=0}^M \mathbb{P}_X(X = k) \geq \frac{1}{2}, \\ \mathbb{P}_X(X \geq M) &= 1 - \mathbb{P}_X(X < M) = 1 - \sum_{k=0}^{M-1} \mathbb{P}_X(X = k) \geq \frac{1}{2}. \end{aligned}$$

Для геометричного розподілу, з урахуванням формули суми геометричної прогресії, маємо

$$\begin{aligned} \sum_{k=0}^M \mathbb{P}_X(X = k) &= \frac{p \cdot (1 - (1 - p)^{M+1})}{1 - (1 - p)} = 1 - (1 - p)^{M+1} \geq \frac{1}{2}, \\ \sum_{k=0}^{M-1} \mathbb{P}_X(X = k) &= \frac{p \cdot (1 - (1 - p)^M)}{1 - (1 - p)} = 1 - (1 - p)^M \leq \frac{1}{2}. \end{aligned}$$

Звідси випливає, що медіаною буде таке ціле число, для якого виконується

$$\frac{-\ln 2}{\ln(1 - p)} - 1 \leq M \leq \frac{-\ln 2}{\ln(1 - p)}.$$

Якщо  $p$  мале, то  $M \approx \frac{1}{2} \ln 2$  (що можна побачити з розвинення в ряд Тейлора). Якщо ж  $p = \frac{1}{2}$ , то  $M \in \{0, 1\}$ .

Нарешті, може постати питання обчислити «типове» значення випадкової величини, тобто його моду  $m = \arg \max_k \mathbb{P}_X (X = k)$ . Для геометричного розподілу, очевидно, модою є  $m = 0$ , оскільки функція ймовірності є строго спадною.  $\square$

### 11.3. Сподівання і медіана як найліпші предиктори

На практиці часто виринає потреба в *прогнозуванні* (prediction) значення випадкової величини  $X$ . Інтуїтивно зрозуміло, що «непоганими» прогнозами значеннями можуть бути як сподівання, так і медіана. Інше питання в тому, що вважати «непоганим» і в яких випадках доречніше застосовувати сподівання, а в яких — медіану.

У статистиці, зокрема регресійному аналізі, та економетриці сподівання відіграє роль найліпшого предиктора значення випадкової величини у розумінні мінімізації *середньоквадратичної похибки* (mean squared error).

**Твердження 11.3.1.** Сподівання інтегровної випадкової величини  $X$ ,  $\mathbb{E}[X]$  є найліпшим предиктором значення  $X$ , у тому сенсі, що воно мінімізує значення середньоквадратичної похибки:

$$\mathbb{E}[X] = \arg \min_{b \in \mathbb{R}} \mathbb{E}[(X - b)^2]. \quad (11.3.1)$$

*Доведення.* Доведення можна теоретично виконати шляхом підрахунку похідної, проте ми розглянемо метод, частіше використовуваний у теорії ймовірностей та статистиці:

$$\begin{aligned} \mathbb{E}[(X - b)^2] &= \mathbb{E}[(X - \mathbb{E}[X] + \mathbb{E}[X] - b)^2] \\ &= \mathbb{E}[(X - \mathbb{E}[X])^2] + 2\mathbb{E}[(X - \mathbb{E}[X])(\mathbb{E}[X] - b)] + (\mathbb{E}[X] - b)^2. \end{aligned}$$

Потрібно помітити, що

$$\mathbb{E}[(X - \mathbb{E}[X])(\mathbb{E}[X] - b)] = (\mathbb{E}[X] - b) \cdot \mathbb{E}[X - \mathbb{E}[X]] = (\mathbb{E}[X] - b) \cdot (\mathbb{E}[X] - \mathbb{E}[X]) = 0,$$

оскільки  $(\mathbb{E}[X] - b)$  є константою і її можна винести за знак сподівання.

Також варто помітити, що вираз  $\mathbb{E}[(X - \mathbb{E}[X])^2]$  взагалі не залежить від  $b$ , а тому не впливає на значення виразу, який ми мінімізуємо. Відтак  $\mathbb{E}[(X - b)^2]$  досягатиме найменшого значення, якщо  $(\mathbb{E}[X] - b)^2$  буде найменшим. Очевидно, це станеться, якщо  $b = \mathbb{E}[X]$ .  $\square$

**Зауваження 11.3.2.** Було б зовсім неправильно шукати таке значення  $b$ , яке мінімізує вираз  $(X - b)^2$ , оскільки цей вираз залежить від  $X$ . Натомість вираз  $\mathbb{E}[(X - b)^2]$  від  $X$  уже не залежить і є константою для будь-якого окремо взятого  $b$ .  $\square$

Сподівання  $\mathbb{E}[X]$  є найліпшим предиктором у тому сенсі, що воно мінімізує (середній, сподіваний) квадрат похибки між справжнім значенням змінної  $X$  та деякою константою  $b$ . Іншими словами, якби ми повторювали прогнозування деякої величини дуже велику кількість разів, щоразу використовуючи її сподівання як прогнозне значення, середній квадрат похибки був би найменший.

Медіана також мінімізує похибку, але не в середньоквадратичному, а в середньому абсолютному сенсі.

**Твердження 11.3.3.** Медіана  $M$  випадкової величини  $X$  є найліпшим предиктором значення  $X$ , у тому сенсі, що вона мінімізує значення *середньої абсолютної похибки* (mean absolute error):

$$M = \arg \min_{b \in \mathbb{R}} \mathbb{E} [|X - b|]. \quad (11.3.2)$$

*Доведення.* Нехай  $a \neq M$ . Потрібно показати, що

$$\mathbb{E} [|X - M|] \leq \mathbb{E} [|X - a|] \quad \Rightarrow \quad \mathbb{E} [|X - a| - |X - M|] \geq 0.$$

Розгляньмо випадок  $M < a$ :

- якщо  $X \leq M$ ,

$$|X - a| - |X - M| = a - X - (M - X) = a - M;$$

- якщо  $X > M$ ,

$$|X - a| - |X - M| = |X - a| - X + M.$$

Якщо  $X \geq a > M$ , то ця різниця обертається в  $M - a$ . Якщо ж  $M < X < a$ , то маємо

$$|X - a| - |X - M| = a + M - 2X \geq M - a.$$

Отже,

$$\begin{aligned} (|X - a| - |X - M|) \cdot \mathbb{1}\{X \leq M\} &= (a - M) \cdot \mathbb{1}\{X \leq M\}, \\ (|X - a| - |X - M|) \cdot \mathbb{1}\{X > M\} &\geq (M - a) \cdot \mathbb{1}\{X > M\}. \end{aligned}$$

Додаючи ці два вирази та беручи сподівання, маємо

$$\mathbb{E} [|X - a| - |X - M|] \geq (a - M) \cdot (\mathbb{P}(X \leq M) - \mathbb{P}(X > M)) = (a - M) \cdot (2\mathbb{P}(X \leq M) - 1).$$

Оскільки  $M$  — медіана,  $\mathbb{P}(X \leq M) \geq \frac{1}{2}$ , а відтак права частина нерівності невід'ємна, тобто  $\mathbb{E} [|X - a|] \geq \mathbb{E} [|X - M|]$ , що й потрібно було довести.

Аналогічно можна довести випадок  $M > a$ .  $\square$

На відміну від сподівання  $\mathbb{E}[X]$ , медіана є найліпшим предиктором у тому сенсі, що вона мінімізує (середню, сподівану) *відстань* між справжнім значенням змінної  $X$  та деякою константою  $b$ . Іншими словами, якби ми повторювали прогнозування деякої величини дуже велику кількість разів, щоразу використовуючи її медіану як прогнозне значення, середня *відстань* від справжнього значення була б найменшою.

**Приклад 11.3.4.** Розгляньмо державну лотерею, у якій щодня незалежно від попередніх у випадковий спосіб генерують три цифри від 0 до 9. Нехай за деякий тривалий період було згенеровано всі можливі комбінації, окрім однієї. Чому дорівнює прогнозне значення кількості днів, які потрібно прождати до появи останньої комбінації?

У цьому випадку  $X$  має геометричний розподіл із параметром  $p = 0.001$ . Сподіванням числом днів до появи останньої комбінації є  $\mathbb{E}[X] = \frac{1-0.001}{0.001} = 999$ . У той же час медіаною є число  $M$ , яке

задовільняє нерівність

$$\frac{-\ln 2}{\ln(1-p)} - 1 = 691.8 \leq M \leq \frac{-\ln 2}{\ln(1-p)} = 692.8,$$

тобто  $M = 692$ . Як можна бачити, для цього прикладу медіана суттєва менша від сподівання, оскільки геометричний розподіл суттєво скосений управо: дуже великі цілі числа мають дуже малу ймовірність, проте вони завищують значення сподівання.  $\square$

**Приклад 11.3.5.** Нехай оцінку деякого студента за екзамен можна виразити як число  $0 \leq X \leq 1$ . Нехай випадкова величина  $X$  має щільність розподілу

$$f_X(x) = \left( x + \frac{1}{2} \right) \cdot \mathbb{1}\{0 \leq x \leq 1\}.$$

Розгляньмо прогнозні значення  $X$  за двома розглянутими вище критеріями.

Середньоквадратичну похибку мінімізує сподівання:

$$\mathbb{E}[X] = \int_0^1 x \left( x + \frac{1}{2} \right) dx = \frac{7}{12} \approx 0.583.$$

Функцією розподілу  $X$  є

$$F_X(x) = \begin{cases} 0, & x < 0 \\ \int_0^x \left( t + \frac{1}{2} \right) dt = \frac{x^2 + x}{2}, & 0 \leq x < 1 \\ 1, & x \geq 1 \end{cases}.$$

Медіана задовільняє як

$$\mathbb{P}_X(X \leq M) = F_X(M) \geq \frac{1}{2},$$

так і

$$\mathbb{P}_X(X \geq M) = 1 - F_X(M) \geq \frac{1}{2} \Rightarrow F_X(M) \leq \frac{1}{2}.$$

Оскільки  $F_X$  строго зростаюча на  $(0; 1)$ , медіана  $M$  задовільняє рівність  $F_X(M) = \frac{1}{2}$ , звідки

$$\frac{M^2 + M}{2} = \frac{1}{2} \Rightarrow M = \frac{\sqrt{5} - 1}{2} \approx 0.618.$$

$\square$

## 11.4. Моменти вищих порядків

Раніше ми з'ясували, що для опису розподілу випадкової величини однієї числової характеристики, наприклад, сподівання, зовсім недостатньо, адже її значення може бути розкидано навколо сподівання. Це мотивувало нас увести поняття дисперсії. Проте, як виявляється, навіть дисперсії може бути недостатньо. На Рис. 11.4.1 зображені щільності двох, очевидно, зовсім різних

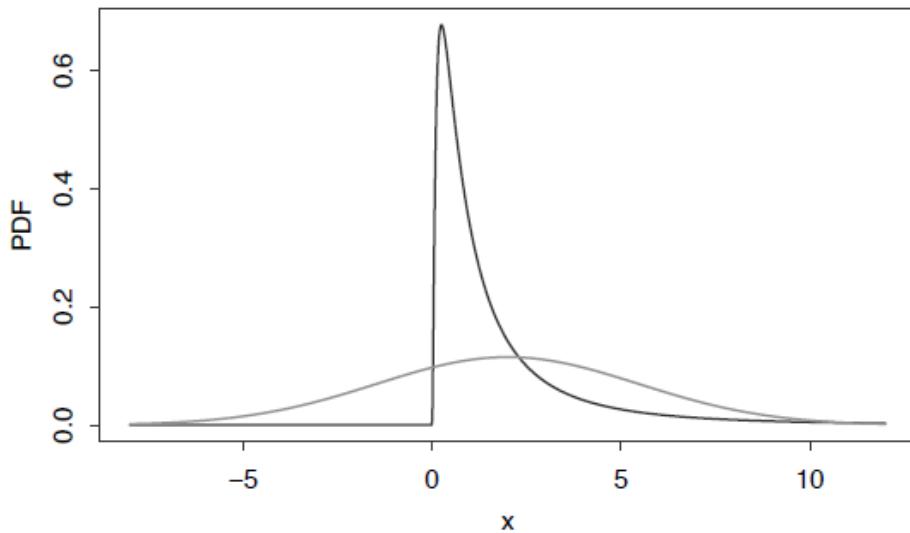


Рис. 11.4.1.: Приклад двох різних розподілів з одинаковим сподіванням 2 та одинаковою дисперсією 12: нормальній (світла крива) та логнормальній (темна крива) ([1], Рис. 6.2)

розподілів, які разом із тим мають одинакові сподівання й дисперсію. Світла крива відповідає нормальному розподілу  $N(2, 12)$ , а темна — логнормальному розподілу. У той час, як сподівання, медіана і мода нормального розподілу усі дорівнюють 2, медіана логнормального розподілу дорівнює 1, що свідчить про його сильну асиметричність.

На Рис. 11.4.2 зображені функції ймовірності випадкових величин  $X \sim \text{Binom}(10, 0.9)$  і  $Y \sim 8 + \text{Binom}(10, 0.1)$ . В обох цих розподілів і сподівання, і медіана, і мода дорівнюють 9, а дисперсія дорівнює 0.9, хоча ці розподіли мають посутні відмінності, зокрема, вони обидва асиметричні, але скошені в різні боки.

Але навіть якщо функції чи щільності розподілів симетричні, вони все одно можуть відрізнятися, навіть маючи одинакові сподівання й дисперсію. На Рис. 11.4.3 зображені щільності випадкових величин  $X \sim N(0, 1)$  і  $Y \sim \frac{1}{\sqrt{3}}t_3$ , які мають сподівання 0 та дисперсію 1, проте суттєво відрізняються в тому, наскільки швидко вони прямують до нуля (наскільки «товстими» є «хвости» відповідних щільностей).

У загальному випадку для будь-якої випадкової величини можна увести поняття моменту.

**Визначення 11.4.1.** Для випадкової величини  $X$  *абсолютним моментом  $k$ -го порядку* ( $k$ -th absolute moment) є

$$\mathbb{E} [|X|^k] = \int |x|^k d\mathbb{P}, \quad k = 1, 2, \dots . \quad (11.4.1)$$

Якщо (11.4.1) скінчений для деякого  $k$ , кажемо, що випадкова величина  $X$  має *момент  $k$ -го порядку* ( $k$ -th absolute moment), який дорівнює

$$\mathbb{E} [X^k] = \int x^k d\mathbb{P}, \quad k = 1, 2, \dots . \quad (11.4.2)$$

□

Очевидно, що відповідні формули для дискретних і (абсолютно) неперервних випадкових ве-

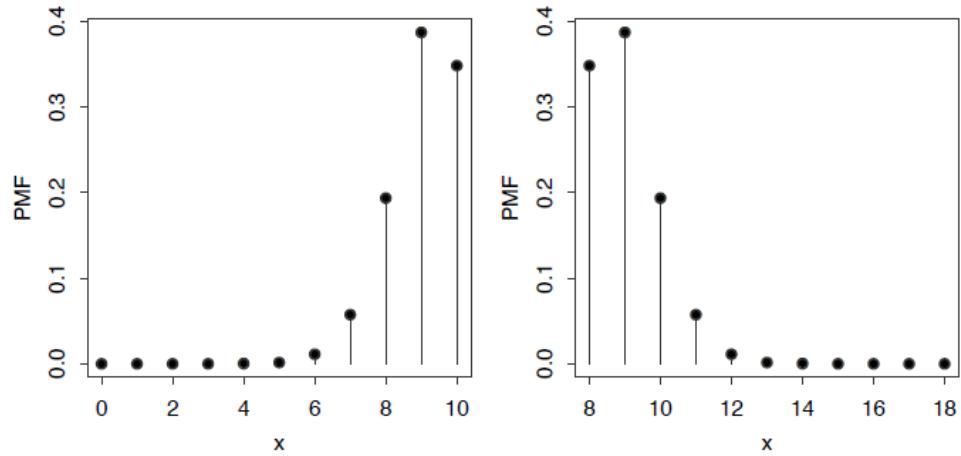


Рис. 11.4.2.: Функції ймовірності випадкових величин  $X \sim \text{Binom}(10, 0.9)$  (зліва) і  $Y \sim 8 + \text{Binom}(10, 0.1)$  (справа) ([1], Рис. 6.3)

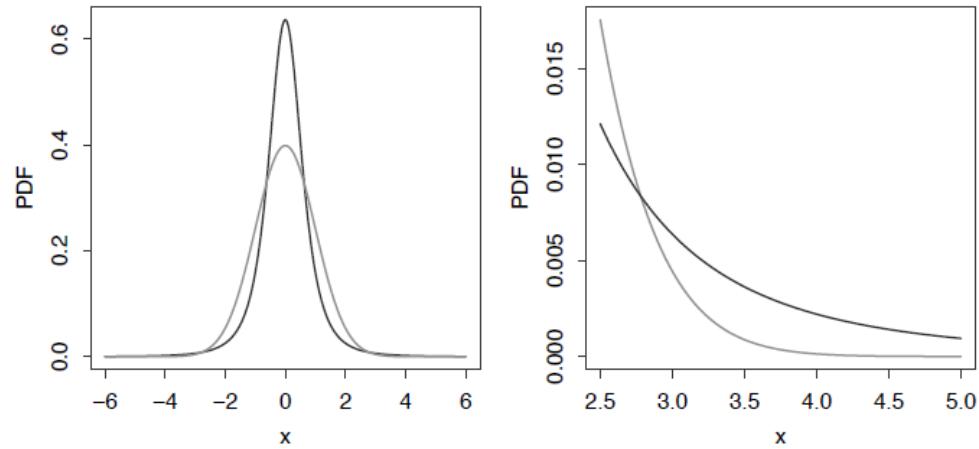


Рис. 11.4.3.: Щільності ймовірності випадкових величин  $X \sim N(0, 1)$  (світла крива) і  $Y \sim \frac{1}{\sqrt{3}}t_3$  (темна крива) та збільшений фрагмент відповідного графіка ([1], Рис. 6.4)

личин з функцією ймовірності  $p_X(x)$  або щільністю  $f_X(x)$  набувають виду

$$\begin{aligned}\mathbb{E}[X^k] &= \sum_{x \in \text{supp}(X)} x^k p_X(x), \\ \mathbb{E}[X^k] &= \int_{-\infty}^{\infty} x^k f_X(x) dx.\end{aligned}\tag{11.4.3}$$

**Визначення 11.4.2.** Простір усіх вимірних функцій  $f : (\Omega, \mathcal{A}, \mathbb{P}) \rightarrow \mathbb{R}$ , для яких існують моменти  $p$ -го порядку, називають *простором*  $L^p(\Omega, \mathcal{A}, \mathbb{P})$ , або просто  $L^p$ .  $\square$

**Твердження 11.4.3.** Якщо абсолютний момент  $k$ -го порядку скінчений, то скінчений також і абсолютний момент  $j$ -го порядку,  $j < k$ .

*Доведення.* Це випливає за монотонністю інтегралу з того факту, що якщо  $0 < j \leq k$ , то для всіх  $x \in \mathbb{R}$  виконується

$$|x|^j \leq \max \{1, |x|^k\} \leq 1 + |x|^k.$$

$\square$

Безпосереднім наслідком цього твердження є той факт, що якщо деякого моменту  $j$ -го порядку не існує, то не існує жодних моментів вищих порядків. Наприклад, відомо, що в розподілі Коші не існує сподівання, відтак у нього не існує жодних інших моментів вищих порядків.

**Приклад 11.4.4.** Нехай маємо  $X \sim N(0, 1)$ . Оскільки для будь-якого  $k \geq 1$

$$\lim_{n \rightarrow \infty} x^k e^{-\frac{x^2}{2}} = 0,$$

моменти всіх порядків існують. Інтегруючи частинами, дістанемо:

$$\frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} x^k e^{-\frac{x^2}{2}} dx = \frac{k-1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} x^{k-2} e^{-\frac{x^2}{2}} dx = (k-1) \mathbb{E}[X^{k-2}], \quad k = 2, 3, \dots.$$

Оскільки  $\mathbb{E}[X] = 0$ , маємо, що  $\mathbb{E}[X^3] = \mathbb{E}[X^5] = \dots = 0$ , тобто всі непарні моменти дорівнюють нулю.

Оскільки  $\mathbb{E}[X^0] = 1$ , маємо, що  $\mathbb{E}[X^2] = (2-1) \cdot 1 = 1$ ,  $\mathbb{E}[X^4] = (4-1) \cdot \mathbb{E}[X^2] = 3$  і т.д., тобто що

$$\mathbb{E}[X^{2n}] = 1 \cdot 3 \cdot 5 \cdot \dots \cdot (2n-1), \quad n = 1, 2, \dots.$$

$\square$

Окрім поняття моменту можна також розглянути поняття центрального моменту.

**Визначення 11.4.5.** Для випадкової величини  $X$  *абсолютним центральним моментом*  $k$ -го порядку ( $k$ -th absolute central moment) є

$$\mathbb{E}[|X - \mathbb{E}[X]|^k] = \int |x - \mathbb{E}[X]|^k d\mathbb{P}, \quad k = 1, 2, \dots.\tag{11.4.4}$$

Якщо (11.4.4) скінчений для деякого  $k$ , кажемо, що випадкова величина  $X$  має *центральний момент  $k$ -го порядку* ( $k$ -th central moment), який дорівнює

$$\mathbb{E}[(X - \mathbb{E}[X])^k] = \int (x - \mathbb{E}[X])^k d\mathbb{P}, \quad k = 1, 2, \dots. \quad (11.4.5)$$

□

Так, зокрема, сподівання є моментом першого порядку, а дисперсія — центральним моментом другого порядку. Центральний момент *третього* порядку має корисну інтерпретацію.

**Визначення 11.4.6.** Момент третього порядку стандартизованої випадкової величини  $X$  (якщо він існує) називають *коєфіцієнтом асиметрії* (skewness):

$$\text{Skew}(X) = \mathbb{E} \left[ \left( \frac{X - \mathbb{E}[X]}{\sqrt{\text{Var}(X)}} \right)^3 \right]. \quad (11.4.6)$$

□

Стандартизація у Визначенні 11.4.6 дає змогу не залежати ні від розташування, ні від масштабування випадкової величини, оскільки асиметричність розподілу від цих параметрів не повинна залежати.

**Визначення 11.4.7.** Розподіл деякої випадкової величини  $X$  є *симетричним* (symmetric) відносно деякого числа  $\mu$ , якщо випадкові величини  $X - \mu$  і  $\mu - X$  мають одинаковий розподіл. □

Для симетричних розподілів справедливо таке:

- $\mu = \mathbb{E}[X]$  (якщо сподівання існує), оскільки однаковість розподілу веде до  $\mathbb{E}[X] - \mu = \mu - \mathbb{E}[X]$ ;
- $M = \mathbb{E}[X]$ , що випливає з того факту, що для одинакових розподілів повинно виконуватися

$$\mathbb{P}(X - \mu \leq 0) = \mathbb{P}(\mu - X \leq 0) \Rightarrow \mathbb{P}(X \leq \mu) = \mathbb{P}(X \geq \mu),$$

звідки

$$\mathbb{P}(X \leq \mu) = 1 - \mathbb{P}(X > \mu) \geq 1 - \mathbb{P}(X \geq \mu) = 1 - \mathbb{P}(X \leq \mu) \Rightarrow \mathbb{P}(X \leq \mu) \geq \frac{1}{2}$$

і аналогічно для  $\mathbb{P}(X \geq \mu) \geq \frac{1}{2}$ .

**Твердження 11.4.8.** Якщо неперервна випадкова величина  $X$  має щільність  $f_X$ , її розподіл буде симетричний відносно  $\mu$  тоді й тільки тоді, коли  $f_X(x) = f_X(2\mu - x)$  для всіх  $x \in \text{supp}(X)$ .

*Доведення.* Через симетрію в розумінні Визначення 11.4.7,

$$F_X(x) = \mathbb{P}(X - \mu \leq x - \mu) = \mathbb{P}(\mu - X \leq x - \mu) = \mathbb{P}(X \geq 2\mu - x) = 1 - F_X(2\mu - x).$$

Беручи похідну, дістаємо

$$f_X(x) = f_X(2\mu - x) .$$

В інший бік доведення таке. Нехай для всіх  $x \in \text{supp}(X)$  виконується  $f_X(x) = f_X(2\mu - x)$ . Тоді:

$$\begin{aligned} \int_{-\infty}^{\mu+t} f_X(x) dx &= \mathbb{P}(X \leq \mu + t) = \mathbb{P}(X - \mu \leq t) , \\ \int_{-\infty}^{\mu+t} f_X(2\mu - x) dx &= \int_{\mu-t}^{\infty} f_X(u) du = \mathbb{P}(X \geq \mu - t) = \mathbb{P}(\mu - X \leq t) . \end{aligned}$$

Оскільки значення цих інтегралів повинні бути однакові, то виходить, що  $\mathbb{P}(X - \mu \leq t) = \mathbb{P}(\mu - X \leq t)$ , тобто відповідні величини мають одинаковий розподіл.  $\square$

Часто кажуть, що розподіл «просто» симетричний, маючи на увазі, що він симетричний відносно 0. У цьому випадку щільність (і функція ймовірності) буде парною функцією:  $f_X(x) = f_X(-x)$ .

Так от, власне, треті центральні моменти для симетричних розподілів дорівнюють нулю.

**Твердження 11.4.9.** Нехай випадкова величина  $X$  має розподіл, симетричний відносно  $\mathbb{E}[X]$ . Тоді будь-який непарний центральний момент  $\mathbb{E}[(X - \mathbb{E}[X])^m] = 0$ ,  $m = 2k + 1$ ,  $k = 0, 1, 2, \dots$  (якщо він існує).

*Доведення.* Оскільки  $X - \mathbb{E}[X]$  і  $\mathbb{E}[X] - X$  мають одинаковий розподіл, вони мають одинаковий непарний  $m$ -ий момент (якщо він існує):

$$\mathbb{E}[(X - \mathbb{E}[X])^m] = \mathbb{E}[(\mathbb{E}[X] - X)^m] .$$

Позначмо  $Y = (X - \mathbb{E}[X])^m$ . Тоді  $(\mathbb{E}[X] - X)^m = (-1)^m Y = -Y$ , отже

$$\mathbb{E}[Y] = \mathbb{E}[-Y] \Rightarrow \mathbb{E}[Y] = \mathbb{E}[(X - \mathbb{E}[X])^m] = 0 .$$

$\square$

На практиці перший центральний момент завжди дорівнює 0, тому третій стандартизований момент  $\text{Skew}(X)$  розглядають як міру асиметричності розподілу. Якщо розподіл симетричний, то згідно з Твердженням 11.4.9,  $\text{Skew}(X) = 0$ . Відтак якщо  $\text{Skew}(X) \neq 0$ , то розподіл не є симетричний, до того ж:

- додатні значення коефіцієнта асиметрії свідчать про скошеність розподілу у правий бік, тобто що його правий хвіст довший від лівого;
- від'ємні значення коефіцієнта асиметрії свідчать про скошеність розподілу у лівий бік, тобто що його лівий хвіст довший від правого.

Варто пам'ятати, що  $\text{Skew}(X) \neq 0$  є тільки достатньою умовою асиметричності: існують асиметричні розподіли, які мають нульовий третій стандартизований момент.

**Приклад 11.4.10.** Коефіцієнти асиметрії для деяких із відомих нам розподілів дорівнюють:

- для біномного розподілу  $\text{Binom}(n, p)$ :  $\text{Skew}(X) = \frac{1-2p}{\sqrt{np(1-p)}}$ . Якщо  $p = 0.5$ , коефіцієнт дорівнює 0, і розподіл симетричний. В інших випадках маємо асиметрію, до того ж якщо  $p < 0.5$ , то коефіцієнт додатний (розподіл скосений управо), а якщо  $p > 0.5$ , то коефіцієнт від'ємний (розподіл скосений уліво);

- для експоненційного розподілу  $\text{Exp}(\lambda)$ :  $\text{Skew}(X) = 2$  незалежно від значення  $\lambda$ . Розподіл завжди скошений управо;
- для Бета-розподілу  $\text{Beta}(a, b)$ :  $\text{Skew}(X) = \frac{2(b-a)\sqrt{a+b+1}}{(a+b+2)\sqrt{ab}}$ . Незважаючи на складність формули, принципово, що знак визначає різниця  $b - a$ . Якщо  $b > a$ , розподіл скошений управо, якщо  $b < a$ , розподіл скошений уліво, для  $a = b$  маємо симетричний (відносно 0.5) розподіл.

□

**Визначення 11.4.11.** Момент четвертого порядку стандартизованої випадкової величини  $X$  (якщо він існує), зсунутий на 3, називають *коєфіцієнтом ексцесу* (kurtosis):

$$\text{Kurt}(X) = \mathbb{E} \left[ \left( \frac{X - \mathbb{E}[X]}{\sqrt{\text{Var}(X)}} \right)^4 \right] - 3. \quad (11.4.7)$$

□

Коефіцієнт ексцесу дає змогу оцінити, наскільки швидко розподіл прямує до нуля. Потреба у відніманні трійки у (11.4.7) виникає з того, що коефіцієнт ексцесу стандартної нормальної випадкової величини, як було встановлено у Прикладі 11.4.4, дорівнює 3. Відповідно, якщо  $\text{Kurt}(X) > 0$ , відповідний розподіл прямує до нуля повільніше від стандартного нормального, а якщо  $\text{Kurt}(X) < 0$ , то швидше.

## 11.5. Квантилі

Окрім уже розглянутого набору числових характеристик випадкових величин розгляньмо додатково таке важливе поняття, як квантиль розподілу.

**Визначення 11.5.1.** Число  $x_p$  називають *p-им квантилем* (pth quantile) випадкової величини  $X$  із розподілом  $\mathbb{P}_X$ , якщо

$$\mathbb{P}_X(X \leq x_p) \geq p, \quad \mathbb{P}_X(X \geq x_p) \geq 1 - p. \quad (11.5.1)$$

□

Очевидно, що медіана є 0.5-квантилем випадкової величини.

Якщо функція розподілу неперервна і строго зростаюча, то  $x_p = F_X^{-1}(p)$  (Рис. 11.5.1). Проте в загальному випадку квантиль, як і медіана, не обов'язково має бути єдиним. Наприклад, для дискретних випадкових величин квантиль визначають так:

- якщо існує значення  $x_k \in \text{supp}(X)$  таке, що  $\mathbb{P}_X(X \leq x_k) = p$ , то  $x_p = (x_k + x_{k+1})/2$ ;
- якщо такого значення не існує, то  $x_p$  визначають як значення таке, що  $\mathbb{P}_X(X \leq x_p) > p$  і  $\mathbb{P}_X(X \geq x_p) > 1 - p$ , або, що еквівалентно,  $\mathbb{P}_X(X < x_p) < p$  і  $\mathbb{P}_X(X \leq x_p) > p$ .

Фактично,  $x_p$  ділить дійсну вісь на два проміжки:

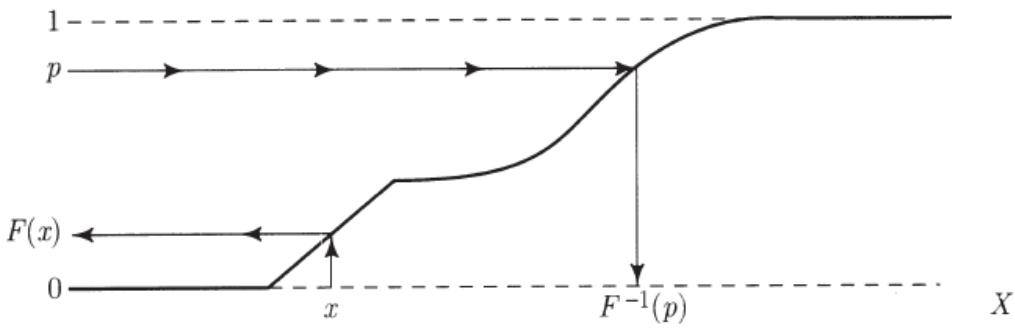


Рис. 11.5.1.: Ілюстрація поняття персентилю ([9], Рис. 4.5)

- $(-\infty; x_p]$ , який містить щонайменше  $100p\%$  маси розподілу;
- $[x_p; \infty)$ , який містить щонайменше  $100(1 - p)\%$  маси розподілу.

Деякі квантилі мають спеціальні назви:

- $x_{0.25}$  називають *першим квартілем* (first quartile);
- $x_{0.5}$ , як уже було зазначено, називають *медіаною*;
- $x_{0.75}$  називають *третім квартілем* (third quartile).

Квантиль  $x_p$  також називають *100p%-им персентилем* (100p%th percentile).

**Приклад 11.5.2.** Велика кількість студентів написали екзамен із певної дисципліни і дістали оцінки від 0 до 100. Нехай  $X$  — оцінка у випадковий спосіб вибраного студента. Якщо апроксимувати розподіл оцінок деяким неперервним розподілом, який має строго зростаючу функцію розподілу  $F_X$ , можемо підрахувати деякі цікаві характеристики розподілу оцінок.

Нехай медіанною оцінкою за екзамен є оцінка  $60 = F_X^{-1}(0.5)$ . Іншими словами, половина студентів дістали оцінку вище за 60, а інша половина — оцінку нижче за 60.

Якщо Іван дістав оцінку 72, його персентилем  $F_X(72)$  буде частка студентів, оцінка яких нижче 72. Оскільки  $72 > 60$ ,  $F_X(72) \in (0.5; 1)$ .

У загальному випадку,  $F_X(X)$  — персентиль випадково вибраного студента. Як ми розглянули в Розд. 8.1, хоча розподіл оцінок  $F_X$  може бути далекий від рівномірного, розподіл *персентилів* є стандартним рівномірним:  $F_X(X) \sim U((0; 1])$ . Наприклад, 10% студентів мають персентиль щонайменше від 0 до 0.1, 20% студентів мають персентиль від 0.7 до 0.9 і т.п.  $\square$

**Приклад 11.5.3.** Для випадкової величини  $X \sim \exp(\lambda)$  персентиль можна дістати з виразу із функції розподілу:

$$F_X(x) = 1 - e^{-\lambda x} = p \quad \Rightarrow \quad x_p = -\frac{1}{\lambda} \ln(1 - p) .$$

Для стандартного нормального розподілу третій квартиль можна дістати з виразу відповідної функції розподілу:  $x_{0.75} = \Phi^{-1}(0.75) \approx 0.675$ .  $\square$

## 12. Умовні сподівання та умовні розподіли

Поняття умовного сподівання і тісно пов'язаного з ним умовного розподілу, як і умовна ймовірність, розглянута в Лекції 3, становлять основу значної частини сучасної теорії ймовірностей.

Ідея умовного сподівання доволі проста. На початку ми не маємо жодної інформації про випадкову величину, а тільки знаємо її розподіл і можемо спрогнозувати її значення за допомогою сподівання. Проте якщо ми дістанемо деяку, нехай навіть часткову, інформацію про випадкову величину, ми фактично зможемо звузити простір  $\Omega$  чи  $\sigma$ -алгебру на ньому до класу множин, які сумісні з новою інформацією. Це вимагає від нас оновлення ймовірнісної міри  $\mathbb{P}$ , яку ми використовуємо для обчислення ймовірностей подій, що у свою чергу дає змогу обчислити сподівання величини з урахуванням додаткової інформації. У цьому випадку сподівання вже не буде сталим, оскільки залежатиме від тієї інформації, яку ми можемо дістати.

### 12.1. Умовні ймовірності і сподівання для дискретних випадкових величин

У Розд. 5.5 ми розглядали Визначення 5.5.1 умовної функції ймовірності. Повторімо його тут для спрощення викладу матеріалу. Отже, для будь-яких дискретних випадкових величин  $X$  і  $Z$  функцію  $p_{Z|X}(z | x) \equiv \mathbb{P}_{Z|X}(Z = z | X = x)$  називають *умовною функцією ймовірності* (conditional probability mass function) величини  $X$  за умови  $X = x$ .

Обчислити цю функцію можна за допомогою (5.5.1) як

$$\mathbb{P}_{Z|X}(Z = z | X = x) = \frac{\mathbb{P}_{Z,X}((Z = z) \cap (X = x))}{\mathbb{P}_X(X = x)}.$$

Якщо  $\mathbb{P}_X(X = x) = 0$  для деякого  $x$ , умовну ймовірність не визначено.

Як можна бачити, фактично ми «вирізаемо» зі спільної функції ймовірності частину, яка відповідає  $X = x$ , і додатково ділимо на  $\mathbb{P}_X(X = x)$  для нормалізації новоутвореної (умовної) функції ймовірності, щоб сума її значень дорівнювала 1. Відповідні міркування проілюстровано на Рис. 12.1.1.

Із цього визначення в повній аналогії з умовними ймовірностями безпосередньо випливає формула Беєса для функцій імовірності дискретних величин  $X$  та  $Y$ :

$$\mathbb{P}_{Z|X}(Z = k | X = n) = \frac{\mathbb{P}_{X|Z}(X = n | Z = k) \mathbb{P}_Z(Z = k)}{\mathbb{P}_X(X = n)}, \quad (12.1.1)$$

якщо, звісно, знаменник не обертається в нуль.

У повній аналогії з Визначенням 6.1.1 можна визначити *умовну функцію розподілу* (conditional distribution function):

$$F_{X|Z}(x | z) = \mathbb{P}_{X|Z}(X \leq x | Z = z). \quad (12.1.2)$$

Щоправда, для дискретних випадкових величин її застосування доволі обмежене, як і безумовної функції розподілу.

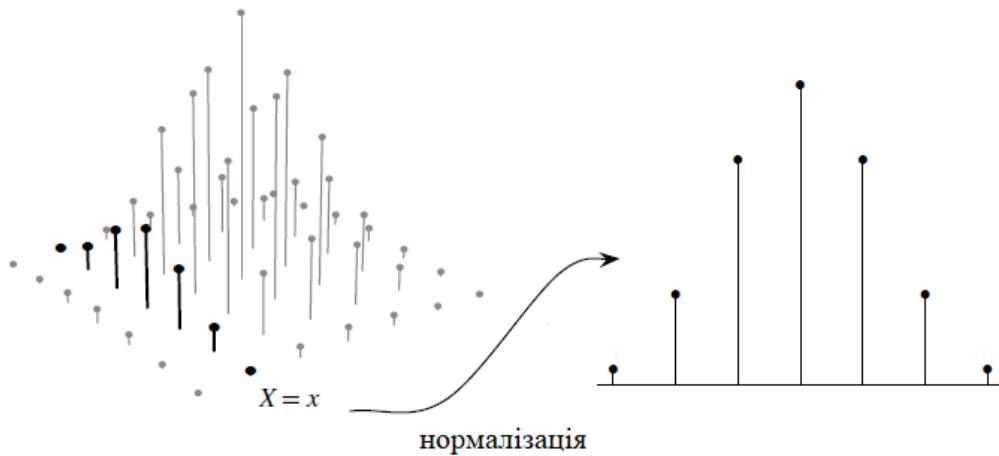


Рис. 12.1.1.: Ілюстрація умовної функції ймовірності ([1], Рис. 7.3)

Табл. 12.1.1.: Спільна функція розподілу для Прикладу 12.1.1

	$Y = 0$	$Y = 1$	$Y = 2$	$Y = 3$	Разом
$X = 0$	0.05	0.21	0	0	0.26
$X = 1$	0.20	0.26	0.08	0	0.54
$X = 2$	0	0.06	0.07	0.02	0.15
$X = 3$	0	0	0.03	0.02	0.05
Разом	0.25	0.53	0.18	0.04	1

Умовне сподівання для дискретних випадкових величин, як і сподівання безумовне, є зваженим середнім:

$$\mathbb{E}[X | Z = z] = \sum_{x \in \text{supp}(X)} x \cdot p_{X|Z}(x | z). \quad (12.1.3)$$

Варто звернути увагу, що умовне сподівання є функцією від  $z$ , а не сталою, тобто умовне сподівання є випадковою величиною.

**Приклад 12.1.1.** Кількість клієнтів у черзі до кожного з двох менеджерів у відділку банку описують дві відповідні випадкові величини  $X$  та  $Y$ . Спільну функцію розподілу відповідних величин наведено в Табл. 12.1.1.

За допомогою (5.5.1) можна обчислити, наприклад,

$$\begin{aligned} \mathbb{P}_{X|Y}(X = 0 | Y = 2) &= \frac{\mathbb{P}_{X,Y}(X = 0, Y = 2)}{\mathbb{P}_Y(Y = 2)} = \frac{0}{0.18} = 0, \\ \mathbb{P}_{X|Y}(X = 1 | Y = 2) &= \frac{\mathbb{P}_{X,Y}(X = 1, Y = 2)}{\mathbb{P}_Y(Y = 2)} = \frac{0.08}{0.18} \approx 0.444, \\ \mathbb{P}_{X|Y}(X = 2 | Y = 2) &= \frac{\mathbb{P}_{X,Y}(X = 2, Y = 2)}{\mathbb{P}_Y(Y = 2)} = \frac{0.07}{0.18} \approx 0.389, \\ \mathbb{P}_{X|Y}(X = 3 | Y = 2) &= \frac{\mathbb{P}_{X,Y}(X = 3, Y = 2)}{\mathbb{P}_Y(Y = 2)} = \frac{0.03}{0.18} \approx 0.167. \end{aligned}$$

Відтак, скажімо,

$$\mathbb{E}[X \mid Y = 2] \approx 0 \cdot 0 + 1 \cdot 0.444 + 2 \cdot 0.389 + 3 \cdot 0.167 \approx 1.723 .$$

□

**Приклад 12.1.2.** У Розд. 5.1 ми зазначали, що розподіл Пуассона є границею біномного розподілу. Між цими розподілами також є й інша форма зв'язку: біномний розподіл випливає з розподілу Пуассона шляхом обумовлення.

Нехай  $X \sim \text{Pois}(\lambda_1)$ ,  $Y \sim \text{Pois}(\lambda_2)$ ,  $X \perp\!\!\!\perp Y$ . За формулою згортки маємо  $X+Y \sim \text{Pois}(\lambda_1 + \lambda_2)$ . Відтак за формулою Беєса маємо

$$\begin{aligned} \mathbb{P}_{X|X+Y}(X = k \mid X + Y = n) &= \frac{\mathbb{P}_{X+Y|X}(X + Y = n \mid X = k) \mathbb{P}_X(X = k)}{\mathbb{P}_{X+Y}(X + Y = n)} \\ &= \frac{\mathbb{P}_Y(Y = n - k) \mathbb{P}_X(X = k)}{\mathbb{P}_{X+Y}(X + Y = n)} \\ &= \frac{n!}{e^{-(\lambda_1 + \lambda_2)} (\lambda_1 + \lambda_2)^n} \cdot \frac{e^{-\lambda_2} \lambda_2^{n-k}}{(n - k)!} \cdot \frac{e^{-\lambda_1} \lambda_1^k}{k!} \\ &= \binom{n}{k} \left( \frac{\lambda_1}{\lambda_1 + \lambda_2} \right)^k \left( \frac{\lambda_2}{\lambda_1 + \lambda_2} \right)^{n-k} . \end{aligned}$$

Іншими словами,  $X \mid X + Y = n \sim \text{Binom}\left(n, \frac{\lambda_1}{\lambda_1 + \lambda_2}\right)$ .

□

## 12.2. Умовна ймовірність у загальному випадку

### 12.2.1. Обумовлення $\sigma$ -алгеброю, породженою розбиттям

Раніше ми розглядали визначення умовної ймовірності відповідно до (3.1.1), де ймовірність події  $A$  з імовірнісного простору  $(\Omega, \mathcal{A})$  за умови настання події  $B$  дорівнює

$$\mathbb{P}(A \mid B) = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)}, \quad \mathbb{P}(B) > 0 .$$

Іншими словами, якщо нам відомо, що експеримент завершився подією  $B$ , то ймовірність події  $A$  можна обчислити за допомогою ймовірнісної міри  $\mathbb{P}(\cdot \mid B)$ . Якщо ж, скажімо, нам стане відомо, що експеримент завершився подією  $B^c$ ,  $\mathbb{P}(B^c) > 0$ , то ми зможемо застосувати іншу міру —  $\mathbb{P}(\cdot \mid B^c)$ . Для компактності можна увести функцію  $f : \Omega \rightarrow [0; 1]$  таку, що

$$f(\omega) = \begin{cases} \mathbb{P}(A \mid B) & , \quad \omega \in B \\ \mathbb{P}(A \mid B^c) & , \quad \omega \in B^c \end{cases} . \quad (12.2.1)$$

Іншими словами, ми спостерігаємо, якою саме подією,  $B$  або  $B^c$ , завершився експеримент, і застосовуємо функцію (12.2.1) залежно від наших спостережень. Варто зазначити, що ми навіть можемо не знати, який *саме* результат експерименту  $\omega$  мав місце: достатньо знати, яка з двох подій сталася.

Оскільки  $B$  і  $B^c$  утворюють розбиття  $\Omega$ , відповідні міркування можна узагальнити на випадок довільного розбиття. Нехай  $B_1, B_2, \dots$  — скінчнене або зліченне розбиття простору  $\Omega$ , яке складається з ненульових подій. Безпосереднім узагальненням (12.2.1) буде функція

$$f(\omega) = \mathbb{P}(A | B_i) = \frac{\mathbb{P}(A \cap B_i)}{\mathbb{P}(B_i)}, \quad \omega \in B_i, \quad i = 1, 2, \dots. \quad (12.2.2)$$

Як і раніше, якщо ми спостерігаємо, якою саме подією  $B_i$  завершився експеримент, ми можемо застосовувати (12.2.2).

Це дає нам змогу розглянути  $\sigma$ -алгебру  $\mathcal{A}_B$  зі зліченних об'єднань подій  $B_i, i = 1, 2, \dots$ <sup>1</sup>, і увести поняття ймовірності, обумовленої  $\sigma$ -алгеброю  $\mathcal{A}_B$ , яку позначають через  $\mathbb{P}(\cdot | \mathcal{A}_B)$  і яка дорівнює функції  $f(\omega)$  з (12.2.2).

У загальному випадку розбиття простору може містити нульові події, тому для повноти визначення потрібно задати значення, яких  $\mathbb{P}(\cdot | \mathcal{A}_B)$  набуває для нульових подій, тобто якщо  $\mathbb{P}(B_i) = 0$  для деякого  $i$ . У цьому випадку можна покласти  $\mathbb{P}(A | \mathcal{A}_B) = c$  для деякої сталої  $c \in \mathbb{R}$ . Іншими словами,  $\mathbb{P}(A | \mathcal{A}_B)$  насправді є не функцією, а сім'єю функцій, які відрізняються конкретним значенням сталої  $c$ . Усі вони, проте, дорівнюють одна одній майже напевно, оскільки відрізняються тільки на нульових подіях.

**Приклад 12.2.1.** Розгляньмо знову процес Пуассона, описаний у Розд. 8.5. Нехай  $0 \leq s \leq t$  і нехай  $A = \{N_s = 0\}, B_i = \{N_t = i\}, i = 0, 1, \dots$ . Через незалежність настання подій у цьому процесі в неперетинані інтервали часу маємо, що

$$\begin{aligned} \mathbb{P}(A | B_i) &= \mathbb{P}(N_s = 0 | N_t = i) \\ &= \frac{\mathbb{P}(\{N_s = 0\} \cap \{N_t = i\})}{\mathbb{P}(N_t = i)} \\ &= \frac{\mathbb{P}(\{N_s = 0\} \cap \{N_t - N_s = i\})}{\mathbb{P}(N_t = i)} \\ &= \frac{\mathbb{P}(N_s = 0) \mathbb{P}(N_t - N_s = i)}{\mathbb{P}(N_t = i)}. \end{aligned}$$

Оскільки число подій в інтервалі довжиною  $x$  має розподіл Пуассона з параметром  $\lambda x$ , маємо:

$$\mathbb{P}(A | B_i) = \frac{(\lambda s)^0 e^{-\lambda s}}{0!} \cdot \frac{(\lambda(t-s))^i e^{-\lambda(t-s)}}{i!} \cdot \frac{i!}{(\lambda t)^i e^{-\lambda t}} = \left(1 - \frac{s}{t}\right)^i.$$

А оскільки на множині  $B_i$  ми маємо  $i = N_t(\omega)$ , остаточно дістаемо

$$\mathbb{P}(N_s = 0 | \mathcal{A}_B) = \left(1 - \frac{s}{t}\right)^{N_t(\omega)}.$$

У цьому прикладі ми спостерігаємо, наприклад, число телефонних дзвінків, що сталися в часовому проміжку  $[0; t]$ . Якщо ми знаємо загальне число дзвінків, але не конкретні часові моменти їх настання, вираз  $\mathbb{P}(N_s = 0 | \mathcal{A}_B)$  дає змогу підрахувати ймовірність того, що *жоден* із дзвінків не стався до моменту  $s$ . І хоча ми не знаємо конкретного значення  $\omega$ , ми знаємо  $N_t(\omega)$ , що дає нам змогу підрахувати відповідну умовну ймовірність.  $\square$

<sup>1</sup>Зверніть увагу, що ми явно не зазначаємо, що ми також використовуємо заперечення цих множин, адже природа розбиття така, що заперечення будь-якого елемента розбиття є об'єднанням усіх інших, наприклад,  $B_1^c = B_2 \cup B_3 \cup \dots$

### 12.2.2. Обумовлення довільною $\sigma$ -алгеброю

Оскільки  $\mathcal{A}_B$  є  $\sigma$ -алгеброю, породженою розбиттям  $B_1, B_2, \dots$  простору  $(\Omega, \mathcal{A}, \mathbb{P})$ , деякий загальний елемент  $E \in \mathcal{A}_B$  буде (не більш ніж зліченним) об'єднанням окремих множин  $B_{i_1} \cup B_{i_2} \cup \dots$  для деякої підмножини індексів  $i_1, i_2, \dots \in \{1, 2, \dots\}$ . Для застосування (12.2.2) потрібно знати, якій саме з множин  $B_i$  належить результат експерименту  $\omega$ . Це те саме, що знати, яким множинам із  $\mathcal{A}_B$  результат  $\omega$  належить, а яким не належить.

Відповідні міркування можна узагальнити на випадок *довільної  $\sigma$ -алгебри*  $\mathcal{A}' \subseteq \mathcal{A}$ , необов'язково породженої розбиттям простору  $\Omega$ .

Уявімо, що експеримент завершився деяким результатом  $\omega$ , який ми *не спостерігаємо*. Натомість для кожної множини  $G \in \mathcal{A}'$  ми знаємо, чи належить їй  $\omega$ , тобто ми знаємо значення  $\mathbb{1}_{\{\omega \in G\}}$  для всіх  $G \in \mathcal{A}'$ . Інтуїтивно ми можемо це інтерпретувати як володіння «інформацією» про  $\omega$ , і цю «інформацію» ми асоціюємо з  $\sigma$ -алгеброю  $\mathcal{A}'$ . Понад те, якщо ми знаємо  $\omega \in G$ , то що *менша*  $G$ , то *більше* інформації ми маємо про  $\omega$ . Але якщо ми знаємо  $\mathbb{1}_{\{\omega \in G\}}$  для всіх  $G \in \mathcal{A}'$ , то що *більша*  $\mathcal{A}'$ , то *більше* інформації ми маємо про  $\omega$ .

Для деякої фіксованої  $A \in \mathcal{A}$  можемо задати міру  $\nu$  на  $\mathcal{A}'$  таку, що

$$\nu(G) = \mathbb{P}(A \cap G), \quad G \in \mathcal{A}'.$$

Звідси випливає  $\mathbb{P}(G) = 0 \Rightarrow \nu(G) = 0$ . Відтак, виконуються всі умови Теореми 6.3.2 для мір  $\nu$  і  $\mathbb{P}$  на вимірному просторі  $(\Omega, \mathcal{A}')$ . Тому існує  $f$ , вимірна відносно  $\mathcal{A}'$  та інтегровна за  $\mathbb{P}$  така, що

$$\mathbb{P}(A \cap G) = \nu(G) = \int_G f d\mathbb{P}, \quad G \in \mathcal{A}'.$$

Це і є умовна ймовірність.

**Визначення 12.2.2.** Нехай маємо вимірні простори  $(\Omega, \mathcal{A})$  і  $(\Omega, \mathcal{A}')$ ,  $\mathcal{A}' \subseteq \mathcal{A}$ . Тоді *умовною ймовірністю за умови  $\mathcal{A}'$*  (conditional probability given  $\mathcal{A}'$ ) є випадкова величина  $\mathbb{P}(A | \mathcal{A}')$  така, що:

(i) вона вимірна відносно  $\mathcal{A}'$  та інтегровна;

(ii) її інтеграл дорівнює

$$\int_G \mathbb{P}(A | \mathcal{A}') d\mathbb{P} = \mathbb{P}(A \cap G), \quad G \in \mathcal{A}'. \quad (12.2.3)$$

□

Визначення 12.2.2 можна порівняти з (3.1.2), де

$$\mathbb{P}(A \cap G) = \mathbb{P}(A | G) \mathbb{P}(G).$$

Тобто якщо ми обумовлюємо однією подією  $G$ , імовірність перетину двох подій дорівнює добутку умовної та безумовної ймовірностей. Якщо ж ми обумовлюємо цілою  $\sigma$ -алгеброю  $\mathcal{A}'$ , імовірність перетину дорівнює інтегралу умовної ймовірності на відповідній множині  $G$ , що можна інтерпретувати як узагальнення суми добутків умовних і безумовних імовірностей для всіх можливих множин із  $\sigma$ -алгебри  $\mathcal{A}'$ , які перетинаються з  $G$ .

Формулювання умовної ймовірності у (12.2.3) корисне тим, що враховує випадок, коли обумовлювальна подія є нульовою. Як ми казали вище, у цьому випадку значення умовної ймовірності

може бути абсолютно довільною сталою. Справді, якщо  $\mathbb{P}(G) = 0$ , матимемо

$$0 = \mathbb{P}(A \cap G) = \mathbb{P}(A | \mathcal{A}') \cdot \int_G 1 d\mathbb{P} = \mathbb{P}(A | \mathcal{A}') \cdot 0 = 0,$$

тобто  $\mathbb{P}(A | \mathcal{A}')$  справді може бути будь-якою сталою для такого випадку.

Звісно, з такого визначення випливає, що умовна ймовірність не єдина, але всі випадкові величини, які задовольняють (12.2.3), рівні майже напевно. Деяка конкретна величина, яка задовольняє (12.2.3), має називу *версії* (version) умовної ймовірності.

**Зауваження 12.2.3.** Умова (i) із Визначення 12.2.2 вимагає, щоб значення  $\mathbb{P}(A | \mathcal{A}')$  залежали тільки від подій в  $\mathcal{A}'$ . Якщо ми спостерігаємо для кожної події  $G \in \mathcal{A}'$ , належить їй результат експерименту чи не належить, то ми знаємо це в тому числі й для множини  $\{\omega : \mathbb{P}(A | \mathcal{A}') = x\}$  для деякого  $x \in [0; 1]$ . Іншими словами, ми знаємо значення умовної ймовірності —  $x$  — навіть не знаючи самого  $\omega$ .

□

**Зауваження 12.2.4.** Нехай  $A \in \mathcal{A}'$ . Зокрема, це завжди виконуватиметься, якщо  $\mathcal{A}' = \mathcal{A}$ . Тоді  $\mathbb{1}_A$  задовольняє обидві умови з Визначення 12.2.2, і  $\mathbb{P}(A | \mathcal{A}') = \mathbb{1}_A$  (майже напевно). Отже інформація про  $\mathcal{A}' = \mathcal{A}$  в цьому випадку тотожна інформації про настання події  $A$ .

Нехай тепер  $\mathcal{A}' = \{\emptyset, \Omega\}$ . Тоді  $\mathbb{P}(A | \mathcal{A}') = \mathbb{P}(A)$  для всіх  $\omega \in \Omega$ , і така  $\mathcal{A}'$  не несе жодної інформації.

Отже, всі можливі умовні ймовірності лежать між найменш інформативною сталою  $\mathbb{P}(A)$  для найменшої  $\sigma$ -алгебри  $\mathcal{A}'$  та найбільш інформативною індикаторною функцією  $\mathbb{1}_A$  для найбільшої  $\sigma$ -алгебри  $\mathcal{A}$ .

□

Потрібно звернути увагу, що вимірна функція, яка задовольняє умови з Визначення 12.2.2, хоч і має називу умовної ймовірності, за великим рахунком не є ймовірнісною мірою, оскільки це просто позначення, за яким ховається ціла сім'я функцій. Проте можемо неформально казати, що умовна ймовірність — це справжня ймовірність, оскільки для неї виконуються всі властивості ймовірнісної міри *майже напевно*.

**Твердження 12.2.5.** Нехай маємо вимірні простори  $(\Omega, \mathcal{A})$  і  $(\Omega, \mathcal{A}')$ ,  $\mathcal{A}' \subseteq \mathcal{A}$ . Тоді умовна ймовірність  $\mathbb{P}(A | \mathcal{A}')$  задовольняє такі властивості майже напевно:

(i)  $0 \leq \mathbb{P}(A | \mathcal{A}') \leq 1$  для будь-якої  $A \in \mathcal{A}$ ;

(ii) якщо  $A_1, A_2, \dots$  є несумісні, то

$$\mathbb{P}\left(\bigcup_{n=1}^{\infty} A_n | \mathcal{A}'\right) = \sum_{n=1}^{\infty} \mathbb{P}(A_n | \mathcal{A}').$$

*Доведення.* Для будь-якої версії умовної ймовірності відповідно до (12.2.3) виконується

$$\int_G \mathbb{P}(A | \mathcal{A}') d\mathbb{P} = \mathbb{P}(A \cap G) \geq 0,$$

оскільки ймовірність як міра невід'ємна. Оскільки функція  $\mathbb{P}(A | \mathcal{A}')$  вимірна відносно  $\mathcal{A}'$ , вона також повинна бути невід'ємна майже напевно. Аналогічно можна довести нерівність з одиницею.

Нехай  $A_1, A_2, \dots$  є несумісні. Якщо  $G \in \mathcal{A}'$ , то з Теореми 11.1.2 випливає

$$\begin{aligned} \int_G \sum_{n=1}^{\infty} \mathbb{P}(A_n | \mathcal{A}') d\mathbb{P} &= \sum_{n=1}^{\infty} \int_G \mathbb{P}(A_n | \mathcal{A}') d\mathbb{P} \\ &= \sum_{n=1}^{\infty} \mathbb{P}(A_n \cap G) \\ &= \mathbb{P}\left(\left(\bigcup_{n=1}^{\infty} A_n\right) \cap G\right). \end{aligned}$$

Отже  $\sum_{n=1}^{\infty} \mathbb{P}(A_n | \mathcal{A}')$  є вимірною відносно  $\mathcal{A}'$  і задовольняє умову (ii) з Визначення 12.2.2 для функції  $\mathbb{P}(\bigcup_{n=1}^{\infty} A_n | \mathcal{A}')$ , тобто дорівнює їй майже напевно.  $\square$

В аналогічний спосіб можна довести низку інших очевидних властивостей імовірностей.

**Твердження 12.2.6.** Нехай маємо вимірні простори  $(\Omega, \mathcal{A})$  і  $(\Omega, \mathcal{A}')$ ,  $\mathcal{A}' \subseteq \mathcal{A}$ . Тоді умовна ймовірність  $\mathbb{P}(A | \mathcal{A}')$  задовольняє такі властивості майже напевно:

- монотонність: якщо  $A \subseteq B$ , то

$$\mathbb{P}(B \setminus A | \mathcal{A}') = \mathbb{P}(B | \mathcal{A}') - \mathbb{P}(A | \mathcal{A}') , \quad \mathbb{P}(A | \mathcal{A}') \leq \mathbb{P}(B | \mathcal{A}') ;$$

- формула включення-виключення:

$$\mathbb{P}\left(\bigcup_{i=1}^n A_i | \mathcal{A}'\right) = \sum_{i=1}^n \mathbb{P}(A_i | \mathcal{A}') - \sum_{i < j} \mathbb{P}(A_i \cap A_j | \mathcal{A}') + \dots + (-1)^{n+1} \mathbb{P}(A_1 \cap \dots \cap A_n | \mathcal{A}') ;$$

- неперервність: якщо  $A_1 \subseteq A_2 \subseteq \dots$ , то

$$\lim_{n \rightarrow \infty} \mathbb{P}(A_n | \mathcal{A}') = \mathbb{P}\left(\bigcup_{n=1}^{\infty} A_n | \mathcal{A}'\right) ,$$

а якщо  $A_1 \supseteq A_2 \supseteq \dots$ , то

$$\lim_{n \rightarrow \infty} \mathbb{P}(A_n | \mathcal{A}') = \mathbb{P}\left(\bigcap_{n=1}^{\infty} A_n | \mathcal{A}'\right) .$$

Раніше ми зазначали, що подія  $A$  незалежна від  $\sigma$ -алгебри  $\mathcal{A}'$ , якщо вона незалежна від будь-якої події  $G$  з цієї  $\sigma$ -алгебри:  $\mathbb{P}(A \cap G) = \mathbb{P}(A) \mathbb{P}(G)$ . Це можна переписати як

$$\mathbb{P}(A \cap G) = \int_G \mathbb{P}(A) d\mathbb{P} ,$$

звідки випливає, що подія  $A$  буде незалежна від  $\sigma$ -алгебри  $\mathcal{A}'$ , якщо  $\mathbb{P}(A | \mathcal{A}') = \mathbb{P}(A)$  (майже напевно).

Це повністю відповідає властивості  $\mathbb{P}(A | B) = \mathbb{P}(A)$  для деяких подій  $A$  і  $B$ , яку ми розглядали в Розд. 3.4.

Хоча формула обчислення умовної ймовірності (3.1.1) незастосовна, якщо обумовлювальна подія має нульову ймовірність, ми можемо обчислити відповідну ймовірність, обумовивши цілою  $\sigma$ -алгеброю.

**Приклад 12.2.7.** Розгляньмо простір  $(\mathbb{R}^2, \mathcal{B}^2)$ . Покладімо за  $\sigma$ -алгебру  $\mathcal{A}'$  клас *вертикальних смужок*, тобто декартових добутків виду  $E \times \mathbb{R}$ ,  $E \in \mathcal{B}$ .

Нехай ми всього лише знаємо, яким саме смужкам належить результат експерименту  $(x, y)^\top \in \mathbb{R}^2$ . Відомо, що одноелементні множини виду  $\{x\}$ ,  $x \in \mathbb{R}$ , належать Борелевій  $\sigma$ -алгебрі  $\mathcal{B}$ . Тому клас вертикальних смужок  $\mathcal{A}'$  містить у тому числі смужки виду  $\{x\} \times \mathbb{R}$ . Це означає, що якщо ми достеменно знаємо, яким саме смужкам належить  $(x, y)$ , то ми достеменно знаємо  $x$ .

Нехай на просторі  $(\mathbb{R}^2, \mathcal{B}^2)$  задано абсолютно неперервну ймовірнісну міру  $\mathbb{P}$  зі щільністю  $f$ :

$$\mathbb{P}(A) = \iint_A f(x, y) dx dy .$$

Розгляньмо деяку *горизонтальну смужку*  $A = \mathbb{R} \times F$ ,  $F \in \mathcal{B}$ , і підрахуймо умовну ймовірність події  $A$ . Іншими словами, ми хочемо підрахувати ймовірність деякої конкретної *горизонтальної* смужки, якщо ми знаємо абсолютно всю інформацію про всі *вертикальні* смужки.

Уведімо функцію

$$\phi(x) = \frac{\int_F f(x, t) dt}{\int_{-\infty}^{\infty} f(x, t) dt}$$

і покладімо, що  $\phi(x) = 0$  для всіх точок, де  $\int_{-\infty}^{\infty} f(x, t) dt = 0$  (множина таких точок являє собою нульову подію).

Функція  $\phi$  вимірна відносно  $\mathcal{A}'$ , оскільки вона є функцією від одного аргументу (тобто будь-який прообраз належить  $\mathcal{B}$ , а відтак і відповідній вертикальній смужці). Нам потрібно показати, що  $\phi(x)$  є умовною ймовірністю  $\mathbb{P}(A | \mathcal{A}')$  (точніше, однією з її версій):

$$\int_{E \times \mathbb{R}} \phi(x) d\mathbb{P} = \mathbb{P}(A \cap (E \times \mathbb{R})) .$$

Оскільки  $A = F \times \mathbb{R}$ , перетин буде дорівнювати  $A \cap (E \times \mathbb{R}) = E \times F$ . Тому права частина рівності має вид  $\mathbb{P}(E \times F)$ .

Згідно з Теоремами 6.4.2 та 9.1.8 маємо:

$$\begin{aligned} \int_E \left( \int_{-\infty}^{\infty} \phi(x) f(x, y) dy \right) dx &= \int_E \left( \int_F f(x, t) dt \right) dx \\ &= \iint_{E \times F} f(x, y) dx dy \\ &= \mathbb{P}(E \times F) . \end{aligned}$$

Отже  $\phi(x)$  справді є версією умовної ймовірності.

Якби ми хотіли підрахувати ймовірність деякої горизонтальної смужки за умови настання події у формі деякої вертикальної смужки, ми повинні були б поділити ймовірність перетину смужок на ймовірність вертикальної смужки. Але оскільки ми володіємо всією інформацією про належність результату експерименту всім вертикальним смужкам, тобто ми достеменно знаємо координату  $x$ , вертикальна смужка фактично перетворюється в лінію. Лінія є множиною міри нуль за двовимірною мірою Лебега, а тому її ймовірність дорівнює нулю. Відповідно, застосувати формулу (3.1.1) неможливо. Натомість обумовлення цілою  $\sigma$ -алгеброю з вертикальних смужок

дає нам змогу обчислити відповідну ймовірність як функцію  $\phi$ , яка задовольняє всі вимоги до умовної ймовірності.

Можна неформально казати, що  $\phi(x)$  є границею умовних імовірностей  $\mathbb{P}(A | B_n)$ , де кожна  $B_n$  — це вертикальна смужка з  $E = (x - h_n; x + h_n)$ , а  $h_n \downarrow 0$ . Для кожної такої смужки ймовірність перетину коректно визначена, а результат граничного переходу, коли вертикальна смужка є лінією, дає функція  $\phi$ .  $\square$

### 12.2.3. Обумовлення випадковою величиною

Розгляньмо деякий випадковий вектор  $\mathbf{X} = (X_1, \dots, X_k) : (\Omega, \mathcal{A}, \mathbb{P}) \rightarrow \mathbb{R}^k$ . Він буде вимірний відносно деякої іншої  $\sigma$ -алгебри  $\mathcal{A}' \subset \mathcal{A}$ , якщо  $\{\omega : \mathbf{X}(\omega) \in H\} \in \mathcal{A}'$  для всіх  $H \in \mathcal{B}^k$ .

**Визначення 12.2.8.** Казатимемо, що  $\sigma$ -алгебру  $\mathcal{A}(\mathbf{X})$  *породжує вектор  $\mathbf{X}$*  ( $\sigma$ -algebra is generated by  $\mathbf{X}$ ), якщо вона є найменшою  $\sigma$ -алгеброю, відносно якої він є вимірний.

Аналогічно, казатимемо, що  $\sigma$ -алгебру  $\mathcal{A}(\{\mathbf{X}_1, \dots, \mathbf{X}_n\})$  породжує множина випадкових векторів, якщо вона є найменшою серед усіх, відносно яких *кожний із них* є вимірний.  $\square$

Тоді умовну ймовірність можна визначити так.

**Визначення 12.2.9.** Нехай маємо вимірний простір  $(\Omega, \mathcal{A})$  і випадкову величину  $X : (\Omega, \mathcal{A}) \rightarrow \mathbb{R}$ , яка породжує  $\sigma$ -алгебру  $\mathcal{A}(X)$ . Тоді *умовною ймовірністю за умови  $X$*  (conditional probability given  $X$ ) є випадкова величина  $\mathbb{P}(A | \mathcal{A}(X)) \equiv \mathbb{P}(A | X)$ .  $\square$

Іншими словами, за умови  $X$  ми, хоча й не знаємо конкретного  $\omega$ , знаємо, яким множинам  $\{\omega' : X(\omega') = x\}$  воно належить, тобто ми фактично знаємо значення  $X(\omega)$ . Тобто можна казати, що ми шукаємо ймовірність деякої події за умови, що  $X = x$ .

Аналогічне визначення можна запропонувати для випадкового вектора  $\mathbf{X} \in \mathbb{R}^k$ .

Ключовим результатом відповідної формальної побудови умовної ймовірності, обумовленої  $\sigma$ -алгеброю, є уможливлення обчислення умовних імовірностей деяких подій за умови, що деяка випадкова величина набула деякого значення  $X(\omega)$ . Оскільки для неперервних випадкових величини подія  $\mathbb{P}(X = x) = 0$ , використання (3.1.1) було б просто неможливим.

**Зауваження 12.2.10.** Потрібно розуміти, що коли ми, наприклад, хочемо порахувати ймовірність, що деяка випадкова величина  $X$ , скажімо, менша 0, за умови, що інша (неперервна) випадкова величина  $Y$ , скажімо, дорівнює 1, ми не можемо обчислити це за визначенням умовної ймовірності, адже  $\mathbb{P}_Y(Y = 1) = 0$ . Проте, обумовлюючи цілою  $\sigma$ -алгеброю, яку породжує  $Y$ , ми можемо обчислити цю ймовірність як границю ймовірностей  $\mathbb{P}_{X|Y}(X \leq 0 | Y \in (1 - h_n; 1 + h_n))$ , де  $h_n \downarrow 0$ .

Тому до позначень на кшталт  $\mathbb{P}_{X|Y}(X \leq 0 | Y = 1)$  потрібно ставитися акуратно, розуміючи, що насправді за ними ховається.  $\square$

## 12.3. Умовні сподівання

### 12.3.1. Сподівання за умови довільної $\sigma$ -алгебри

Умовне сподівання можна визначити у спосіб, схожий на визначення умовної ймовірності.

**Визначення 12.3.1.** Нехай  $X : (\Omega, \mathcal{A}, \mathbb{P}) \rightarrow \mathbb{R}$  — інтегровна випадкова величина, а  $\mathcal{A}' \subseteq \mathcal{A}$ . Тоді існує випадкова величина  $\mathbb{E}[X | \mathcal{A}']$ , яка має назву *умовного сподівання за умови  $\mathcal{A}'$*  (conditional expectation of  $X$  given  $\mathcal{A}'$ ), якщо:

(i)  $\mathbb{E}[X | \mathcal{A}']$  вимірна відносно  $\mathcal{A}'$  та інтегровна;

(ii) її інтеграл дорівнює

$$\int_G \mathbb{E}[X | \mathcal{A}'] d\mathbb{P} = \int_G X d\mathbb{P}, \quad G \in \mathcal{A}'. \quad (12.3.1)$$

□

Фактично справа стоїть інтеграл випадкової величини  $X$  за деякою множиною  $G$ , тобто неформально можна стверджувати, що це «сподівання» величини  $X$ , якщо звузити простір до однієї множини  $G$ .

**Зауваження 12.3.2.** На перший погляд може здатися, що достатньо просто підставити в (12.3.1)  $X$  замість  $\mathbb{E}[X | \mathcal{A}']$  і дістати тривіальну тотожність. Проте  $X$  може не задовольняти умову (i), а саме бути вимірною відносно  $\mathcal{A}'$ .

Якщо ж випадкова величина  $X$  вимірна відносно  $\mathcal{A}'$ , то справді  $\mathbb{E}[X | \mathcal{A}'] = X$  майже напевно. Інтерпретація цього факту така, що якщо ми як спостерігачі знаємо достеменно, яким множинам із  $\mathcal{A}'$  належить  $\omega$ , тобто фактично якого значення набула  $X$ , то найліпшим прогнозним значенням  $X$  є сама  $X$ . □

Те, що така випадкова величина справді існує, випливає з таких міркувань. Нехай  $X \geq 0$  майже напевно. Нехай міра  $\nu$  на  $\mathcal{A}'$  така, що  $\nu(G) = \int_G X d\mathbb{P}$ . Оскільки  $X$  інтегровна, така міра буде скінченною. Окрім цього, вона абсолютно неперервна відносно  $\mathbb{P}$ . Згідно з Теоремою 6.3.2, існує щільність така, що  $\nu(G) = \int_G f d\mathbb{P}$ . Ця щільність має всі властивості з Визначення 12.3.1.

Якщо ж  $X$  загальна, то можна розглянути функцію  $\mathbb{E}[X^+ | \mathcal{A}'] - \mathbb{E}[X^- | \mathcal{A}']$ , яка буде мати відповідні властивості.

У повній аналогії з наведеними вище властивостями умовної ймовірності можна зазначити таке:

- будь-яку величину  $\mathbb{E}[X | \mathcal{A}']$ , яка задовольняє (12.3.1), називають версією умовного сподівання, усі з яких рівні майже напевно;
- $\mathbb{E}[X | \{\emptyset, \Omega\}] = \mathbb{E}[X]$  майже напевно;
- $\mathbb{E}[X | \mathcal{A}] = X$  майже напевно;
- якщо  $X$  незалежна від  $\sigma$ -алгебри  $\mathcal{A}'$ , то  $\mathbb{E}[X | \mathcal{A}'] = \mathbb{E}[X]$  майже напевно;
- умова (i) із Визначення 12.3.1 стверджує, що ми можемо підрахувати умовне сподівання, навіть не знаючи  $\omega$ , а тільки знаючи, для кожної події  $G \in \mathcal{A}'$ , належить їй  $\omega$  чи не належить.

Нехай події  $B_1, B_2, \dots$  утворюють не більш ніж зліченне розбиття  $\Omega$ , яке породжує  $\sigma$ -алгебру  $\mathcal{A}'$ . Оскільки ми знаємо, якій події  $B_i, i = 1, 2, \dots$ , з розбиття належить  $\omega$ , найліпшим прогнозним значенням  $X$  повинно бути середнє значення  $X$  по відповідній події  $B_i$ . Іншими словами, умовне сподівання  $\mathbb{E}[X | \mathcal{A}']$  як функція від  $\omega$  повинно бути сталою для  $\omega \in B_i$  і дорівнювати

$$\mathbb{E}[X | \mathcal{A}'] = \frac{1}{\mathbb{P}(B_i)} \int_{B_i} X d\mathbb{P}, \quad \omega \in B_i, \quad \mathbb{P}(B_i) > 0. \quad (12.3.2)$$

Якщо  $\mathbb{P}(B_i) = 0$ , то (стале) значення  $\mathbb{E}[X | \mathcal{A}']$  може бути абсолютно довільне.

Для перевірки того факту, що це справді умовне сподівання, потрібно помітити, що стала є вимірна відносно  $\mathcal{A}'$ , і тому виконується умова (i) із Визначення 12.3.1. Умова (ii) виконується також, оскільки

$$\int_{B_i} \left( \frac{1}{\mathbb{P}(B_i)} \int_{B_i} X d\mathbb{P} \right) d\mathbb{P} = \int_{B_i} X d\mathbb{P} \cdot \frac{\mathbb{P}(B_i)}{\mathbb{P}(B_i)} = \int_{B_i} X d\mathbb{P}, \quad B_i \in \mathcal{A}'.$$

Кажучи трішки неформально, ми показали, що умовне сподівання «за умови, що сталася подія  $B$ », дорівнює

$$\mathbb{E}[X | B] = \frac{\int_B X d\mathbb{P}}{\mathbb{P}(B)}.$$

**Приклад 12.3.3.** Для дуже простої ілюстрації цих міркувань розгляньмо ймовірнісний простір  $(\Omega, \mathcal{A}, \mathbb{P})$ , де  $\Omega = \{a, b, c, d, e, f\}$ ,  $\mathcal{A} = 2^\Omega$ , а  $\mathbb{P}$  відповідає рівномірному розподілу. Розгляньмо випадкову величину  $X$  таку, що

$$X(\omega) = \begin{cases} 1, & \omega = a \\ 3, & \omega = b \\ 3, & \omega = c \\ 5, & \omega = d \\ 5, & \omega = e \\ 7, & \omega = f \end{cases}.$$

Простір  $\Omega$  можна розбити на три множини:  $\{B_1 = \{a, b\}, B_2 = \{c, d\}, B_3 = \{e, f\}\}$ . Тоді

$$\mathbb{E}[X | B_1] = \frac{\int_{B_1} X d\mathbb{P}}{\mathbb{P}(B_1)} = \frac{1 \cdot 1/6 + 3 \cdot 1/6}{2/6} = 2,$$

де ми обчислили інтеграл як інтеграл простої функції.

Аналогічно маємо  $\mathbb{E}[X | B_2] = 4$ ,  $\mathbb{E}[X | B_3] = 6$ . Отже умовне сподівання  $\mathbb{E}[X | \mathcal{A}_B]$  є випадковою величиною

$$\mathbb{E}[X | \mathcal{A}_B](\omega) = \begin{cases} 2, & \omega \in \{a, b\} \\ 4, & \omega \in \{c, d\} \\ 6, & \omega \in \{e, f\} \end{cases}.$$

Як і було вказано вище, вона є сталою на множинах із розбиття.  $\square$

Як і у випадку з умовними ймовірностями, якщо подія  $B$  нульова, можемо покласти  $\mathbb{E}[X | B]$  рівним будь-якій довільній сталій. У зв'язку з цим потрібно бути дуже обережним, обумовлюючи

нульовою подією, адже результат відповідних обчислень не має практичної інтерпретації, через що виникають різні парадокси.

Із розгляду сподівання за умови подій із деякого розбиття безпосередньо випливає дуже важливе правило обчислення безумовних сподівань через сподівання умовні.

**Твердження 12.3.4** (Закон повного сподівання (Law of total expectation)). Якщо події  $B_1, B_2, \dots$  утворюють не більш ніж зліченне розбиття  $\Omega$ , яке породжує  $\sigma$ -алгебру  $\mathcal{A}'$ , то безумовне сподівання випадкової величини  $X$  дорівнює

$$\mathbb{E}[X] = \sum_i \mathbb{E}[X | B_i] \mathbb{P}(B_i) . \quad (12.3.3)$$

*Доведення.* Доведення безпосередньо випливає з того факту, що випадкову величину  $X$  можна розписати як

$$X(\omega) = \sum_i X(\omega) \cdot \mathbb{1}\{X(\omega) \in B_i\} ,$$

а

$$\int X(\omega) \cdot \mathbb{1}\{X(\omega) \in B_i\} d\mathbb{P} = \int_{B_i} X(\omega) d\mathbb{P} = \mathbb{E}[X | \mathcal{A}'](\omega) \cdot \mathbb{P}(B_i)$$

згідно з (12.3.2), а  $\mathbb{E}[X | \mathcal{A}'](\omega)$  має інтерпретацію  $\mathbb{E}[X | B_i]$ , якщо  $\omega \in B_i$ .  $\square$

Для індикаторної функції  $\mathbb{1}_A(\omega)$  умови Визначення 12.2.2 і 12.3.1 збігаються, а тому  $\mathbb{E}[\mathbb{1}_A(\omega) | \mathcal{A}'] = \mathbb{P}(A | \mathcal{A}')$  майже напевно.

**Зауваження 12.3.5.** Закон повної ймовірності із Теореми 3.3.2 є частковим випадком Твердження 12.3.4, якщо використати випадкову величину  $\mathbb{1}_A$ :

$$\mathbb{P}(A) = \mathbb{E}[\mathbb{1}_A] = \sum_i \mathbb{E}[\mathbb{1}_A | B_i] \mathbb{P}(B_i) = \sum_{i=1}^n \mathbb{P}(A | A_i) \mathbb{P}(A_i) . \quad \square$$

**Приклад 12.3.6.** Нехай підкидають монетку з імовірністю випаду герба  $p$ . Нехай  $X = \text{«кількість підкидань до випаду першого герба»}$ . Із Лекції 5 ми знаємо, що  $X \sim \text{Geom}(p)$ . Підрахунок  $\mathbb{E}[X]$  можна вести не тільки згідно з міркуваннями, викладеними в Твердженні 5.2.5, а й помітивши, що події  $B_1 = \text{«перший випад був герб»}$  та  $B_2 = \text{«перший випад було число»}$  утворюють розбиття простору. Тоді можна застосувати (12.3.3):

$$\mathbb{E}[X] = \mathbb{E}[X | B_1] \mathbb{P}(B_1) + \mathbb{E}[X | B_2] \mathbb{P}(B_2) = 0 \cdot p + (1 + \mathbb{E}[X]) \cdot (1 - p) ,$$

звідки можна виразити  $\mathbb{E}[X] = \frac{1-p}{p}$ .  $\square$

**Визначення 12.3.7.** Нехай маємо вимірний простір  $(\Omega, \mathcal{A})$  і випадкову величину  $X : (\Omega, \mathcal{A}) \rightarrow \mathbb{R}$ , яка породжує  $\sigma$ -алгебру  $\mathcal{A}(X)$ . Тоді для випадкової величини  $Y$  *умовним сподіванням за умови  $X$*  (conditional expectation given  $X$ ) є випадкова величина  $\mathbb{E}[Y | \mathcal{A}(X)] \equiv \mathbb{E}[Y | X]$ .  $\square$

**Приклад 12.3.8.** Палицю довжиною 1 переламують у випадковій точці  $X$ . Ми знаємо реалізацію випадкової величини  $X = x$ . Нехай  $Y$  — випадкова точка з проміжку  $[0; x]$ .

Якщо ми знаємо, що  $X = x$ , то випадкова точка  $Y$  матиме розподіл  $U([0; x])$ . Відтак сподівання буде дорівнювати  $x/2$ . Звідси випливає, що

$$\mathbb{E}[Y | X] = \frac{X}{2}.$$

Звісно, трішки неформально можна казати, що в цьому випадку, наприклад,  $\mathbb{E}[Y | X = 0.5] = 0.25$ . Проте, за аналогією з [Зауваженням 12.2.10](#), потрібно розуміти, що це просто позначення і ми не обумовлюємо  $Y$  нульовою подією. Насправді коректно казати про випадкову величину  $\mathbb{E}[Y | X]$ , яка поводить себе як  $\frac{X}{2}$ , і в тому числі може набувати значення  $\frac{0.5}{2} = 0.25$ . Ми можемо дістати такий результат саме завдяки тому, що обумовлюємо цілою  $\sigma$ -алгеброю, яку породжує  $X$ , а не однією нульовою подією  $X = 0.5$ .

Оскільки  $\mathbb{E}[Y | X]$  є випадковою величиною, можемо дістати її сподівання та дисперсію:

$$\begin{aligned}\mathbb{E}[\mathbb{E}[Y | X]] &= \mathbb{E}\left[\frac{X}{2}\right] = \frac{1}{4}, \\ \text{Var}(\mathbb{E}[Y | X]) &= \text{Var}\left(\frac{X}{2}\right) = \frac{1}{48}.\end{aligned}$$
□

### 12.3.2. Властивості умовних сподівань

Використовуючи відповідні визначення, а також властивості інтегралу, можна довести такі очевидні властивості умовного сподівання.

**Теорема 12.3.9.** Нехай випадкові величини  $X, Y$  і  $X_n, n = 1, 2, \dots$  інтегровні. Тоді

- (i) якщо  $X = a$  майже напевно, то  $\mathbb{E}[X | \mathcal{A}'] = a$  майже напевно;
- (ii) для деяких  $a, b \in \mathbb{R}$ ,  $\mathbb{E}[aX + bY | \mathcal{A}'] = a\mathbb{E}[X | \mathcal{A}'] + b\mathbb{E}[Y | \mathcal{A}']$  майже напевно;
- (iii) якщо  $X \leq Y$  майже напевно, то  $\mathbb{E}[X | \mathcal{A}'] \leq \mathbb{E}[Y | \mathcal{A}']$  майже напевно;
- (iv) якщо  $\lim_{n \rightarrow \infty} X_n = X$  майже напевно,  $|X_n| < Y$  майже напевно,  $Y$  інтегровна, тоді

$$\lim_{n \rightarrow \infty} \mathbb{E}[X_n | \mathcal{A}'] = \mathbb{E}[X | \mathcal{A}']$$

майже напевно;

- (v) якщо  $\phi$  — опукла функція, то  $\mathbb{E}[\phi(X) | \mathcal{A}'] \geq \phi(\mathbb{E}[X | \mathcal{A}'])$  майже напевно.

**Зауваження 12.3.10.** Надзвичайно важливо пам'ятати, що лінійність застосовна тільки для випадкових величин, сподівання яких шукаємо, а не для обумовлювальних величин. Так, у загальному випадку  $\mathbb{E}[Y | X_1 + X_2] \neq \mathbb{E}[Y | X_1] + \mathbb{E}[Y | X_2]$ .

□

Оскільки  $\mathbb{P}(A | \mathcal{A}') = \mathbb{E}[\mathbb{1}_A(\omega) | \mathcal{A}']$ , усі раніше згадані властивості умовної ймовірності можна вивести з [Теореми 12.3.9](#).

**Твердження 12.3.11.** Якщо  $X$  вимірна відносно  $\mathcal{A}'$ , а величини  $Y$  і  $XY$  інтегровні, то

$$\mathbb{E}[XY | \mathcal{A}'] = X\mathbb{E}[Y | \mathcal{A}'] \quad (12.3.4)$$

майже напевно.

Зверніть увагу, що  $X$  сама по собі не обов'язково повинна бути інтегровною.

*Доведення.* Як завжди, доведення будуватимемо покроково, починаючи з індикаторних функцій.

Нехай  $X = \mathbb{1}_{G_0}$ ,  $G_0 \in \mathcal{A}'$ . Величина  $\mathbb{1}_{G_0} \cdot \mathbb{E}[Y | \mathcal{A}']$  вимірна відносно  $\mathcal{A}'$ , оскільки  $\mathbb{E}[Y | \mathcal{A}']$  також вимірна відносно  $\mathcal{A}'$  за визначенням умовного сподівання. Понад те,

$$\int_G \mathbb{1}_{G_0} \cdot \mathbb{E}[Y | \mathcal{A}'] d\mathbb{P} = \int_{G \cap G_0} \mathbb{E}[Y | \mathcal{A}'] d\mathbb{P} = \int_{G \cap G_0} Y d\mathbb{P} = \int_G \mathbb{1}_{G_0} \cdot Y d\mathbb{P},$$

де ми використали той факт, що  $\mathbb{E}[Y | \mathcal{A}']$  задовольняє умову (ii) з Визначення 12.3.1 для  $G \cap G_0$ .

Відтак, ми показали, що (12.3.4) виконується для індикаторної функції.

За властивістю (ii) з Теореми 12.3.9 маємо, що (12.3.4) виконується також для простих функцій.

Для загальної випадкової величини  $X$ , вимірної відносно  $\mathcal{A}'$ , існує послідовність простих величин  $X_n$ , вимірних відносно  $\mathcal{A}'$ , і таких, що  $|X_n| \leq |X|$  і  $\lim_{n \rightarrow \infty} X_n = X$ . Оскільки  $|X_n Y| \leq |XY|$  і  $|XY|$  інтегровна (бо  $XY$  інтегровна), із властивості (iv) з Теореми 12.3.9 випливає, що

$$\lim_{n \rightarrow \infty} \mathbb{E}[X_n Y | \mathcal{A}'] = \mathbb{E}[XY | \mathcal{A}']$$

майже напевно. Але оскільки  $\mathbb{E}[X_n Y | \mathcal{A}'] = X_n \mathbb{E}[Y | \mathcal{A}']$ , бо ми вже це довели для простих функцій, маємо

$$\lim_{n \rightarrow \infty} X_n \mathbb{E}[Y | \mathcal{A}'] = X \mathbb{E}[Y | \mathcal{A}'] .$$

Отже (12.3.4) виконується в загальному випадку.  $\square$

**Приклад 12.3.12.** Нехай  $X, Y \sim \text{Exp}(\lambda)$ ,  $X \perp\!\!\!\perp Y$ . Нехай  $M = \max\{X, Y\}$ ,  $L = \min\{X, Y\}$ . Як ми зазначали в Прикладі 8.5.3,  $L \sim \text{Exp}(2\lambda)$ , а  $M - L \perp\!\!\!\perp L$  відповідно до властивості не мати пам'яти. Відтак за властивістю лінійності умовного сподівання маємо

$$\mathbb{E}[M | L] = \mathbb{E}[L | L] + \mathbb{E}[M - L | L] = L + \mathbb{E}[M - L] = L + \frac{1}{\lambda} .$$

**Приклад 12.3.13.** Нехай  $Z \sim N(0, 1)$ , а  $Y = Z^2$ . Тоді можемо обчислити

$$\mathbb{E}[Y | Z] = \mathbb{E}[Z^2 | Z] = Z^2 ,$$

оскільки  $Y = Z^2$  вимірна відносно  $\mathcal{A}(Z)$ .

Щоб дістати  $\mathbb{E}[Z | Y]$ , потрібно помітити, що  $Y = y$  тоді й тільки тоді, коли  $Z = \pm\sqrt{y}$ . Відтак інформація про  $Y$  як  $\sigma$ -алгебра  $\mathcal{A}(Y)$  містить усі такі події  $G$ , на яких  $Z$  симетрична, тобто на яких  $\int_G Z d\mathbb{P} = 0$ . Відтак  $\mathbb{E}[Z | Y] = 0$  майже напевно.  $\square$

Умовне сподівання (як і звичайне сподівання) можна розглядати як зважене середнє. Інтуїція підказує, що якщо спочатку взяти середнє відносно  $\mathcal{A}_2$ , а потім — відносно меншої  $\mathcal{A}_1$ , то результат повинен бути той же, що й усереднення відносно  $\mathcal{A}_1$  з самого початку.

**Теорема 12.3.14** (Закон ітерованих сподівань (Law of iterated expectations)). Якщо  $X$  інтегровна, а  $\mathcal{A}_1$  і  $\mathcal{A}_2$  —  $\sigma$ -алгебри такі, що  $\mathcal{A}_1 \subset \mathcal{A}_2$ , то

$$\mathbb{E}[\mathbb{E}[X | \mathcal{A}_2] | \mathcal{A}_1] = \mathbb{E}[X | \mathcal{A}_1]. \quad (12.3.5)$$

Зокрема, якщо  $\mathcal{A}_1 = \{\emptyset, \Omega\}$ , а  $\mathcal{A}_2 = \mathcal{A}$ , то

$$\mathbb{E}[\mathbb{E}[X | \mathcal{A}_2] | \mathcal{A}_1] = \mathbb{E}[\mathbb{E}[X | \mathcal{A}]] = \mathbb{E}[X]. \quad (12.3.6)$$

*Доведення.* Оскільки, за визначенням, лівий бік рівності вимірний відносно  $\mathcal{A}_1$ , показати, що це — версія  $\mathbb{E}[X | \mathcal{A}_1]$ , достатньо, перевіривши виконання (12.3.1):

$$\int_G \mathbb{E}[\mathbb{E}[X | \mathcal{A}_2] | \mathcal{A}_1] d\mathbb{P} = \int_G X d\mathbb{P}, \quad G \in \mathcal{A}_1.$$

По-перше, за визначенням,

$$\int_G \mathbb{E}[\mathbb{E}[X | \mathcal{A}_2] | \mathcal{A}_1] d\mathbb{P} = \int_G \mathbb{E}[X | \mathcal{A}_2] d\mathbb{P}, \quad G \in \mathcal{A}_1.$$

Але з  $G \in \mathcal{A}_1$  випливає  $G \in \mathcal{A}_2$ , а отже

$$\int_G \mathbb{E}[X | \mathcal{A}_2] d\mathbb{P} = \int_G X d\mathbb{P}. \quad \square$$

Можна помітити, що Твердження 12.3.4 є частковим випадком Теореми 12.3.14.

Також можна помітити, що оскільки  $\mathbb{P}_X(X \in A) = \mathbb{E}[\mathbb{1}\{X \in A\}]$ , закон ітерованих сподівань дає підстави вивести формулу повної ймовірності для неперервного випадку:

$$\begin{aligned} \mathbb{P}_X(X \in A) &= \mathbb{E}[\mathbb{1}\{X \in A\}] = \mathbb{E}[\mathbb{E}[\mathbb{1}\{X \in A\} | Y]] = \mathbb{E}[\mathbb{P}_{X|Y}(X \in A | Y)] \\ &= \int_{-\infty}^{\infty} \mathbb{P}_{X|Y}(X \in A | Y = y) \cdot f_Y(y) dy. \end{aligned} \quad (12.3.7)$$

Варто звернути увагу на умову інтегровності  $X$  для застосування (12.3.6). Наприклад, для розподілу Коші, як відомо з Прикладу 6.4.8,  $\mathbb{E}[X]$  не існує, хоча сподівання  $\mathbb{E}[X | X > 0]$  та  $\mathbb{E}[X | X < 0]$  існують і дорівнюють  $\infty$  і  $-\infty$  відповідно.

**Приклад 12.3.15.** Доведімо варіант *тотожності Волда* (Wald's equation)<sup>2</sup>. Нехай  $X_1, X_2, \dots$  — незалежні випадкові величини з однаковим розподілом і скінчненим сподіванням  $\mu_X$ . Нехай  $N$  — дискретна випадкова величина, незалежна від  $X_i$ ,  $i = 1, 2, \dots$ , носій якої  $\text{supp}(N) = \mathbb{N}$ . Тоді можна розглянути величину  $S_N = X_1 + X_2 + \dots + X_N$ .

Наприклад,  $X_i$  може відповідати сумі грошей, витрачених покупцем  $i$  у деякому магазині за день, а  $N$  — загальному числу покупців. Тоді  $S_N$  буде загальною сумаю, витраченою всіма покупцями за день.

Нехай відомо, що  $\mathbb{E}[X_i] = 8$ , а  $\mathbb{E}[N] = 50$ . Підрахуймо  $\mathbb{E}[S_N]$ . Можна помітити, що

$$\mathbb{E}[S_N | N = n] = \mathbb{E}[X_1 + \dots + X_N | N = n] = \sum_{i=1}^N \mathbb{E}[X_i | N = n] = \sum_{i=1}^N \mathbb{E}[X_i] = n\mu_X,$$

<sup>2</sup>Названа так на честь угорського математика Абрагама Волда (Abraham Wald, 1902–1950).

де ми використали той факт, що  $X_i \perp\!\!\!\perp N$ ,  $i = 1, 2, \dots$ . Іншими словами,  $\mathbb{E}[S_N | N] = \mu_X N$ .

Тепер можемо застосувати закон ітерованих сподівань (12.3.6), врахувавши незалежність  $S_N$  та  $N$ :

$$\mathbb{E}[S_N] = \mathbb{E}[\mathbb{E}[S_N | N]] = \mathbb{E}[\mu_X N] = \mu_X \mathbb{E}[N].$$

Варто звернути увагу, що ми змогли обчислити сподівання  $\mathbb{E}[S_N]$ , не маючи абсолютно жодної інформації про розподіл відповідних величин, окрім відповідних сподівань. Ми просто скористалися загальними властивостями умовних сподівань, справедливих у будь-яких випадках.  $\square$

### 12.3.3. Умовні сподівання й медіана як найліпші предиктори

У Розд. 11.3 ми стверджували, що сподівання  $\mathbb{E}[X]$  випадкової величини  $X$  є її найліпшим предиктором у тому сенсі, що воно мінімізує значення середньоквадратичної похибки  $\mathbb{E}[(X - b)^2]$  серед усіх  $b$ . Аналогічну властивість має й умовне сподівання.

Розгляньмо задачу прогнозування значення деякої величини  $Y$  на основі спостережень за значенням іншої величини  $X$ . Позначмо предиктор величини  $Y$  через  $g(X)$ , підкresливши, що він є функцією від  $X$ . Іншими словами, якщо ми спостерігаємо, що  $X = x$ , то нашим прогнозом значення  $Y$  буде число  $g(x)$ . Вочевидь, хотілося б мати предиктор, який був би в певному сенсі близький до  $Y$ . Таким предиктором і є умовне сподівання.

**Твердження 12.3.16.** Умовне сподівання випадкової величини  $Y$  зі скінченим моментом другого порядку за умови іншої величини  $X$ ,  $\mathbb{E}[Y | X]$ , є найліпшим предиктором значення  $Y$ , у тому сенсі, що воно мінімізує значення середньоквадратичної похибки:

$$\mathbb{E}[Y | X] = \arg \min_g \mathbb{E}[(Y - g(X))^2]. \quad (12.3.8)$$

*Доведення.* За законом ітерованих сподівань із Теореми 12.3.14,

$$\mathbb{E}[(Y - g(X))^2] = \mathbb{E}[\mathbb{E}[(Y - g(X))^2 | X]].$$

Далі доведення аналогічне доведенню для Твердження 11.3.1:

$$\begin{aligned} \mathbb{E}[\mathbb{E}[(Y - g(X))^2 | X]] &= \mathbb{E}[(Y - \mathbb{E}[Y | X] + \mathbb{E}[Y | X] - g(X))^2] \\ &= \mathbb{E}[(Y - \mathbb{E}[Y | X])^2 | X] \\ &\quad + 2\mathbb{E}[(Y - \mathbb{E}[Y | X])(\mathbb{E}[Y | X] - g(X)) | X] \\ &\quad + \mathbb{E}[(\mathbb{E}[Y | X] - g(X))^2 | X]. \end{aligned}$$

Потрібно помітити, що

$$\begin{aligned} \mathbb{E}[(Y - \mathbb{E}[Y | X])(\mathbb{E}[Y | X] - g(X)) | X] &= (\mathbb{E}[Y | X] - g(X)) \cdot \mathbb{E}[Y - \mathbb{E}[Y | X] | X] \\ &= (\mathbb{E}[Y | X] - g(X)) \cdot (\mathbb{E}[Y | X] - \mathbb{E}[Y | X]) = 0, \end{aligned}$$

оскільки  $(\mathbb{E}[Y | X] - g(X))$  за умови  $X$  можна винести за сподівання згідно з Твердженням 12.3.11.

Також варто помітити, що вираз  $\mathbb{E}[(Y - \mathbb{E}[Y | X])^2 | X]$  взагалі не залежить від  $g(X)$ , а тому не впливає на значення виразу, який ми мінімізуємо. Відтак  $\mathbb{E}[(\mathbb{E}[Y | X] - g(X))^2 | X]$  досягатиме найменшого значення, якщо  $(\mathbb{E}[Y | X] - g(X))^2$  буде найменшим, тобто якщо  $g(X) = \mathbb{E}[Y | X]$ .

Отже ми показали, що  $g(X)$  мінімізує вираз  $\mathbb{E}[(Y - g(X))^2 | X]$ , а відтак і вираз  $\mathbb{E}[(Y - g(X))^2]$ .  $\square$

Можна довести, що *умовна медіана* (conditional median)  $M_{Y|X}$  також є найліпшим предиктором, але в розумінні мінімізації *середньої абсолютної похибки* (mean absolute error):

$$M_{Y|X} = \arg \min_g \mathbb{E} [|Y - g(X)|] . \quad (12.3.9)$$

#### 12.3.4. Умовні дисперсії

За аналогією з умовними сподіванням також можна визначити умовні моменти довільного порядку. Зокрема, *умовною дисперсією* (conditional variance) величини  $Z$  за умови величини  $X$  є

$$\text{Var}(Z | X) = \mathbb{E} [(Z - \mathbb{E}[Z | X])^2 | X] . \quad (12.3.10)$$

**Приклад 12.3.17.** Нехай, як і в Прикладі 12.3.13,  $Z \sim N(0, 1)$ , а  $Y = Z^2$ . Тоді можемо обчислити

$$\text{Var}(Y | Z) = \mathbb{E} [(Y - \mathbb{E}[Y | Z])^2 | Z] = \mathbb{E} [(Z^2 - \mathbb{E}[Z^2 | Z])^2 | Z] = \mathbb{E} [(Z^2 - Z^2)^2 | Z] = 0 ,$$

що зрозуміло навіть без обчислень, адже якщо ми знаємо все про  $Z$ , то  $Y = Z^2$  поводить себе як стала. В інший бік маємо

$$\text{Var}(Z | Y) = \mathbb{E} [Z^2 | Y] - (\mathbb{E}[Z | Y])^2 = Z^2 - 0^2 = Z^2 = Y .$$

Для умовних дисперсій справедлива така властивість.

**Теорема 12.3.18** (Закон повної дисперсії (Law of total variance)). Нехай  $X$  та  $Z$  — дві випадкові величини. Тоді дисперсію  $\text{Var}(X)$  можна розписати так:

$$\text{Var}(X) = \mathbb{E}[\text{Var}(X | Z)] + \text{Var}(\mathbb{E}[X | Z]) . \quad (12.3.11)$$

*Доведення.* Доведення цього твердження безпосередньо випливає, якщо розписати всі його складові та скористатися (12.3.6):

$$\begin{aligned} \mathbb{E}[\text{Var}(X | Z)] &= \mathbb{E}[\mathbb{E}[(X - \mathbb{E}[X | Z])^2 | Z]] \\ &= \mathbb{E}[\mathbb{E}[X^2 - 2X\mathbb{E}[X | Z] + (\mathbb{E}[X | Z])^2 | Z]] \\ &= \mathbb{E}[\mathbb{E}[X^2 | Z]] - 2\mathbb{E}[\mathbb{E}[X\mathbb{E}[X | Z] | Z]] + \mathbb{E}[\mathbb{E}[(\mathbb{E}[X | Z])^2 | Z]] \\ &= \mathbb{E}[X^2] - 2\mathbb{E}[(\mathbb{E}[X | Z])^2] + \mathbb{E}[(\mathbb{E}[X | Z])^2] \\ &= \mathbb{E}[X^2] - \mathbb{E}[(\mathbb{E}[X | Z])^2] . \end{aligned}$$

З іншого боку,

$$\begin{aligned} \text{Var}(\mathbb{E}[X | Z]) &= \mathbb{E}[(\mathbb{E}[X | Z] - \mathbb{E}[\mathbb{E}[X | Z]])^2] \\ &= \mathbb{E}[(\mathbb{E}[X | Z] - \mathbb{E}[X])^2] \\ &= \mathbb{E}[(\mathbb{E}[X | Z])^2] - 2\mathbb{E}[\mathbb{E}[X | Z]\mathbb{E}[X]] + \mathbb{E}[(\mathbb{E}[X])^2] \\ &= \mathbb{E}[(\mathbb{E}[X | Z])^2] - 2(\mathbb{E}[X])^2 + (\mathbb{E}[X])^2 \\ &= \mathbb{E}[(\mathbb{E}[X | Z])^2] - (\mathbb{E}[X])^2 . \end{aligned}$$

Отже,

$$\begin{aligned}\mathbb{E}[\text{Var}(X | Z)] + \text{Var}(\mathbb{E}[X | Z]) &= \mathbb{E}[X^2] - \mathbb{E}[(\mathbb{E}[X | Z])^2] + \mathbb{E}[(\mathbb{E}[X | Z])^2] - (\mathbb{E}[X])^2 \\ &= \mathbb{E}[X^2] - (\mathbb{E}[X])^2 = \text{Var}(X).\end{aligned}$$

□

Інтерпретація цього закону має велику практичну цінність. Нехай маємо популяцію, у якій кожний об'єкт має дві властивості,  $X$  і  $Y$ . Наприклад, множина осіб може мати різний вік  $X$  і зріст  $Y$ . Для того, щоб підрахувати загальну дисперсію  $Y$ , ми можемо згрупувати осіб за віком, і тоді  $\text{Var}(Y)$  — середній розкид значень зросту — складатиметься з двох компонент:

- у кожній віковій групі буде свій розкид значень зросту  $\text{Var}(Y | X)$ , і середнє значення таких розкидів дорівнюватиме  $\mathbb{E}[\text{Var}(Y | X)]$ . Відповідний показник називають *внутрішньогруповою варіацією* (within-group variation);
- у кожній віковій групі буде своє середнє значення зросту  $\mathbb{E}[Y | X]$ . Розкид таких середніх значень дорівнюватиме  $\text{Var}(\mathbb{E}[Y | X])$ . Відповідний показник називають *міжгруповою варіацією* (between-group variation).

Сума внутрішньогрупової і міжгрупової варіації становить повну дисперсію.

У контексті прогнозування, наприклад, прогнозування значень зросту  $Y$  на основі віку  $X$  людини, ідеальною була б ситуація, якби всі люди одного віку мали одинаковий зріст. Тоді ми б могли спрогнозувати зріст людини певного віку з абсолютною точністю. Іншими словами, внутрішньогрупова варіація повинна бути якнайменшою, оскільки її неможливо спрогнозувати на основі значень  $X$ . Тому часто таку варіацію називають *непоясненою дисперсією* (unexplained variance), тоді як міжгрупова варіація має також назву *поясненої дисперсії* (explained variance).

## 12.4. Умовні щільність і функція розподілу

Виклад значної частки матеріалу в попередніх лекціях було присвячено тому, що ми воліли знайти спосіб обчислювати ймовірності за участю деяких випадкових величин не за Визначенням 4.1.6, а з допомогою функцій розподілу, а також функцій імовірності та щільностей. Так само і для обчислення умовних імовірностей за участю випадкових величин корисно мати можливість застосовувати простіші способи, аніж Визначення 12.2.9. Для дискретних випадкових величин умовні функції ймовірності та розподілу можна визначити доволі просто, як це було видно в Розд. 12.1. Виявляється, для (абсолютно) неперервних випадкових величин також можна розглянути поняття умовної щільності розподілу.

**Твердження 12.4.1.** Нехай випадкові величини  $X$  та  $Z$  мають абсолютно неперервний спільний розподіл зі щільністю  $f_{Z,X}$ . Нехай функція  $g(Z)$  інтегровна. Тоді умовне сподівання  $\mathbb{E}[g(Z) | X]$  можна обчислити за допомогою *умовної щільності розподілу  $Z$  за умови  $X$*  (conditional probability density of  $Z$  given  $X$ )

$$f_{Z|X}(z | x) = \begin{cases} \frac{f_{Z,X}(z, x)}{f_X(x)}, & f_X(x) \neq 0 \\ \xi(z), & f_X(x) = 0 \end{cases} \quad (12.4.1)$$

у тому розумінні, що

$$\mathbb{E}[g(Z) | X] = h(X),$$

де

$$h(x) = \int_{-\infty}^{\infty} g(z) \cdot f_{Z|X}(z | x) dz. \quad (12.4.2)$$

У визначені умовної щільності  $\xi$  є деякою абсолютно довільною щільністю розподілу.

*Доведення.* Покажімо, що

$$h(x) = \frac{\int_{-\infty}^{\infty} g(z) f_{Z,X}(z, x) dz}{f_X(x)}.$$

Якщо ми це покажемо, то ми фактично покажемо, що

$$\mathbb{E}[g(Z) | X = x] = \frac{\int_{-\infty}^{\infty} g(z) f_{Z,X}(z, x) dz}{f_X(x)} = \int_{-\infty}^{\infty} g(z) \frac{f_{Z,X}(z, x)}{f_X(x)} dz = \int_{-\infty}^{\infty} g(z) \cdot f_{Z|X}(z | x) dz,$$

а відтак  $f_{Z|X}(z | x)$  матиме інтерпретацію умовної щільності розподілу.

Для того, щоб показати, що  $h(x)$  є версією умовного сподівання, потрібно перевірити виконання обох умов із Визначення 12.3.1. По-перше, функція  $h(x)$  вимірна відносно  $\mathcal{A}(X)$  та інтегровна (оскільки  $f(z, x)$  вимірна та інтегровна). Тому потрібно показати виконання умови

$$\int_G h(X) d\mathbb{P} = \int_G g(Z) d\mathbb{P}, \quad G \in \mathcal{A}(X).$$

Потрібно помітити, що якщо  $G \in \mathcal{A}(X)$ , то  $G = \{\omega : X(\omega) \in B\}$  для деякого  $B \in \mathcal{B}$ . Оскільки  $Z$  і  $X$  мають спільну щільність,  $X$  має маржинальну щільність  $f_X$ , і тому, згідно з Теоремами 6.4.2 та 9.1.8,

$$\begin{aligned} \int_G h(X) d\mathbb{P} &= \int_B h(x) f_X(x) dx \\ &= \int_B \frac{\int_{-\infty}^{\infty} g(z) f_{Z,X}(z, x) dz}{f_X(x)} \cdot f_X(x) dx \\ &= \int_B \int_{-\infty}^{\infty} g(z) f_{Z,X}(z, x) dz dx \\ &= \iint_{\mathbb{R}^2} g(z) f_{Z,X}(z, x) \cdot \mathbb{1}\{x \in B\} dz dx \\ &= \int g(Z) \cdot \mathbb{1}\{X \in B\} d\mathbb{P} = \int_G g(Z) d\mathbb{P}. \end{aligned}$$

□

Варто звернути увагу, що у випадку, коли  $f_X(x) = 0$  у (12.4.1), ми можемо дозволити собі покласти  $f_{Z|X}(z | x) = \xi(z)$ , де  $\xi$  — абсолютно довільна щільність, оскільки все одно

$$\int_B h(x) f_X(x) dx = 0 = \int_B \int_{-\infty}^{\infty} g(z) \xi(z) f_X(x) dz dx.$$

До речі, нескладно бачити, що умовна щільність розподілу (12.4.1) має всі властивості будь-якої щільності, зокрема, вона очевидно є невід'ємною, а її інтеграл дорівнює 1:

$$\int_{-\infty}^{\infty} f_{Z|X}(z | x) dz = \int_{-\infty}^{\infty} \frac{f_{Z,X}(z, x)}{f_X(x)} dz = \frac{1}{f_X(x)} \cdot \int_{-\infty}^{\infty} f_{Z,X}(z, x) dz = \frac{f_X(x)}{f_X(x)} = 1,$$

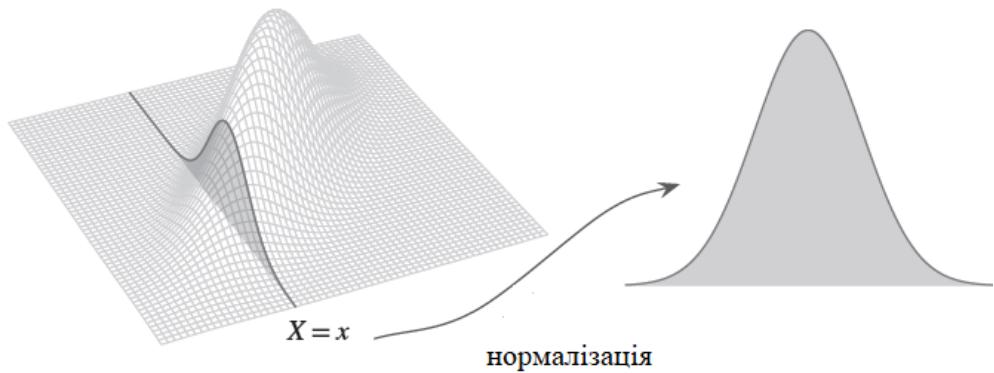


Рис. 12.4.1.: Ілюстрація умовної щільності розподілу ([1], Рис. 7.3)

якщо, звісно,  $f_X(x) \neq 0$ . Хоча якщо  $f_X(x) = 0$ , то ми взагалі маємо

$$\int_{-\infty}^{\infty} f_{Z|X}(z|x) dz = \int_{-\infty}^{\infty} \xi(z) dz = 1 ,$$

оскільки ми поклали з самого початку, що  $\xi$  є деякою довільною, але все ж таки щільністю.

**Зауваження 12.4.2.** Умовна щільність (12.4.1) має інтерпретацію щільності ще й тому, що за її допомогою можна обчислювати умовні ймовірності. Справді, якщо покласти  $g(Z) = \mathbb{1}\{Z \in A\}$ , згідно з (12.4.2) матимемо

$$\mathbb{P}(A | X = x) = \mathbb{E}[\mathbb{1}\{Z \in A\} | X = x] = \int_A f_{Z|X}(z|x) dz .$$

□

Як і у випадку умовної функції ймовірності, ми «вирізаємо» зі спільної щільності розподілу частину, яка відповідає  $X = x$ , і додатково ділимо на  $f_X(x)$  для нормалізації новоутвореної (умовної) щільності розподілу, щоб її інтеграл дорівнював 1. Відповідні міркування проілюстровано на Рис. 12.4.1.

**Приклад 12.4.3.** Розгляньмо нормальній вектор  $\mathbf{X} = (X_1, \dots, X_k)^\top \sim N(\boldsymbol{\mu}, \Sigma)$ . Його особливістю є не тільки той факт, що будь-яка підмножина його координат має багатовимірний нормальній розподіл, а й те, що умовні розподіли одних координат за умови інших також є нормальними. Так, зокрема, нехай

$$\mathbf{X} = (\mathbf{X}_1^\top, \mathbf{X}_2^\top)^\top \sim N\left((\boldsymbol{\mu}_1^\top, \boldsymbol{\mu}_2^\top)^\top, \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix}\right) ,$$

де  $\mathbf{X}_1 = (X_1, \dots, X_k)^\top$ ,  $\mathbf{X}_2 = (X_{k+1}, \dots, X_n)^\top$ ,  $\boldsymbol{\mu}_1 = (\mu_1, \dots, \mu_k)^\top$ ,  $\boldsymbol{\mu}_2 = (\mu_{k+1}, \dots, \mu_n)^\top$ , а  $\Sigma_{ij}$ ,  $i, j = 1, 2$  — це відповідні блоки матриці  $\Sigma$ .

Тоді простими, але тривалими алгебричними маніпуляціями спільної та маржинальної щіль-

ностей можна показати, що

$$\mathbf{X}_2 \mid \mathbf{X}_1 = \mathbf{x}_1 \sim N(\boldsymbol{\mu}_2 + \Sigma_{21}\Sigma_{11}^{-1}(\mathbf{x}_1 - \boldsymbol{\mu}_1), \Sigma_{22} - \Sigma_{21}\Sigma_{11}^{-1}\Sigma_{12}) .$$

□

**Приклад 12.4.4.** Повернімося до Прикладу 9.2.6, у якому ми розглядали спільний розподіл тиску в шинах автомобіля  $X$  та продуктивності пробігу  $Y$

$$f_{X,Y}(x, y) = \frac{21}{2}x^2y \cdot \mathbb{1}\{0 < x^2 \leq y < 1\} \cdot \mathbb{1}\{x > 0\} .$$

Спочатку дістаньмо маржинальні щільності розподілів:

$$\begin{aligned} f_X(x) &= \int_{-\infty}^{\infty} \frac{21}{2}x^2y \cdot \mathbb{1}\{0 < x^2 \leq y < 1\} \cdot \mathbb{1}\{x > 0\} dy \\ &= \frac{21}{2}x^2 \int_{x^2}^1 y \cdot \mathbb{1}\{0 < x < 1\} dy = \frac{21}{4}x^2(1 - x^4) \cdot \mathbb{1}\{0 < x < 1\} , \\ f_Y(y) &= \int_{-\infty}^{\infty} \frac{21}{2}x^2y \cdot \mathbb{1}\{0 < x^2 \leq y < 1\} \cdot \mathbb{1}\{x > 0\} dx \\ &= \frac{21}{2}y \int_0^{\sqrt{y}} x^2 \cdot \mathbb{1}\{0 < y < 1\} dx = \frac{7}{2}y^2\sqrt{y} \cdot \mathbb{1}\{0 < y < 1\} . \end{aligned}$$

Ми можемо обчислити умовну щільність розподілу  $X$  за умови  $Y$  і навпаки:

$$\begin{aligned} f_{X|Y}(x \mid y) &= \frac{f_{X,Y}(x, y)}{f_Y(y)} = \frac{\frac{21}{2}x^2y \cdot \mathbb{1}\{0 < x^2 \leq y < 1\} \cdot \mathbb{1}\{x > 0\}}{\frac{7}{2}y^2\sqrt{y} \cdot \mathbb{1}\{0 < y < 1\}} \\ &= \frac{3x^2}{y\sqrt{y}} \cdot \mathbb{1}\{0 < x \leq \sqrt{y}\} \cdot \mathbb{1}\{0 < y < 1\} , \\ f_{Y|X}(y \mid x) &= \frac{f_{X,Y}(x, y)}{f_X(x)} = \frac{\frac{21}{2}x^2y \cdot \mathbb{1}\{0 < x^2 \leq y < 1\} \cdot \mathbb{1}\{x > 0\}}{\frac{21}{4}x^2(1 - x^4) \cdot \mathbb{1}\{0 < x < 1\}} \\ &= \frac{2y}{1 - x^4} \cdot \mathbb{1}\{x^2 \leq y < 1\} \cdot \mathbb{1}\{0 < x < 1\} . \end{aligned}$$

Використовуючи ці умовні щільності, можемо, наприклад, обчислити таку умовну ймовірність ( $x^2 < a < b < 1$ ):

$$\mathbb{P}(a < Y < b \mid X = x) = \int_a^b \frac{2y}{1 - x^4} dy = \frac{b^2 - a^2}{1 - x^4} .$$

Наприклад, імовірність того, що продуктивність пробігу буде лежати між  $a = \frac{3}{8}$  та  $b = \frac{3}{4}$  за умови, що тиск у шинах дорівнює  $\frac{1}{2}$ , становить 0.45.

Також можемо обчислити умовне сподівання

$$\begin{aligned} \mathbb{E}[Y \mid X = x] &= \int_{-\infty}^{\infty} y \cdot \frac{2y}{1 - x^4} \cdot \mathbb{1}\{x^2 \leq y < 1\} \cdot \mathbb{1}\{0 < x < 1\} dy \\ &= \int_{x^2}^1 y \cdot \frac{2y}{1 - x^4} \cdot \mathbb{1}\{0 < x < 1\} dy = \frac{2(1 - x^6)}{3(1 - x^4)} \cdot \mathbb{1}\{0 < x < 1\} . \end{aligned}$$

Наприклад, якщо тиск у шинах дорівнює  $X = 0.5$ , то середня продуктивність пробігу буде 0.7.  $\square$

Для повноти картини ми також можемо розглянути поняття умовної функції розподілу.

**Визначення 12.4.5.** Якщо випадкові величини  $X$  і  $Z$  мають абсолютно неперервний спільний розподіл зі щільністю  $f_{Z,X}$ , то *умовна функція розподілу  $Z$  за умови  $X$*  (conditional distribution function of  $Z$  given  $X$ ) дорівнює

$$F_{Z|X}(z | x) = \int_{-\infty}^z f_{Z|X}(u | x) du = \int_{-\infty}^z \frac{f_{Z,X}(u, x)}{f_X(x)} du. \quad (12.4.3)$$

$\square$

**Зауваження 12.4.6.** Потрібно розуміти, що обумовлення в записі умовної щільності  $f_{Z|X}$  здійснюють цілою випадковою величиною  $X$  із відповідною  $\sigma$ -алгеброю, а не конкретною подією  $X = x$ , яка є нульовою, оскільки  $\mathbb{P}_X(X = x) = 0$  для неперервних величин.

За аналогією з Зауваженням 12.2.10 певну інтуїтивну інтерпретацію умовної щільності можна подати, розглянувши її не як похідну умовної функції розподілу, обчисленої для конкретної  $X = x$ ,  $F_{Z|X}(z | x) = \mathbb{P}_{Z|X}(Z \leq z | X = x)$ , а як границю таких похідних:

$$f_{Z|X}(z | x) = \lim_{h_n \downarrow 0} \frac{\partial}{\partial z} \mathbb{P}_{Z|X}(Z \leq z | X \in (x - h_n; x + h_n)).$$

У цьому випадку обумовлювальна подія завжди має ненульову ймовірність, якщо маржинальна щільність  $f_X(x) > 0$ .  $\square$

Для умовної щільності також, із цілком очевидних міркувань, справедлива теорема Беєса:

$$f_{Z|X}(z | x) = \frac{f_{X|Z}(x | z) f_Z(z)}{f_X(x)}, \quad f_X(x) > 0. \quad (12.4.4)$$

На практиці інколи виринають ситуації, коли потрібно дістати спільний розподіл величин, одна з яких дискретна, а інша — неперервна.

**Приклад 12.4.7.** Нехай дві компанії виробляють лампочки, до того ж лампочки компанії 0 працюють  $T_0 \sim \text{Exp}(\lambda_0)$  часу, а лампочки компанії 1 працюють  $T_1 \sim \text{Exp}(\lambda_1)$  часу. На ринку частка  $p_0$  лампочок належить компанії 0, а частка  $p_1 = 1 - p_0$  — компанії 1. Щоправда, на лампочках не нанесено жодного маркування, тому без їх випробування неможливо сказати, яка компанія їх виробила.

Нехай  $T$  = «тривалість роботи лампочки»,  $I$  = «компанія-виробник», де  $I = 1$ , якщо лампочку виробила компанія 1, і  $0$  — якщо компанія 0. Іншими словами, маржинально  $I \sim \text{Bern}(p_1)$ . Особливість спільного розподілу полягає в тому, що величина  $T$  має всього дві умовні щільності розподілу, адже  $T | I = 0 \sim \text{Exp}(\lambda_0)$  і  $T | I = 1 \sim \text{Exp}(\lambda_1)$ .

Спільна щільність розподілу величин  $T$  та  $I$  буде мати вид, зображеній на Рис. 12.4.2. Потрібно звернути увагу, що в цьому випадку добуток мір, для якого визначено відповідну щільність, є добутком мірі Лебега та лічної міри.

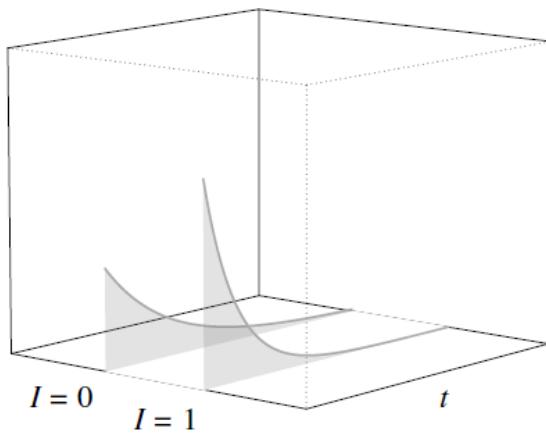


Рис. 12.4.2.: Спільний розподіл дискретної величини  $I$  та неперервної величини  $T$  з Прикладу 12.4.7 ([1], Рис. 7.8)

Звідси, наприклад, випливає маржинальна щільність розподілу величини  $T$  (для  $t > 0$ ):

$$\begin{aligned} f_T(t) &= \sum_i f_{T,I}(t, i) = f_{T,I}(t, 1) + f_{T,I}(t, 0) = f_{T|I}(t | 1)\mathbb{P}_I(I = 1) + f_{T|I}(t | 0)\mathbb{P}_I(I = 0) \\ &= p_1\lambda_1 e^{-\lambda_1 t} + p_0\lambda_0 e^{-\lambda_0 t}. \end{aligned}$$

Неважко бачити, що маржинальний розподіл  $T$  *не є* експоненційним.

Нарешті, маючи відповідні маржинальні щільності, можемо підрахувати ймовірність того, що деяку лампочку виробила конкретна компанія, *за умови*, що лампочка пропрацювала  $t$  часу. Наприклад,

$$\begin{aligned} \mathbb{P}_{I|T}(I = 1 | T = t) &= \frac{f_{T,I}(t, 1)}{f_T(t)} = \frac{f_{T|I}(t | 1)\mathbb{P}_I(I = 1)}{f_T(t)} \\ &= \frac{p_1\lambda_1 e^{-\lambda_1 t}}{p_1\lambda_1 e^{-\lambda_1 t} + p_0\lambda_0 e^{-\lambda_0 t}} = \frac{p_1\lambda_1}{p_1\lambda_1 + p_0\lambda_0 e^{\lambda_1 - \lambda_0 t}}, \end{aligned}$$

тобто

$$I | T = t \sim \text{Bern}\left(\frac{p_1\lambda_1}{p_1\lambda_1 + p_0\lambda_0 e^{\lambda_1 - \lambda_0 t}}\right).$$

□

### Частина III.

## Послідовності випадкових величин

## 13. Збіжність послідовностей випадкових величин

Чи не найважливішою частиною теорії ймовірностей є вивчення поведінки деякої *послідовності* випадкових величин. Цей розділ називають *асимптотичною теорією* (asymptotic theory), або інколи *теорією великих вибірок* (large sample theory) чи *граничною теорією* (limit theory). У рамках цієї теорії зasadниче питання, на яке ми прагнемо дати відповідь, таке: як себе поводить послідовність випадкових величин  $X_1, X_2, \dots$ , коли  $n \rightarrow \infty$ , тобто коли  $n$  «дуже велике»? Оскільки статистика та наука про дані в ширшому сенсі працюють із великими наборами даних, нас безперечно цікавить, що відбувається зі збільшенням кількості даних, доступних досліднику.

Двома ключовими результатами цього розділу для нас будуть закон великих чисел та центральна гранична теорема. Але почати ми повинні з розгляду базових понять і властивостей, пов'язаних із послідовностями подій, випадкових величин та їхніх сподівань.

### 13.1. Послідовності подій

За аналогією з числовими послідовностями (Визначення 7.3.11) можна визначити поняття верхньої і нижньої границі деякої послідовності подій.

**Визначення 13.1.1.** Нехай маємо деяку послідовність подій  $A_1, A_2, \dots$  у деякому ймовірнісному просторі  $(\Omega, \mathcal{A}, \mathbb{P})$ . *Нижньою границею* (limit inferior) цієї послідовності є подія

$$\liminf_{n \rightarrow \infty} A_n = \bigcup_{n=1}^{\infty} \bigcap_{k=n}^{\infty} A_k. \quad (13.1.1)$$

Аналогічно, *верхньою границею* (limit superior) цієї послідовності є

$$\limsup_{n \rightarrow \infty} A_n = \bigcap_{n=1}^{\infty} \bigcup_{k=n}^{\infty} A_k. \quad (13.1.2)$$

□

Словесна інтерпретація відповідних граничних подія така:

- ми кажемо, що  $\omega \in \liminf_{n \rightarrow \infty} A_n$ , якщо  $\omega \in A_n$  рано чи пізно (eventually). Іншими словами, існує таке  $n$ , що для всіх  $m \geq n$  матимемо  $\omega \in A_m$ , тобто елемент належатиме всім множинам послідовності, починаючи з деякої;
- ми кажемо, що  $\omega \in \limsup_{n \rightarrow \infty} A_n$ , якщо  $\omega \in A_n$  нескінченно часто (infinitely often). Іншими словами, для будь-якого  $n$ , яке б велике воно не було, існує таке  $m \geq n$ , що  $\omega \in A_m$ , тобто  $\omega$  належить нескінченному числу множин із послідовності. Часто позначають  $\limsup_{n \rightarrow \infty} A_n = \{\omega : \omega \in A_n \text{ н.ч.}\}$ .

Нижня й верхня границі послідовності подій вкладені одна в одну:  $\liminf_{n \rightarrow \infty} A_n \subseteq \limsup_{n \rightarrow \infty} A_n$ . Це випливає з відповідної інтерпретації: якщо елемент  $\omega$  належить усім множинам рано чи пізно, то зрозуміло, що він належить нескінченному числу подій. Зворотне вкладення в загальному випадку не справедливе: якщо елемент  $\omega$  належить нескінченному числу подій, немає жодної гарантії, що він належатиме усім із них, окрім перших  $m$  штук.

**Приклад 13.1.2.** На перший погляд може здатися, що ці визначення доволі накручені. Проте їх інтерпретація в простих випадках інтуїтивна зрозуміла. Нехай ми підкидаємо гральну кісточку нескінченну кількість разів. Очевидно, що цифра 1 випаде *нескінченно часто*, тобто  $1 \in \limsup_{n \rightarrow \infty} A_n$  або  $1 \in \{A_n \text{ н.ч.}\}$ , де  $A_i \in \{\{1\}, \{2\}, \{3\}, \{4\}, \{5\}, \{6\}\}$ ,  $i > 1$ . Звісно, ми не можемо гарантувати, що 1 буде випадати *постійно*, починаючи з деякого  $m$ , тому ми не можемо казати, що 1 належить  $\liminf_{n \rightarrow \infty} A_n$ .

З іншого боку, можна стверджувати, що середнє значення випалих чисел  $\bar{X}_n = \frac{\sum_{i=1}^n X_i}{n}$ , де  $X_i = \text{«число, яке випало на гральній кісточці на } i\text{-ому підкиданні»}$ , буде *рано чи пізно* відхилятися від 3.5 менше від  $\varepsilon$  для будь-якого  $\varepsilon > 0$  (принаймні ми це доведемо в наступній лекції). Тобто  $\left\{ \omega : \left| \bar{X}(\omega) - 3.5 \right| < \varepsilon \right\} \in \liminf_{n \rightarrow \infty} B_n$  для будь-якого  $\varepsilon > 0$ , де, наприклад,  $B_n = \left\{ \left| \bar{X} - 3.5 \right| < \frac{1}{n} \right\}$   $\square$

Ми кажемо, що існує границя подій  $A$ , якщо нижня й верхня границі дорівнюють одна одній:

$$A = \lim_{n \rightarrow \infty} A_n = \liminf_{n \rightarrow \infty} A_n = \limsup_{n \rightarrow \infty} A_n. \quad (13.1.3)$$

Для спрощення позначень писатимемо  $A_n \rightarrow A$ .

**Твердження 13.1.3.** Нехай маємо деяку послідовність подій  $A_1, A_2, \dots$  у деякому ймовірнісному просторі  $(\Omega, \mathcal{A}, \mathbb{P})$ . Тоді

$$\mathbb{P} \left( \liminf_{n \rightarrow \infty} A_n \right) \leq \liminf_{n \rightarrow \infty} \mathbb{P}(A_n) \leq \limsup_{n \rightarrow \infty} \mathbb{P}(A_n) \leq \mathbb{P} \left( \limsup_{n \rightarrow \infty} A_n \right). \quad (13.1.4)$$

До того ж, якщо  $A_n \rightarrow A$ , то  $\mathbb{P}(A_n) \rightarrow \mathbb{P}(A)$ .

Доведення цього твердження наведено в Розд. 13.4.

**Приклад 13.1.4.** Розгляньмо експеримент із підкиданням правильної монетки нескінченну кількість разів, який ми вперше описали в Прикладі 1.2.7. У цьому випадку простір елементарних подій  $\Omega = \{\omega = (d_1, d_2, \dots), d_j \in \{0, 1\}, j \geq 1\}$  ізоморфний інтервалу  $(0; 1]$ , оскільки кожне дійсне число з цього інтервалу має двійкове розвинення

$$x = \sum_{n=1}^{\infty} \frac{d_n}{2^n} = 0.d_1d_2\dots, \quad d_n \in \{0, 1\}, n \in \mathbb{N}.$$

Розгляньмо деяку серію підкідів монетки і позначмо через  $l_n(\omega)$  число нулів, починаючи з позиції  $n$ . Тобто  $l_n(\omega) = k$ , якщо  $\omega = \{d_1, \dots, d_n = 0, d_{n+1} = 0, \dots, d_{n+(k-1)} = 0, d_{n+k} = 1, \dots\}$ . Також вважатимемо, що  $l_n(\omega) = 0$ , якщо відповідний  $d_n = 1$ . Імовірність множини  $L_{n,k}$  всіх таких  $\omega$ , для яких  $l_n(\omega) = k$ , можна обчислити, помітивши, що позиція  $n$  через незалежність підкідів

не має значення: ми можемо покласти  $n = 1$ . Як було показано в Прикладі 1.3.3, імовірність дорівнюватиме

$$\mathbb{P}(L_{n,k}) = \frac{1}{2^{k+1}},$$

оскільки  $k$  позицій дорівнюють 0, а  $(k+1)$ -а дорівнює 1.

Розгляньмо тепер множину  $A_n = \{\omega : l_n(\omega) \geq r\}$ , тобто множину всіх таких серій підкідів, у яких, починаючи з  $n$ -ої позиції, мають місце  $r$  нулів і більше. Імовірність можна обчислити як суму нескінченно спадної геометричної прогресії:

$$\mathbb{P}(\{\omega : l_n(\omega) \geq r\}) = \sum_{k=r}^{\infty} \mathbb{P}(L_{n,k}) = \sum_{k=r}^{\infty} \frac{1}{2^{k+1}} = \frac{1}{2^{r+1}} \frac{1}{1 - 1/2} = \frac{1}{2^r}.$$

Тоді подія  $\limsup_{n \rightarrow \infty} A_n$  має інтерпретацію множини всіх серій підкідів  $\omega$  таких, що в них нескінченно часто зустрічаються ситуації, коли послідовність нулів має довжину, більшу або рівну до  $r$ .

Відповідно до Твердження 13.1.3, імовірність цього

$$\mathbb{P}(\{A_n \text{ н.ч.}\}) \geq \frac{1}{2^r}.$$

□

Можемо розглянути дві дуже важливі властивості послідовностей подій, які мають назви лем Бореля-Кантеллі.

**Теорема 13.1.5** (Перша лема Бореля-Кантеллі (First Borel-Cantelli lemma)<sup>1</sup>). Нехай маємо деяку послідовність подій  $A_1, A_2, \dots$  у деякому ймовірнісному просторі  $(\Omega, \mathcal{A}, \mathbb{P})$ . Якщо ряд  $\sum_{n=1}^{\infty} \mathbb{P}(A_n)$  збіжний, то  $\mathbb{P}(\{A_n \text{ н.ч.}\}) \equiv \mathbb{P}\left(\limsup_{n \rightarrow \infty} A_n\right) = 0$ .

*Доведення.* Нехай  $C_n = \bigcup_{k=n}^{\infty} A_k$ . Тоді  $C_n \rightarrow \limsup_n A_n$ , оскільки  $C_n$  утворюють незростаючу послідовність множин. Відтак

$$\mathbb{P}(\{A_n \text{ н.ч.}\}) \equiv \mathbb{P}\left(\bigcap_{n=1}^{\infty} \bigcup_{k=n}^{\infty} A_k\right) = \mathbb{P}\left(\lim_{n \rightarrow \infty} C_n\right) = \lim_{n \rightarrow \infty} \mathbb{P}(C_n) \leq \lim_{n \rightarrow \infty} \sum_{k=n}^{\infty} \mathbb{P}(A_k),$$

де ми використали неперервність міри та нерівність Була (2.2.2).

Якщо  $\sum_{n=1}^{\infty} \mathbb{P}(A_n) < \infty$ , то  $\lim_{n \rightarrow \infty} \sum_{k=n}^{\infty} \mathbb{P}(A_k) = 0$ , звідки й випливає результат леми. □

Суть цієї леми полягає в тому, що якщо ймовірності подій наближаються до нуля, і відповідний ряд збіжний, то події  $A_n$  обов'язково припиняться.

<sup>1</sup>Названа так на честь Еміля Бореля та італійського математика Франческо Кантеллі (Francesco Paolo Cantelli, 1875–1966).

**Приклад 13.1.6.** Продовжмо розгляд Прикладу 13.1.4. Цього разу вважатимемо, що маємо не стало  $r$ , а послідовність  $\{r_n\}$  додатних цілих чисел. Тоді, якщо

$$\sum_{n=1}^{\infty} \frac{1}{2^{r_n}} < \infty,$$

згідно з першою лемою Бореля-Кантеллі маємо

$$\mathbb{P}(\{\omega : l_n(\omega) \geq r_n \text{ н.ч.}\}) = 0.$$

Щоб відповідний ряд був збіжний,  $r_n$  повинні постійно збільшуватися. Тобто ми довели, що серії підкідів, у яких нескінченну кількість разів зустрічаються нерозривні послідовності нулів усе більшої й більшої довжини, є неймовірні (тобто становлять собою множину міри нуль).  $\square$

**Приклад 13.1.7.** Повернімося до обчислення ймовірності події (1.3.3), що в нескінченній послідовності підкідів монетки з Прикладу 1.2.7 міститься «однакова кількість» нулів і одиниць у розумінні, що границя емпіричних частот зустрічання нулів або одиниць дорівнює  $\frac{1}{2}$ :

$$N = \left\{ \omega : \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{i=1}^n d_i(\omega) = \frac{1}{2} \right\},$$

де

$$\Omega = (0; 1] = \{\omega = (d_1(\omega), d_2(\omega), \dots), d_j(\omega) \in \{0, 1\}, j \geq 1\}.$$

Числа з множини  $N$  називають *нормальними* (normal).

Можна показати, що  $N \in \mathcal{B}(0; 1]$ , тобто належить відповідній Борелевій  $\sigma$ -алгебрі. Ми цього робити не будемо, а приймемо як даність.

Покажімо, що заперечення множини  $N$  є множиною міри нуль. Це не так очевидно, адже множина  $N^c$  є незліченою. Наприклад, розгляньмо точку  $\omega = (1, 1, u_3, 1, 1, u_6, \dots)$ ,  $u_i \in \{0, 1\}$ , тобто таку точку, де  $d_i(\omega) = 1$  якщо  $i$  не кратне 3. Оскільки в цьому випадку

$$\frac{1}{n} \sum_{i=1}^n d_i(\omega) \geq \frac{2}{3},$$

така точка не є нормальнюю. Проте таких точок є незлічена кількість, оскільки незліченим є число послідовностей  $(u_3, u_6, u_9, \dots)$ .

Розгляньмо замість послідовності із 0 і 1 еквівалентну їй послідовність із  $-1$  та  $1$  таку, що  $r_i = 2d_i - 1$ ,  $i = 1, 2, \dots$ . Уведімо позначення часткових сум:

$$s_n(\omega) = \sum_{i=1}^n r_i(\omega).$$

Тоді множину нормальніх чисел можна переписати як

$$N = \left\{ \omega : \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{i=1}^n s_i(\omega) = 0 \right\}.$$

Згідно з нерівністю Маркова (11.1.5), маємо

$$\mathbb{P}(|s_n(\omega)| \geq n\varepsilon_n) \leq \frac{\mathbb{E}[s_n^4]}{n^4\varepsilon_n^4}$$

для деякого  $\varepsilon_n > 0$ .

Відповідний момент  $s_n(\omega)$  можна обчислити, помітивши, що  $d_i(\omega) \sim \text{Bern}(0.5)$ , а, відповідно,  $D \equiv \sum_{i=1}^n d_i(\omega) \sim \text{Binom}(n, 0.5)$ . Тоді можна показати, що  $\mathbb{E}[s_n^4] = n + 3n(n-1) \leq 3n^2$ , звідки випливає

$$\mathbb{P}\left(\left|\frac{1}{n}s_n(\omega)\right| \geq \varepsilon_n\right) \leq \frac{3}{n^2\varepsilon_n^4}.$$

Зокрема, якщо покласти  $\varepsilon_n = n^{-1/8}$ , щоб був збіжний ряд

$$\sum_{n=1}^{\infty} \frac{3}{n^2\varepsilon_n^4} = \sum_{n=1}^{\infty} \frac{3}{n^2(n^{-1/8})^4} = \sum_{n=1}^{\infty} \frac{3}{n^{-3/2}} < \infty,$$

то виконуватимуться умови першої леми Бореля-Кантеллі 13.1.5, а відтак  $\mathbb{P}(\{A_n \text{ н.ч.}\}) = 0$ , де

$$A_n = \left\{ \omega : \left| \frac{1}{n}s_n(\omega) \right| \geq n^{-\frac{1}{8}} \right\}.$$

Іншими словами, ми показали, що множина  $\{A_n \text{ н.ч.}\}$  є множиною міри нуль. Ця множина є нічим іншим, як  $N^c$ . Справді, якщо серія підкідів  $\omega \in \{A_n \text{ н.ч.}\}$ , то це означає, що існує нескінчена кількість  $n$  таких, що середньоаритметичне  $s_n$  не менше від  $n^{-1/8} > 0$ . Жодне нормальне число не може належати такій множині, оскільки тоді границя середньоаритметичних  $s_n$  ніколи не наблизиться до 0. Тому  $\mathbb{P}(N) = 1$ , тобто серія підкідів матиме «однакову» кількість одиниць і нулів майже напевно.  $\square$

Розгляньмо другу лему Бореля-Кантеллі, яка в певному сенсі є дзеркальним твердженням до леми першої, проте висуває додаткову вимогу до відповідних подій.

**Теорема 13.1.8** (Друга лема Бореля-Кантеллі (Second Borel-Cantelli lemma)). Нехай маємо деяку послідовність *незалежних* подій  $A_1, A_2, \dots$  у деякому ймовірнісному просторі  $(\Omega, \mathcal{A}, \mathbb{P})$ . Тоді якщо ряд  $\sum_{n=1}^{\infty} \mathbb{P}(A_n)$  розбіжний, то  $\mathbb{P}(\{A_n \text{ н.ч.}\}) \equiv \mathbb{P}\left(\limsup_{n \rightarrow \infty} A_n\right) = 1$ .

*Доведення.* Очевидно, що можна довести еквівалентне твердження, що  $\mathbb{P}(\{A_n \text{ н.ч.}\}^c) = 0$ , тобто що множина

$$\left(\limsup_{n \rightarrow \infty} A_n\right)^c = \left(\bigcap_{n=1}^{\infty} \bigcup_{k=n}^{\infty} A_k\right)^c = \bigcup_{n=1}^{\infty} \bigcap_{k=n}^{\infty} A_k^c$$

є множиною міри нуль. Для цього достатньо показати, що множинами міри нуль є всі множини

$\bigcap_{k=n}^{\infty} A_k^c$ . Справді,

$$\begin{aligned}\mathbb{P}\left(\bigcap_{k=n}^{\infty} A_k^c\right) &= \lim_{r \rightarrow \infty} \mathbb{P}\left(\bigcap_{k=n}^r A_k^c\right) = \lim_{r \rightarrow \infty} \prod_{k=n}^r \mathbb{P}(A_k^c) \\ &= \lim_{r \rightarrow \infty} \prod_{k=n}^r (1 - \mathbb{P}(A_k)) \leq \lim_{r \rightarrow \infty} \prod_{k=n}^r e^{-\mathbb{P}(A_k)} \\ &= \lim_{r \rightarrow \infty} e^{-\sum_{k=n}^r \mathbb{P}(A_k)} \\ &= e^{-\sum_{k=n}^{\infty} \mathbb{P}(A_k)} = 0,\end{aligned}$$

де ми використали неперервність міри, незалежність подій для переходу від перетину до добутку, властивість  $1 - x \leq e^{-x}$  для  $x \geq 0$ , та той факт, що відповідний ряд є розбіжний.  $\square$

**Приклад 13.1.9.** Продовжмо розгляд Прикладів 13.1.4–13.1.6. Розгляньмо події

$$A_n = \{\omega : l_n(\omega) = 0\} = \{\omega : d_n(\omega) = 1\}, \quad n = 1, 2, \dots.$$

Ці події, очевидно, незалежні, оскільки підкиди монетки незалежні. Понад те, оскільки  $\mathbb{P}(A_n) = \frac{1}{2}$  незалежно від  $n$ , відповідний ряд буде розбіжний. Отже виконуються всі умови другої леми Бореля-Кантеллі, а відтак  $\mathbb{P}(\{A_n \text{ н.ч.}\}) = 1$ , тобто майже напевно 1 буде зустрічатися на нескінченній кількості позицій.

Розгляньмо тепер події

$$A_n = \{\omega : l_n(\omega) = 1\} = \{\omega : d_n(\omega) = 0, d_{n+1}(\omega) = 1\}, \quad n = 1, 2, \dots.$$

Ці події *не є незалежні*, оскільки якщо ми знаємо, що сталася подія  $A_1$ , то подія  $A_2$  ніяк не може статися, адже ми знаємо, що на другій позиції уже стоїть 1, хоча для події  $A_2$  там повинен стояти 0. Тобто умови другої леми Бореля-Кантеллі не виконуються.

Проте можна помітити, що незалежними є події  $A_2, A_4, A_6, \dots$ , оскільки в цьому випадку ділянки серії підкідів не будуть накладатися одна на одну. До того ж  $\mathbb{P}(A_n) = \frac{1}{4}$ ,  $n = 2, 4, \dots$ , незалежно від  $n$ , тобто відповідний ряд із імовірностей знову буде розбіжний. Тоді з другої леми Бореля-Кантеллі випливає  $\mathbb{P}(\{A_n \text{ н.ч.}\}) = 1$ ,  $n = 2, 4, \dots$ . Звідси випливає, що  $\mathbb{P}(\{A_n \text{ н.ч.}\}) = 1$  для всіх  $n = 1, 2, \dots$ . Іншими словами, майже напевно будемо зустрічати нескінченну кількість ділянок виду  $\dots, 0, 1, \dots$

Аналогічні міркування можна продовжувати для  $l_n(\omega) = 2, 3, \dots$   $\square$

**Зауваження 13.1.10.** Можна помітити, що якщо події  $A_1, A_2, \dots$  незалежні, то виконуються обидві леми Бореля-Кантеллі. Оскільки ряд із  $\mathbb{P}(A_n)$  може бути або збіжний, або розбіжний (третього не дано), то  $\mathbb{P}(\{A_n \text{ н.ч.}\})$  може бути *тільки або 0, або 1*.  $\square$

## 13.2. Послідовності випадкових величин

Винятково важливим для статистики та теорії випадкових процесів є поняття збіжності послідовності випадкових величин. Якби ми нічого не знали про поведінку послідовностей випадкових

величин зі збільшенням числа таких величин у послідовності, ми б зовсім не могли аналізувати дані, оцінювати параметри відповідних розподілів, які ці дані породжують, та робити на основі цих даних прогнози.

Проте якщо границею деякої числової послідовності є число, а границею деякої послідовності подій — подія, то в контексті випадкових величин можна говорити про різні види збіжностей. Наприклад, можна розглядати збіжність послідовності випадкових величин як функцій до деякої функції (до деякої випадкової величини), а можна говорити про збіжність послідовності випадкових величин до деякого невипадкового числа, у тому розумінні, що зі збільшенням числа випадкових величин їх розподіл концентрується навколо конкретного значення з нульовою дисперсією. Тому варто розглянути, які в літературі виділяють види збіжностей.

### 13.2.1. Збіжність майже напевно

Розгляньмо послідовність випадкових величин  $X_n : (\Omega, \mathcal{A}, \mathbb{P}) \rightarrow (\mathbb{R}, \mathcal{B})$ ,  $n = 1, 2, \dots$ , та деяку випадкову величину  $X : (\Omega, \mathcal{A}, \mathbb{P}) \rightarrow (\mathbb{R}, \mathcal{B})$ .

Оскільки випадкова величина — це вимірна функція, найперший вид збіжності, який спадає на думку, полягає в тому, що послідовність  $X_n(\omega) \rightarrow X(\omega)$  у кожній точці  $\omega$  як числові послідовності.

**Визначення 13.2.1.** Послідовність випадкових величин  $X_n$  збігається до випадкової величини  $X$  поточково (converges pointwise), якщо для кожного елемента простору виконується

$$\lim_{n \rightarrow \infty} X_n(\omega) = X(\omega), \quad \omega \in \Omega. \quad (13.2.1)$$

□

Зверніть увагу, що границя відповідної послідовності буде вимірною функцією, що ми показали в Твердженні 7.3.12.

**Приклад 13.2.2.** Нехай  $\Omega = [0; 1]$  і

$$X_n(\omega) = \left(1 + \frac{\omega}{n}\right)^n.$$

Тоді  $X_n \rightarrow X$  поточково, де

$$X(\omega) = e^\omega,$$

оскільки для всіх  $\omega$  виконується друга важлива границя.

Нехай тепер

$$X_n(\omega) = \omega^n.$$

У цьому випадку маємо таку поточкову збіжність:

$$X_n(\omega) \rightarrow X(\omega) = \begin{cases} 0, & \omega \in [0; 1) \\ 1, & \omega = 1 \end{cases}.$$

Іншими словами,  $X_n \rightarrow X$  поточково, проте не можна сказати, що  $X_n \rightarrow 0$  поточково, оскільки в одній точці збіжність до 0 не виконується. □

Як можна побачити з цього прикладу, поточкова збіжність може не виконуватися в деякій конкретній точці, або ширше — на деякій зліченні множині точок, або ще ширше — не деякій множині міри нуль. Як ми вже звикли в цьому курсі, якщо якась властивість не виконується на множині міри нуль, нею можна знехтувати, адже від цього не зміниться, наприклад, значення інтегралу випадкової величини.

Із цих міркувань випливає таке посилене визначення збіжності.

**Визначення 13.2.3.** Послідовність випадкових величин  $X_n$  збігається до випадкової величини  $X$  *майже напевно* (converges almost surely) або з *імовірністю 1* (with probability 1), що позначають як  $X_n \xrightarrow{\text{м.н.}} X$ , якщо виконується

$$\mathbb{P} \left( \left\{ \omega \in \Omega : \lim_{n \rightarrow \infty} X_n(\omega) = X(\omega) \right\} \right) = 1, \quad (13.2.2)$$

або, що те ж саме,

$$\mathbb{P} \left( \left\{ \omega \in \Omega : \exists \varepsilon > 0 \quad \forall N_\varepsilon(\omega) \in \mathbb{N}_\varepsilon(\omega) \quad \exists n \geq N \quad |X_n(\omega) - X(\omega)| \geq \varepsilon \right\} \right) = 0. \quad (13.2.3)$$

Якщо замість випадкових величин ми маємо послідовність випадкових векторів  $\mathbf{X}_1, \mathbf{X}_2, \dots$ , то  $\mathbf{X}_n \xrightarrow{\text{м.н.}} \mathbf{X}$ , якщо виконується покоординатна збіжність.  $\square$

**Зауваження 13.2.4.** Можна помітити, що якщо  $X_n \xrightarrow{\text{м.н.}} X$ , то це те саме, ніби  $X_n - X \xrightarrow{\text{м.н.}} 0$ . Це ж міркування справедливе для всіх інших видів збіжності, які розглядалимо нижче.  $\square$

**Приклад 13.2.5.** Нехай маємо послідовність незалежних випадкових величин  $X_1, X_2, \dots, X_n \sim U((0; 1])$ ,  $n = 1, 2, \dots$ . Нехай  $Y_n = \min \{X_1, \dots, X_n\}$ . Можна показати, що  $Y_n \xrightarrow{\text{м.н.}} 0$ .

Цілком очевидно, що  $Y_n \geq Y_{n+1}$  для будь-якого  $n$ . Оскільки така послідовність не може впасти нижче 0, вона повинна мати границю, яку позначимо через  $Y$ . Нехай  $\varepsilon > 0$ . Тоді  $Y_n \geq \varepsilon$  тоді й тільки тоді, коли всі  $X_i \geq \varepsilon$ :

$$\mathbb{P}_{Y_n} (Y_n \geq \varepsilon) = \mathbb{P}_X (X_1 \geq \varepsilon) \cdot \dots \cdot \mathbb{P}_X (X_n \geq \varepsilon) = (1 - \varepsilon)^n.$$

Оскільки це твердження справедливе для будь-якого  $n$ ,

$$\mathbb{P}_Y (Y \geq \varepsilon) \leq \lim_{n \rightarrow \infty} (1 - \varepsilon)^n = 0.$$

Отже ми показали, що  $\mathbb{P}_Y (Y \geq \varepsilon) = 0$ , але оскільки це справедливо для будь-якого  $\varepsilon > 0$ ,  $\mathbb{P}_Y (Y > 0) = 0$ . Іншими словами,  $\mathbb{P}_Y (Y = 0) = 1$ . Але оскільки ми поклали, що  $Y$  — це границя  $Y_n$ , то фактично  $Y_n \xrightarrow{\text{м.н.}} 0$ .  $\square$

Існують два альтернативні формулювання збіжності майже напевно.

**Твердження 13.2.6.** Нехай маємо послідовність випадкових величин  $X_n$ ,  $n = 1, 2, \dots$ , та випадкову величину  $X$ . Тоді  $X_n \xrightarrow{\text{м.н.}} X$ , якщо для будь-якого  $\varepsilon > 0$  виконується

$$\mathbb{P} \left( \left\{ \omega \in \Omega : |X_n(\omega) - X(\omega)| \geq \varepsilon \text{ н.ч.} \right\} \right) = 0.$$

Доведення цього твердження наведено в Розд. 13.4.

Особлива принада Твердження 13.2.6 полягає в тому, що збіжність майже напевно матиме місце, якщо для будь-якого  $\varepsilon > 0$  та для будь-якого  $\omega$  множина  $n$  таких, що  $|X_n(\omega) - X(\omega)| \geq \varepsilon$  є скінченою.

Твердження 13.2.6 у поєднанні з першою лемою Бореля-Кантеллі 13.1.5 дає таку характеристику збіжності майже напевно, яку приймемо без доведення.

**Твердження 13.2.7.** Нехай маємо послідовність випадкових величин  $X_n, n = 1, 2, \dots$ , та випадкову величину  $X$ . Тоді  $X_n \xrightarrow{\text{М.Н.}} X$ , якщо для будь-якого  $\varepsilon > 0$  виконується

$$\sum_{n=1}^{\infty} \mathbb{P}(\{\omega \in \Omega : |X_n(\omega) - X(\omega)| \geq \varepsilon\}) < \infty.$$

### 13.2.2. Збіжність за ймовірністю

Окрім збіжності майже напевно можна також розглянути варіант збіжності, за якого ми міняємо місцями взяття границі та обчислення ймовірності. Якщо для збіжності майже напевно ми вимагали, щоб множиною міри нуль була множина  $A$  всіх точок, у яких не виконується поточкова збіжність, то можемо розглянути вид збіжності, для якого ми будуємо аналогічну множину точок  $A_n$  для кожного  $n$  окремо. Тоді ми кажемо, що послідовність збігається, якщо послідовність відповідних множин прямує до множини міри нуль.

**Визначення 13.2.8.** Послідовність випадкових величин  $X_n$  збігається до випадкової величини  $X$  за ймовірністю (converges in probability), що позначають як  $X_n \xrightarrow{p} X$ , якщо виконується

$$\forall \varepsilon > 0 \quad \lim_{n \rightarrow \infty} \mathbb{P}(\{\omega \in \Omega : |X_n(\omega) - X(\omega)| > \varepsilon\}) = 0. \quad (13.2.4)$$

Якщо замість випадкових величин ми маємо послідовність випадкових векторів  $\mathbf{X}_1, \mathbf{X}_2, \dots$ , то  $\mathbf{X}_n \xrightarrow{p} \mathbf{X}$ , якщо виконується покоординатна збіжність.  $\square$

**Приклад 13.2.9.** Нехай  $X \sim \text{Bern}\left(\frac{1}{3}\right)$ , а

$$X_n = \left(1 + \frac{1}{n}\right) X.$$

Можна помітити, що в цьому випадку

$$|X_n - X| = \left| \frac{1}{n} X \right| = \frac{1}{n} X,$$

оскільки  $X \geq 0$ .

Тоді для деякого  $\varepsilon > 0$  маємо

$$\begin{aligned}\mathbb{P}(|X_n - X| > \varepsilon) &= \frac{2}{3}\mathbb{P}(|X_n - X| > \varepsilon \mid X = 0) + \frac{1}{3}\mathbb{P}(|X_n - X| > \varepsilon \mid X = 1) \\ &= \frac{2}{3}\mathbb{P}(0 > \varepsilon) + \frac{1}{3}\mathbb{P}\left(\frac{1}{n} > \varepsilon\right) \\ &= \frac{1}{3}\mathbb{P}\left(n < \frac{1}{\varepsilon}\right) = \frac{1}{3}\mathbb{1}\left\{n < \frac{1}{\varepsilon}\right\},\end{aligned}$$

що явно прямує до 0, коли  $n \rightarrow \infty$ .

Отже ми показали, що  $X_n \xrightarrow{p} X$ . □

**Зауваження 13.2.10.** Варто помітити, що для перевірки збіжності за ймовірністю потрібно знати спільний розподіл  $X_n$  та  $X$ . Проте якщо ми розглядаємо збіжність до деякого детермінованого числа  $c$ ,  $X_n \xrightarrow{p} c$ , достатньо перевірити виконання

$$\forall \varepsilon > 0 \quad \lim_{n \rightarrow \infty} \mathbb{P}(\{\omega \in \Omega : |X_n(\omega) - c| > \varepsilon\}) = 0,$$

що вимагає тільки знання розподілу  $X_n$ . □

**Приклад 13.2.11.** Нехай  $Y \sim \text{Exp}(1)$ . Визначмо послідовність випадкових величин як  $Y_n = Y/n$  для всіх  $n \in \mathbb{N}$ . Тоді для будь-якого  $\varepsilon > 0$  маємо

$$\mathbb{P}(\{\omega \in \Omega : |Y_n(\omega) - 0| \geq \varepsilon\}) = \mathbb{P}_Y(Y \geq n\varepsilon) = e^{-n\varepsilon} \rightarrow 0.$$

Отже можемо стверджувати, що  $Y_n \xrightarrow{p} 0$ . □

Ключову відмінність збіжності за ймовірністю від збіжності майже напевно можна проілюструвати так. Нехай маємо деяке фіксоване  $\varepsilon > 0$ . Тоді збіжність майже напевно гарантує, що дляожної точки  $\omega$  (окрім як із множини міри нуль) знайдеться таке  $n$ , починаючи з якого випадкова величина відхилятиметься від своєї границі менше від  $\varepsilon$ .

Натомість збіжність за ймовірністю нічого такого не гарантує. Єдине, що вона гарантує — що ймовірності відповідних відхилень прямуватимуть до нуля. Тобто множина таких  $\omega$ , для яких випадкова величина відхиляється від своєї границі більше від  $\varepsilon$ , зі збільшенням  $n$  стає все менш імовірною, проте вона в загальному випадку матиме ненульову ймовірність для будь-якого  $n$ .

Можна розглянути корисну теорему, яка дає змогу зберігати збіжність випадкових величин під дією неперервної функції.

**Теорема 13.2.12** (Теорема про неперервне відображення (Continuous mapping theorem)). Нехай маємо послідовність випадкових векторів  $\mathbf{X}_n \xrightarrow{p} \mathbf{X}$ . Нехай  $g : \mathbb{R}^k \rightarrow \mathbb{R}^n$  неперервна майже напевно. Тоді  $g(\mathbf{X}_n) \xrightarrow{p} g(\mathbf{X})$ .

Аналогічний висновок справедливий для збіжності майже напевно.

Приймімо цю теорему без доведення.

Теорема 13.2.12, зокрема, дає змогу додавати, множити, ділити випадкові величини тощо, дістаючи їхні границі, застосовуючи аналогічну функцію. Наприклад, якщо  $X_n \xrightarrow{p} X$ , а  $Y_n \xrightarrow{p} Y$ , то  $X_n \cdot Y_n \xrightarrow{p} X \cdot Y$ , а, скажімо,  $\frac{X_n}{X_n + Y_n} \xrightarrow{p} \frac{X}{X + Y}$ .

### 13.2.3. Збіжність у середньому

Однією з мотивацій розгляду збіжності майже напевно було спостереження, що інтегралі двох функцій, які відрізняються тільки на деякій множині міри нуль, однакові. Відповідно, можемо розглянути вид збіжності випадкових величин, визначений через їх інтеграли.

**Визначення 13.2.13.** Послідовність випадкових величин  $X_n$  збігається до випадкової величини  $X$  у середньому порядку  $p$  (збігається за  $L^p$ , converges in  $p$ th mean, converges in  $L^p$ ), що позначають як  $X_n \xrightarrow{L^p} X$ , якщо виконується

$$\lim_{n \rightarrow \infty} \mathbb{E}[|X_n - X|^p] = 0, \quad (13.2.5)$$

якщо кожна випадкова величина належить простору  $L^p$ :  $\mathbb{E}[|X|^p] < \infty$ ,  $\mathbb{E}[|X_n|^p] < \infty$ ,  $n = 1, 2, \dots$

Якщо замість випадкових величин ми маємо послідовність випадкових векторів  $\mathbf{X}_1, \mathbf{X}_2, \dots$ , то  $\mathbf{X}_n \xrightarrow{L^p} \mathbf{X}$ , якщо виконується покоординатна збіжність.  $\square$

Особливої ваги має випадок  $p = 2$ . Тоді кажуть, що послідовність випадкових величин збігається в середньому квадратичному, відтак можна стверджувати, що випадкові величини «блізькі одна до одної», якщо квадрат їхньої різниці, у середньому, малий. Якщо ж  $p = 1$ , просто кажуть, що послідовність збігається в середньому.

**Зауваження 13.2.14.** Раніше ми зазначали (Твердження 11.4.3), що якщо існує момент деякого одного порядку, то існують і моменти усіх менших порядків. З аналогічних міркувань випливає, що якщо послідовність збігається в середньому порядку  $p$ , то вона збігається і в середньому порядку  $q < p$ .  $\square$

## 13.3. Зв'язок між видами збіжності випадкових величин

Зв'язок між різними видами збіжності схематично зображене на Рис. 13.3.1. Збіжність  $X_n \xrightarrow{d} X$  ми вивчатимемо в наступних лекціях.

### 13.3.1. Збіжність майже напевно і збіжність за ймовірністю

Збіжність майже напевно є сильнішою від збіжності за ймовірністю.

**Твердження 13.3.1.** Якщо  $X_n \xrightarrow{\text{м.н.}} X$ , то  $X_n \xrightarrow{p} X$ .

Якщо  $X_n \xrightarrow{p} X$ , то існує підпослідовність  $n_k$  така, що  $X_{n_k} \xrightarrow{\text{м.н.}} X$  за  $k \rightarrow \infty$ .

*Доведення.* Доведімо першу частину твердження. Розгляньмо множину  $A$  всіх точок, у яких виконується поточкова збіжність:

$$A = \left\{ \omega \in \Omega : \lim_{n \rightarrow \infty} |X_n(\omega) - X(\omega)| = 0 \right\}.$$

Розгляньмо для деякого фіксованого  $\varepsilon > 0$  таку множину:

$$N_\varepsilon(n) = \{ \omega \in \Omega : |X_n(\omega) - X(\omega)| \geq \varepsilon \}.$$

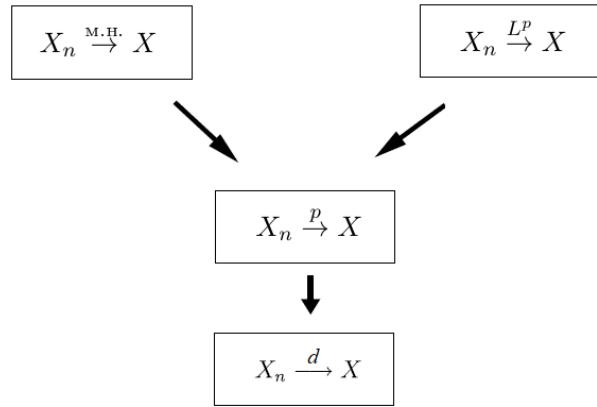


Рис. 13.3.1.: Взаємозв'язок між різними видами збіжності: стрілками показано, які види випливають із яких у найбільш загальному випадку

Нехай

$$M_k = \left( \bigcup_{n \geq k} N_\varepsilon(n) \right)^c = \bigcap_{n \geq k} N_\varepsilon^c(n) .$$

Можна помітити, що  $M_k \subseteq M_{k+1}$ , оскільки

$$M_k = N_\varepsilon^c(n) \cap M_{k+1} .$$

Нехай  $\omega \in M_k$ . За побудовою, для всіх  $n \geq k$  точка  $\omega \in N_\varepsilon^c(n)$ , тобто  $|X_n(\omega) - X(\omega)| < \varepsilon$ . Ще ніщо інше, як умова збіжності  $X_n(\omega)$  до  $X(\omega)$  у точці  $\omega$ .

Звідси випливає, що  $M_k \subseteq A$  для всіх  $k$ , тобто

$$\bigcup_{k=1}^{\infty} M_k \subseteq A .$$

З іншого боку, можна взяти деяку точку  $\omega \in A$  та показати, що існує  $k_0$  таке, що  $\omega \in M_{k_0}$ . Звідси випливає

$$A \subseteq \bigcup_{k=1}^{\infty} M_k .$$

Остаточно маємо

$$\mathbb{P} \left( \bigcup_{k=1}^{\infty} M_k \right) = \mathbb{P}(A) = 1 .$$

Оскільки  $M_k$  утворюють неспадну послідовність множин, за неперервністю міри маємо

$$1 = \mathbb{P} \left( \bigcup_{k=1}^{\infty} M_k \right) = \lim_{k \rightarrow \infty} \mathbb{P}(M_k) = \lim_{k \rightarrow \infty} \mathbb{P} \left( \left( \bigcup_{n \geq k} N_\varepsilon(n) \right)^c \right) ,$$

звідки

$$\lim_{k \rightarrow \infty} \mathbb{P} \left( \bigcup_{n \geq k} N_\varepsilon(n) \right) = 0 ,$$

звідки, у свою чергу,

$$\lim_{k \rightarrow \infty} \mathbb{P}(N_\varepsilon(k)) = 0 ,$$

що є нічим іншим, як збіжністю за ймовірністю.

Доведімо тепер другу частину твердження. Нехай виконується збіжність за ймовірністю. Розгляньмо  $\varepsilon_k > 0$ . Тоді можна підібрати<sup>2</sup>  $n_k$  таке, що

$$\forall n \geq n_k \quad \mathbb{P}(A_k) \equiv \mathbb{P}(\{\omega \in \Omega : |X_n(\omega) - X(\omega)| > \varepsilon_k\}) < \frac{1}{2^k} .$$

Отже, ми показали, що

$$\sum_{k=1}^{\infty} \mathbb{P}(A_k) < \sum_{k=1}^{\infty} \frac{1}{2^k} < \infty .$$

тобто виконуються умови першої леми Бореля-Кантеллі 13.1.5, а відтак

$$\mathbb{P}(\{A_k \text{ н.ч.}\}) = 0 ,$$

що за Твердженням 13.2.6 має наслідком збіжність майже напевно для відповідної підпослідовності з індексами  $k$ .  $\square$

У загальному випадку збіжність за ймовірністю не має наслідком збіжності майже напевно.

**Приклад 13.3.2.** Розгляньмо ймовірнісний простір  $((0; 1], \mathcal{B}(0; 1], \lambda)$ . Нехай маємо *незалежні* величини Бернуллі  $X_n \sim \text{Bern}\left(\frac{1}{n}\right)$ . Наприклад (Рис. 13.3.2),

$$X_1 = \mathbb{1} \left\{ \left(0; \frac{1}{2}\right] \right\} , \quad X_2 = \mathbb{1} \left\{ \left(\frac{3}{8}; \frac{5}{8}\right] \right\} .$$

Тоді

$$\mathbb{P}_{X_1, X_2}((X_1 = 1) \cap (X_2 = 1)) = \frac{1}{8} = \frac{1}{2} \cdot \frac{1}{4} = \mathbb{P}_{X_1}(X_1 = 1) \mathbb{P}_{X_2}(X_2 = 1) .$$

У схожий спосіб можна продовжити будувати цю послідовність, щоб усі величини були незалежні.

Нескладно бачити, що має місце збіжність за ймовірністю до 0. Нехай  $1 > \varepsilon > 0$ , тоді

$$\lim_{n \rightarrow \infty} \mathbb{P}(\{\omega \in \Omega : |X_n(\omega) - X(\omega)| > \varepsilon\}) = \lim_{n \rightarrow \infty} \mathbb{P}(X_n > \varepsilon) = \lim_{n \rightarrow \infty} \mathbb{P}(X_n = 1) = \lim_{n \rightarrow \infty} \frac{1}{n} = 0 .$$

Для  $\varepsilon \geq 1$  також маємо

$$\lim_{n \rightarrow \infty} \mathbb{P}(X_n > \varepsilon) = 0 .$$

У той же час, із другої леми Бореля-Кантеллі 13.1.8 випливає  $\mathbb{P}(\{X_n = 1 \text{ н.ч.}\}) = 1$ , оскільки ряд імовірностей  $\mathbb{P}(X_n = 1)$  розбіжний як гармонійний ряд. Отже

$$\mathbb{P} \left( \left\{ \omega \in \Omega : \lim_{n \rightarrow \infty} X_n(\omega) = 0 \right\} \right) = 0 ,$$

<sup>2</sup>Можна, але ми на цьому зупиняємося не будемо.

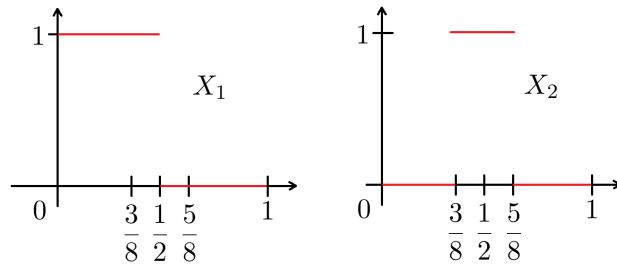


Рис. 13.3.2.: Ілюстрація до Прикладу 13.3.2

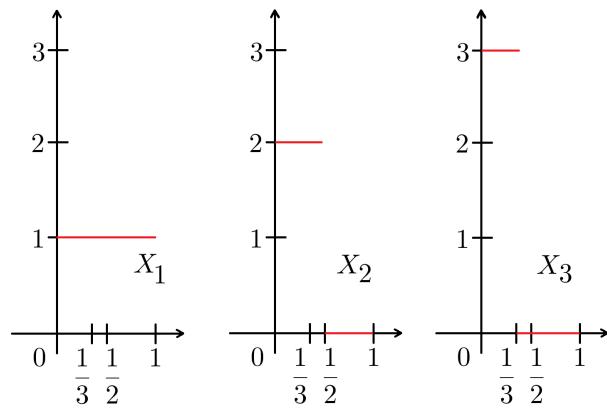


Рис. 13.3.3.: Ілюстрація до Прикладу 13.3.4

тобто збіжність майже напевно не виконується.  $\square$

### 13.3.2. Збіжність у середньому і збіжність за ймовірністю

Також можна показати, що збіжність за ймовірністю випливає зі збіжності в середньому.

**Твердження 13.3.3.** Якщо  $X_n \xrightarrow{L^p} X$ , то  $X_n \xrightarrow{p} X$ .

**Доведення.** Нехай  $g(x) = |x|^p$ , а  $Y_n = |X_n - X|$ . Тоді для деякого  $\varepsilon > 0$  можна застосувати нерівність Маркова (11.1.5):

$$\mathbb{P}_X(Y_n \geq \varepsilon) \leq \frac{\mathbb{E}[|Y_n|^p]}{\varepsilon^p}.$$

Очевидно, якщо виконується збіжність у середньому, тобто якщо  $\lim_{n \rightarrow \infty} \mathbb{E}[|Y_n|^p] = 0$ , то виконується і збіжність за ймовірністю.  $\square$

Проте обернене твердження в загальному випадку не виконується.

**Приклад 13.3.4.** Розгляньмо ймовірнісний простір  $((0; 1], \mathcal{B}(0; 1], \lambda)$ . Нехай  $X_n(\omega) = n \cdot \mathbb{1} \{[0; \frac{1}{n}] \}$  (Рис. 13.3.3).

Можемо показати, що має місце збіжність за ймовірністю до  $0^3$ , якщо помітити, що  $\frac{X_n}{n} \sim \text{Bern}\left(\frac{1}{n}\right)$ , тобто  $X_n$  є дискретною величиною з носієм  $\text{supp}(X_n) = \{0, n\}$ . Тоді доведення повторює доведення з Прикладу 13.3.2.

З іншого боку,

$$\mathbb{E}[|X_n - 0|^p] = \mathbb{E}[|X_n|^p] = 0^p \cdot \mathbb{P}_{X_n}(X_n = 0) + n^p \cdot \mathbb{P}_{X_n}(X_n = n) = \frac{n^p}{n} ,$$

що зовсім не прямує нуля для будь-якого  $p \geq 1$ .  $\square$

**Зауваження 13.3.5.** Оскільки зі збіжності в середньому випливає збіжність за ймовірністю, маємо, що якщо  $\mathbb{E}[X_n] = 0$ , а  $\text{Var}(X_n) = \mathbb{E}[X_n^2] \rightarrow 0$ , то  $X_n \xrightarrow{p} 0$ .

Проте оскільки зі збіжності за ймовірністю не випливає збіжність у середньому, навіть якщо  $X_n \xrightarrow{p} c$ , де  $c \in \mathbb{R}$  — деяка константа, не можна казати, що  $\text{Var}(X_n) \rightarrow 0$ . Скажімо, у Прикладі 13.3.4  $X_n \xrightarrow{p} 0$ , проте  $\mathbb{E}[X_n] = 1$ ,  $\mathbb{E}[X_n^2] = n$ , а відтак  $\text{Var}(X_n) = n - 1$ , що не прямує до 0.  $\square$

Щоправда, за певних додаткових умов можна показати, що збіжність у середньому випливає зі збіжності за ймовірністю.

**Твердження 13.3.6.** Нехай послідовність  $X_n \xrightarrow{p} X$  і нехай існує випадкова величина  $Y \in L^p$  така, що  $|X_n| < Y$  майже напевно для всіх  $n$ . Тоді  $X_n \xrightarrow{L^p} X$ .

Доведення цього твердження наведено в Розд. 13.4.

### 13.3.3. Збіжність у середньому і збіжність майже напевно

Нарешті, між збіжністю майже напевно та збіжністю в середньому прямого зв'язку немає. Точніше, якщо виконуються умови теореми про монотонну збіжність 11.1.2 та теореми про мажоровану збіжність 11.1.3, то збіжність майже напевно матиме наслідком збіжність у середньому для  $p = 1$ . Проте в загальному випадку цей результат не випливає.

**Приклад 13.3.7.** Розгляньмо ймовірнісний простір  $((0; 1], \mathcal{B}(0; 1], \lambda)$ . Нехай  $X_n(\omega) = n \cdot \mathbb{1}\left\{(\frac{1}{n}; \frac{2}{n})\right\}$ ,  $n \geq 2$  (Рис. 13.3.4).

Послідовність таких величин збігається до 0 майже напевно. Справді, нехай  $\varepsilon > 0$ . Тоді для будь-якого  $\omega \in (0; 1]$  знайдеться  $n_\varepsilon(\omega)$  таке, що  $\omega \notin (\frac{1}{n}; \frac{2}{n})$ , тобто для всіх  $n \geq n_\varepsilon(\omega)$ , послідовність  $X_n(\omega) = 0$ . У цьому випадку маємо збіжність не просто майже напевно, а навіть поточкову.

У той же час

$$\mathbb{E}[X_n] = 0 \cdot \mathbb{P}_{X_n}(X_n = 0) + n \cdot \mathbb{P}_{X_n}(X_n = n) = n \cdot \lambda\left(\frac{2}{n} - \frac{1}{n}\right) = n \cdot \frac{1}{n} = 1 ,$$

що зовсім не прямує до 0. Відтак послідовність не збігається до 0 в середньому для  $p \geq 1$ .  $\square$

<sup>3</sup>Понад те, така збіжність буде навіть майже напевно, щоправда, не поточковою: у точці 0 збіжність не виконуватиметься.

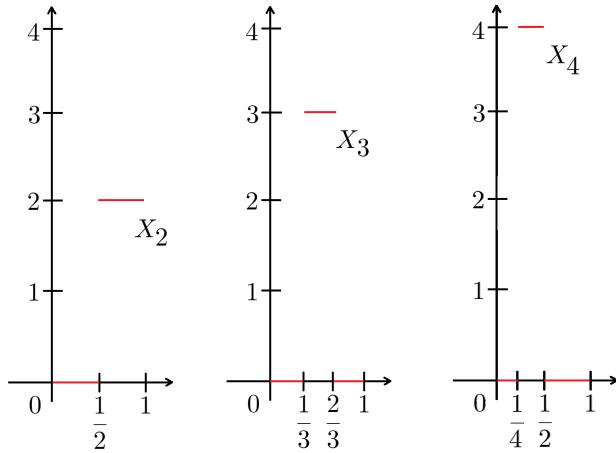


Рис. 13.3.4.: Ілюстрація до Прикладу 13.3.7

**Приклад 13.3.8.** Розгляньмо ймовірнісний простір  $((0; 1], \mathcal{B}(0; 1], \lambda)$ . Нехай

$$\begin{aligned} X_1(\omega) &= \omega + \mathbb{1}\{\omega \in (0; 1)\} , \\ X_{2,1}(\omega) &= \omega + \mathbb{1}\left\{\omega \in \left(0; \frac{1}{2}\right)\right\} , \quad X_{2,2}(\omega) = \omega + \mathbb{1}\left\{\omega \in \left(\frac{1}{2}; 1\right)\right\} , \\ X_{3,1}(\omega) &= \omega + \mathbb{1}\left\{\omega \in \left(0; \frac{1}{3}\right)\right\} , \quad X_{3,2}(\omega) = \omega + \mathbb{1}\left\{\omega \in \left(\frac{1}{3}; \frac{2}{3}\right)\right\} , \quad X_{3,3}(\omega) = \omega + \mathbb{1}\left\{\omega \in \left(\frac{2}{3}; 1\right)\right\} , \end{aligned}$$

і т.д. (Рис. 13.3.5).

Можна показати, що  $X_n$  збігається до  $X(\omega) = \omega$  і за ймовірністю, і в середньому, але не майже напевно. Справді, для деякого  $\varepsilon > 0$  маємо

$$\mathbb{P}(\{\omega : |X_{n,k}(\omega) - X(\omega)| > \varepsilon\}) = \mathbb{P}\left(\left\{\omega : \mathbb{1}\left\{\omega \in \left(\frac{k-1}{n}; \frac{k}{n}\right)\right\} > \varepsilon\right\}\right) = \frac{1}{n} .$$

Очевидно, що ця ймовірність прямує до нуля, а тому  $X_n \xrightarrow{p} X$ .

Також можемо обчислити відповідне сподівання:

$$\mathbb{E}[|X_{n,k} - X|^p] = \mathbb{E}\left[\left(\mathbb{1}\left\{\omega \in \left(\frac{k-1}{n}; \frac{k}{n}\right)\right\}\right)^p\right] = \mathbb{E}\left[\mathbb{1}\left\{\omega \in \left(\frac{k-1}{n}; \frac{k}{n}\right)\right\}\right] = \frac{1}{n} .$$

Очевидно, що це сподівання прямує до нуля, а тому  $X_n \xrightarrow{L^p} X$ .

Проте, для будь-якого  $\omega$  і будь-якого  $n$  знайдеться таке  $m > n$  і таке  $k$ , що  $X_{m,k}(\omega) - X(\omega) = 1$ . Тобто для кожного  $\omega$  послідовність  $X_{m,k}(\omega) - X(\omega)$  буде коливатися між 0 і 1 і не буде прямувати до нуля. Відтак збіжність майже напевно не виконується.  $\square$

## 13.4. Доведення окремих тверджень

Доведімо Твердження 13.1.3.

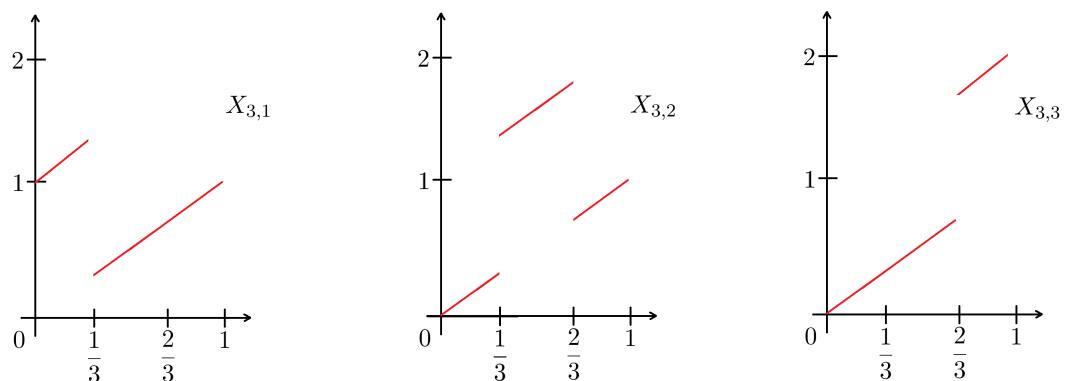
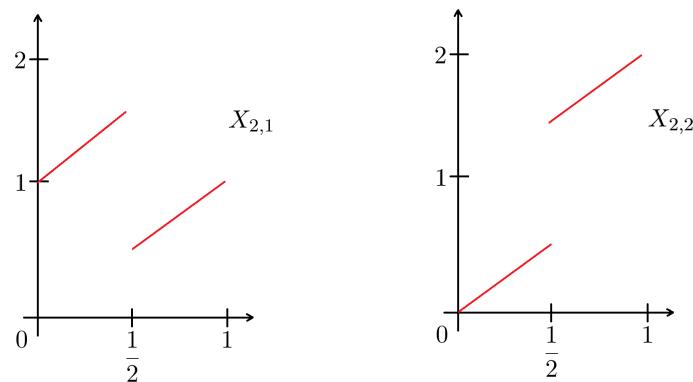
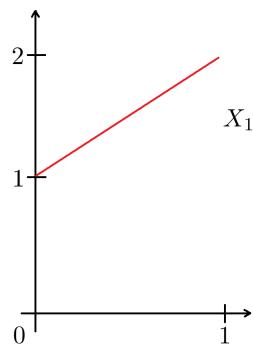


Рис. 13.3.5.: Ілюстрація до Прикладу 13.3.7

**Доведення.** Нехай  $B_n = \bigcap_{k=n}^{\infty} A_k$ , а  $C_n = \bigcup_{k=n}^{\infty} A_k$ . Очевидно, що  $B_n \rightarrow \liminf_n A_n$ , оскільки  $B_n$  утворюють неспадну послідовність множин. Аналогічно  $C_n \rightarrow \limsup_n A_n$ , оскільки  $C_n$  утворюють незростаючу послідовність множин. Відтак, за властивістю неперервності міри з Теореми 2.2.8,

$$\mathbb{P} \left( \liminf_n A_n \right) = \mathbb{P} \left( \bigcup_{n=1}^{\infty} \bigcap_{k=n}^{\infty} A_k \right) = \lim_{n \rightarrow \infty} \mathbb{P} \left( \bigcap_{k=n}^{\infty} A_k \right) = \liminf_{n \rightarrow \infty} \mathbb{P} \left( \bigcap_{k=n}^{\infty} A_k \right) \leq \liminf_{n \rightarrow \infty} \mathbb{P} (A_n) ,$$

де ми використали той факт, що якщо границя існує, то вона дорівнює нижній границі. Останню нерівність можна довести з властивості монотонності міри. Нарешті, для (обмежених) числових послідовностей відомо, що  $\liminf_n \mathbb{P} (A_n) \leq \limsup_n \mathbb{P} (A_n)$ .

Якщо  $A_n \rightarrow A$ , то  $\liminf_n A_n = \limsup_n A_n$ , звідки безпосередньо випливає  $\mathbb{P} (A_n) \rightarrow \mathbb{P} (A)$ , оскільки в цьому випадку  $\liminf_n \mathbb{P} (A_n) = \limsup_n \mathbb{P} (A_n)$ .  $\square$

Доведімо Твердження 13.2.6.

**Доведення.** Використаймо твердження про числові послідовності, згідно з яким  $x_n \rightarrow x$  тоді й тільки тоді, коли для будь-якого  $\varepsilon > 0$  множина  $\{n \in \mathbb{N} : |x_n - x| \geq \varepsilon\}$  є скінченою.

Тоді множина всіх точок  $\omega \in \Omega$ , у яких  $X_n(\omega) \rightarrow X(\omega)$  поточково, є множиною точок  $\omega \in \Omega$ , у яких виконується твердження, що для будь-якого  $\varepsilon > 0$  множина  $\{n \in \mathbb{N} : |X_n(\omega) - X(\omega)| \geq \varepsilon\}$  є скінченою. Іншими словами, це така множина точок, у яких виконується обернене твердження, а саме що *знаходитьсья* таке  $\varepsilon > 0$ , що множина  $\{n \in \mathbb{N} : |X_n(\omega) - X(\omega)| < \varepsilon\}$  є *некінченою*. Це ніщо інше, як визначення верхньої границі послідовності подій:

$$\mathbb{P} (\{\omega \in \Omega : X_n(\omega) \rightarrow X(\omega)\}) = 1 - \mathbb{P} (\{\omega \in \Omega : \exists \varepsilon > 0 \quad |X_n(\omega) - X(\omega)| \geq \varepsilon \text{ н.ч.}\}) .$$

Якщо  $\varepsilon$  раціональне, то можемо використати нерівність Була (2.2.2):

$$\mathbb{P} (\{\omega \in \Omega : \exists \varepsilon \in \mathbb{Q}^+ \quad |X_n(\omega) - X(\omega)| \geq \varepsilon \text{ н.ч.}\}) \leq \sum_{\varepsilon \in \mathbb{Q}^+} \mathbb{P} (\{\omega \in \Omega : |X_n(\omega) - X(\omega)| \geq \varepsilon \text{ н.ч.}\}) = 0$$

за умовами теореми.

Якщо ж  $\varepsilon > 0$  не є раціональним, то існує раціональне  $\varepsilon' > 0$  таке, що  $\varepsilon' < \varepsilon$ . Відтак

$$\{\omega \in \Omega : |X_n(\omega) - X(\omega)| \geq \varepsilon \text{ н.ч.}\} \subseteq \{\omega \in \Omega : |X_n(\omega) - X(\omega)| \geq \varepsilon' \text{ н.ч.}\} ,$$

звідки за монотонністю міри випливає аналогічне співвідношення ймовірностей:

$$\mathbb{P} (\{\omega \in \Omega : \exists \varepsilon > 0 \quad |X_n(\omega) - X(\omega)| \geq \varepsilon \text{ н.ч.}\}) \leq \mathbb{P} (\{\omega \in \Omega : \exists \varepsilon' \in \mathbb{Q}^+ \quad |X_n(\omega) - X(\omega)| \geq \varepsilon' \text{ н.ч.}\}) = 0 .$$

$\square$

Доведімо Твердження 13.3.6.

**Доведення.** Оскільки  $X_n \xrightarrow{p} X$ , згідно з Твердженням 13.3.1 існує така підпослідовність  $X_{n_k}$ , яка збігається до  $X$  майже напевно. А оскільки  $|X_n| < Y$ , за властивістю монотонності інтегралу маємо

$$\sup_n \mathbb{E} [|X_n|] \leq \mathbb{E} [Y] ,$$

звідки випливає  $X \in L^1$ , а  $|X| \leq Y$ .

Нехай  $\varepsilon > 0$ . Тоді

$$\begin{aligned}\mathbb{E}[|X_n - X|] &= \mathbb{E}[|X_n - X| \cdot \mathbb{1}\{|X_n - X| \leq \varepsilon\}] + \mathbb{E}[|X_n - X| \cdot \mathbb{1}\{|X_n - X| > \varepsilon\}] \\ &= \int_{\{\omega:|X_n(\omega)-X(\omega)|\leq\varepsilon\}} |X_n - X| d\mathbb{P} + \int_{\{\omega:|X_n(\omega)-X(\omega)|>\varepsilon\}} |X_n - X| d\mathbb{P} \\ &\leq \varepsilon + \int_{\{\omega:|X_n(\omega)-X(\omega)|>\varepsilon\}} |X_n - X| d\mathbb{P} \\ &\leq \varepsilon + \int_{\{\omega:|X_n(\omega)-X(\omega)|>\varepsilon\}} 2Y d\mathbb{P},\end{aligned}$$

де остання нерівність випливає з нерівності трикутника  $\sup_n |X_n - X| \leq 2Y$ .

Нехай  $c > 0$ . Тоді

$$\begin{aligned}\mathbb{E}[|X_n - X|] &\leq \varepsilon + \int_{\{\omega:(|X_n(\omega)-X(\omega)|>\varepsilon)\cap(|Y(\omega)|\leq c)\}} 2Y d\mathbb{P} + \int_{\{\omega:(|X_n(\omega)-X(\omega)|>\varepsilon)\cap(|Y(\omega)|>c)\}} 2Y d\mathbb{P} \\ &\leq \varepsilon + 2c \cdot \mathbb{P}(|X_n - X| > \varepsilon) + 2 \int_{\{\omega:|Y(\omega)|\geq c\}} Y d\mathbb{P},\end{aligned}$$

оскільки в першому випадку  $|Y| \leq c$ , а відтак залишається інтеграл від індикаторної функції за множиною  $|X_n - X| > \varepsilon$ , а в другому випадку ми перейшли за монотонністю до інтегралу за меншою множиною.

Можна помітити, що послідовність  $|Y| \cdot \mathbb{1}\{|Y| \geq c\}$  прямує до 0, коли  $c \rightarrow \infty$ . Також можна помітити, що

$$\mathbb{E}[|Y| \cdot \mathbb{1}\{|Y| \geq c\}] \leq \mathbb{E}[Y] < \infty.$$

Можна застосувати Теорему про мажоровану збіжність 11.1.3 та дістати  $\mathbb{E}[|Y| \cdot \mathbb{1}\{|Y| \geq c\}] \rightarrow 0$ , тобто серед іншого

$$\int_{\{\omega:|Y(\omega)|\geq c\}} Y d\mathbb{P} \leq \varepsilon$$

для деякого достатньо великого  $c$ .

Що стосується  $2c \cdot \mathbb{P}(|X_n - X| > \varepsilon)$ , то ця величина також прямує до 0, оскільки виконується збіжність за ймовірністю.

Відтак справа від знаку нерівності всі три члени прямують до 0, коли  $\varepsilon \rightarrow 0$ , а відтак  $\mathbb{E}[|X_n - X|] \rightarrow 0$ .  $\square$

## 14. Закон великих чисел

Середнє вибіркове значення, яке ми розглянули в Прикладі 9.4.16, є чи не найважливішою чи-слововою характеристикою деякої вибірки даних. Середнє вибіркове значення подібне до сподівання випадкової величини в тому розумінні, що як перше, так і друге дають уявлення про певне «найбільш репрезентативне» значення вибірки та випадкової величини відповідно. Виявляється, для цього спостереження є фундаментальне теоретичне підґрунтя, яке є предметом поточної лекції.

### 14.1. Слабкий закон великих чисел

Можна сформулювати різні версії слабкого закону великих чисел (ЗВЧ), висуваючи різні вимоги до послідовності випадкових величин. Доведімо спочатку просту версію ЗВЧ.

**Теорема 14.1.1** ((Слабкий) Закон великих чисел<sup>1</sup> ((Weak) Law of large numbers)). Нехай маємо послідовність незалежних випадкових величин  $X_1, X_2, \dots$  таких, що їхні сподівання існують, скінченні й дорівнюють  $\mathbb{E}[X_i] = \mu_i$ ,  $i \geq 1$ , а дисперсії також існують, скінченні й рівномірно обмежені, тобто  $\text{Var}(X_i) \leq C < \infty$ ,  $i \geq 1$ . Тоді

$$\frac{1}{n} \sum_{i=1}^n (X_i - \mathbb{E}[X_i]) \xrightarrow{p} 0. \quad (14.1.1)$$

*Доведення.* Позначмо кожну часткову суму через  $S_n = X_1 + \dots + X_n$ . Тоді  $\mathbb{E}[S_n] = \mathbb{E}[X_1] + \dots + \mathbb{E}[X_n]$  за лінійністю сподівання, а з урахуванням незалежності маємо

$$\text{Var}(S_n) = \text{Var}\left(\sum_{i=1}^n X_i\right) = \sum_{i=1}^n \text{Var}(X_i) \leq nC.$$

Із нерівності Чебишова (11.1.7) випливає, що для будь-якого  $c > 0$

$$\mathbb{P}_X\left(\left|\frac{1}{n}S_n - \frac{1}{n}\mathbb{E}[S_n]\right| \geq c\right) \leq \frac{\text{Var}\left(\frac{1}{n}S_n\right)}{c^2} \leq \frac{nC}{n^2c^2}.$$

Вочевидь, коли  $n \rightarrow \infty$ , ця ймовірність прямує до 0. Відтак, за визначенням збіжності за ймовірністю,  $\frac{1}{n}(S_n - \mathbb{E}[S_n]) \xrightarrow{p} 0$ , а відтак виконується (14.1.1).  $\square$

**Зауваження 14.1.2.** Як можна бачити з доведення Теореми 14.1.1, випадкові величини насправді необов'язково повинні бути незалежні. Достатньо вимагати їхньої некорельованості.

Також нескладно бачити, що за умов Теореми 14.1.1 можна довести не тільки збіжність за ймовірністю, а й збіжність у середньоквадратичному. Справді, оскільки

$$\mathbb{E}\left[\left(\frac{1}{n}(S_n - \mathbb{E}[S_n])\right)^2\right] = \frac{1}{n^2} \text{Var}(S_n) \leq \frac{C}{n} \rightarrow 0,$$

<sup>1</sup>ЗВЧ у такому формулуванні також називають теоремою Чебишова.

маємо збіжність у середньоквадратичному.

Нарешті, можна завжди розглянути *ценетровані* величини  $X_i^* = X_i - \mathbb{E}[X_i]$ ,  $i \geq 1$ , такі, що  $\mathbb{E}[X_i^*] = 0$ , і розглянути ЗВЧ, у якому  $\frac{S_n^*}{n} \xrightarrow{p} 0$ . Тому часто ЗВЧ формулюють для випадкових величини, які мають однакове сподівання.  $\square$

**Приклад 14.1.3.** Найпростішим застосуванням ЗВЧ є той факт, що для схеми Бернуллі, де  $X_n$  = «індикатор того, що випробування  $n$  завершилося успіхом»,  $S_n$  = «число успіхів із-посеред  $n$  випробувань» і ймовірність успіху дорівнює  $p$ , маємо

$$\frac{1}{n} (S_n - \mathbb{E}[S_n]) = \frac{1}{n} (S_n - np) = \frac{1}{n} S_n - p \xrightarrow{p} 0 ,$$

тобто зі збільшенням числа випробувань відносна частка успіхів прямує (за ймовірністю) до ймовірності успіху.

Справді, виконуються всі умови Теореми 14.1.1, адже  $\frac{1}{n} S_n = \frac{1}{n} \sum_{i=1}^n X_i$ ,  $\mathbb{E}[X_i] = p < \infty$ ,  $\text{Var}(X_i) = p(1-p) \leq \frac{1}{4} < \infty$ , і всі  $X_i$  незалежні між собою,  $i \geq 1$ .  $\square$

## 14.2. Посилений закон великих чисел

Розгляньмо тепер *посилений* закон великих чисел (ПЗВЧ). Із-посеред різних версій цього закону ми доведемо тільки одну, яку часто використовують на практиці.

**Теорема 14.2.1** (Посилений закон великих чисел (Strong law of large numbers)). Нехай маємо послідовність незалежних випадкових величин  $X_1, X_2, \dots$  таких, що вони мають *однаковий розподіл* зі скінченим сподіванням  $\mathbb{E}[X_i] = \mu$ ,  $i \geq 1$ . Тоді

$$\frac{1}{n} \sum_{i=1}^n (X_i - \mu) \xrightarrow{\text{М.Н.}} 0 . \quad (14.2.1)$$

*Доведення.* Для доведення цієї теореми ми скористуємося підходом, який полягає в «обрізанні» випадкової величини у такий спосіб, що «обрізані» величини матимуть моменти вищих порядків.

Для початку можемо показати, що достатньо розглянути тільки невід'ємні випадкові величини. Справді, позначмо кожну часткову суму через  $S_n = X_1 + \dots + X_n$ . Тоді  $\mathbb{E}[S_n] = \mathbb{E}[X_1] + \dots + \mathbb{E}[X_n] = n\mu$  за лінійністю сподівання. Якби ми показали виконання ПЗВЧ для невід'ємних випадкових величин, то ми б могли миттєво зробити висновок, що з  $\frac{1}{n} S_n^+ - \mu^+ \xrightarrow{\text{М.Н.}} 0$  і  $\frac{1}{n} S_n^- - \mu^- \xrightarrow{\text{М.Н.}} 0$  випливає  $\frac{1}{n} S_n - \mu \xrightarrow{\text{М.Н.}} 0$ .

Тому нехай  $X_i \geq 0$ . Тоді розгляньмо їхні обрізані версії  $Y_i = X_i \cdot \mathbb{1}\{X_i \leq i\}$ . Оскільки  $0 \leq Y_i \leq i$ , для відповідних моментів за монотонністю сподівання маємо таку оцінку зверху:

$$\mathbb{E}[Y^k] = \int Y^k d\mathbb{P} \leq \int i^k d\mathbb{P} = i^k \int 1 d\mathbb{P} = i^k < \infty .$$

Оскільки всі  $X_i$  незалежні між собою,  $i \geq 1$ , то і  $Y_i$  незалежні між собою за Теоремою 9.4.9, адже вони є функціями від  $X_i$ . Понад те, за Теоремою 11.1.2 маємо

$$\mathbb{E}[Y_i] = \mathbb{E}[X_i \cdot \mathbb{1}\{X_i \leq i\}] = \mathbb{E}[X_1 \cdot \mathbb{1}\{X_1 \leq i\}] \xrightarrow{i \rightarrow \infty} \mathbb{E}[X_1] = \mu ,$$

оскільки всі  $X_i$  мають одинаковий розподіл,  $i \geq 1$ ,  $X_1 \cdot \mathbb{1}\{X_1 \leq i\}$  утворюють неспадну послідовність випадкових величин, і  $X_1 \cdot \mathbb{1}\{X_1 \leq i\} \rightarrow X_1$ .

Розгляньмо також часткові суми  $S_n^* = Y_1 + \dots + Y_n$ . Обчислімо їхні дисперсії з урахуванням незалежності та однаковості розподілів:

$$\begin{aligned} \text{Var}(S_n^*) &= \text{Var}\left(\sum_{i=1}^n Y_i\right) = \sum_{i=1}^n \text{Var}(Y_i) \\ &\leq \sum_{i=1}^n \mathbb{E}[Y_i^2] \\ &= \sum_{i=1}^n \mathbb{E}[X_i^2 \cdot \mathbb{1}\{X_i \leq i\}] = \sum_{i=1}^n \mathbb{E}[X_1^2 \cdot \mathbb{1}\{X_1 \leq i\}] \\ &\leq n \mathbb{E}[X_1^2 \cdot \mathbb{1}\{X_1 \leq n\}] \leq n^3 < \infty. \end{aligned}$$

Тобто відповідні дисперсії існують і скінченні.

Розгляньмо тепер  $\alpha > 1$  і позначмо цілу частину її ступеня через  $u_n = \lfloor \alpha^n \rfloor$ ,  $n \in \mathbb{N}$ . Вочевидь,  $u_n \leq \alpha^n$ . Понад те, оскільки  $\alpha^n > 1$ , маємо просту нерівність

$$u_n \geq \frac{\alpha^n}{2} \Rightarrow \frac{1}{u_n} \leq \frac{2}{\alpha^n}.$$

Відтак для будь-якого  $x > 0$  маємо, за формулою нескінченно спадної геометричної прогресії:

$$\begin{aligned} \sum_{\substack{n=1 \\ u_n \geq x}}^{\infty} \frac{1}{u_n} &\leq \sum_{\substack{n=1 \\ \alpha^n \geq x}}^{\infty} \frac{1}{u_n} \leq \sum_{\substack{n=1 \\ \alpha^n \geq x}}^{\infty} \frac{2}{\alpha^n} \\ &\leq \sum_{k=\log_{\alpha} x}^{\infty} \frac{2}{\alpha^k} = \frac{2}{\alpha^{\log_{\alpha} x}} \sum_{j=0}^{\infty} \frac{1}{\alpha^j} = \frac{2/x}{1 - 1/\alpha}. \end{aligned}$$

Зафіксуємо деяке  $\varepsilon > 0$ . Тоді за нерівністю Чебишова (11.1.7) маємо такий результат:

$$\sum_{n=1}^{\infty} \mathbb{P}\left(\left|\frac{S_{u_n}^* - \mathbb{E}[S_{u_n}^*]}{u_n}\right| \leq \varepsilon\right) = \sum_{n=1}^{\infty} \mathbb{P}\left(\left|\frac{S_{u_n}^*}{u_n} - \frac{\mathbb{E}[S_{u_n}^*]}{u_n}\right| \geq \varepsilon\right) \leq \sum_{n=1}^{\infty} \frac{\text{Var}(S_{u_n}^*/u_n)}{\varepsilon^2} = \sum_{n=1}^{\infty} \frac{\text{Var}(S_{u_n}^*)}{u_n^2 \varepsilon^2}.$$

Можемо продовжувати оцінювати цей вираз зверху. Дисперсію  $\text{Var}(S_{u_n}^*)$  ми вже оцінили вище:

$$\sum_{n=1}^{\infty} \frac{\text{Var}(S_{u_n}^*)}{u_n^2 \varepsilon^2} \leq \sum_{n=1}^{\infty} \frac{u_n \mathbb{E}[X_1^2 \mathbb{1}\{X_1 \leq u_n\}]}{u_n^2 \varepsilon^2} = \frac{1}{\varepsilon^2} \mathbb{E}\left[X_1^2 \sum_{n=1}^{\infty} \frac{1}{u_n} \cdot \mathbb{1}\{u_n \geq X_1\}\right].$$

Для оцінки зверху цього виразу скористаймося попереднім результатом із сумою геометричної прогресії:

$$\frac{1}{\varepsilon^2} \mathbb{E}\left[X_1^2 \sum_{n=1}^{\infty} \frac{1}{u_n} \cdot \mathbb{1}\{u_n \geq X_1\}\right] \geq \frac{1}{\varepsilon^2} \mathbb{E}\left[X_1^2 \cdot \frac{2/X_1}{1 - 1/\alpha}\right] = \frac{2}{\varepsilon^2 (1 - 1/\alpha)} \mathbb{E}[X_1] = \frac{2\mu}{\varepsilon^2 (1 - 1/\alpha)} < \infty.$$

Отже, що ми показали? Ми показали, що

$$\frac{S_{u_n}^* - \mathbb{E}[S_{u_n}^*]}{u_n} \xrightarrow{\text{м.н.}} 0 ,$$

оскільки виконалися всі умови Твердження 13.2.7. Для завершення доведення нам потрібно замінити  $\mathbb{E}\left[\frac{S_{u_n}^*}{u_n}\right]$  на  $\mu$ ,  $S_{u_n}^*$  — на  $S_{u_n}$ , і, зрештою,  $u_n$  — на  $k$ .

Що стосується заміни сподівань, то ми вище показали, що  $\mathbb{E}[Y_i] \rightarrow \mu$ . Також цілком очевидно, що  $u_n \rightarrow \infty$ . Відтак

$$\mathbb{E}\left[\frac{S_{u_n}^*}{u_n}\right] = \frac{u_n \mathbb{E}[Y_1]}{u_n} \rightarrow \mu .$$

Тому можемо стверджувати, що

$$\frac{S_{u_n}^*}{u_n} - \mu \xrightarrow{\text{м.н.}} 0 .$$

Тепер можемо розглянути  $S_{u_n}^*$  і показати, що

$$\frac{S_{u_n}}{u_n} \xrightarrow{\text{м.н.}} \mu ,$$

тобто що відповідні границі однакові.

Справді, розгляньмо

$$\begin{aligned} \sum_{n=1}^{\infty} \mathbb{P}(X_n \neq Y_n) &= \sum_{n=1}^{\infty} \mathbb{P}(X_n \neq X_n \cdot \mathbb{1}\{X_n \leq n\}) = \sum_{n=1}^{\infty} \mathbb{P}(X_n > n) \\ &\leq \sum_{n=1}^{\infty} \mathbb{P}(X_n \geq n) = \sum_{n=1}^{\infty} \mathbb{P}(X_1 \geq n) , \end{aligned}$$

де в останній рівності ми використали однаковість розподілів  $X_i$ ,  $i \geq 1$ .

Обчислити цей ряд можна, виходячи з таких міркувань:

$$\begin{aligned} \sum_{n=1}^{\infty} \mathbb{P}(X_1 \geq n) &= \sum_{n=1}^{\infty} (\mathbb{P}(n \leq X_1 < n+1) + \mathbb{P}(n+1 \leq X_1 < n+2) + \dots) \\ &= \sum_{k=1}^{\infty} k \cdot \mathbb{P}(k \leq X_1 < k+1) = \sum_{k=1}^{\infty} k \cdot \mathbb{P}(\lfloor X_1 \rfloor = k) = \mathbb{E}[\lfloor X_1 \rfloor] . \end{aligned}$$

А оскільки за монотонністю сподівання  $\mathbb{E}[\lfloor X_1 \rfloor] \leq \mathbb{E}[X_1] = \mu < \infty$ , маємо, що

$$\sum_{n=1}^{\infty} \mathbb{P}(X_n \neq Y_n) < \infty .$$

Із першої леми Бореля-Кантеллі 13.1.5 безпосередньо випливає, що  $\mathbb{P}(\{X_n \neq Y_n \text{ н.ч.}\}) = 0$ , тобто що  $\mathbb{P}(\{X_n = Y_n \text{ рано чи пізно}\}) = 1$ . Відтак для деякого дуже великого  $n$  границі  $S_{u_n}/u_n$  та  $S_{u_n}^*/u_n$  будуть однакові.

Нарешті можемо позбутися  $u_n$ . Для будь-якого натурального  $k$  можемо знайти таке натуральне

число  $n = n_k$ , що  $u_n \leq k < u_{n+1}$ . Тоді

$$\frac{u_n}{u_{n+1}} \cdot \frac{S_{u_n}}{u_n} = \frac{S_{u_n}}{u_{n+1}} \leq \frac{S_k}{k} \leq \frac{S_{u_{n+1}}}{u_n} = \frac{u_{n+1}}{u_n} \cdot \frac{S_{u_{n+1}}}{u_{n+1}}.$$

Спрямувавши  $k \rightarrow \infty$ , маємо  $n = n_k \rightarrow \infty$ , звідки

$$\frac{u_n}{u_{n+1}} \rightarrow \frac{1}{\alpha}, \quad \frac{u_{n+1}}{u_n} \rightarrow \alpha.$$

Раніше ми показали, що  $\frac{S_{u_n}}{u_n} \xrightarrow{\text{М.Н.}} \mu$ , а відтак для будь-якого  $\alpha > 1$  та будь-якого  $\delta > 0$  маємо

$$\frac{\mu}{(1 + \delta)\alpha} \leq \frac{S_k}{k} \leq (1 + \delta)\alpha\mu$$

майже напевно для деякого достатньо великого  $k$ . Тоді для будь-якого  $\varepsilon > 0$  можна підібрати  $\alpha > 1$  і  $\delta > 0$  такі, що

$$\frac{\mu}{(1 + \delta)\alpha} > \mu - \varepsilon, \quad (1 + \delta)\alpha\mu < \mu + \varepsilon,$$

звідки випливає

$$\mu - \varepsilon < \frac{S_k}{k} < \mu + \varepsilon,$$

тобто

$$\mathbb{P} \left( \left\{ \left| \frac{S_k}{k} - \mu \right| \geq \varepsilon \text{ н.ч.} \right\} \right) = 0.$$

Отже за Твердженням 13.2.6 робимо висновок, що

$$\frac{S_k}{k} - \mu \xrightarrow{\text{М.Н.}} 0.$$

□

**Зауваження 14.2.2.** У доведенні Теореми 14.2.1 ми використали незалежність величин  $X_i, i \geq 1$ , у двох місцях. Спочатку ми показали, що незалежними є величини  $Y_i, i \geq 1$ , а потім розписали дисперсію суми таких величин як суму дисперсій. Із цього випливає, що доведення теореми залишиться справедливим, навіть якщо величини  $X_i, i \geq 1$ , будуть усього лише *попарно* незалежними. □

**Зауваження 14.2.3.** Із Твердження 13.3.1 відразу випливає, що за умов Теореми 14.2.1 виконується не тільки ПЗВЧ, а й ЗВЧ. На практиці в більшості випадків збіжності за ймовірністю цілком достатньо, проте приємно усвідомлювати, що збіжність майже напевно для однаково розподілених випадкових величин не вимагає додаткових посиленіх вимог до відповідної послідовності. □

**Приклад 14.2.4.** Як і в випадку ЗВЧ, найпростішим застосуванням ПЗВЧ є той факт, що для схеми Бернуллі, де  $X_n = \text{«індикатор того, що випробування } n \text{ завершилося успіхом»}$ ,  $S_n = \text{«число}$

успіхів із-посеред  $n$  випробувань», і ймовірність успіху дорівнює  $p$ , маємо

$$\frac{1}{n} (S_n - \mathbb{E}[S_n]) = \frac{1}{n} (S_n - np) = \frac{1}{n} S_n - p \xrightarrow{\text{М.Н.}} 0 ,$$

тобто яку б послідовність результатів випробувань Бернуллі ми не взяли, рано чи пізно відносна частка успіхів буде скільки завгодно близька до ймовірності успіху.

Справді, виконуються всі умови Теореми 14.2.1, адже кожне випробування Бернуллі завершується величиною Бернуллі з одним і тим самим параметром  $p$ , яка має скінченне сподівання  $p$ , і всі ці величини незалежні.  $\square$

Існують і декілька інших формулювань ПЗВЧ, які відрізняються умовами, що накладають на послідовність випадкових величин. Ми їх доводити не будемо, але згадаємо в ознайомчих цілях.

**Твердження 14.2.5.** Нехай маємо послідовність незалежних випадкових величин  $X_1, X_2, \dots$  таких, що їхні сподівання існують, скінченні і дорівнюють  $\mathbb{E}[X_i] = \mu_i$ ,  $i \geq 1$ , а четверті центральні моменти також існують, скінченні і рівномірно обмежені, тобто  $\mathbb{E}[(X_i - \mathbb{E}[X_i])^4] \leq C < \infty$ ,  $i \geq 1$ . Тоді

$$\frac{1}{n} \sum_{i=1}^n (X_i - \mathbb{E}[X_i]) \xrightarrow{\text{М.Н.}} 0 .$$

Відмінність такого формулювання полягає в тому, що ми не вимагаємо, щоб випадкові величини з послідовності мали одинаковий розподіл. Проте як компенсацію за це ми повинні перевірити, що існують скінченні четверті центральні моменти (а відтак і дисперсії).

Нарешті, розгляньмо ще одну версію ПЗВЧ, яку інколи називають ПЗВЧ Колмогорова.

**Твердження 14.2.6.** Нехай маємо послідовність незалежних випадкових величин  $X_1, X_2, \dots$  таких, що вони мають скінченні дисперсії, і до того ж

$$\sum_{n=1}^{\infty} \frac{\text{Var}(X_n)}{n^2} < \infty .$$

Тоді

$$\frac{1}{n} \sum_{i=1}^n (X_i - \mathbb{E}[X_i]) \xrightarrow{\text{М.Н.}} 0 .$$

Знову ж таки, позбувшись вимоги щодо однаковості розподілів, ми вимагаємо скінченності дисперсій, до того ж такої, щоб відповідний ряд був збіжний. Зрозуміло, що якщо всі дисперсії одинакові і скінченні, то цей ряд буде збіжний, бо ряд  $\sum_{i=1}^{\infty} 1/n^2$  збіжний.

Існують і версії ПЗВЧ для величин, які не є незалежні, але вони виходять за межі нашого курсу.

### 14.3. Різниця між слабким та посиленім законами великих чисел

Різниця між ЗВЧ та ПЗВЧ полягає у видах збіжності, які застосовуються у відповідних формулах. Різниця у збіжностях майже напевно та за ймовірністю посутня, про що було сказано в Лекції 13, хоча й не завжди очевидна у першому прочитанні.

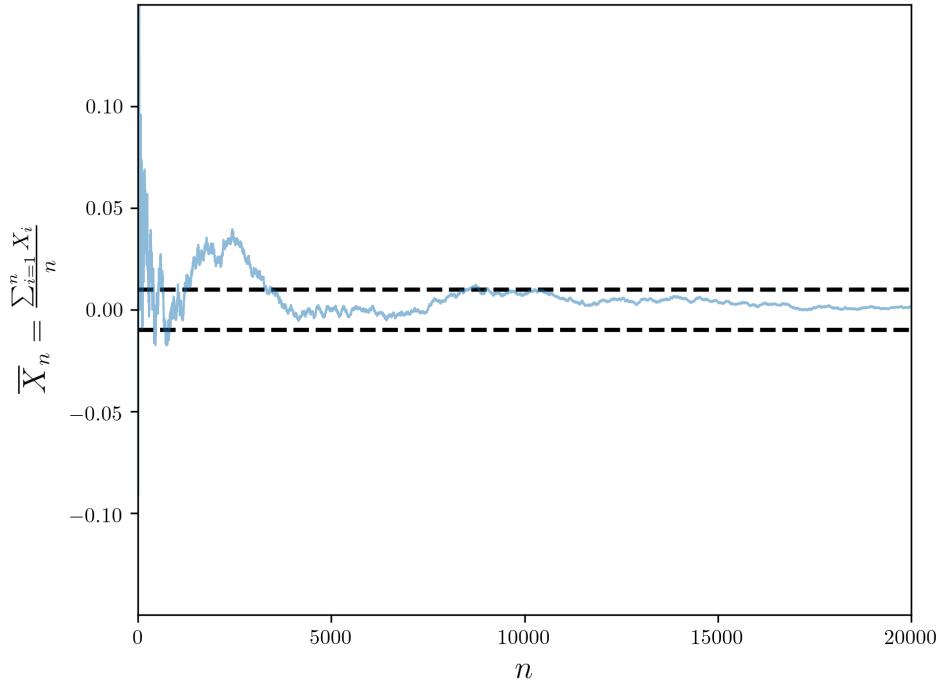


Рис. 14.3.1.: Графік однієї послідовності реалізацій  $\bar{X}_n$ ,  $n = 1, \dots, 20\,000$ , випадкових величин  $\bar{X}_n$ . Штрихованими лініями окреслено окіл  $(-0.01; 0.01)$

Розгляньмо послідовність незалежних випадкових величин  $X_n$ ,  $n \geq 1$ , де

$$p_{X_n}(x) = \begin{cases} 0.5, & x = 1 \\ 0.5, & x = -1 \end{cases}.$$

Нескладно бачити, що  $\mathbb{E}[X_n] = 0$  для всіх  $n$ .

Утворімо середньоаритметичні значень відповідних величин:

$$\bar{X}_n = \frac{1}{n} \sum_{i=1}^n X_i, \quad n = 1, 2, \dots$$

Згідно з ЗВЧ та ПЗВЧ, такі середньоаритметичні прямують до  $\mathbb{E}[X_n] = 0$  як за ймовірністю, так і майже напевно. Розберімося, у чому полягає різниця інтерпретацій цих двох законів.

Той факт, що послідовність  $\bar{X}_n \xrightarrow{\text{М.Н.}} 0$ , означає, що для будь-якого  $\varepsilon > 0$  і для будь-якого вектора  $\omega \in \{-1, 1\}^n$  знайдеться таке  $n$ , після якого  $|\bar{X}_n| < \varepsilon$ . Іншими словами, якщо ми розглянемо деяку послідовність реалізацій  $\bar{X}_n$ ,  $n \geq 1$ , відповідних випадкових величин, то її графік рано чи пізно потрапить в окіл  $(-\varepsilon; \varepsilon)$  для будь-якого  $\varepsilon$ . Це буде справедливо для всіх можливих реалізацій. Це проілюстровано на Рис. 14.3.1.

Той факт, що послідовність  $\bar{X}_n \xrightarrow{p} 0$ , означає, що для будь-якого  $\varepsilon > 0$  ймовірності того, що  $|\bar{X}_n| > \varepsilon$ , прямують до нуля. Іншими словами, якщо ми розглянемо множину послідовностей реалізацій  $\bar{X}_n$ ,  $n \geq 1$ , відповідних випадкових величин, то частка графіків, які виходять за межі околу  $(-\varepsilon; \varepsilon)$ , буде прямувати до нуля. При цьому не існує гарантії, що деяка конкретна послідовність  $\bar{X}_n$  завжди лежатиме у відповідному околі. Це проілюстровано на Рис. 14.3.2.

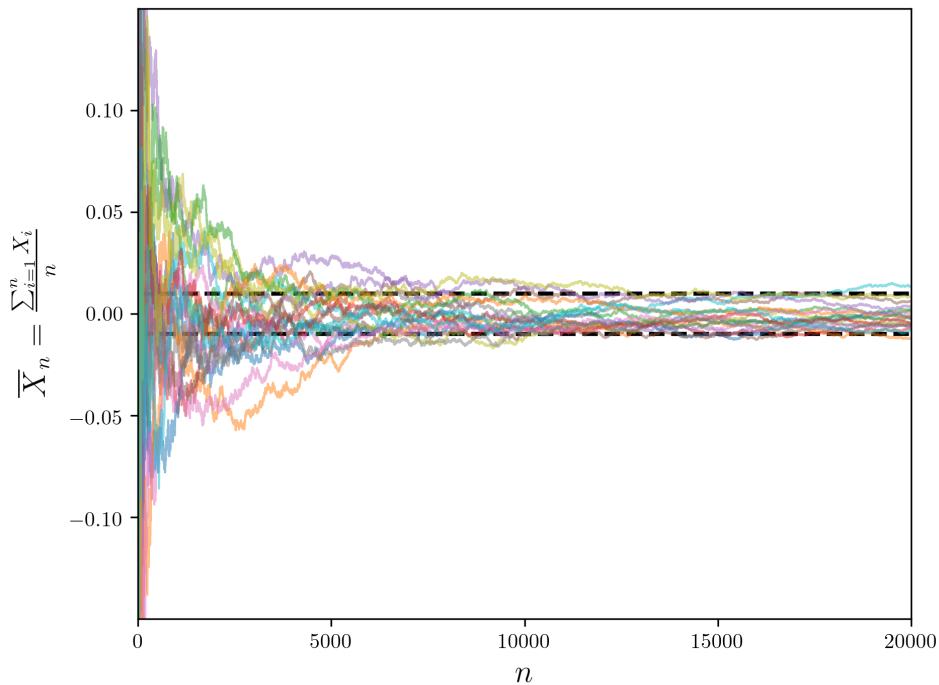


Рис. 14.3.2.: Графіки 20 послідовностей реалізацій  $\bar{X}_n$ ,  $n = 1, \dots, 20\ 000$ , випадкових величин  $\bar{X}_n$ . Штрихованими лініями окреслено окіл  $(-0.01; 0.01)$

Відповідні графіки було згенеровано за допомогою коду на Лістингу 14.3.1.

Лістинг 14.3.1: Код для інтерпретації ЗВЧ та ПЗВЧ

```

import matplotlib.pyplot as plt
import numpy as np
from numpy.random import seed, random

5
def plot_lln(n, m=1, ep=0.01):
    plt.rcParams['text.usetex'] = True

    s = np.cumsum(2 * ((random((n, m)) < 0.5) - 0.5), axis=0) \
        /\ \
        (np.arange(1, n + 1).reshape(-1, 1))

    plt.plot(np.arange(1, n + 1), s, '-', lw=1, alpha=0.5)
    plt.hlines(ep, 0, n, linestyle='dashed', lw=2, alpha=1)
    plt.hlines(-ep, 0, n, linestyle='dashed', lw=2, alpha=1)
    plt.xlim(0, n)
    plt.ylim(-0.15, 0.15)
    plt.xticks([0, 5000, 10000, 15000, 20000], fontsize=10)
    plt.yticks([-0.1, -0.05, 0, 0.05, 0.1], fontsize=10)
    plt.xlabel(r"\$n\$", fontsize=15)
    plt.ylabel(r"\$\overline{X}_n = \frac{\sum_{i=1}^n X_i}{n}\$", fontsize=15)
    plt.tight_layout()
    plt.savefig(
        "../images/Lecture 14/convergence_{}.png".format(m),
        dpi=300
25

```

```

)
plt.close()

30 seed(100)

plot_11n(20000)
plot_11n(20000, m=20)

```

## 14.4. Приклади застосування законів великих чисел

Практичних ситуацій, де застосовні ЗВЧ та ПЗВЧ, існує безліч. Наприклад, страхова компанія визначає вартість полісу з міркувань максимізації свого прибутку на основі історичних статистичних даних у такий спосіб. Імовірність виплати деякої конкретної суми кладуть рівною частці реальних страхових випадків, які завершилися виплатою такої суми. Якщо страхових випадків у даних доволі багато, ЗВЧ дає підстави вважати таку рівність майже достовірною.

За великим рахунком, без ЗВЧ неможливо було б уявити собі статистику як галузь, присвячену оцінюванню параметрів невідомих розподілів на основі наявних даних. Саме справедливість ЗВЧ дає підстави очікувати, що збільшення числа доступних даних веде до поліпшення якості оцінок параметрів та, відповідно, прогнозування майбутніх значень випадкових величин.

У сучасну еру великих даних кількість доступних досліднику реалізацій випадкових величин настільки велика, що на практиці ЗВЧ можна застосовувати в усіх випадках, коли виконуються відповідні умови теореми, фактично кладучи рівними емпіричні частоти та відповідні їм імовірності.

У цьому розділі розглянемо декілька конкретних прикладів, де застосування ЗВЧ дає змогу досягати цікавих і корисних результатів.

### 14.4.1. Спроможні оцінки параметрів розподілів

Як зазначалося деінде раніше, суть статистики полягає в тому, що на основі наявних даних ми намагаємося встановити, згідно з яким розподілом їх було згенеровано. Із знання дають нам змогу ліпше зрозуміти природу цих даних, обчислити ймовірності, пов'язані з відповідними величинами, та спрогнозувати майбутні значення з цього розподілу.

Якщо ми підозрюємо, що розподіл даних належить до деякої загальновідомої сім'ї розподілів (наприклад, є нормальним), то достатньо визначити параметри відповідного розподілу (наприклад, сподівання й дисперсію). Такого роду задачі є предметом так званої *параметричної статистики* (parametric statistics).

Нехай маємо вибірку з  $n$  елементів, кожен із яких є реалізацією однієї з незалежних випадкових величини  $X_i$ ,  $i = 1, \dots, n$ , усі з яких мають одинаковий розподіл. Нехай ми припускаємо, що цей розподіл належить деякій сім'ї розподілів із параметром  $\theta \in A \subseteq \mathbb{R}^k$ . Наприклад, якщо це сім'я експоненційних розподілів, то  $\theta = \lambda \in \mathbb{R}^+$ , а якщо нормальніх, то  $\theta = (\mu, \sigma^2)^\top \in \mathbb{R} \times \mathbb{R}^+$ . Тоді на основі значень вибірки можна обчислити так звану *оцінку* (estimator):

$$\hat{\theta} = g(X_1, \dots, X_n). \quad (14.4.1)$$

За своюю природою, оцінка параметра є *випадковою величиною*, а для кожної конкретної вибірки ми можемо обчислити значення її *реалізації*<sup>2</sup>.

<sup>2</sup> В англомовній термінології чітко розрізняють поняття оцінки як випадкової величини, яку називають estimator,

Звісно, ключове питання, яке цікавить нас як дослідників — наскільки близьким є значення підрахованої оцінки до справжнього значення параметра? Відповідь на це питання залежить від того, що ми розуміємо під близькістю. Наприклад, одна з властивостей, яку може мати деяка оцінка  $\hat{\theta}$  — бути *незміщеною* (unbiased), тобто щоб  $\mathbb{E}[\hat{\theta}] = \theta$ .

Проте навіть зміщені оцінки можуть бути корисні, а в деяких випадках і привабливіші (наприклад, якщо вони мають нижчу дисперсію), якщо можна показати, що зі збільшенням розміру вибірки значення оцінки прямує до справжнього значення параметра. Оцінку називають *спроможною* (consistent), якщо  $\hat{\theta} \xrightarrow{p} \theta$ <sup>3</sup>.

Сам (П)ЗВЧ свідчить про те, що середнє вибіркове є спроможною оцінкою для сподівання, адже з (14.2.1) випливає

$$\bar{X}_n = \frac{1}{n} \sum_{i=1}^n X_i \xrightarrow{\text{М.Н.}} \mathbb{E}[X] .$$

А з урахуванням Теореми 13.2.12 маємо, що якщо параметр  $\theta$  можна подати як деяку неперервну майже напевно функцію від  $\mathbb{E}[X]$ , тобто  $\theta = g(\mathbb{E}[X])$ , то відповідною спроможною оцінкою буде

$$\hat{\theta} = g(\bar{X}_n) \xrightarrow{\text{М.Н.}} g(\mathbb{E}[X]) = \theta .$$

**Приклад 14.4.1.** Чи не найдавнішим і найпростішим методом оцінювання параметрів є так званий *метод моментів* (method of moments). Нехай маємо вибірку незалежних  $X_1, \dots, X_n$ , де  $X_i$  мають деякий одинаковий розподіл із вектором параметрів  $\theta = (\theta_1, \dots, \theta_k)^\top$ ,  $i = 1, \dots, n$ . Тоді оцінку такого вектора параметрів можна здійснити, розв'язавши систему рівнянь

$$\begin{aligned} \bar{X}_n &= \mathbb{E}[X_i] = g_1(\theta_1, \dots, \theta_k) , \\ &\vdots \\ \bar{X}^k_n &= \mathbb{E}[X_i^k] = g_k(\theta_1, \dots, \theta_k) , \end{aligned}$$

де  $g_1, \dots, g_k$  — деякі функції.

Наприклад, нехай  $X_i \sim N(\mu, \sigma^2)$ ,  $i = 1, \dots, n$ . Тоді  $\theta_1 = \mu$ ,  $\theta_2 = \sigma^2$ , і можна утворити таку систему рівнянь:

$$\begin{aligned} \bar{X}_n &= \frac{1}{n} \sum_{i=1}^n X_i = \mu , \\ \bar{X}^2_n &= \frac{1}{n} \sum_{i=1}^n X_i^2 = \mu^2 + \sigma^2 , \end{aligned}$$

звідки

$$\hat{\mu} = \bar{X}_n , \quad \hat{\sigma}^2 = \bar{X}^2_n - \bar{X}_n = \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X}_n)^2 .$$

Оцінки  $\hat{\mu}$  та  $\hat{\sigma}^2$  є спроможними. Проте можна показати, що оцінка  $\hat{\sigma}^2$  є зміщеною. Незміщена оцінка дисперсії дорівнює  $\frac{n}{n-1} \hat{\sigma}^2$ , але це вивчають у курсі «Математичної статистики».  $\square$

та її реалізації, яку називають *estimate*.

<sup>3</sup>Звісно, якщо можна показати, що  $\hat{\theta} \xrightarrow{\text{М.Н.}} \theta$ , то це ще ліпше, проте на практиці цілком достатньо збіжності за ймовірністю.

#### 14.4.2. Емпіричні функції розподілів

На відміну від згаданої вище параметричної статистики, у рамках *непараметричної статистики* (nonparametric statistics) ми намагаємося аналізувати дані, не висуваючи припущені щодо конкретного класу розподілів, до якого можуть належати наявні дані.

Одним із методів непараметричної статистики є оцінювання функції розподілу даних без жодних ап'юорних припущенів. Відповідну оцінку називають *емпіричною функцією розподілу* (empirical distribution function). Нехай маємо вибірку з  $n$  елементів, кожен із яких є реалізацією однієї з незалежних випадкових величини  $X_i, i = 1, \dots, X_n$ , усі з яких мають однакову функцію розподілу  $F$ . Тоді для будь-якого  $x$  можна порахувати, скільки величин набули значення, що не перевищують  $x$ :

$$R_n(x) = \sum_{j=1}^n \mathbb{1}\{X_j \leq x\} .$$

Кожна з індикаторних величин має розподіл Бернуллі з параметром  $p = F(x)$ , а тому  $R_n(x) \sim \text{Binom}(n, F(x))$ .

Емпіричну функцію розподілу тоді можна дістати як

$$\hat{F}_n(x) = \frac{R_n(x)}{n} . \quad (14.4.2)$$

Варто розуміти, що для кожного конкретного  $x$  оцінка  $\hat{F}_n(x)$  є випадковою величиною, оскільки вона є функцією від індикаторних випадкових величин. Для кожного конкретного набору даних ми маємо справу з *реалізацією* цієї випадкової величини, тобто з конкретними значеннями, які можна обчислити на основі наявної вибірки.

Як себе поводить  $\hat{F}_n(x)$ , коли даних є дуже багато, тобто коли  $n \rightarrow \infty$ ? Згідно з ПЗВЧ, для кожного  $x$  випадкова величина  $R_n(x)$  є сумаю індикаторних величин з однаковим розподілом  $\text{Bern}(F(x))$  та скінченим сподіванням, яке дорівнює

$$\mathbb{E}[\hat{F}_n(x)] = \frac{1}{n} \sum_{i=1}^n \mathbb{E}[R_n(x)] = \frac{1}{n} \cdot n \cdot F(x) = F(x) .$$

Відтак за Теоремою 14.2.1  $\hat{F}_n(x) \xrightarrow{\text{М.Н.}} F(x)$ .

Більше того, цей результат можна посилити.

**Теорема 14.4.2** (Теорема Глівенка-Кантеллі (Glivenko-Cantelli theorem)<sup>4</sup>). Нехай маємо послідовність незалежних випадкових величин  $X_1, X_2, \dots$  таких, що вони мають однакову функцію розподілу  $F$ . Тоді

$$\sup_x |F(x) - \hat{F}_n(x)| \xrightarrow{\text{М.Н.}} 0 . \quad (14.4.3)$$

Результат цієї теореми сильніший від раніше зазначеного, оскільки тут збіжність є рівномірною для всіх  $x$ .

**Приклад 14.4.3.** Розглянемо простий приклад, у якому вибірка  $X_1, \dots, X_n$  містить реалізації незалежних випадкових величин із розподілом  $N(20, 25)$ . На Рис. 14.4.1 зображені графіки емпіричних функцій розподілу, обчислені за допомогою (14.4.2), та відповідних теоретичних

<sup>4</sup>Названа так на честь українського математика Валерія Глівенка (1896–1940) та італійського математика Франческо Кантеллі (Francesco Paolo Cantelli, 1875–1966).

функцій розподілу, тобто  $F(x) = \Phi\left(\frac{x-20}{5}\right)$ . На панелі (а) зображене ситуацію для  $n = 100$ , на панелі (б) — для  $n = 1000$ . Як можна бачити, збільшення вибірки веде до суттєвого наближення емпіричної функції розподілу до теоретичної. В умовах наявності великих даних ПЗВЧ гарантує дуже високу якість такої оцінки функції розподілу. Відповідний код наведено на Лістингу 14.4.1.

Лістинг 14.4.1: Код для Прикладу 14.4.3

```

import numpy as np
from numpy.random import seed
from scipy import stats
import matplotlib.pyplot as plt
from matplotlib import rc

5 def empirical_df(n_sample, n_plot=1000, loc=20, scale=5):
    sample = stats.norm.rvs(loc=loc, scale=scale, size=n_sample)
    sample = np.sort(sample)

    x = np.linspace(sample[0], sample[-1], n_plot)
    F_hat = np.array([sum(sample <= x[i]) / n_sample for i in range(n_plot)])
    F = stats.norm.cdf(x, loc=loc, scale=scale)

10   return x, F_hat, F

15 def plot_df(x, F_hat, F, n_sample):
    rc('text', usetex=True)
    rc('text.latex', unicode=True)
    rc('text.latex', preamble=r'\usepackage[utf8]{inputenc}')
    rc('text.latex', preamble=r'\usepackage[ukrainian]{babel}')

20   plt.plot(x, F, label="Теоретична функція розподілу")
    plt.plot(x, F_hat, label="Емпірична функція розподілу")
    plt.xlabel(r"$x$", fontsize=15)
    plt.ylabel(r"$F(x)$", fontsize=15)
    plt.xticks(fontsize=10)
    plt.yticks(fontsize=10)
    plt.legend(fontsize=15, loc="upper left")

25   plt.savefig(
        "../../images/Lecture 14/empiricalcdf_{}.png".format(n_sample),
        dpi=300
    )
    plt.close()

30   seed(100)
n_sample_1 = 100
n_sample_2 = 1000

35   plot_df(*empirical_df(n_sample_1), n_sample_1)
plot_df(*empirical_df(n_sample_2), n_sample_2)

```



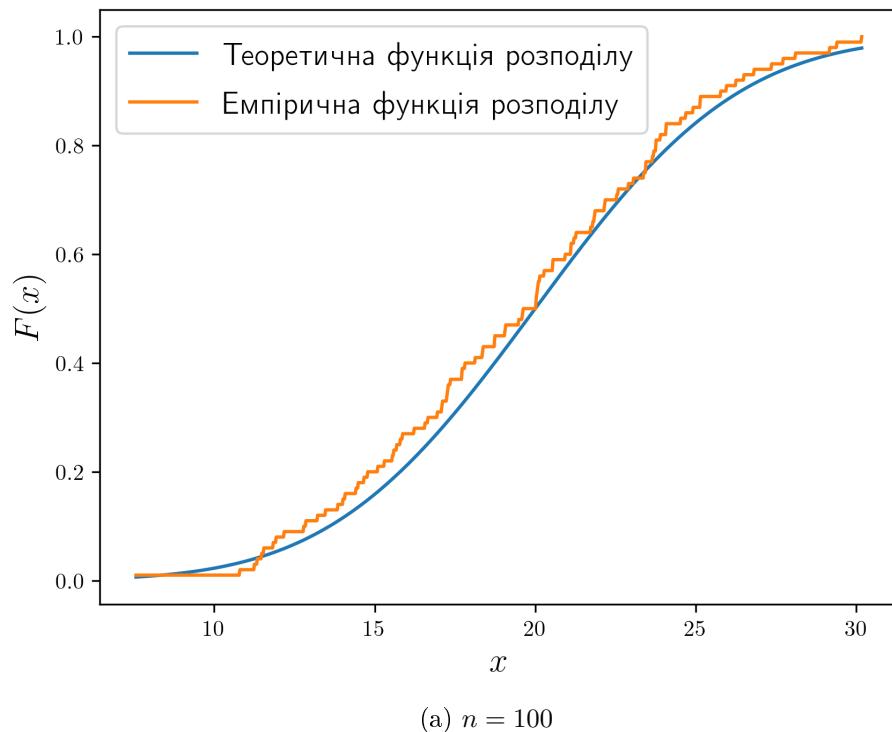
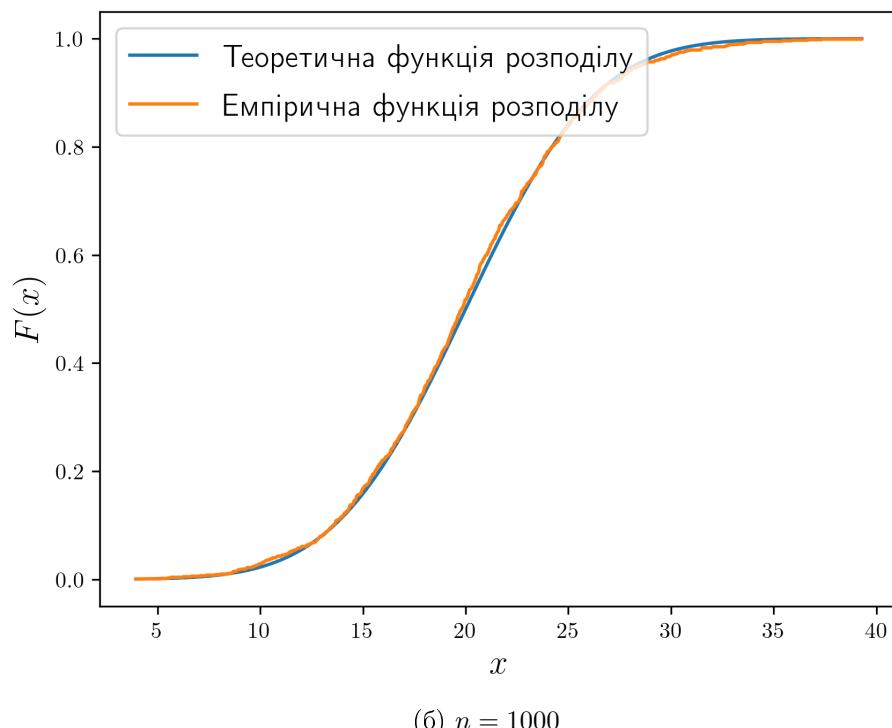
(a)  $n = 100$ (b)  $n = 1000$ 

Рис. 14.4.1.: Графіки емпіричних та теоретичних функцій розподілу для Прикладу 14.4.3

### 14.4.3. Гістограми розподілів

Розгляньмо ще один надзвичайно поширений метод непараметричної статистики, пов'язаний із оцінкою щільності розподілу випадкової величини, — побудову *гістограм* (histograms) розподілів. Насправді ми ним користувалися й раніше через його інтуїтивну зрозумілість, але зараз у нас достатньо теоретичного апарату, щоб довести його адекватність.

Згадаймо, що ми називаємо гістограмою. Нехай маємо набір значень  $x_1, \dots, x_n$  таких, що  $a \leq x_i \leq b$ ,  $i = 1, \dots, n$ . Нехай  $k$  — деяке ціле число. Інтервал  $[a; b]$  можна розбити на  $k$  підінтервалів однакової довжини  $(b - a)/k^5$ . Для кожного підінтервалу можна обчислити число елементів  $c_i$  нашого набору, які в нього потрапили. Стовпчикова діаграма, де кожний стовпчик відповідає своєму підінтервалу, а висота стовпчика дорівнює  $c_i$ , і має назву гістограми. Ми бачили такого роду графік на Рис. 8.1.2. Принцип побудови гістограми можна адаптувати для апроксимації щільності розподілу.

Розгляньмо тепер вибірку з  $n$  елементів, кожен із яких є реалізацією однієї з незалежних випадкових величини  $X_i$ ,  $i = 1, \dots, n$ , усі з яких мають однакові функцію розподілу  $F$  та щільність розподілу  $f$ . Розгляньмо індикаторну випадкову величину  $Y_i = \mathbb{1}_{\{c_1 \leq X_i < c_2\}}$  для деяких дійсних чисел  $c_1 < c_2$ . Сподівання  $Y_i$  скінченне і дорівнює  $\mathbb{E}[Y_i] = \mathbb{P}_{Y_i}(Y_i = 1) = \mathbb{P}_{X_i}(c_1 \leq X_i < c_2) = F(c_2) - F(c_1)$ .

Тоді, згідно з ПЗВЧ,  $\bar{Y} \xrightarrow{m.n.} F(c_2) - F(c_1)$ . Іншими словами, якщо побудувати гістограму таку, що площа (не висота, а площа) кожного стовпчика над інтервалом  $[c_1; c_2]$  дорівнює відносній частці  $\bar{Y}$  елементів вибірки, які в нього потрапили, то для дуже великих вибірок ( $n \rightarrow \infty$ ) площа відповідного стовпчика дорівнюватиме площі відповідної криволінійної трапеції під графіком щільності розподілу (адже функція розподілу є первісною для щільності).

Альтернативно, зі збільшенням числа елементів вибірки  $n \rightarrow \infty$  та звуженням довжини стовпчиків до нескінченно малих величин матимемо графік щільності відповідного розподілу, адже за визначенням похідної маємо

$$f(c_1) = \lim_{c_2 \rightarrow c_1} \frac{F(c_2) - F(c_1)}{c_2 - c_1},$$

тобто площа відповідного стовпчика, поділена на довжину його основи, у граничному переході дорівнюватиме значенню щільності у відповідній точці.

**Приклад 14.4.4.** Як конкретний приклад розгляньмо дві вибірки значень, згенерованих відповідно до розподілу  $\text{Exp}(1)$ . Перша вибірка налічує  $n = 100$  елементів, а друга —  $n = 1000$ . Розбиймо носій першої вибірки на  $k = 10$  підінтервалів однакової довжини, а носій другої — на  $k = 30$ . Для кожного підінтервалу обчислімо відповідні частки  $\bar{Y}$  елементів вибірки, які в нього потрапили. Висоти відповідних стовпчиків можна визначити, виходячи зі значення  $\bar{Y}$  та довжини підінтервалу.

На Рис. 14.4.2 зображено відповідні гістограми та накладені на них графіки щільності розподілу  $\text{Exp}(1)$ . На панелі (а) зображено ситуацію для  $n = 100$ , на панелі (б) — для  $n = 1000$ . Як можна бачити, збільшення вибірки веде до вищої якості апроксимації щільності розподілу відповідною гістограмою. Відповідний код наведено на Лістингу 14.4.2.

Лістинг 14.4.2: Код для Прикладу 14.4.4

```
import numpy as np
from numpy.random import seed
```

<sup>5</sup> В англомовній літературі такі підінтервали називають *bins* (дослівно сміттєвий бак).

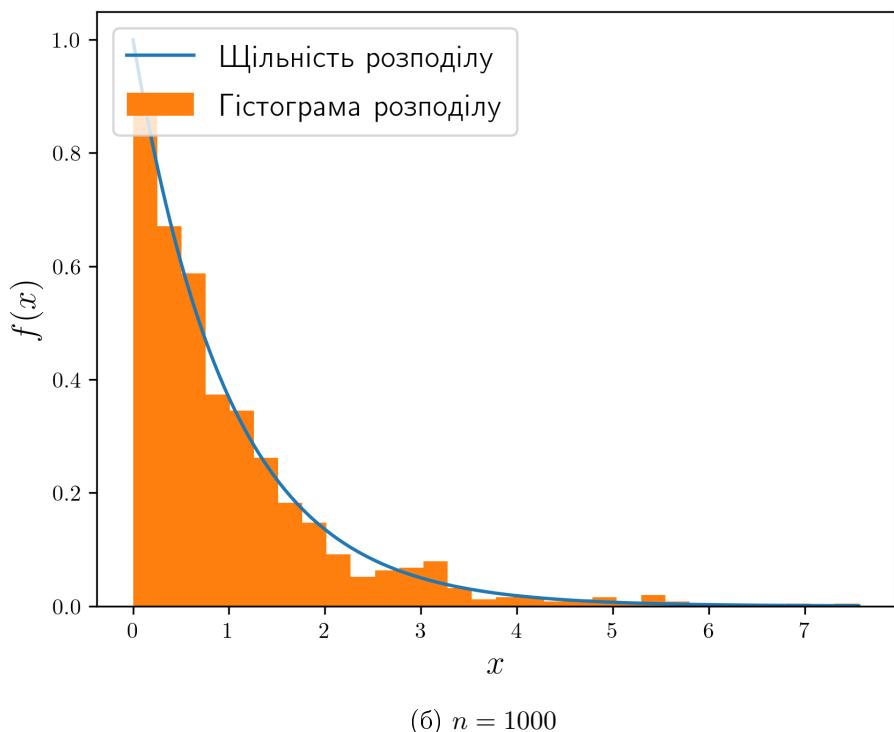
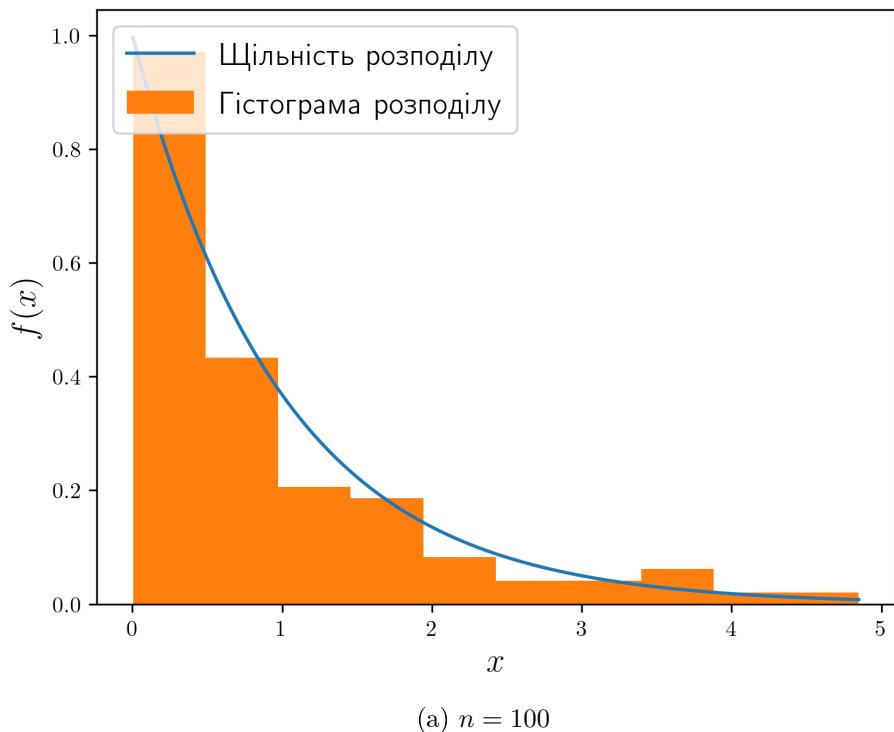


Рис. 14.4.2.: Гістограми та щільність розподілу для Прикладу 14.4.4

```

from scipy import stats
import matplotlib.pyplot as plt
from matplotlib import rc

5   def plot_hist(n_sample, n_bins, n_plot=10**3):
    rc('text', usetex=True)
10   rc('text.latex', unicode=True)
    rc('text.latex', preamble=r'\usepackage[utf8]{inputenc}')
    rc('text.latex', preamble=r'\usepackage[ukrainian]{babel}')

15   sample = stats.expon.rvs(size=n_sample)
    sample = np.sort(sample)
    x = np.linspace(sample[0], sample[-1], n_plot)
    f = stats.expon.pdf(x)

20   plt.plot(x, f, label="Щільність розподілу")
    plt.hist(sample, bins=n_bins,
              label="Гістограма розподілу", density=True)
    plt.xlabel(r"$x$", fontsize=15)
    plt.ylabel(r"$f(x)$", fontsize=15)
    plt.xticks(fontsize=10)
25   plt.yticks(fontsize=10)
    plt.legend(fontsize=15, loc="upper left")

    plt.savefig(
        "../images/Lecture 14/histogram_{}.png".format(n_sample),
30   dpi=300
    )
    plt.close()

35   seed(100)
n_sample_1 = 100
n_sample_2 = 1000

n_bins_1 = 10
40   n_bins_2 = 30

plot_hist(n_sample_1, n_bins_1)
plot_hist(n_sample_2, n_bins_2)

```

□

#### 14.4.4. Інтегрування методом Монте-Карло

Нехай  $f$  — деяка функція  $k$  змінних, інтеграл якої за деякою множиною  $A$ ,  $\int_A f(\mathbf{x}) d\mathbf{x}$ , непросто обчислити. Наприклад, функція може взагалі не мати аналітичного запису своєї первісної, або  $k$  є дуже великим. На перший погляд може здатися, що обчислення такого інтегралу в жодний спосіб не пов'язано з теорією ймовірностей. Виявляється, у цьому випадку для наближеного обчислення інтегралу можна застосувати ЗВЧ. Такий метод інтегрування називають *інтегруванням методом Монте-Карло* (Monte Carlo integration).

Нехай  $A$  — деяка компактна множина в просторі  $\mathbb{R}^k$ . Суть методу полягає в генеруванні незалежних випадкових векторів  $\mathbf{X}_i$ , рівномірно розподілених на  $A$ ,  $i = 1, \dots, n$ . Тоді інтеграл можна

переписати так:

$$\int_A f(\mathbf{x}) d\mathbf{x} = \lambda_k(A) \int_{\mathbb{R}^k} f(\mathbf{x}) \cdot \mathbb{1}\{\mathbf{x} \in A\} \cdot \frac{1}{\lambda_k(A)} d\mathbf{x} = \lambda_k(A) \cdot \mathbb{E}[f(\mathbf{X})] ,$$

де  $\mathbf{X} \sim U(A)$ . Виконуються всі умови для застосування ПЗВЧ у версії Теореми 14.2.1, адже величини  $f(\mathbf{X}_i)$ ,  $i = 1, \dots, n$ , незалежні, мають одинаковий розподіл та скінченне сподівання. Тому

$$\int_A f(\mathbf{x}) d\mathbf{x} \approx \lambda_k(A) \cdot \frac{1}{n} \sum_{i=1}^n f(\mathbf{X}_i) . \quad (14.4.4)$$

Особливу цінність метод Монте-Карло має для інтегрування функції від дуже великого числа аргументів (декілька сотень чи тисяч), адже в цьому випадку складність інтегрування зростає як показникова функція від розмірності простору інтегрування  $k$ . У той же час складність генерування випадкових чисел у  $k$ -вимірному просторі зростає лінійно. Такий підхід до інтегрування можна застосовувати навіть у нескінченно-вимірних просторах.

**Приклад 14.4.5.** Хрестоматійним прикладом застосування методу Монте-Карло є обчислення константи  $\pi$ . Розгляньмо одиничний круг  $D = \{(x, y)^\top : x^2 + y^2 \leq 1\}$ , уписаний у квадрат  $A = [-1; 1] \times [-1; 1]$ , площа якого дорівнює 4. Розгляньмо площу круга як інтеграл від одиниці за множиною  $D$ :

$$\iint_D 1 dx dy = \pi .$$

Його можна розглядати як інтеграл від індикаторної функції:

$$\pi = \iint_{\mathbb{R}^2} \mathbb{1}\{(x, y)^\top \in D\} dx dy .$$

Тоді наближене значення цього інтегралу можна дістати з (14.4.4):

$$\pi \approx 4 \cdot \frac{1}{n} \sum_{i=1}^n \mathbb{1}\{(X_i, Y_i)^\top : X_i^2 + Y_i^2 \leq 1\} .$$

Код на Листингу 14.4.3 генерує 1 000 000 випадкових точок із прямокутника  $[-1; 1] \times [-1; 1]$  та обчислює частку точок, які потрапили в одиничний круг. На Рис. 14.4.3 зображено відповідну ілюстрацію. Результати роботи програми такі:

Справжнє значення константи дорівнює 3.14159

Наближене значення константи дорівнює 3.14193

Отже ми бачимо, що результат доволі близький, хоча й неідеальний.

Листинг 14.4.3: Код для Прикладу 14.4.5

```
import numpy as np
from numpy.random import seed
from scipy.stats import uniform

seed(100)
n = 10**6

x = uniform.rvs(size=n, loc=-1, scale=2)
```

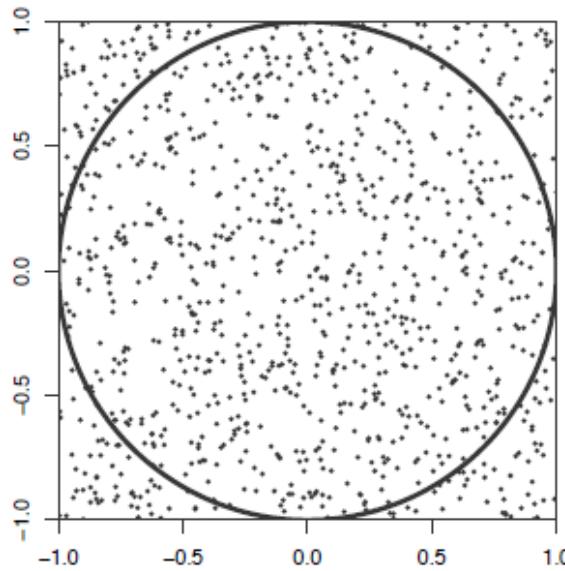


Рис. 14.4.3.: Точки, згенеровані випадково з рівномірним розподілом на прямокутнику  $[-1; 1] \times [-1; 1]$ , а також одиничне коло для Прикладу 14.4.5 ([1], Рис. 10.8)

```

10 y = uniform.rvs(size=n, loc=-1, scale=2)
  p = sum(x**2 + y**2 < 1) / n
  pi_approx = 4 * p
15 print("Справжнє значення константи дорівнює {:.6}".format(np.pi))
  print("Наближене значення константи дорівнює {:.6}".format(pi_approx))

```

□

## 15. Твірні функції моментів та характеристичні функції

У математиці поняття твірної функції використовують на позначення деякого способу компактного опису нескінченної послідовності дійсних чисел. Це уможливлює аналіз властивостей таких послідовностей за допомогою аналітичних методів. У теорії ймовірностей використовують декілька різних видів таких функцій. На цій лекції ми розглянемо твірну функцію моментів випадкової величини та пов'язані з нею властивості. Наприкінці лекції ми також коротко ознайомимося з узагальненням твірної функції моментів, яке має назву характеристичної функції.

### 15.1. Визначення та основні властивості

**Визначення 15.1.1.** *Твірною функцією моментів* (moment generating function) деякої випадкової величини  $X$  є

$$M_X(t) = \mathbb{E}[e^{tX}] = \int e^{tX} d\mathbb{P}, \quad (15.1.1)$$

яку визначено в усіх точках  $t$ , де  $M_X(t) < \infty$ .  $\square$

Аргумент  $t$  у цьому визначенні не має самостійного значення.

Якщо з контексту зрозуміло, про яку саме величину  $X$  мова, то відповідний індекс опускати- memo. Можна помітити, що твірна функція моментів для різних випадкових величин з однаковим розподілом однаакова, оскільки це сподівання, яке для величин  $X \xrightarrow{d} Y$  однаакове. Відтак можна говорити про твірну функцію моментів не окремої випадкової величини, а цілого розподілу.

Нескладно бачити, що  $M_X(0) = 1 < \infty$ , тобто в точці 0 твірна функція моментів завжди визначена.

Існує дуже простий результат для обчислення твірної функції моментів деякого лінійного петретворення випадкової величини.

**Твердження 15.1.2.** Нехай  $M_X$  — твірна функція моментів деякої величини  $X$ . Тоді для довільних  $a, b \in \mathbb{R}$  для випадкової величини  $Y = aX + b$  існує твірна функція моментів, яка дорівнює  $M_Y(t) = e^{tb} M_X(at)$ .

*Доведення.* Це можна показати безпосереднім обчисленням:

$$M_Y(t) = \mathbb{E}[e^{t(aX+b)}] = \mathbb{E}[e^{tb} \cdot e^{atX}] = e^{tb} M_X(at).$$

Така твірна функція моментів справді визначена, адже  $e^{tb}$  скінченне в усіх  $t$ , у тому числі тих, для яких скінчена  $M_X(t)$ .  $\square$

Апарат твірних функцій моментів дає змогу суттєво спростити обчислення розподілів сум незалежних випадкових величин.

**Твердження 15.1.3.** Нехай  $X$  та  $Y$  — дві незалежні випадкові величини. Тоді

$$M_{X+Y}(t) = M_X(t)M_Y(t) . \quad (15.1.2)$$

*Доведення.* Це можна довести безпосередньо з визначення:

$$M_{X+Y}(t) = \mathbb{E} [e^{t(X+Y)}] = \mathbb{E} [e^{tX}e^{tY}] = \mathbb{E} [e^{tX}] \mathbb{E} [e^{tY}] = M_X(t)M_Y(t) ,$$

де ми врахували незалежність  $X \perp\!\!\!\perp Y$  та використали Теорему 9.4.11.  $\square$

Цілком очевидно, що за індукцією можна довести відповідний результат для суми  $n$  незалежних величин:

$$M_{X_1+\dots+X_n}(t) = M_{X_1}(t) \dots M_{X_n}(t) .$$

Обчислімо твірні функції моментів для окремих відомих розподілів.

**Твердження 15.1.4.** Твірні функції моментів деяких відомих розподілів дорівнюють:

- (i) якщо  $X \sim \text{Bern}(p)$ , то  $M_X(t) = 1 + p(e^t - 1)$  для всіх  $t \in \mathbb{R}$ ;
- (ii) якщо  $X \sim \text{Binom}(n, p)$ , то  $M_X(t) = (1 + p(e^t - 1))^n$  для всіх  $t \in \mathbb{R}$ ;
- (iii) якщо  $X \sim \text{Geom}(p)$ , то  $M_X(t) = \frac{p}{1 - (1-p)e^t}$  для всіх  $t \in (-\infty; \ln \frac{1}{1-p})$ ;
- (iv) якщо  $X \sim \text{Pois}(\lambda)$ , то  $M_X(t) = e^{\lambda(e^t - 1)}$  для всіх  $t \in \mathbb{R}$ ;
- (v) якщо  $X \sim \text{U}((a; b))$ , то  $M_X(t) = \frac{e^{tb} - e^{ta}}{t(b-a)}$  для  $t \neq 0$  і  $M_X(0) = 1$ ;
- (vi) якщо  $X \sim N(\mu, \sigma^2)$ , то  $M_X(t) = e^{t\mu + \frac{1}{2}\sigma^2 t^2}$  для всіх  $t \in \mathbb{R}$ ;
- (vii) якщо  $X \sim \text{Exp}(\lambda)$ , то  $M_X(t) = \frac{\lambda}{\lambda - t}$  для  $t < \lambda$ .

*Доведення.* (i) Нехай  $X \sim \text{Bern}(p)$ . Тоді її твірна функція моментів дорівнює

$$M_X(t) = \mathbb{E} [e^{tX}] = e^{t \cdot 0} \mathbb{P}_X(X = 0) + e^{t \cdot 1} \mathbb{P}_X(X = 1) = 1 - p + e^t \cdot p = 1 + p(e^t - 1) .$$

Оскільки ця функція завжди скінчена, то  $M_X$  визначена для всіх  $t \in \mathbb{R}$ ;

- (ii) нехай  $X \sim \text{Binom}(n, p)$ . Як ми показали в Твердженні 9.5.4,  $X$  можна подати як суму  $n$  незалежних величин Бернуллі  $Y_i \sim \text{Bern}(p)$ ,  $i = 1, \dots, n$ . Відтак, за Твердженням 15.1.3,

$$M_X(t) = M_{Y_1+\dots+Y_n}(t) = (M_{Y_1}(t))^n = (1 + p(e^t - 1))^n .$$

Оскільки ця функція завжди скінчена, то  $M_X$  визначена для всіх  $t \in \mathbb{R}$ ;

- (iii) нехай  $X \sim \text{Geom}(p)$ . Тоді

$$M_X(t) = \mathbb{E} [e^{tX}] = \sum_{k=0}^{\infty} e^{tk} (1-p)^k p = p \sum_{k=0}^{\infty} ((1-p)e^t)^k = \frac{p}{1 - (1-p)e^t} ,$$

що випливає з суми нескінченно спадної геометричної прогресії, але тільки для таких  $t$ , що  $(1-p)e^t < 1$ , інакше ряд буде розбіжний, і  $M_X$  не буде скінченою. Тобто твірна функція моментів визначена для  $t \in (-\infty; \ln \frac{1}{1-p})$ ;

(iv) нехай  $X \sim \text{Pois}(\lambda)$ . Тоді

$$M_X(t) = \mathbb{E}[e^{tX}] = \sum_{k=0}^{\infty} e^{tk} \cdot \frac{e^{\lambda} \lambda^k}{k!} = e^{-\lambda} e^{\lambda e^t} = e^{\lambda(e^t - 1)},$$

що випливає з розвинення експоненти в ряд Тейлора. Оскільки ця функція завжди скінчена, то  $M_X$  визначена для всіх  $t \in \mathbb{R}$ ;

(v) нехай  $X \sim \text{U}((a; b))$ . Тоді

$$\begin{aligned} M_X(t) &= \int_a^b \frac{1}{b-a} e^{tx} dx \\ &= \frac{1}{t(b-a)} e^{tx} \Big|_a^b = \frac{e^{tb} - e^{ta}}{t(b-a)} \end{aligned}$$

для всіх  $t \neq 0$ . Якщо  $t = 0$ , то  $M(0) = \mathbb{E}[1] = 1$ ;

(vi) нехай  $Z \sim N(0, 1)$ . Тоді

$$M_Z(t) = \int_{-\infty}^{\infty} e^{tz} \cdot \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}} dz.$$

Виділивши повний квадрат, маємо

$$\begin{aligned} tz - \frac{z^2}{2} &= - \left( \left( \frac{z}{\sqrt{2}} \right)^2 - 2 \cdot \frac{z}{\sqrt{2}} \cdot \frac{t}{\sqrt{2}} + \frac{t^2}{2} - \frac{t^2}{2} \right) \\ &= - \left( \frac{z}{\sqrt{2}} - \frac{t}{\sqrt{2}} \right)^2 + \frac{t^2}{2}. \end{aligned}$$

Іншими словами,

$$M_Z(t) = e^{\frac{t^2}{2}} \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{(z-t)^2}{2}} dz = e^{\frac{t^2}{2}},$$

оскільки під інтегралом стоїть щільність нормального розподілу  $N(t, 1)$ , інтеграл якої дорівнює 1.

Для нормального розподілу  $X \sim N(\mu, \sigma^2)$  маємо результат за Твердженням 15.1.2:

$$M_X(t) = M_{\sigma Z + \mu}(t) = e^{t\mu + \frac{1}{2}\sigma^2 t^2}.$$

Оскільки ця функція завжди скінчена, то  $M_X$  визначена для всіх  $t \in \mathbb{R}$ ;

(vii) нехай  $Z \sim \text{Exp}(1)$ . Тоді

$$M_Z(t) = \int_{-\infty}^{\infty} e^{tz} \cdot e^{-z} \cdot \mathbb{1}\{z > 0\} dz = \int_0^{\infty} e^{-z(1-t)} dz = \frac{1}{1-t}$$

для  $t < 1$ , оскільки для  $t > 1$  інтеграл буде розбіжний.

Для експоненційного розподілу  $X \sim \text{Exp}(\lambda)$  маємо результат за Твердженням 15.1.2:

$$M_X(t) = M_{\frac{Z}{\lambda}}(t) = \frac{1}{1 - \frac{t}{\lambda}} = \frac{\lambda}{\lambda - t}$$

для  $t < \lambda$ .

□

## 15.2. Зв'язок між моментами та твірною функцією моментів

Між твірною функцією моментів деякої величини та її моментами існує тісний зв'язок. Зокрема, якщо твірна функція моментів визначена на множині з певними характеристиками, то існують моменти всіх порядків, а їх обчислення суттєво спрощується.

Ключовою особливістю такої множини є те, що це повинен бути відкритий інтервал, що містить 0. Справді, розгляньмо  $M_X(t) = \mathbb{E}[e^{tX}]$  для деякої невід'ємної величини  $X \geq 0$  майже напевно. Оскільки  $e^{-x} \leq 1$  для всіх  $x \geq 0$ , за властивістю монотонності сподівання маємо, що

$$M_X(t) = \mathbb{E}[e^{tX}] \leq \mathbb{E}[1] = 1 < \infty$$

для всіх  $t \leq 0$ . Нехай тепер  $M_X(t_p) < \infty$  для деякого  $t_p > 0$ . Тоді для всіх  $0 < t'_p < t_p$  маємо  $e^{t'_p} < e^{t_p}$ , а відтак знову за монотонністю сподівання випливає  $M_X(t'_p) < \infty$  для всіх  $t \in [0; t_p)$ .

Іншими словами, для  $X \geq 0$  маємо, що якщо  $M_X$  визначена на деякій множині, то це проміжок від  $-\infty$  до деякого  $t_p \geq 0$  (можливо, включно), яке може бути  $\infty$ . За аналогією для  $X \leq 0$  маємо, що якщо  $M_X$  визначена на деякій множині, то це проміжок від деякого  $t_n \leq 0$  (можливо, включно), яке може бути  $-\infty$ , до  $\infty$ .

Відтак якщо  $X$  є деякою загальною величиною, розписавши її у звичний спосіб через  $X = X^+ - X^-$ , робимо висновок, що якщо твірна функція моментів  $M_X$  визначена, то вона визначена для деякого проміжку, що містить 0. Єдиною проблемою може бути тільки те, що цей проміжок або складається з однієї точки 0, або має 0 як свою ліву або праву межу.

Виявляється, що якщо твірна функція моментів визначена на деякому *відкритому* інтервалі  $(-s_0; s_0)$ ,  $s_0 > 0$ , то випадкова величина має скінченні моменти всіх порядків.

**Теорема 15.2.1.** Нехай твірна функція моментів  $M_X$  випадкової величини  $X$  визначена на відкритому інтервалі  $(-s_0; s_0)$ , що містить 0. Тоді моменти  $X$  довільного порядку є скінченні, і вони дорівнюють

$$M_X^{(k)}(0) = \mathbb{E}[X^k]. \quad (15.2.1)$$

*Доведення.* Можна розвинути твірну функцію моментів у ряд Тейлора. Спочатку потрібно помітити, що

$$M_X(s) = \mathbb{E}[e^{sX}] = \mathbb{E}\left[\sum_{k=0}^{\infty} \frac{s^k X^k}{k!}\right].$$

Оскільки  $e^{|sx|} \leq e^{sx} + e^{-sx}$ , і сума цих експонент є інтегровною на проміжку  $|s| < s_0$ , маємо

$$\sum_{k=0}^n \frac{|s^k x^k|}{k!} \leq e^{|sx|},$$

тобто що всі скінченні суми відповідного розвинення ряду Тейлора обмежено зверху інтегровною функцією, а самі скінченні суми прямують до  $e^{sx}$ . За теоремою про мажоровану збіжність 11.1.3 можна помінити місцями інтегрування (ряд) і сподівання:

$$M_X(s) = \mathbb{E}\left[\sum_{k=0}^{\infty} \frac{s^k X^k}{k!}\right] = \sum_{k=0}^{\infty} \frac{s^k \mathbb{E}[X^k]}{k!}. \quad (15.2.2)$$

оскільки  $M_X(s) < \infty$  для  $s \in (-s_0; s_0)$ , то відповідний ступеневий ряд збіжний, а відтак можна взяти похідну зліва і справа та підставити нуль:

$$M_X^{(k)}(0) = \mathbb{E}[X^k] .$$

□

**Приклад 15.2.2.** Використаймо Теорему 15.2.1 для обчислення моментів окремих відомих розподілів.

Нехай  $X \sim N(0, 1)$ . Тоді, згідно з Твердженням 15.1.4,

$$M_X(t) = e^{\frac{t^2}{2}} = \sum_{k=0}^{\infty} \frac{t^{2k}}{2^k k!} = \sum_{k=0}^{\infty} \frac{(2k)!}{2^k \cdot k!} \cdot \frac{t^{2k}}{(2k)!} < \infty$$

для всіх  $t \in \mathbb{R}$ , тобто відкритим інтервалом, що містить 0, у цьому випадку є вся дійсна вісь. Порівнюючи цей ряд із виразом розвинення твірної функції моментів у ряд Тейлора, бачимо, що

$$\mathbb{E}[X^{2k}] = \frac{(2k)!}{2^k \cdot k!}$$

і що  $\mathbb{E}[X^{2k+1}] = 0$ . Аналогічний результат ми дістали в Прикладі 11.4.4, але тепер нам не доведено виконувати інтегрування частинами.

Аналогічно можна обчислити моменти експоненційного розподілу. Нехай  $X \sim \text{Exp}(1)$ . Тоді, згідно з Твердженням 15.1.4,

$$M_X(t) = \frac{1}{1-t} = \sum_{k=0}^{\infty} t^k = \sum_{k=0}^{\infty} k! \cdot \frac{t^k}{k!}$$

для  $|t| < 1$ , тобто маємо відкритий інтервал  $(-1; 1)$ , який містить точку 0. Порівнюючи цей ряд із виразом розвинення твірної функції моментів у ряд Тейлора, бачимо, що

$$\mathbb{E}[X^k] = k! .$$

А якщо  $Y \sim \text{Exp}(\lambda)$ , тобто якщо  $Y = \frac{X}{\lambda}$ , то

$$\mathbb{E}[Y^k] = \frac{k!}{\lambda^k} .$$

Зокрема,  $\mathbb{E}[Y] = \frac{1}{\lambda}$ , а, наприклад,

$$\text{Var}(Y) = \mathbb{E}[Y^2] - (\mathbb{E}[Y])^2 = \frac{2}{\lambda^2} - \frac{1}{\lambda^2} = \frac{1}{\lambda^2} .$$

Нехай  $X \sim \text{Pois}(\lambda)$ . Тоді, згідно з Твердженням 15.1.4,

$$M_X(t) = e^{\lambda(e^t - 1)}$$

для всіх  $t \in \mathbb{R}$ , тобто відкритим інтервалом, що містить 0, у цьому випадку є вся дійсна вісь. Тоді

сподівання та дисперсію можна обчислити, взявши похідні цього виразу:

$$\begin{aligned}\mathbb{E}[X] &= M'_X(0) = e^{\lambda(e^t-1)} \cdot \lambda e^t \Big|_{t=0} = \lambda, \\ \mathbb{E}[X^2] &= M''_X(0) = \left( e^{\lambda(e^t-1)} \cdot \lambda^2 e^{2t} + e^{\lambda(e^t-1)} \cdot \lambda e^t \right) \Big|_{t=0} = \lambda^2 + \lambda, \\ \text{Var}(X) &= \mathbb{E}[X^2] - (\mathbb{E}[X])^2 = \lambda^2 + \lambda - \lambda^2 = \lambda.\end{aligned}$$

□

Теорема 15.2.1 дає змогу швидко встановити, чи існують моменти деякої випадкової величини  $X$ . За принципом контрапозиції, якщо хоча б один момент не існує, то твірна функція моментів не може бути визначена на відкритому інтервалі навколо 0. Такою є, зокрема, ситуація у випадку розподілу Коші. Ми вже знаємо, що його сподівання не існує, а відтак не повинна бути скінченою й твірна функція моментів. Справді, оскільки для довільного  $x > 0$  з розвинення експоненти в ряд Тейлора випливає  $e^x \geq \frac{x^3}{6}$ , твірна функція моментів для величини  $X$  з розподілом Коші для  $t > 0$  є нескінченою:

$$\begin{aligned}M_X(t) &= \mathbb{E}[e^{tX}] = \int_0^\infty \frac{e^{tx}}{\pi(1+x^2)} dx \\ &\geq \int_1^\infty \frac{t^3 x^3}{6\pi(1+x^2)} dx \\ &\geq \frac{t^3}{12\pi} \int_1^\infty x dx = \infty.\end{aligned}$$

Аналогічно це справедливо до  $t < 0$ . Тобто  $M_X(t) < \infty$  тільки в точці 0.

Проте потрібно розуміти, що Теорема 15.2.1 дає тільки достатні, але не неодмінні умови існування моментів.

**Приклад 15.2.3.** Розгляньмо логнормальний розподіл. Існують його моменти довільних порядків, хоча його твірна функція моментів не задовольняє умов Теореми 15.2.1.

Справді, нехай  $Y \sim N(\mu, \sigma^2)$  і  $X = e^Y$ . Тоді  $X \geq 0$  майже напевно. Як ми встановили вище, твірна функція моментів скінчена для  $t \leq 0$ . Проте для  $t > 0$  вона завжди дорівнює нескінчності. Справді,

$$M_X(t) = \mathbb{E}[e^{tX}] = \mathbb{E}[e^{te^Y}] = \int_{-\infty}^{\infty} e^{te^y} \cdot \frac{1}{\sqrt{2\pi}} e^{-\frac{y^2}{2}} dy = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{te^y - \frac{y^2}{2}} dy.$$

Для довільного  $x > 0$  з розвинення експоненти в ряд Тейлора випливає, що

$$e^x \geq 1 + x + \frac{1}{2}x^2 + \frac{1}{6}x^3.$$

Для  $t > 0$  можна підібрати такий достатньо великий  $y$ , що

$$te^y - \frac{1}{2}y^2 \geq t + ty.$$

Але тоді

$$M_X(t) \geq \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{t+ty} dy \geq \int_a^{\infty} e^{t+ty} dy = \infty$$

для будь-якого  $a > 0$ . Відтак  $M_X$  не є визначена на відкритому інтервалі навколо 0.

З іншого боку, моменти логнормального розподілу можна обчислити через твірну функцію моментів розподілу нормального:

$$\mathbb{E}[X^k] = \mathbb{E}[e^{kY}] = M_Y(k) = e^{k\mu + \frac{1}{2}k^2\sigma^2}.$$

Зокрема,

$$\begin{aligned} \mathbb{E}[X] &= e^{\mu + \frac{1}{2}\sigma^2}, \\ \mathbb{E}[X^2] &= e^{2\mu + 2\sigma^2}, \\ \text{Var}(X) &= \mathbb{E}[X^2] - (\mathbb{E}[X])^2 = e^{2\mu + \sigma^2} (e^{\sigma^2} - 1). \end{aligned}$$

Ми знаємо, що логнормальний розподіл скошений управо, і це можна підтвердити, обчисливши для нього коефіцієнт асиметрії. Для спрощення розглянемо величину  $\ln X = Y \sim N(0, \sigma^2)$ . Тоді

$$\begin{aligned} \mathbb{E}[X] &\equiv \mu = e^{\frac{1}{2}\sigma^2}, \\ \mathbb{E}[X^k] &= e^{\frac{1}{2}k^2\sigma^2} = \mu^{k^2}, \\ \text{Var}(X) &= \mu^4 - \mu^2 = \mu^2(\mu^2 - 1), \\ \mathbb{E}[(X - \mu)^3] &= \mathbb{E}[X^3 - 3\mu X^2 + 3\mu^2 X - \mu^3] = \mu^9 - 3\mu^5 + 2\mu^3. \end{aligned}$$

Тоді коефіцієнт асиметрії дорівнює

$$\begin{aligned} \text{Skew}(X) &= \mathbb{E}\left[\left(\frac{(X - \mu)}{\sqrt{\text{Var}(X)}}\right)^3\right] = \frac{\mu^9 - 3\mu^5 + 2\mu^3}{\sqrt{\mu^6(\mu^2 - 1)^3}} \\ &= \frac{\mu^6 - 3\mu^2 + 2}{(\mu^2 - 1)^{3/2}} = \frac{(\mu^2 + 2)(\mu - 1)^2(\mu + 1)^2}{(\mu^2 - 1)^{3/2}} = (\mu^2 + 2)\sqrt{\mu^2 - 1} > 0, \end{aligned}$$

оскільки  $\mu > 1$ . □

Можна довести таке твердження, яке задає неодмінні й достатні умови скінченності твірної функції моментів деякої випадкової величини. Відповідне доведення наведено в Розд. 15.6.

**Твердження 15.2.4.** Твірна функція моментів випадкової величини  $X$  скінчена в деякому околі нуля  $(-s_0; s_0)$ ,  $s_0 > 0$ , тоді й тільки тоді, коли її функція розподілу  $F_X$  експоненційно обмежена, тобто  $\mathbb{P}_X(|X| > x) \leq Ce^{-sx}$  для деяких  $C > 0$ ,  $s > 0$ .

### 15.3. Оцінка Черноффа

У Розд. 11.1.4 ми розглядали нерівності Маркова й Чебишова і зазначали, що друга значно точніша від першої, оскільки окрім існування сподівання вона також вимагає існування дисперсії. Виявляється, можна вивести ще точнішу оцінку ймовірності, що випадкова величина відхиляється від свого сподівання, якщо знати твірну функцію моментів.

**Теорема 15.3.1** (Оцінка Черноффа<sup>1</sup> (Chernoff bound)). Нехай випадкова величина  $X$  має твірну функцію моментів  $M_X$ . Тоді для будь-якого  $t \in \mathbb{R}$  справедливо

$$\mathbb{P}_X (X \geq t) \leq \min_{s>0} e^{-st} M_X(s) . \quad (15.3.1)$$

*Доведення.* Нерівність очевидно виконується, якщо  $M_X(s) = \infty$  для  $s > 0$ . Нехай тоді  $M_X(s) < \infty$  для всіх  $s \in [0; s_0)$ . Тоді за нерівністю Маркова (11.1.6)

$$\mathbb{P}_X (X \geq t) = \mathbb{P}_X (e^{sX} \geq e^{st}) \leq \frac{\mathbb{E}[e^{sX}]}{e^{st}} = e^{-st} M_X(s) .$$

Цілком очевидно, що найточніша оцінка досягається, якщо останній вираз мінімізувати за всіма  $s > 0$ .  $\square$

**Приклад 15.3.2.** Нехай  $X_1, \dots, X_n$  — незалежні випадкові величини з розподілом  $\text{Exp}(\lambda)$ , проте параметр  $\lambda > 0$  невідомий. Нехай ми хочемо з'ясувати, якого розміру  $n$  повинна бути вибірка, щоб середнє вибіркове значення  $\bar{X}$  відхилялося від свого сподівання  $\frac{1}{\lambda}$  на деяке невелике значення з деякою невеликою ймовірністю.

Згадуючи, що

$$\mathbb{E}[\bar{X}] = \mathbb{E}[X_1] = \frac{1}{\lambda} , \quad \text{Var}(\bar{X}) = \frac{\text{Var}(X_1)}{n} = \frac{1}{n\lambda^2} ,$$

можна використати нерівність Чебишова (11.1.7):

$$\mathbb{P}_{\bar{X}} \left( \left| \bar{X} - \frac{1}{\lambda} \right| > \varepsilon \frac{1}{\lambda \sqrt{n}} \right) \leq \frac{1}{\varepsilon^2} .$$

Можна переписати цю ймовірність, помітивши, що

$$\begin{aligned} \mathbb{E}[\lambda \bar{X}] &= \lambda \mathbb{E}[\bar{X}] = 1 , \\ \text{Var}(\lambda \bar{X}) &= \lambda^2 \text{Var}(\bar{X}) = \frac{1}{n^2} . \end{aligned}$$

Справді, матимемо

$$\mathbb{P}_{\bar{X}} \left( \left| \lambda \bar{X} - 1 \right| > \frac{\varepsilon}{\sqrt{n}} \right) \leq \frac{1}{\varepsilon^2} ,$$

або ж

$$\mathbb{P}_{\bar{X}} \left( \left| \lambda \bar{X} - 1 \right| > \varepsilon \right) \leq \frac{1}{n\varepsilon^2} ,$$

Наприклад, якщо ми хочемо, щоб  $\bar{X}$  відхилялося від свого сподівання  $\frac{1}{\lambda}$  більше від 0.01 середньоквадратичних відхилень з імовірністю не більше від 0.05, нам потрібно  $n$  таке, що

$$\frac{1}{n \cdot 0.01^2} \leq 0.05 \quad \Rightarrow \quad n \geq 200\,000 .$$

Аналогічну оцінку можна дістати методом Черноффа. Справді, для  $t < \lambda$  твірна функція

<sup>1</sup>Названа так на честь американського математика Германа Черноффа (Herman Chernoff, нар. 1923), який опублікував її в 1952 р.

моментів кожного окремого  $X_i, i = 1, \dots, n$  дорівнює, згідно з Твердженням 15.1.4,

$$M_{X_i}(t) = \frac{\lambda}{\lambda - t},$$

звідки за Твердженням 15.1.3, для  $t < \lambda$ ,

$$M_{\sum_{i=1}^n X_i}(t) = \left( \frac{\lambda}{\lambda - t} \right)^n.$$

Остаточно за Твердженням 15.1.2 маємо

$$M_{\lambda \bar{X}} = M_{\sum_{i=1}^n X_i} \left( \frac{\lambda}{n} t \right) = \left( \frac{\lambda}{\lambda - \frac{\lambda}{n} t} \right)^n = \left( \frac{1}{1 - \frac{t}{n}} \right)^n.$$

За Теоремою 15.3.1 випливає оцінка Черноффа для відхилення  $\lambda \bar{X}$  від свого сподівання, 1. Справді,

$$\mathbb{P}_{\bar{X}}(\lambda \bar{X} - 1 > \varepsilon) = \mathbb{P}_{\bar{X}}(\lambda \bar{X} > \varepsilon + 1) \leq \min_{s>0} e^{-s(\varepsilon+1)} \left( \frac{1}{1 - \frac{s}{n}} \right)^n.$$

Виконати мінімізацію можна за допомогою умови першого порядку, проте в такому записі її вивести не дуже просто. Натомість можемо взяти логарифм від цього виразу, помітивши, що оскільки логарифм є строго зростаючою функцією, то його мінімум досягатиметься в тій же точці. Відтак потрібно мінімізувати вираз

$$\min_{s>0} \left( -s(\varepsilon + 1) - n \ln \left( 1 - \frac{s}{n} \right) \right).$$

Умова першого порядку дорівнює

$$-(\varepsilon + 1) + \frac{n}{1 - \frac{s}{n}} \cdot \frac{1}{n} = 0,$$

звідки  $s = \frac{n\varepsilon}{1+\varepsilon}$ . Остаточно оцінка Черноффа дорівнює

$$\mathbb{P}_{\bar{X}}(\lambda \bar{X} - 1 > \varepsilon) \leq e^{-\frac{n\varepsilon}{1+\varepsilon} \cdot (\varepsilon+1)} \left( \frac{1}{1 - \frac{1}{n} \cdot \frac{n\varepsilon}{1+\varepsilon}} \right)^n = e^{-n\varepsilon} (1 + \varepsilon)^n.$$

З іншого боку,

$$\mathbb{P}_{\bar{X}}(\lambda \bar{X} - 1 < -\varepsilon) = \mathbb{P}_{\bar{X}}(-\lambda \bar{X} > \varepsilon - 1) \leq \min_{s>0} e^{-s(\varepsilon-1)} \left( \frac{1}{1 + \frac{s}{n}} \right)^n,$$

оскільки  $M_{-X}(t) = M_X(-t)$ .

Логаритмуючи, маємо вираз

$$\min_{s>0} \left( -s(\varepsilon - 1) - n \ln \left( 1 + \frac{s}{n} \right) \right).$$

Умова першого порядку дорівнює

$$-(\varepsilon - 1) - \frac{n}{1 + \frac{s}{n}} \cdot \frac{1}{n} = 0,$$

звідки  $s = \frac{n\varepsilon}{1-\varepsilon}$ . Цей вираз додатний, тільки якщо  $\varepsilon < 1$ . Остаточно оцінка Черноффа в цьому випадку дорівнює

$$\mathbb{P}_{\bar{X}}(\lambda\bar{X} - 1 < -\varepsilon) \leq e^{-\frac{n\varepsilon}{1-\varepsilon}(\varepsilon-1)} \left( \frac{1}{1 + \frac{1}{n} \cdot \frac{n\varepsilon}{1-\varepsilon}} \right)^n = e^{n\varepsilon}(1-\varepsilon)^n.$$

Якщо ж  $\varepsilon \geq 1$ , то показник експоненти в оцінці Черноффа буде від'ємний, і мінімум буде досягатися в  $s = \infty$ , і в цьому випадку оцінка буде, що ймовірність дорівнює 0. Це цілком очікувано, адже якщо  $\varepsilon \geq 1$ , то нас цікавить імовірність, що додатна випадкова величина буде менше від'ємного числа, а така подія неймовірна.

Остаточно оцінка Черноффа дорівнює сумі двох оцінок:

$$\mathbb{P}_X(|\lambda\bar{X} - 1| > \varepsilon) \leq \begin{cases} e^{-n\varepsilon}(1+\varepsilon)^n + e^{n\varepsilon}(1-\varepsilon)^n, & \varepsilon < 1, \\ e^{-n\varepsilon}(1+\varepsilon)^n, & \varepsilon \geq 1 \end{cases}.$$

Зокрема, якщо ми хочемо, щоб  $\bar{X}$  відхилялося від свого сподівання  $\frac{1}{\lambda}$  більше від 0.01 середньо-квадратичних відхилень з імовірністю не більше від 0.05, нам потрібно  $n$  таке, що

$$e^{-0.01n} \cdot 1.01^n + e^{0.01n} \cdot 0.99^n \leq 0.05.$$

Це можна розв'язати хіба що чисельно і дістати  $n \geq 73\,780$ . Як можна бачити, розмір вибірки повинен бути більш ніж удвічі менший, ніж випливає з нерівності Чебишова. Отже оцінка Черноффа є значно точнішою, але платою за це є потреба знати твірну функцію моментів відповідних випадкових величин.  $\square$

## 15.4. Єдиність твірної функції моментів

Властивості, наведені в Твердженнях 15.1.2 та 15.1.3, як такі дають змогу тільки обчислювати моменти лінійних перетворень та сум незалежних випадкових величин. Значно більшу користь відповідні твердження мали б, якби їх можна було застосувати для *визначення розподілів* утворюваних у такий спосіб випадкових величин. Але для цього потрібно мати гарантію, що одна й та ж твірна функція моментів не відповідає декільком різним розподілам, адже тоді неможливо буде встановити, який саме розподіл має, скажімо, сума деяких незалежних випадкових величин.

Виявляється, така гарантія існує, але в дешто специфічному сенсі. Приймімо спочатку без доведення таку теорему.

**Теорема 15.4.1.** Нехай випадкова величина  $X$  має розподіл  $\mathbb{P}_X$ , а всі її моменти довільних порядків існують і є скінченні. Якщо ступеневий ряд (відносно  $r$ )

$$\sum_{k=1}^{\infty} \frac{\mathbb{E}[X^k] r^k}{k!}$$

має додатний радіус збіжності, то  $\mathbb{P}_X$  — це *єдиний* розподіл, який має моменти  $\mathbb{E}[X^k]$ ,  $k = 1, 2, \dots$

Розподіл, який задовольняє умови Теореми 15.4.1, має назву *розподілу, який визначають його моменти* (distribution determined by its moments).

**Приклад 15.4.2.** Нехай  $X \sim N(0, 1)$ . У Прикладі 11.4.4 ми вивели формули для обчислення моментів  $X$  довільних порядків. Нескладно бачити, що  $|\mathbb{E}[X^k]| \leq k!$ ,  $k \in \mathbb{N}$ . А відтак маємо ступеневий ряд

$$\sum_{k=1}^{\infty} \frac{\mathbb{E}[X^k] r^k}{k!} \leq \sum_{k=1}^{\infty} r^k,$$

який є збіжний для всіх  $0 \leq r < 1$ . Відтак можемо стверджувати, що нормальній розподіл повністю визначають його моменти.

Цього не можна сказати про логнормальний розподіл. Справді, нехай  $Y$  має стандартний логнормальний розподіл, тобто  $X = \ln Y \sim N(0, 1)$  і

$$f_1(y) = \frac{1}{\sqrt{2\pi}y} e^{-\frac{(\ln y)^2}{2}} \cdot \mathbb{1}\{y \geq 0\}.$$

У Прикладі 15.2.3 ми показали, що існують і є скінченні всі моменти довільного порядку:

$$\mathbb{E}[Y^k] = e^{\frac{1}{2}k^2}.$$

Розгляньмо тепер випадкову величину  $W$ , яка має таку щільність:

$$f_2(w) = f_1(w)(1 + \sin(2\pi \ln w)) \cdot \mathbb{1}\{w \geq 0\}.$$

Тоді нескладно обчислити відповідні моменти:

$$\begin{aligned} \mathbb{E}[W^k] &= \int_0^{\infty} w^k f_1(w)(1 + \sin(2\pi \ln w)) dw \\ &= \mathbb{E}[Y^k] + \int_0^{\infty} w^k f_1(w) \sin(2\pi \ln w) dw \\ &= \mathbb{E}[Y^k] + \int_{-\infty}^{\infty} e^{uk+k^2} \cdot \frac{1}{\sqrt{2\pi}} e^{-(u+k)} e^{-\frac{(u+k)^2}{2}} \sin(2\pi u) e^{u+k} du \\ &= \mathbb{E}[Y^k] + \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{u^2}{2}} \sin(2\pi u) du = \mathbb{E}[Y^k], \end{aligned}$$

оскільки останній інтеграл дорівнює 0 як інтеграл від непарної функції за симетричним проміжком<sup>2</sup>. Відтак випадкова величина  $W$  має ті ж моменти, що й величина  $Y$ , хоча їхні щільності суттєво відмінні (Рис. 15.4.1). □

Отже можемо зробити висновок, що можуть існувати різні розподіли, які мають однакові моменти всіх порядків. Проте за певних умов твірна функція моментів все ж таки однозначно задає відповідний їй розподіл.

**Теорема 15.4.3.** Якщо твірні функції моментів двох випадкових величин  $X$  та  $Y$  є скінченні та дорівнюють одна одній на деякому відкритому інтервалі  $(-s_0, s_0)$ , що містить 0, то  $X \stackrel{d}{=} Y$ .

**Доведення.** Оскільки твірні функції моментів скінченні на деякому відкритому інтервалі  $(-s_0, s_0)$ , що містить 0, то згідно з Теоремою 15.2.1 існують і є скінченні моменти довільних порядків і для

<sup>2</sup>Можна легко показати, що на нескінченостях підінтегральна функція прямує до нуля.

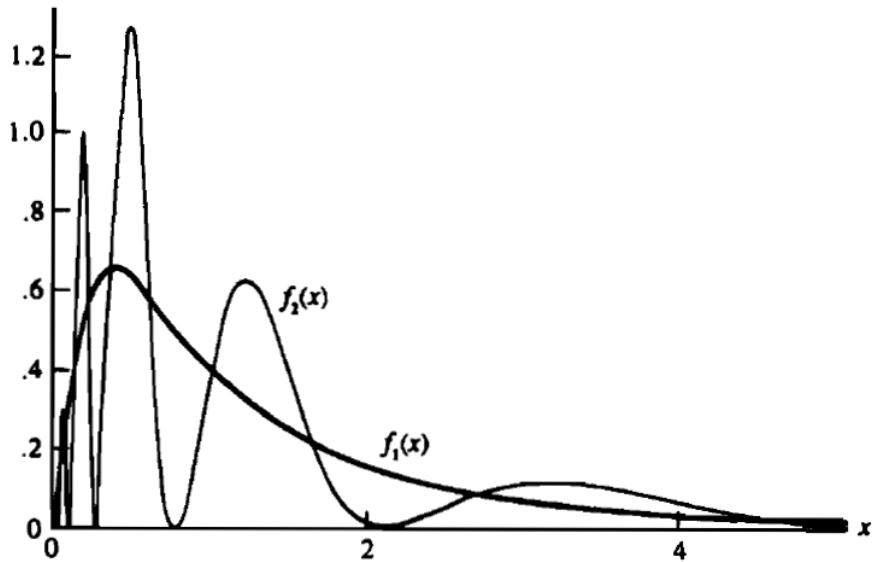


Рис. 15.4.1.: Ілюстрація щільностей  $f_1$  для випадкової величини  $Y$  та  $f_2$  для випадкової величини  $W$  з Прикладу 15.4.2 ([12], Рис. 2.3.2)

величини  $X$ , і для величини  $Y$ . Але оскільки твірні функції моментів дорівнюють одна одній, то дорівнюють і всі моменти однакових порядків  $\mathbb{E}[X^k] = \mathbb{E}[Y^k]$ ,  $k \geq 1$ .

Згідно з (15.2.2),

$$M_X(s) = \sum_{k=0}^{\infty} \frac{s^k \mathbb{E}[X^k]}{k!},$$

а отже виконуються всі вимоги Теореми 15.4.1. Відтак розподіли  $\mathbb{P}_X$  та  $\mathbb{P}_Y$  однозначно задають їхні відповідні моменти, але оскільки моменти однакових порядків однакові, то й розподіли повинні бути однакові, тобто  $X \stackrel{d}{=} Y$ .  $\square$

**Приклад 15.4.4.** Використовуючи Твердження 15.1.3 та Теорему 15.4.3, можемо повторно довести результати з Твердження 9.5.4.

Зокрема, якщо  $X \sim \text{Pois}(\lambda_1)$ ,  $Y \sim \text{Pois}(\lambda_2)$ ,  $X \perp\!\!\!\perp Y$ , то з урахуванням Твердження 15.1.4 маємо

$$M_{X+Y}(t) = M_X(t)M_Y(t) = e^{\lambda_1(e^t-1)}e^{\lambda_2(e^t-1)} = e^{(\lambda_1+\lambda_2)(e^t-1)},$$

звідки очевидно випливає, що  $X + Y \sim \text{Pois}(\lambda_1 + \lambda_2)$ .

А, наприклад, якщо  $X_1 \sim N(\mu_1, \sigma_1^2)$ ,  $X_2 \sim N(\mu_2, \sigma_2^2)$ ,  $X \perp\!\!\!\perp Y$ , то

$$M_{X_1+X_2}(t) = M_{X_1}(t)M_{X_2}(t) = e^{t\mu_1 + \frac{1}{2}\sigma_1^2 t^2}e^{t\mu_2 + \frac{1}{2}\sigma_2^2 t^2} = e^{t(\mu_1+\mu_2) + \frac{1}{2}(\sigma_1^2 + \sigma_2^2)t^2},$$

звідки очевидно випливає, що  $X + Y \sim N(\mu_1 + \mu_2, \sigma_1^2 + \sigma_2^2)$ .  $\square$

## 15.5. Поняття про характеристичну функцію

**Визначення 15.5.1.** *Характеристичною функцією* (characteristic function) деякої випадкової величини  $X$  є

$$\varphi_X(t) = \mathbb{E}[e^{itX}] = \int e^{itX} d\mathbb{P}, \quad (15.5.1)$$

де  $i$  — *уявна одиниця* (imaginary unit), тобто  $i^2 = -1$ .  $\square$

Як і в контексті твірних функцій моментів, можемо говорити про характеристичну функцію не окремої випадкової величини, а цілого розподілу. Якщо з контексту зрозуміло, про яку саме величину  $X$  мова, то відповідний індекс опускатимемо.

**Зауваження 15.5.2.** Якщо випадкова величина  $X$  має щільність  $f_X$ , то її характеристична функція

$$\varphi_X(t) = \int_{-\infty}^{\infty} e^{itx} f_X(x) dx$$

є нічим іншим, як загальновідомим *перетворенням Фур'є* (Fourier transform)<sup>3</sup>.

До речі, твірна функція моментів також за межами теорії ймовірностей відома під іншим іменем — *перетворення Лапласа* (Laplace transform)<sup>4</sup>.  $\square$

Як можна бачити з визначення, характеристична функція дуже подібна до твірної функції моментів, єдина відмінність полягає в присутності уявної одиниці. Для того, щоб розглянути деякі властивості характеристичних функцій, згадаймо основні властивості комплексних чисел.

### 15.5.1. Комплексні числа та їхні властивості

Згадаймо, що *комплексним числом* (complex number) називають число

$$z = a + ib, \quad a, b \in \mathbb{R}, \quad i^2 = -1. \quad (15.5.2)$$

Число  $a$  називають *дійсною частиною* (real part) комплексного числа і позначають  $\Re(z)$ , а число  $b$  — *уявною частиною* (imaginary part) і позначають  $\Im(z)$ .

Корисною властивістю уявної одиниці є

$$\frac{1}{i} = -i.$$

Комплексне число  $z$  має геометричну інтерпретацію: його можна зобразити на координатній площині, де вісь абсцис відповідає дійсній частині, а вісь ординат — уявній частині. Таку площину називають комплексною площинною, вісь абсцис — дійсною, а вісь ординат — уявною.

Будь-яке комплексне число  $z$  має *спряжене* (complex conjugate) до нього, яке позначають через  $\bar{z} = a - ib$ .

<sup>3</sup>Назване так на честь французького математика Жозефа Фур'є (Jean-Baptiste Joseph Fourier, 1768–1830), який уперше згадав про нього в 1822 р.

<sup>4</sup>Назване так на честь французького математика П'єра-Симона де Лапласа (Pierre-Simon, marquis de Laplace, 1749–1827).

Комплексне число  $z$  також можна подати в тригонометричній формі, здійснивши перехід від прямокутної системи координат до полярної:

$$z = r(\cos \varphi + i \cdot \sin \varphi), \quad (15.5.3)$$

де  $r$  називають *модулем* (modulus) і позначають  $|z|$ , а  $\varphi$  — *аргументом* (argument) і позначають  $\arg z$ . Серед іншого, можна помітити, що

$$r = |z| = \sqrt{a^2 + b^2}. \quad (15.5.4)$$

Нарешті, третьою формою запису комплексного числа є показникова форма, яка випливає з *формули Ойлера* (Euler's formula)<sup>5</sup>:

$$e^{i\varphi} = \cos \varphi + i \cdot \sin \varphi. \quad (15.5.5)$$

Тому експоненційна форма комплексного числа має вид

$$z = r e^{i\varphi}. \quad (15.5.6)$$

### 15.5.2. Властивості характеристичних функцій

Розглянемо деякі очевидні властивості характеристичних функцій.

**Твердження 15.5.3.** Нехай  $\varphi$  — характеристична функція деякої величини  $X$ . Тоді

- (i)  $\varphi(0) = 1$ ;
- (ii)  $|\varphi(t)| \leq 1$  для всіх  $t \in \mathbb{R}$ ;
- (iii)  $\varphi(-t) = \overline{\varphi(t)}$ ;
- (iv) для довільних  $a, b \in \mathbb{R}$  випадкова величина  $Y = aX + b$  має характеристичну функцію  $\varphi_Y(t) = e^{itb}\varphi_X(at)$ ;
- (v)  $\varphi$  рівномірно неперервна на  $\mathbb{R}$ ;
- (vi)  $\varphi$  невід'ємно визначена, тобто для будь-якого  $n \in \mathbb{N}$  і будь-яких  $t_1, \dots, t_n \in \mathbb{R}$  і  $c_1, \dots, c_n \in \mathbb{C}$

$$\sum_{k=1}^n \sum_{m=1}^n \varphi(t_k - t_m) c_k \bar{c}_m \geq 0.$$

*Доведення.* (i) Очевидно, адже  $e^{iX \cdot 0} = 1$ ;

(ii) випливає з нерівності Єнсена (11.1.2), формули Ойлера (15.5.5), формули для модуля (15.5.4) та основної тригонометричної тотожності:

$$|\varphi(t)| = |\mathbb{E}[e^{itX}]| \leq \mathbb{E}[|e^{itX}|] = \mathbb{E}[|\cos tX + i \cdot \sin tX|] = \mathbb{E}\left[\sqrt{\cos^2 tX + \sin^2 tX}\right] = 1;$$

(iii) очевидно випливає з формули Ойлера (15.5.5), адже  $e^{-iz} = \overline{e^{iz}}$ ;

<sup>5</sup>Названа так на честь швейцарського математика Леонарда Ойлера (Leonhard Euler, 1707–1783), який опублікував її в 1748 р.

(iv) це можна показати безпосереднім обчисленням:

$$\varphi_Y(t) = \mathbb{E} \left[ e^{it(aX+b)} \right] = \mathbb{E} \left[ e^{itb} \cdot e^{iatX} \right] = e^{itb} \varphi_X(at);$$

(v) без доведення;

(vi) без доведення.

□

Виявляється, що деякі з наведених вище умов є достатніми для того, щоб функція була характеристичною.

**Теорема 15.5.4** (Теорема Бохнера-Хінчина (Bochner's theorem)<sup>6</sup>). Функція  $\varphi$  буде характеристичною функцією деякого розподілу тоді й тільки тоді, коли вона невід'ємно визначена, неперервна і  $\varphi(0) = 1$ .

Той факт, що  $|\varphi(t)| \leq 1$  для всіх  $t \in \mathbb{R}$  означає, що, на відміну від твірних функцій моментів, характеристичні функції визначені *завжди*. Понад те, усі ключові властивості твірних функцій моментів справедливі й для характеристичних функцій, але без жодних додаткових умов на кшталт скінченності в деякому околі точки 0. Платою за потужність відповідних результатів є вміння працювати з функціями комплексної змінної, тому ми розглянемо ці результати без доведення.

**Твердження 15.5.5.** Якщо для випадкової величини  $X$  існує твірна функція моментів  $M_X$ , скінчена в деякому інтервалі  $(-s_0; s_0)$ , що містить 0, то її характеристична функція дорівнює  $\varphi_X(t) = M_X(it)$ .

Зокрема, характеристичні функції деяких відомих розподілів дорівнюють, для всіх  $t \in \mathbb{R}$ :

- (i) якщо  $X \sim \text{Bern}(p)$ , то  $\varphi_X(t) = 1 + p(e^{it} - 1)$ ;
- (ii) якщо  $X \sim \text{Binom}(n, p)$ , то  $\varphi_X(t) = (1 + p(e^{it} - 1))^n$ ;
- (iii) якщо  $X \sim \text{Geom}(p)$ , то  $\varphi_X(t) = \frac{p}{1 - (1-p)e^{it}}$ ;
- (iv) якщо  $X \sim \text{Pois}(\lambda)$ , то  $\varphi_X(t) = e^{\lambda(e^{it} - 1)}$ ;
- (v) якщо  $X \sim \text{U}((a; b))$ , то  $\varphi_X(t) = \frac{e^{itb} - e^{ita}}{it(b-a)}$ ;
- (vi) якщо  $X \sim N(\mu, \sigma^2)$ , то  $\varphi_X(t) = e^{it\mu - \frac{1}{2}\sigma^2 t^2}$ ;
- (vii) якщо  $X \sim \text{Exp}(\lambda)$ , то  $\varphi_X(t) = \frac{\lambda}{\lambda - it}$ .

Як і для твірних функцій моментів, можемо вивести формулу для обчислення характеристичної функції добутку незалежних випадкових величин.

**Твердження 15.5.6.** Нехай  $X$  та  $Y$  — дві незалежні випадкові величини. Тоді

$$\varphi_{X+Y}(t) = \varphi_X(t)\varphi_Y(t). \quad (15.5.7)$$

<sup>6</sup>Названа так на честь австрійського математика Саломона Бохнера (Salomon Bochner, 1899–1982) і радянського математика Олександра Хінчина (1894–1959).

*Доведення.* Це можна довести безпосередньо з визначення:

$$\varphi_{X+Y}(t) = \mathbb{E} [e^{it(X+Y)}] = \mathbb{E} [e^{itX} e^{itY}] = \mathbb{E} [e^{itX}] \mathbb{E} [e^{itY}] = \varphi_X(t) \varphi_Y(t),$$

де ми врахували незалежність  $X \perp\!\!\!\perp Y$  та використали Теорему 9.4.11.  $\square$

Цілком очевидно, що за індукцією можна довести відповідний результат для суми  $n$  незалежних величин:

$$\varphi_{X_1+\dots+X_n}(t) = \varphi_{X_1}(t) \dots \varphi_{X_n}(t).$$

Одне з корисних застосувань характеристичних функцій на практиці полягає в обчисленні моментів довільних порядків. У цьому сенсі вони виконують ту ж роль, що й твірні функції моментів, із тією відмінністю, що характеристичні функції існують для будь-яких розподілів.

Відповідний результат сформульовано в наступній теоремі, яку приймемо без доведення, оскільки воно використовує факти з диференціювання та інтегрування функцій комплексної змінної, які виходять за межі нашого курсу.

**Теорема 15.5.7.** Нехай випадкова величина  $X$  має момент порядку  $k$ . Тоді для всіх  $0 \leq j \leq k$  характеристична функція має неперервну похідну  $j$ -ого порядку, яка дорівнює

$$\varphi_X^{(j)}(t) = \mathbb{E} [(iX)^j e^{itX}].$$

Зокрема,

$$\varphi_X^{(j)}(0) = i^j \mathbb{E} [X^j]. \quad (15.5.8)$$

На відміну від твірних функцій моментів, які однозначно задають відповідні їм розподіли тільки за виконання певних умов, *коєсна* характеристична функція однозначно відповідає конкретному розподілу.

**Теорема 15.5.8** (Формула обернення (Inversion formula)). Якщо випадкова величина  $X$  має функцію розподілу  $F_X$  і характеристичну функцію  $\varphi_X$ , то для  $a$  і  $b$ , у яких  $F_X(a) = F_X(b) = 0$  справедливо

$$F_X(b) - F_X(a) = \lim_{T \rightarrow \infty} \frac{1}{2\pi} \int_{-T}^T \frac{e^{-ita} - e^{-itb}}{it} \varphi_X(t) dt. \quad (15.5.9)$$

Різні розподіли не можуть мати однакової характеристичної функції.

Також можна показати сильніший результат.

**Твердження 15.5.9.** Якщо характеристична функція  $\varphi_X$  випадкової величини  $X$  інтегровна, тобто

$$\int_{-\infty}^{\infty} |\varphi_X(t)| dt < \infty,$$

то  $X$  має щільність

$$f_X(x) = \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{-itx} \varphi_X(t) dt. \quad (15.5.10)$$

## 15.6. Доведення окремих тверджень

Доведімо Твердження 15.2.4.

*Доведення.* Нехай  $M_X(s) < \infty$  для деякого  $s > 0$ . Тоді нерівність Маркова (11.1.6) дає оцінку

$$\mathbb{P}_X(X > x) = \mathbb{P}_X(e^{tX} > e^{tx}) \leq \frac{\mathbb{E}[e^{tX}]}{e^{tx}} = e^{-tx} M_X(t).$$

Поклавши  $C = M_X(t)$  й  $b = s$ , бачимо, що функція розподілу експоненційно обмежена.

Нехай тепер існують такі  $C > 0$ ,  $s > 0$ , що виконується  $\mathbb{P}_X(X > x) \leq Ce^{-sx}$ . Тоді для деякого  $t > 0$  можемо застосувати властивість

$$\mathbb{E}[X] = \int_0^\infty (1 - F(t)) dt, \quad (15.6.1)$$

справедливу для  $X \geq 0$ . Справді, для будь-якого невід'ємного  $x \geq 0$  маємо

$$x = \int_0^x dt = \int_0^\infty \mathbb{1}\{x > t\} dt,$$

відтак

$$X = \int_0^\infty \mathbb{1}\{X > t\} dt.$$

Візьмімо сподівання з обох боків та використаймо той факт, що сподівання — це також інтеграл:

$$\mathbb{E}[X] = \mathbb{E}\left[\int_0^\infty \mathbb{1}\{X > t\} dt\right] = \int_0^\infty \mathbb{E}[\mathbb{1}\{X > t\}] dt = \int_0^\infty \mathbb{P}_X(X > t) dt = \int_0^\infty (1 - F(t)) dt.$$

Відтак у нашому випадку

$$\mathbb{E}[e^{tX}] = \int_0^\infty \mathbb{P}_X(e^{tX} > y) dy \leq 1 + \int_1^\infty \mathbb{P}_X(e^{tX} > y) dy \leq 1 + \int_1^\infty Cy^{-\frac{s}{t}} dy.$$

Узявши будь-яке  $0 < t < s$ , бачимо, що інтеграл буде скінчений, відтак  $M_X(t)$  буде скінченою для таких  $t$ .  $\square$

## 16. Центральна гранична теорема

У цій лекції ми розглянемо один із найважливіших результатів у теорії ймовірностей — центральну граничну теорему (ЦГТ). Деяко спрощуючи, можемо стверджувати, що сума «великої» кількості випадкових величин має приблизно нормальній розподіл. Зокрема, це пояснює, чому так часто на практиці розподіл даних має вигляд нормального. Власне, тому нормальній розподіл і називають нормальним, що він зустрічається повсюдно.

Звісно, якщо ми будемо додавати нескінченну кількість разів деякі загальні числа, то ми можемо дістати й нескінченність. Тому ЦГТ застосовна в тому випадку, коли випадкові величини, які додаються, є малими з великою ймовірністю (щонайменше, якщо їхні дисперсії скінчені, але цього також може бути замало). Два типові приклади, які можна зустріти на практиці — вимірювання деякого процесу та аналіз статистичних даних, зібраних про деяких осіб або про деякі компанії. В обох випадках існує безліч факторів, які неможливо врахувати, але які впливають на точність вимірювання або на якість оцінювання статистичних параметрів на основі наявних даних:

- вимірювання процесу може супроводжуватися похибками як об'єктивного характеру (механічні пошкодження приладу чи атмосферні явища), так і суб'єктивного (поганий зір дослідника, його втомлюваність тощо);
- статистичні дані про деяких осіб, наприклад, для аналізу впливу деяких ліків на перебіг хвороби, можуть містити стати пацієнтів, їхній вік, параметри здоров'я до вживання нових ліків, параметри після вживання нових ліків і т.п. Вочевидь, які б якіні не були дані, врахувати абсолютно все неможливо, або з технічних причин (наприклад, неможливо провести досконале дослідження всіх медичних показників), або з концептуальних (наприклад, неможливо врахувати абсолютно все генетичні особливості кожного пацієнта, адже вони не є досліджені в повному обсязі).

Звісно, якби ці невраховані фактори були абсолютно довільні, то ті показники, які ми могли б дістати, не мали б жодного сенсу. Але якщо самі по собі впливи цих факторів не є великими, то їх сума може розглядатися як випадкова величина з нормальним розподілом.

Для того, щоб коректно сформулювати результат ЦГТ, потрібно увести ще один вид збіжності послідовності випадкових величин.

### 16.1. Збіжність за розподілом

#### 16.1.1. Визначення та інтерпретація

**Визначення 16.1.1.** Послідовність функцій розподілу  $F_n$  збігається до функції розподілу  $F$  слабко (converges weakly), що позначають як  $F_n \Rightarrow F$ , якщо

$$\lim_{n \rightarrow \infty} F_n(x) = F(x) \quad (16.1.1)$$

для всіх точок  $x$ , у яких  $F$  неперервна. □

Іншими словами, маємо збіжність послідовності випадкових величин  $X_1, X_2, \dots$  у тому розумінні, що їхні функції розподілів прямають до деякої граничної функції розподілу. Або, кажучи неформально, можемо говорити про «приблизний» розподіл  $X_n$  для деякого «великого»  $n$ .

Вимога про неперервність є критичною, що видно з такого дуже простого прикладу. Нехай  $X$  має функцію розподілу  $F$ , а члени послідовності дорівнюють  $X_n = X + \frac{1}{n}$ ,  $n = 1, 2, \dots$ . Тоді

$$F_n(x) \equiv \mathbb{P}_X \left( X + \frac{1}{n} \leq x \right) = F \left( x - \frac{1}{n} \right).$$

Якщо  $n \rightarrow \infty$ , то  $F_n(x) \rightarrow F(x-)$ , тобто прямує до лівої границі в точці  $x$ . Відтак збіжність матиме сенс тільки тоді, коли  $F(x-) = F(x)$ , тобто якщо  $x$  є точкою неперервності  $F$ .

Для дослідження відповідних приблизних розподілів часто здійснюють деяке нормування випадкових величин.

**Приклад 16.1.2.** Нехай маємо процес Пуассона, де подіями можуть бути, наприклад, дзвінки на кол-центр. Нехай інтенсивність процесу дорівнює  $\lambda$  (дзвінків на хвилину). Розгляньмо максимальний часовий проміжок  $M_n$  серед часових проміжків між початком відліку та подією 1, між подією 1 та подією 2 і т.д. до події  $n$ . Оскільки всі часові проміжки мають одинаковий експоненційний розподіл і є незалежні, маємо

$$F_{M_n}(x) = \left(1 - e^{-\lambda x}\right)^n \cdot \mathbb{1}\{x \geq 0\}.$$

Для кожного  $x \in \mathbb{R}$  послідовність  $F_{M_n}(x)$ , вочевидь, збігається до 0. Іншими словами,  $M_n \rightarrow \infty$ , оскільки вся маса ймовірностей стає все більше сконцентровано навколо все більших значень.

З іншого боку, розгляньмо

$$\mathbb{P}_{M_n} \left( M_n - \frac{\ln n}{\lambda} \leq x \right) = F_{M_n} \left( x + \frac{\ln n}{\lambda} \right).$$

Ця функція розподілу має таку границю:

$$F_{M_n} \left( x + \frac{\ln n}{\lambda} \right) = \left(1 - e^{-(\lambda x + \ln n)}\right)^n \rightarrow e^{-e^{-\lambda x}},$$

що випливає з другої важливої границі. Отже для великих  $n$  маємо приблизний розподіл нормалізованої випадкової величини  $M_n - \frac{\ln n}{\lambda}$ .  $\square$

**Приклад 16.1.3.** Вимога щодо збіжності в точках неперервності з  $F$  у Визначенні 16.1.1 посутня. Справді, нехай  $X_1, X_2, \dots$  незалежні і мають одинаковий дискретний розподіл із функцією ймовірності  $p_X(1) = p_X(-1) = \frac{1}{2}$ . Нехай  $S_n = X_1 + \dots + X_n$ . Оскільки  $\mathbb{E}[S_n] = 0$ , за ЗВЧ маємо, що для всіх  $\varepsilon > 0$

$$\mathbb{P}_{S_n} \left( \left| \frac{S_n}{n} \right| > \varepsilon \right) \rightarrow 0.$$

Нехай  $F_n$  — функція розподілу  $n^{-1}S_n$ . Якщо  $x > 0$ , то

$$F_n(x) = 1 - \mathbb{P}_{S_n} \left( \frac{S_n}{n} > x \right) \rightarrow 1.$$

Якщо ж  $x < 0$ , то

$$F_n(x) \leq \mathbb{P}_{S_n} \left( \left| \frac{S_n}{n} \right| \geq |x| \right) \rightarrow 0.$$

Якщо ми розглянемо функцію розподілу

$$F(x) = \begin{cases} 0, & x < 0 \\ 1, & x \geq 0 \end{cases},$$

то ми помітимо, що в нашому випадку  $F_n \Rightarrow F$ , адже збіжність виконується в усіх точках неперервності  $F$ . Відтак ми робимо висновок, що ЗВЧ еквівалентний слабкій збіжності функцій розподілу арифметичних середніх до  $F$ .

Якщо  $n$  непарне,  $S_n = 0$  ніколи не може статися. Тоді події  $S_n \leq 0$  і  $S_n \geq 0$  мають однакову ймовірність, а відтак  $F_n(0) = \frac{1}{2}$ . Отже  $F_n(0)$  ніяк не збігається до  $F(0) = 1$ . Але оскільки  $F$  у точці 0 не є неперервною, це зовсім не впливає на той факт, що  $F_n \Rightarrow F$ .  $\square$

Отже, як можна бачити, вимога про збіжність тільки в точках неперервності граничної функції  $F$  дає змогу розглядати ЗВЧ в контексті слабкої збіжності. Проте природно виникає питання, чи є слабка границя єдиною. Наприклад, може бути таке, що  $F_n \Rightarrow F$  і  $F_n \Rightarrow G$ , якщо  $F(x) = G(x)$  в усіх точках своєї неперервності, але якщо вони різняться в усіх інших точках. Проте ми зазначали в Твердженні 6.1.8, що у функції розподілу кількість точок розриву може бути не більш ніж зліченою, а це значить, що множина точок розриву є щільною в множині дійсних чисел, і тому  $F$  і  $G$  є однаковими, оскільки обидві є неперервними справа. Отже слабка границя, якщо вона існує, є єдина.

У контексті послідовностей випадкових величин слабка збіжність має окрему назву.

**Визначення 16.1.4.** Послідовність випадкових величин  $X_n$  збігається до випадкової величини  $X$  за розподілом (converges in distribution), що позначають як  $X_n \Rightarrow X$  або  $X_n \xrightarrow{d} X$ , якщо  $F_n \Rightarrow F$  для відповідних функцій розподілу.

Якщо замість випадкових величин ми маємо послідовність випадкових векторів  $\mathbf{X}_1, \mathbf{X}_2, \dots$ , то  $\mathbf{X}_n \xrightarrow{d} \mathbf{X}$ , якщо виконується слабка збіжність для спільних функцій розподілу (а не маржинальних окремо!).  $\square$

Як правило, на практиці нас цікавить доведення слабкої збіжності функцій розподілу для деякої послідовності випадкових величин, і відповідний результат найпростіше сформулювати в термінах саме випадкових величин, а не їхніх розподілів. Так, наприклад, у Прикладі 16.1.2 ми можемо казати, що  $M_n - \frac{\ln n}{\lambda} \xrightarrow{d} X$ , де  $X$  має відповідну функцію розподілу. Ця величина є штучною і розглядається тільки для того, щоб можна було компактно записати відповідний результат.

На відміну від видів збіжності, які ми розглядали до цього, збіжність за розподілом можна сформулювати навіть для випадкових величин, які визначено на різних імовірнісних просторах. Ми оперуємо тільки функціями розподілу, а природа походження випадкових величин для нас не-цікава. Вочевидь, це зовсім неможливо для, скажімо, збіжності за ймовірністю, адже ймовірність відхилення  $|X_n - X|$  має сенс, тільки якщо обидві функції діють з одного простору.

### 16.1.2. Зв'язок з іншими видами збіжності

**Твердження 16.1.5.** Якщо  $X_n \xrightarrow{p} X^1$ , то  $X_n \xrightarrow{d} X$ .

*Доведення.* Нехай  $X_n \xrightarrow{p} X$ . Тоді для кожного  $\varepsilon > 0$  маємо

$$\lim_{n \rightarrow \infty} \mathbb{P}_{X, X_n} (|X_n - X| > \varepsilon) = 0.$$

Нехай  $F_n$  і  $F$  — функції розподілів випадкових величин  $X_n$  та  $X$  відповідно. Тоді

$$\begin{aligned} F_n(x) &= \mathbb{P}_{X_n} (X_n \leq x) = \mathbb{P}_{X_n, X} (X_n \leq x, X \leq x + \varepsilon) + \mathbb{P}_{X_n, X} (X_n \leq x, X > x + \varepsilon) \\ &\leq \mathbb{P}_X (X \leq x + \varepsilon) + \mathbb{P}_{X_n, X} (X_n - X \leq x - X, X - x > \varepsilon) \\ &= F(x + \varepsilon) + \mathbb{P}_{X_n, X} (X_n - X \leq x - X, x - X < -\varepsilon) \\ &\leq F(x + \varepsilon) + \mathbb{P}_{X_n, X} (X_n - X < -\varepsilon) \\ &\leq F(x + \varepsilon) + \mathbb{P}_{X_n, X} (X_n - X < -\varepsilon) + \mathbb{P}_{X_n, X} (X_n - X > \varepsilon) \\ &= F(x + \varepsilon) + \mathbb{P}_{X_n, X} (|X_n - X| > \varepsilon). \end{aligned}$$

Можна провести аналогічні викладки, показавши

$$F(x - \varepsilon) \leq F_n(x) + \mathbb{P}_{X_n, X} (|X_n - X| > \varepsilon).$$

Об'єднуючи ці дві нерівності, маємо

$$F(x - \varepsilon) - \mathbb{P}_{X_n, X} (|X_n - X| > \varepsilon) \leq F_n(x) \leq F(x + \varepsilon) + \mathbb{P}_{X_n, X} (|X_n - X| > \varepsilon).$$

Нехай  $n \rightarrow \infty$ . Оскільки ми не можемо бути впевнені, що існують границі відповідних послідовностей, ми розглянемо нижні та верхні границі. Оскільки виконується збіжність за ймовірністю, то якраз для ймовірностей і нижня, і верхня границі будуть дорівнювати 0. Враховуючи, що нижня границя не перевищує верхньої, маємо

$$F(x - \varepsilon) \leq \liminf_{n \rightarrow \infty} F_n(x) \leq \limsup_{n \rightarrow \infty} F_n(x) \leq F(x + \varepsilon).$$

Оскільки нас цікавить збіжність тільки в точках неперервності  $F$ , а тому, якщо  $x$  є такою точкою, можемо взяти з обох боків  $\varepsilon \downarrow 0$ , і тоді з обох боків будемо мати  $F(x)$ , що означатиме рівність нижньої й верхньої границь, а отже що  $F_n(x) \rightarrow F(x)$ .  $\square$

У загальному випадку зі збіжності за розподілом збіжність за ймовірністю не випливає (особливо якщо величини визначено на різних просторах!). Це було відображенено на Рис. 13.3.1.

**Приклад 16.1.6.** Нехай маємо дві незалежні величини Бернуллі  $X, Y \sim \text{Bern}(0.5)$ . Тоді нехай  $X_1 = Y, X_2 = Y, \dots$ . Вочевидь, суть формально,  $X_n \xrightarrow{d} Y$ , проте  $X_n$  не прямує до  $Y$  за ймовірністю, адже  $\mathbb{P}_{X, Y} (|X - Y| > \varepsilon) = \mathbb{P}_{X, Y} (|X - Y| = 1) = \frac{1}{2}$ , і до 0 це зовсім не прямує.  $\square$

Проте існує частковий випадок, коли зворотне твердження також виконується, навіть якщо випадкові величини визначено на різних просторах.

**Твердження 16.1.7.** Якщо  $X_n \xrightarrow{d} c, c \in \mathbb{R}$ , то  $X_n \xrightarrow{p} c$ .

<sup>1</sup>І всі випадкові величини визначено на одному й тому ж імовірнісному просторі.

*Доведення.* Якщо гранична величина є виродженою і фактично є константою  $c$ , то її функція розподілу дорівнює

$$F_c(x) = \mathbb{P}_c(c \leq x) = \mathbb{1}\{x \in [c, \infty)\}.$$

Тоді

$$\begin{aligned} \mathbb{P}_{X_n}(|X_n - c| > \varepsilon) &= \mathbb{P}_{X_n}(X_n > c + \varepsilon) + \mathbb{P}_{X_n}(X_n < c - \varepsilon) \\ &= 1 - \mathbb{P}_{X_n}(X_n \leq c + \varepsilon) + \mathbb{P}_{X_n}(X_n < c - \varepsilon) \\ &\leq 1 - F_n(c + \varepsilon) + F_n(c - \varepsilon). \end{aligned}$$

Оскільки  $F_n \Rightarrow F_c$  у всіх точках  $x \neq c$  (де має місце розрив), то  $\mathbb{P}_{X_n}(|X_n - c| > \varepsilon) \rightarrow 0$  в усіх таких точках. Відтак ми довели, що  $X_n \xrightarrow{p} c$ .  $\square$

**Приклад 16.1.8.** Нехай  $X_n \sim N(0, \frac{1}{n})$ ,  $n = 1, 2, \dots$ . Тоді можна показати, що  $X_n \xrightarrow{d} 0$ . Справді, переходячи до стандартної нормальної величини, маємо

$$F_n(x) = \mathbb{P}_{X_n}(X_n < x) = \mathbb{P}_Z\left(Z \leq \frac{x}{\sqrt{n}}\right) = \Phi(\sqrt{n}x).$$

Якщо  $x < 0$ ,  $F_n(x) \rightarrow 0$ . Якщо ж  $x > 0$ ,  $F_n(x) \rightarrow 1$ . Іншими словами, граничною величиною для  $X_n$  є константа 0:  $X_n \xrightarrow{d} 0$ . Зверніть увагу, що в точці  $x = 0$  маємо  $F_n(0) \rightarrow \frac{1}{2}$ , проте ця точка не є точкою неперервності для «функції розподілу» нуля, а тому нею можна знехтувати.

Згідно з Твердженням 16.1.7,  $X_n \xrightarrow{p} 0$ . Власне, це можна було б показати явно, адже для будь-якого  $\varepsilon > 0$  через симетрію маємо

$$\mathbb{P}_{X_n}(|X_n - 0| > ep) = 2\mathbb{P}_{X_n}(X_n > \varepsilon) = 2(1 - \Phi(\sqrt{n}\varepsilon)) = 0.$$

$\square$

Для збіжності за розподілом також справедлива теорема про неперервне відображення.

**Теорема 16.1.9** (Теорема про неперервне відображення (Continuous mapping theorem)). Нехай маємо послідовність випадкових векторів  $\mathbf{X}_n \xrightarrow{d} \mathbf{X}$ . Нехай  $g: \mathbb{R}^k \rightarrow \mathbb{R}^n$  неперервна майже напевно. Тоді  $g(\mathbf{X}_n) \xrightarrow{d} g(\mathbf{X})$ .

Також у статистиці, економетриці та інших прикладних галузях широко використовують інші один схожий результат.

**Теорема 16.1.10** (Теорема Слуцького (Slutsky's Theorem)<sup>2</sup>). Нехай маємо дві послідовності випадкових величин  $X_n \xrightarrow{d} X$  і  $Y_n \xrightarrow{p} c$ . Тоді  $X_n + Y_n \xrightarrow{d} X + c$ ,  $X_n Y_n \xrightarrow{d} Xc$ , і, якщо  $c \neq 0$ ,  $\frac{X_n}{Y_n} \xrightarrow{d} \frac{X}{c}$ .

Для випадкових векторів існує дуже корисний результат, який дає змогу встановити відповідність між збіжністю за розподілом вектора та його координат.

<sup>2</sup>Названа так на честь українського та радянського математика, економіста та статистика Євгена Слуцького (1880–1948).

**Теорема 16.1.11** (Теорема Крамера-Волда (Cramér–Wold theorem)<sup>3</sup>). Неодмінною є достатньою умовою збіжності за розподілом  $\mathbf{X}_n = (X_{n1}, \dots, X_{nk})^\top \xrightarrow{d} \mathbf{X} = (X_1, \dots, X_k)$  є збіжність за розподілом  $\sum_{j=1}^k t_j X_{nj} \xrightarrow{d} \sum_{j=1}^k t_j X_j$  для всіх  $(t_1, \dots, t_k)^\top \in \mathbb{R}^k$ .

### 16.1.3. Зв'язок із характеристичними функціями

Для того, щоб довести ЦГТ, нам потрібно буде використати допоміжний результат, який дає змогу спростити перевірку збіжності за розподілом. Ми його викладемо в такому формулуванні<sup>4</sup>.

**Теорема 16.1.12** (Теорема Леві (Lévy's theorem)<sup>5</sup>). Нехай маємо послідовність випадкових величин  $X_1, X_2, \dots$  із відповідними характеристичними функціями  $\varphi_1, \varphi_2, \dots$ . Нехай випадкова величина  $X$  має характеристичну функцію  $\varphi$ . Тоді:

- якщо  $X_n \xrightarrow{d} X$ , то  $\varphi_n(t) \rightarrow \varphi(t)$  поточково;
- якщо  $\varphi_n(t) \rightarrow g(t)$  поточково, і  $g$  неперервна в нулі, то  $X_n \xrightarrow{d} X$ , а  $g$  є характеристичною функцією для  $X$ .

Доводити цю теорему ми не можемо, але згадати про неї ми повинні були, адже вона грає важливу роль у повноцінному доведенні різних версій ЦГТ. Власне, наявність цього теоретичного результату є чи не найважливішою причиною, чому в принципі розглядають таке поняття, як характеристична функція.

**Приклад 16.1.13.** Нехай  $X_n \sim U((-n; n))$ ,  $n = 1, 2, \dots$ . Згідно з Твердженням 15.5.5, за формулою Ойлера (15.5.5) для  $t \neq 0$  маємо

$$\varphi_n(t) = \frac{e^{itn} - e^{-itn}}{2itn} = \frac{\cos tn + i \sin tn - (\cos tn - i \sin tn)}{2itn} = \frac{\sin tn}{tn}.$$

У точці  $t = 0$  всі характеристичні функції дорівнюють 1. Якщо ж  $t \neq 0$ , то коли  $n \rightarrow \infty$ , ми маємо збіжність до 0. Отже остаточно  $\varphi_n \rightarrow \mathbb{1}\{t = 0\}$  поточково. Проте така функція не є неперервною в нулі, а отже *не можна казати*, що  $X_n$  збігається за розподілом до деякого  $X$ .  $\square$

## 16.2. Центральна гранична теорема для незалежних однаково розподілених величин

### 16.2.1. Формулування та приклади

З усіх версій ЦГТ, які доведені в літературі, ми можемо дозволити собі розглянути доведення тільки найпростішої — коли випадкові величини незалежні і мають однаковий розподіл. Інші версії ми просто зафіксуємо без доведення.

<sup>3</sup>Названа так на честь шведського математика та статистика Гаральда Крамера (Harald Cramér, 1893–1985) та норвезького економетриста й статистика Германа Волда (Herman Ole Andreas Wold, 1908–1992).

<sup>4</sup>Існують і дещо відмінні.

<sup>5</sup>Названа так на честь французького математика Поля Леві (Paul Pierre Lévy, 1886–1971).

**Теорема 16.2.1** (Центральна гранична теорема Ліндеберга-Леві (Lindeberg-Lévy central limit theorem)<sup>6</sup>). Нехай маємо послідовність  $X_1, X_2, \dots$  незалежних величин з однаковим розподілом зі скінченими сподіванням  $\mu$  та дисперсією  $\sigma^2$ . Тоді

$$\frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \xrightarrow{d} Z \sim N(0, 1). \quad (16.2.1)$$

*Доведення.* Доведення ЦГТ базується на використанні теореми Леві 16.1.12. Але оскільки ми не вміємо працювати з характеристичними функціями, ми розглянемо тільки варіант, коли всі випадкові величини мають твірні функції моментів, визначені у відкритому околі нуля. Справді, оскільки в цьому випадку характеристичні функції і твірні функції моментів тісно пов'язані ( $M_{X_n}(it) = \varphi_{X_n}(t)$ ), теорема Леві буде справедлива.

Отже доведення ЦГТ побудуємо в такому ключі. Якщо довести, що твірні функції моментів випадкових величин прямують до твірної функції моментів стандартного нормальногорозподілу, то ми в такий спосіб доведемо збіжність за розподілом (16.2.1).

Нехай для простоти  $\mu = 0$ ,  $\sigma^2 = 1$ . Тоді відповідно до Твердження 15.1.2,

$$M_{X_n/\sqrt{n}}(t) = M_{X_n}\left(\frac{t}{\sqrt{n}}\right).$$

А за Твердженням 15.1.3, з урахуванням незалежності величин із послідовності,

$$M_{\sum_{k=1}^n X_k/\sqrt{n}}(t) = \left(M_{X_n}\left(\frac{t}{\sqrt{n}}\right)\right)^n.$$

Розгляньмо  $L(t) = \ln M_{X_n}(t)$ . Цілком очевидно, що  $L(0) = \ln M_{X_n}(0) = \ln 1 = 0$ . Також можна бачити, що з Теореми 15.2.1 випливає

$$\begin{aligned} L'(0) &= \frac{M'_{X_n}(0)}{M_{X_n}(0)} = \mu = 0, \\ L''(0) &= \frac{M_{X_n}(0)M''_{X_n}(0) - (M'_{X_n}(0))^2}{(M_{X_n}(0))^2} = \mathbb{E}[X_n^2] = 1. \end{aligned}$$

Отже нам потрібно показати, що

$$nL\left(\frac{t}{\sqrt{n}}\right) \rightarrow \frac{t^2}{2}$$

поточково, адже  $M_Z(t) = e^{\frac{t^2}{2}}$ , як було встановлено у Твердженні 15.1.4. Справді, застосовуючи

<sup>6</sup>Названа так на честь фінського математика Ярла Ліндеберга (Jarl Waldemar Lindeberg, 1876–1932) та французького математика Поля Леві (Paul Pierre Lévy, 1886–1971).

правило Л'Опіталя двічі, маємо

$$\begin{aligned} \lim_{n \rightarrow \infty} \frac{L\left(\frac{t}{\sqrt{n}}\right)}{n^{-1}} &= \lim_{n \rightarrow \infty} \frac{-L'\left(\frac{t}{\sqrt{n}}\right) n^{-\frac{3}{2}} t}{-2n^{-2}} \\ &= \lim_{n \rightarrow \infty} \frac{L'\left(\frac{t}{\sqrt{n}}\right) t}{2n^{-\frac{1}{2}}} \\ &= \lim_{n \rightarrow \infty} \frac{-L''\left(\frac{t}{\sqrt{n}}\right) n^{-\frac{3}{2}} t^2}{-2n^{-\frac{3}{2}}} \\ &= \lim_{n \rightarrow \infty} L''\left(\frac{t}{\sqrt{n}}\right) \frac{t^2}{2} = \frac{t^2}{2}. \end{aligned}$$

Отже ми довели ЦГТ для випадку  $\mu = 0$ ,  $\sigma^2 = 1$ . Доведення загального випадку тривіальне, якщо розглянути стандартизовані випадкові величини

$$X_i^* = \frac{X_i - \mu}{\sigma},$$

адже  $\mathbb{E}[X^*] = 0$ ,  $\text{Var}(X^*) = 1$ , а

$$\frac{\bar{X} - \mu}{\sigma/\sqrt{n}} = \frac{\bar{X}^*}{1/\sqrt{n}}.$$

□

Деяло неформально можна стверджувати, що для «дуже великого  $n$ » випадкова величина  $\frac{\bar{X} - \mu}{\sigma/\sqrt{n}}$  має приблизно стандартний нормальній розподіл, або, що те саме, що  $\bar{X}$  має приблизно нормальній розподіл  $N\left(\mu, \frac{\sigma^2}{n}\right)$ . Такий розподіл називають *асимптомотичним* (asymptotic distribution), і ми позначатимемо його через  $\stackrel{a}{\sim}$ . Згідно з ПЗВЧ,  $\bar{X} \xrightarrow{\text{М.Н.}} \mu$ , а ЦГТ додатково вказує, який розподіл має  $\bar{X}$  «на шляху до»  $\mu$ .

**Приклад 16.2.2.** Повернімося до класичного прикладу з підкиданням правильної монетки, тобто до послідовності незалежних величин  $X_n \sim \text{Bern}(0.5)$ . Згідно з ПЗВЧ,  $\bar{X} \xrightarrow{\text{М.Н.}} 0.5$ . Проте для будь-якої конкретної вибірки ми можемо сказати значно більше, використавши ЦГТ у формулюванні Теореми 16.2.1. Справді,  $\mathbb{E}[X_n] = 0.5$ ,  $\text{Var}(X_n) = 0.25$ , а відтак

$$\frac{\bar{X} - 0.5}{0.25/n} \xrightarrow{d} N(0, 1),$$

тобто для «великих»  $n$

$$\frac{\bar{X} - 0.5}{0.25/n} \stackrel{a}{\sim} N(0, 1),$$

або ж

$$\bar{X} \stackrel{a}{\sim} N\left(0.5, \frac{0.25}{n}\right).$$

Така додаткова інформація дає підстави стверджувати, наприклад, що для  $n = 100$  середньоквадратичне відхилення дорівнюватиме  $\sigma_{\bar{X}} = 0.05$ , а відтак можемо казати, що згідно з правилом трьох сигм, з імовірністю 95%  $\bar{X} \in [0.5 - 2 \cdot 0.05; 0.5 + 2 \cdot 0.05] = [0.4; 0.6]$ . □

**Приклад 16.2.3.** Для графічної ілюстрації ЦГТ розгляньмо три послідовності незалежних величин:

- $X_1, X_2, \dots \sim \text{Binom}(10, 0.9)$ ;
- $Y_1, Y_2, \dots \sim \text{Beta}(0.8, 0.8)$ ;
- $Z_1, Z_2, \dots \sim \text{Exp}(1)$ .

Згенеруймо по  $N = 10\,000$  разів вибірки з відповідних розподілів розміром  $n = 1, 10, 100$ . На Рис. 16.2.1 наведено гістограми розподілів середніх вибіркових, обчислені для кожної вибірки, та накладені на них щільноті нормального розподілу  $N\left(\mu, \frac{\sigma^2}{n}\right)$ , обчислені з урахуванням сподівання та дисперсії відповідних розподілів. Як можна бачити, зі збільшенням  $n$  розподіл середніх вибіркових все більше наближається до нормального. Проте швидкість збіжності різна, залежно від того, наскільки асиметричним або багатомодальним є відповідний розподіл.

Відповідні графіки було згенеровано за допомогою коду на Лістингу 16.2.1.

Лістинг 16.2.1: Код для ілюстрації ЦГТ

```

import numpy as np
from numpy.random import seed
from scipy import stats
import matplotlib.pyplot as plt
from matplotlib import rc

5

def latex_float(f):
    if abs(f) < 0.01:
        float_str = "{0:.2e}".format(f)
        base, exponent = float_str.split("e")
        return r"\times 10^{{{}}}".format(base, int(exponent))
    else:
        return "{0:.2f}".format(f)

10

def plot_sample_mean_dist(sample, mu, sigma, title, n_bins=100, n_plot=10**3):
    rc('text', usetex=True)
    rc('text.latex', unicode=True)
    rc('text.latex', preamble=r'\usepackage[utf8]{inputenc}')
    rc('text.latex', preamble=r'\usepackage[ukrainian]{babel}')

20

    x = np.linspace(np.min(sample), np.max(sample), n_plot)
    f = stats.norm.pdf(x, loc=mu, scale=sigma)

25

    plt.plot(
        x, f, linewidth=2, label=r"\u041f\u0435\u0437\u043b\u0438\u043d\u0430\u043f\u0438\u0442\u0430\u043d\u0438\u044f \u043d\u0430\u0437\u0430\u043d\u0430\u043b\u0438\u043d\u0430\u043f\u0438\u0442\u0430\u043d\u0438\u044f $N(\mu, \sigma^2/n)$".
        format(mu, latex_float(sigma**2 / n_sample)))
    )

30

    plt.hist(sample, bins=n_bins,
            label=r"\u0413\u0438\u0441\u0442\u0430\u043d\u0430\u043f\u0438\u0442\u0430\u043d\u0438\u044f \u043d\u0430\u0437\u0430\u043d\u0430\u043b\u0438\u043d\u0430\u043f\u0438\u0442\u0430\u043d\u0438\u044f", density=True)
    plt.xlabel(r"$x$", fontsize=15)
    plt.ylabel(r"$f(x)$", fontsize=15)
    plt.xticks(fontsize=10)
    plt.yticks(fontsize=10)
    plt.legend(fontsize=15, loc="upper left")

35

    plt.savefig(
        "../../../images/Lecture 16/clt_{0}_{1}.png".format(title, n_sample),

```

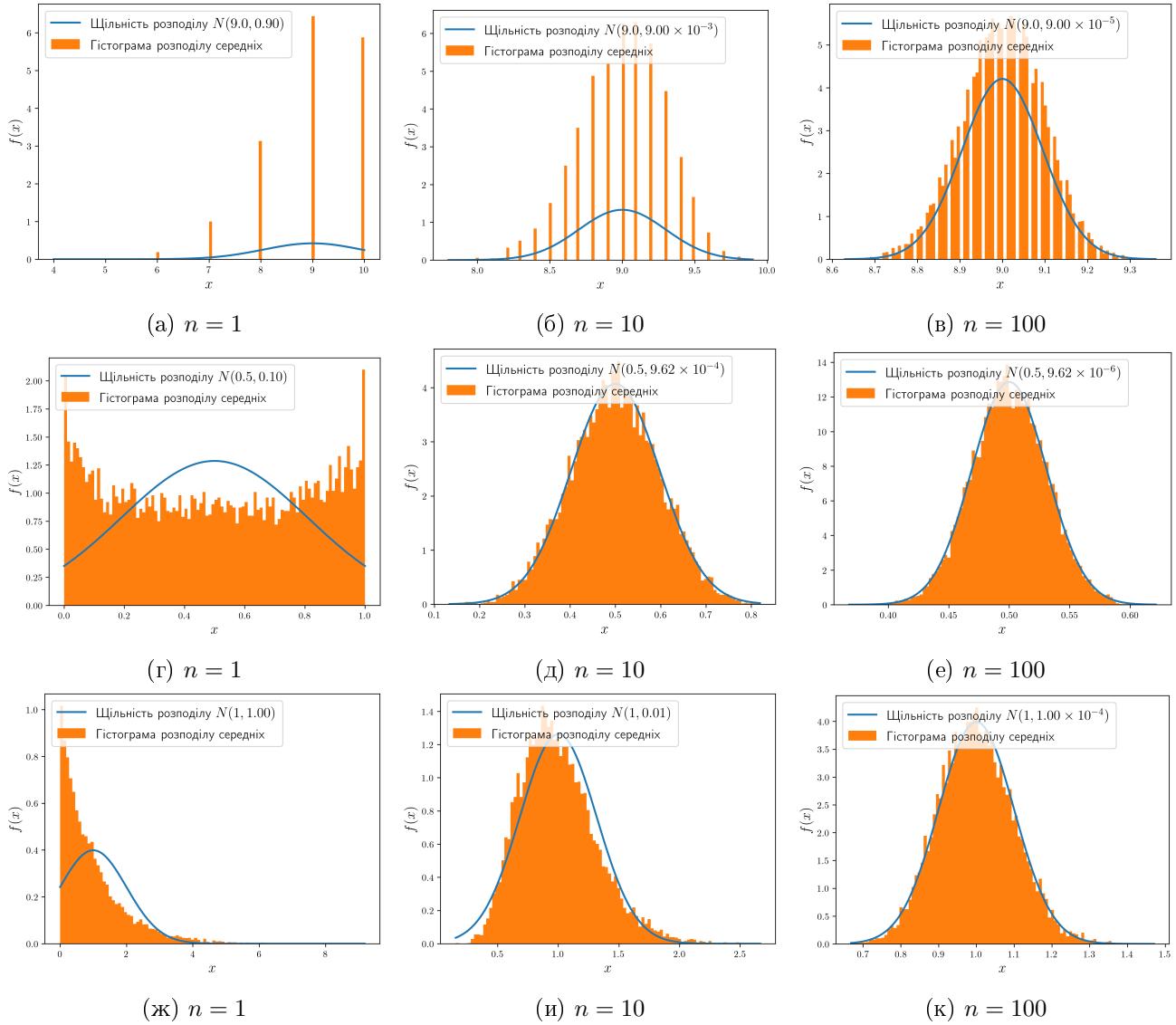


Рис. 16.2.1.: Ілюстрація ЦГТ для Прикладу 16.2.3: (а–в) величини з біноміальним розподілом  $\text{Binom}(10, 0.9)$ ; (г–е) величини з Бета-розподілом  $\text{Beta}(0.8, 0.8)$ ; (ж–к) величини з експоненційним розподілом  $\text{Exp}(1)$

```

40     dpi=300,
41     bbox_inches="tight"
42 )
43 plt.close()

44
45 seed(100)
46 n_samples = [1, 10, 100]
47 N = 10000

48 n_bin = 10
49 p_bin = 0.9
50 a_bet = 0.8
51 b_bet = 0.8
52 lambda_expo = 1

53 for n_sample in n_samples:
54     sample_binom = [
55         stats.binom.rvs(size=n_sample, n=n_bin, p=p_bin).mean() for i in range(N)
56     ]

57     plot_sample_mean_dist(
58         sample_binom,
59         mu=n_bin*p_bin,
60         sigma=np.sqrt(n_bin*p_bin*(1 - p_bin) / n_sample),
61         title="binom"
62     )

63     sample_beta = [
64         stats.beta.rvs(size=n_sample, a=a_bet, b=b_bet).mean() for i in range(N)
65     ]

66     plot_sample_mean_dist(
67         sample_beta,
68         mu=a_bet / (a_bet + b_bet),
69         sigma=np.sqrt(
70             a_bet * b_bet / (a_bet + b_bet) ** 2 /
71             (a_bet + b_bet + 1) / n_sample
72         ),
73         title="beta"
74     )

75     sample_expo = [
76         stats.expon.rvs(size=n_sample).mean() for i in range(N)
77     ]

78     plot_sample_mean_dist(
79         sample_expo,
80         mu=1,
81         sigma=np.sqrt(1 / n_sample),
82         title="expo"
83     )

```

□

Як можна бачити з Теореми 16.2.1, саме правильне нормування випадкових величин уможливлює встановити відповідний результат. Якби замість  $\sqrt{n}$  ми використали, наприклад,  $n$ ,

то  $\frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \xrightarrow{d} 0$ , хоча б тому, що  $\frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \xrightarrow{p} 0$ , що можна легко показати. Відтак коректне нормування є ключовим для використання ЦГТ, що ілюструє такий приклад.

**Приклад 16.2.4.** З'ясуймо асимптотичний розподіл для послідовності незалежних величин  $X_n \sim \text{Binom}(n, p)$ ,  $n = 1, 2, \dots$ ,  $p \in (0; 1)$ . Для того, щоб застосувати ЦГТ у формулюванні Теореми 16.2.1, потрібно подати  $X_n$  як у відповідний спосіб нормоване середнє вибіркове. Справді, використаймо той факт, що  $X_n = \sum_{i=1}^n Y_i$ , де  $Y_i \sim \text{Bern}(p)$ , і всі вони незалежні. Оскільки  $\mathbb{E}[Y_n] = p$ , а  $\text{Var}(Y_n) = p(1-p)$ , маємо, що для послідовності  $\bar{Y}_n$  виконується ЦГТ:

$$\frac{\bar{Y} - p}{\sqrt{p(1-p)/n}} \xrightarrow{d} N(0, 1) ,$$

звідки

$$\frac{\frac{X_n}{n} - p}{\sqrt{p(1-p)/n}} \xrightarrow{a} N(0, 1) ,$$

або ж

$$X_n \xrightarrow{a} N(np, np(1-p)) .$$

У схожий спосіб, помітивши, що сума пуассонівських величин є пуассонівською, можна показати, що для послідовності незалежних  $X_n \sim \text{Pois}(n\lambda)$  асимптотичний розподіл буде

$$\frac{\frac{X_n}{n} - \lambda}{\sqrt{\lambda/n}} \xrightarrow{a} N(0, 1) ,$$

або ж

$$X_n \xrightarrow{a} N(n\lambda, n\lambda) .$$

□

Розгляньмо ще декілька прикладів застосування ЦГТ.

**Приклад 16.2.5.** Астроном хоче виміряти відстань (у світлових роках) до деякої далекої зірки. Він свідомий того, що через змінні атмосферні умови та похибки вимірювального пристрою не варто сподіватися, що результат, який він дістане, буде істинним. Для підвищення впевненості у своїх вимірюваннях астроном виконує заміри багато разів і бере їх середнє значення.

Нехай  $d$  — справжня відстань до зірки. Якщо вважати, що результати  $n$  вимірювань,  $X_1, \dots, X_n$ , незалежні та мають однаковий розподіл зі сподіванням  $d$  та дисперсією 4, то  $\bar{X}$  матиме асимптотичний розподіл

$$\frac{\bar{X} - d}{\sqrt{4/n}} \xrightarrow{a} N(0, 1) .$$

Нехай астроном хоче бути впевнений, що його оцінка відстани відхилятиметься від справжнього  $d$  не більше від 0.5 світлового року з імовірністю, більшою від 95%. Тоді

$$\begin{aligned} \mathbb{P}_{X_1, \dots, X_n}(-0.5 \leq \bar{X} - d \leq 0.5) &= \mathbb{P}_{Z_n}\left(-0.5 \frac{\sqrt{n}}{2} \leq Z_n \leq 0.5 \frac{\sqrt{n}}{2}\right) \\ &\approx \Phi\left(\frac{\sqrt{n}}{4}\right) - \Phi\left(-\frac{\sqrt{n}}{4}\right) = 2\Phi\left(\frac{\sqrt{n}}{4}\right) - 1 , \end{aligned}$$

а отже

$$2\Phi\left(\frac{\sqrt{n}}{4}\right) - 1 \geq 0.95 \quad \Rightarrow \quad \Phi\left(\frac{\sqrt{n}}{4}\right) \geq 0.975 ,$$

звідки чисельно можна визначити  $n = 61.47$ , тобто потрібно здійснити щонайменше  $n = 62$  замірів.

Звісно, для того, щоб використовувати цей результат на практиці, потрібно бути переконаним, що для  $n = 62$  апроксимація ЦГТ є достатньо точною. У загальному випадку це часто так і є, що видно з Рис. 16.2.1. Проте якщо дослідника беруть сумніви, він завжди може використати, наприклад, нерівність Чебишова.  $\square$

**Приклад 16.2.6.** Викладач повинен перевірити 50 екзаменаційних робіт. На кожну роботу, незалежно від інших, потрібно витратити певну кількість часу. Тривалість перевірки однієї роботи має деякий розподіл<sup>7</sup> зі сподіванням 20 (хвилин) та середньоквадратичним відхиленням 4 (хвилини).

Навіть за наявності такої обмеженої інформації можна підрахувати ймовірність того, що, наприклад, за 450 хвилин викладач устигне перевірити щонайменше 25 робіт. Справді, нехай  $X_i = \text{«тривалість перевірки роботи } i\text{»}$ ,  $i = 1, \dots, 25$ . Тоді нас цікавить імовірність того, що  $X = \sum_{i=1}^{25} X_i$  не перевищує 450. За ЦГТ,

$$\frac{\frac{X}{25} - 20}{4/\sqrt{25}} = \frac{X - 500}{20} \stackrel{\text{a}}{\sim} N(0, 1) ,$$

або ж

$$X \stackrel{\text{a}}{\sim} N(500, 400) .$$

Відтак

$$\mathbb{P}_X (X \leq 450) \approx \mathbb{P}_Z \left( Z \leq \frac{450 - 500}{\sqrt{400}} \right) = 1 - \Phi(2.5) \approx 0.006 .$$

$\square$

**Приклад 16.2.7.** Часто для наближеного обчислення факторіалів дуже великих чисел використовують формулу Стирлінга (Stirling's formula)<sup>8</sup>:

$$n! \approx \sqrt{2\pi n} \left( \frac{n}{e} \right)^n . \quad (16.2.2)$$

Вивести цю формулу можна за допомогою ЦГТ в такий спосіб. Нехай  $X_1, X_2, \dots \sim \text{Exp}(1)$ , і всі незалежні. Тоді з (16.2.1) випливає

$$\mathbb{P}_{X_1, \dots, X_n} \left( \frac{\bar{X} - 1}{1/\sqrt{n}} \leq x \right) \rightarrow \mathbb{P}_Z (Z \leq x) .$$

<sup>7</sup>Можливо, експоненційний, але це не точно.

<sup>8</sup>Названа так на честь шотландського математика Джеймса Стирлінга (James Stirling, 1692–1770).

Візьмімо похідні з обох боків, помітивши, що  $W \equiv \sum_{i=1}^n X_i \sim \text{Gamma}(n, 1)$ :

$$\frac{d}{dx} \mathbb{P}_Z(Z \leq x) = \phi(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}},$$

$$\frac{d}{dx} \mathbb{P}_{X_1, \dots, X_n} \left( \frac{\bar{X} - 1}{1/\sqrt{n}} \leq x \right) = f_W(x\sqrt{n} + n) \cdot \sqrt{n} = \frac{1}{\Gamma(n)} (x\sqrt{n} + n)^{n-1} e^{-(x\sqrt{n} + n)} \cdot \sqrt{n}.$$

Зокрема, якщо  $x = 0$ , маємо

$$\frac{1}{\sqrt{2\pi}} \approx \frac{e^{-n} n^{n-\frac{1}{2}}}{\Gamma(n)},$$

або ж

$$n\Gamma(n) \approx \sqrt{2\pi n} \left( \frac{n}{e} \right)^n.$$

Оскільки  $\Gamma(n) = (n-1)!$ , зліва стоїть  $n!$ , отже ми вивели формулу Стирлінга.  $\square$

### 16.2.2. Інтегральна теорема де Муавра-Лапласа

Одним із найпоширеніших часткових випадків Теореми 16.2.1 є теорема, яка визначає асимптотичний розподіл для біномної величини. Ми її вже показали в Прикладі 16.2.4, проте варто сформулювати її так, як це прийнято в літературі.

**Теорема 16.2.8** (Інтегральна теорема де Муавра-Лапласа (DeMoivre-Laplace limit theorem)<sup>9</sup>). Нехай маємо схему Бернуллі з імовірністю успіху  $p \in (0; 1)$ . Тоді для числа успіхів  $X \sim \text{Binom}(n, p)$  для великих  $n$  можна приблизно обчислити таку ймовірність:

$$\mathbb{P}_X \left( a \leq \frac{X - np}{\sqrt{np(1-p)}} \leq b \right) \approx \Phi(b) - \Phi(a). \quad (16.2.3)$$

**Зауваження 16.2.9.** Зверніть увагу, що з урахуванням Теорем 5.1.4 і 16.2.8 ми маємо два різні способи апроксимувати біномні ймовірності. Апроксимація за допомогою розподілу Пуассона працює, коли  $n$  «велике»,  $p$  «мале», а  $np$  залишається сталим. Натомість апроксимація за допомогою нормальногорозподілу працює, коли «великим» є  $np(1-p)$  (на практиці достатньо мати  $np(1-p) \geq 10$ ).  $\square$

**Зауваження 16.2.10.** Оскільки біномний розподіл дискретний, а нормальній — неперервний, апроксимувати ймовірність того, що  $X \sim \text{Binom}(n, p)$  набула деякого конкретного значення,  $\mathbb{P}_X(X = k)$ , нормальним розподілом неможливо, адже вона дорівнює 0. Тому на практиці використовують так звану *корекцію неперервності* (continuity correction) і апроксимують не  $\mathbb{P}_X(X = k)$ , а  $\mathbb{P}_X(k - 0.5 < X < k + 0.5)$ . Таку корекцію корисно застосовувати і для інших дискретних розподілів.  $\square$

<sup>9</sup>Названа так на честь французьких математиків Авраама де Муавра (Abraham de Moivre, 1667–1754) та П'єра-Симона де Лапласа (Pierre-Simon, marquis de Laplace, 1749–1827).

**Приклад 16.2.11.** Практика показує, що на деякий захід справді приходять приблизно 30% запрошених на нього. Бажаючи зібрати приблизно 150 осіб, організатори розіслали запрошення 450 особам. Підрахуймо ймовірність того, що на захід з'являться більше 150 осіб (враховуючи, звісно, що кожен запрошений ухвалює рішення прийти на захід незалежно від інших).

Формально  $X = \text{«число осіб, що з'явилися на захід»}$ , має біномний розподіл  $\text{Binom}(450, 0.3)$ . Проте обчислювати відповідну ймовірність напряму не є можливим, адже для цього потрібно було б рахувати біномні коефіцієнти з дуже великими числами. Натомість можемо скористуватися результатом Теореми 16.2.8 і підрахувати

$$\begin{aligned}\mathbb{P}_X(X > 150) &= \mathbb{P}_X(X \geq 150 + 0.5) = \mathbb{P}_X\left(\frac{150.5 - 450 \cdot 0.3}{\sqrt{450 \cdot 0.3 \cdot 0.7}} \leq \frac{X - 450 \cdot 0.3}{\sqrt{450 \cdot 0.3 \cdot 0.7}} < \infty\right) \\ &\approx \Phi(\infty) - \Phi(1.59) = 1 - \Phi(1.59) \approx 0.056.\end{aligned}$$

□

**Приклад 16.2.12.** Для визначення ефективності нового методу лікування проводять клінічні випробування зі 100 учасниками. Дослідник ухвалить позитивне рішення щодо продовження роботи над новим методом, тільки якщо принаймні 65% пацієнтів відчувають поліпшення свого стану. Нехай новий метод лікування насправді не має жодного ефекту. Порахуймо ймовірність того, що дослідник помилково вирішить, що цей метод заслуговує на дальші випробування.

Якщо метод не має жодного ефекту, кожний пацієнт може одужати з імовірністю 0.5. Тоді  $X = \text{«загальна кількість одужалих пацієнтів»}$  має біномний розподіл  $\text{Binom}(100, 0.5)$ . Відтак шукана ймовірність фактично дорівнює  $\mathbb{P}_X(X \geq 65)$ , або, за Теоремою 16.2.8,

$$\mathbb{P}_X(X \geq 65 - 0.5) \approx \mathbb{P}_X\left(\frac{X - 100 \cdot 0.5}{\sqrt{100 \cdot 0.5 \cdot 0.5}} \geq 2.9\right) \approx 1 - \Phi(2.9) \approx 0.0019.$$

□

### 16.2.3. Дельта-метод

Розгляньмо апарат, за допомогою якого можна встановлювати асимптотичний розподіл функцій від випадкових величин.

**Теорема 16.2.13** (Дельта-метод (Delta method)). Нехай  $X_1, X_2, \dots$  такі, що їхні сподівання та дисперсії кожної з них скінченні й дорівнюють  $\mu$  і  $\sigma^2$  відповідно, і виконується

$$\frac{X_n - \mu}{\sigma/\sqrt{n}} \xrightarrow{d} N(0, 1).$$

Нехай функція  $g(X_n)$  диференційовна, і її похідна в точці  $\mu$  не дорівнює 0. Тоді

$$\frac{g(X_n) - g(\mu)}{\sigma g'(\mu)/\sqrt{n}} \xrightarrow{d} N(0, 1). \quad (16.2.4)$$

*Доведення.* Функцію  $g$  можна розвинути в ряд Тейлора:

$$g(X_n) = g(\mu) + g'(\mu)(X_n - \mu) + (X_n - \mu)R(X_n - \mu),$$

де  $R(X_n)$  — залишковий член ряду, який із визначення похідної можна записати як

$$R(h) = \frac{g(\mu + h) - g(\mu)}{h} - g'(\mu) ,$$

якщо  $x \neq 0$ .

Можна показати, що  $X_n \xrightarrow{p} \mu$ . Справді,

$$\mathbb{P}_{X_n} (|X_n - \mu| < \varepsilon) = \mathbb{P}_{X_n} (|\sqrt{n}(X_n - \mu)| < \sqrt{n}\varepsilon) .$$

Оскільки  $\sqrt{n}(X_n - \mu) \xrightarrow{d} N(0, \sigma^2)$ ,

$$\lim_{n \rightarrow \infty} \mathbb{P}_{X_n} (|\sqrt{n}(X_n - \mu)| < \sqrt{n}\varepsilon) = \Phi(\infty) = 1 .$$

Відтак  $R(X_n - \mu) \xrightarrow{p} 0$  за теоремою про неперервне відображення (оскільки  $g$  неперервна).

Можемо застосувати теорему Слуцького 16.1.10. Справді, оскільки  $\sqrt{n}(X_n - \mu) \xrightarrow{d} N(0, \sigma^2)$ , а  $g'(\theta) \xrightarrow{p} g'(\theta)$ , маємо

$$g'(\theta) \cdot \sqrt{n}(X_n - \mu) \xrightarrow{d} g'(\theta) \cdot N(0, 1) = N\left(0, \sigma^2 (g'(\theta))^2\right) .$$

Тоді маємо

$$\sqrt{n}(g(X_n) - g(\mu)) = g'(\mu) \cdot \sqrt{n}(X_n - \mu) + \sqrt{n}(X_n - \mu) R(X_n - \mu) .$$

Знову ж таки за теоремою Слуцького,

$$\sqrt{n}(X_n - \mu) R(X_n - \mu) \xrightarrow{p} 0 .$$

Остаточно

$$\sqrt{n}(g(X_n) - g(\mu)) = g'(\mu) \cdot \sqrt{n}(X_n - \mu) + \sqrt{n}R(X_n) \xrightarrow{d} N\left(0, \sigma^2 (g'(\theta))^2\right) .$$

□

Найпоширеніше застосування дельта-методу полягає у визначенні асимптотичного розподілу функції від середнього вибіркового.

**Приклад 16.2.14.** Нехай у деяку чергу прибувають клієнти, і на опрацювання  $i$ -го клієнта потрібно витратити  $X_i$  часу. Нехай маємо вибірку незалежних  $X_1, \dots, X_n$ , усі з яких мають сподівання  $\mu$  та дисперсію  $\sigma^2$ . Нехай нас цікавить швидкість опрацювання клієнтів на одиницю часу, яку логічно оцінити за допомогою  $\frac{1}{\bar{X}}$ .

Для цього випадку застосовний дельта-метод, оскільки функція  $g(x) = \frac{1}{x}$  диференційовна для  $x > 0$ , і її похідна не дорівнює нулю в жодній точці  $x > 0$ . Тоді за ЦГТ маємо

$$\frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \xrightarrow{d} N(0, 1) ,$$

а отже виконуються всі умови Теореми 16.2.13. Тоді за (16.2.4) маємо

$$\frac{\frac{1}{\bar{X}} - \frac{1}{\mu}}{\sigma \cdot \left(-\frac{1}{\mu^2}\right) / \sqrt{n}} \xrightarrow{d} N(0, 1) ,$$

звідки

$$\frac{1}{\bar{X}} \xrightarrow{a} N\left(\frac{1}{\mu}, \frac{\sigma^2}{n\mu^4}\right) .$$

□

Якщо  $g'(\mu) = 0$ , можна продовжувати розвивати цю функцію в ряд Тейлора (якщо друга похідна існує і не дорівнює 0 в точці  $\mu$ ), і тоді (16.2.4) набуде вигляду

$$n(g(X_n) - g(\mu)) \xrightarrow{d} \sigma^2 \frac{g''(\mu)}{2} \chi_1^2 .$$

### 16.3. Інші варіанти центральної граничної теореми

Для узагальнення ЦГТ потрібно позбутися деяких умов, викладених у Теоремі 16.2.1. Найпростіше можна позбутися умови про однаковість розподілу випадкових величин у послідовності. Розглянемо послідовність величин  $X_1, X_2, \dots$ , які незалежні і мають скінченні сподівання  $\mathbb{E}[X_n] = \mu_n$  та дисперсії  $\text{Var}(X_n) = \sigma_n^2$ ,  $n = 1, 2, \dots$ . Потрібно помітити, що (16.2.1) можна переписати як

$$\frac{\bar{X} - \mu}{\sigma/\sqrt{n}} = \frac{\sum_{k=1}^n X_k - n\mu}{\sqrt{n}\sigma} = \frac{\sum_{k=1}^n X_k - \sum_{k=1}^n \mu}{\sqrt{\sum_{k=1}^n \sigma_k^2}} \xrightarrow{d} N(0, 1) .$$

Якщо випадкові величини мають різні розподіли, відповідні суми не спрощуються, тому шукаємо граничний розподіл такої випадкової величини:

$$\frac{\sum_{k=1}^n X_k - \sum_{i=k}^n \mu_k}{\sqrt{\sum_{k=1}^n \sigma_k^2}} .$$

Для спрощення дальших викладок можемо вважати, що випадкові величини мають нульові сподівання:  $\mu_k = 0$  для всіх  $k$ .

Для того, щоб вищевказана випадкова величина прямувала за розподілом до стандартної нормальні, на члени послідовності  $X_1, X_2, \dots$  потрібно накласти додаткову умову, щоб відповідна сума не зростала необмежено. Щією умовою є *умова Ліндеберга* (Lindeberg condition), яка для будь-якого  $\varepsilon > 0$  має таке формулювання:

$$\lim_{n \rightarrow \infty} \sum_{k=1}^n \frac{1}{s_n^2} \int_{|X_k| \geq \varepsilon s_n} X_k^2 d\mathbb{P} = 0 , \quad (16.3.1)$$

де  $s_n^2 = \sum_{k=1}^n \sigma_k^2$ .

У чому полягає принада саме такої умови? Зафіксуймо деяке  $\varepsilon > 0$ . Розглянемо ймовірність того, що хоча б одна з величин  $\frac{X_k}{s_n}$  для  $k = 1, \dots, n$  перевищить  $\varepsilon$  за модулем:

$$\mathbb{P}_{X_k} \left( \max_k \{|X_k|\} > \varepsilon s_n \right) = \mathbb{P}(A_1 \cup \dots \cup A_n) ,$$

де  $A_k = \{\omega : |X_k(\omega)| > \varepsilon s_n\}$ ,  $k = 1, \dots, n$ . Оскільки ці події в загальному випадку не є незалежні,

$$\mathbb{P}(A_1 \cup \dots \cup A_n) \leq \sum_{k=1}^n \mathbb{P}(A_k) .$$

З іншого боку,

$$\mathbb{P}(A_k) = \int_{|X_k| \geq \varepsilon s_n} 1 d\mathbb{P} \leq \frac{1}{(\varepsilon s_n)^2} \int_{|X_k| \geq \varepsilon s_n} X_k^2 d\mathbb{P} ,$$

а тому

$$\mathbb{P}_{X_k} \left( \max_k \{|X_k|\} > \varepsilon s_n \right) \leq \frac{1}{\varepsilon^2 s_n^2} \sum_{k=1}^n \int_{|X_k| \geq \varepsilon s_n} X_k^2 d\mathbb{P} .$$

З умови (16.3.1) випливає, що вираз справа прямує до 0 для будь-якого  $\varepsilon > 0$ , отже ймовірність того, що випадкові величини будуть необмежено зростати, прямує до 0.

З урахуванням цих міркувань можна сформулювати таку версію ЦГТ.

**Теорема 16.3.1** (Центральна гранична теорема Ліндеберга (Lindeberg central limit theorem)). Нехай маємо послідовність  $X_1, X_2, \dots$  незалежних величин зі скінченними сподіваннями  $\mathbb{E}[X_n] = \mu_n$  та дисперсіями  $\text{Var}(X_n) = \sigma_n^2$ . Нехай виконується умова Ліндеберга (16.3.1). Тоді

$$\frac{\sum_{k=1}^n (X_k - \mu_k)}{s_n} \xrightarrow{d} Z \sim N(0, 1) , \quad (16.3.2)$$

де  $s_n^2 = \sum_{k=1}^n \sigma_k^2$ .

Деяло відмінним є таке формулювання ЦГТ.

**Теорема 16.3.2** (Центральна гранична теорема Ляпунова (Lyapunov central limit theorem)<sup>10</sup>). Нехай маємо послідовність  $X_1, X_2, \dots$  незалежних величин зі скінченними сподіваннями  $\mathbb{E}[X_n] = \mu_n$  та дисперсіями  $\text{Var}(X_n) = \sigma_n^2$ . Нехай виконується умова Ляпунова (Lyapunov condition)

$$\lim_{n \rightarrow \infty} \sum_{k=1}^n \frac{1}{s_n^{2+\delta}} \mathbb{E} [|X_k|^{2+\delta}] = 0 , \quad (16.3.3)$$

де  $\delta > 0$ . Тоді

$$\frac{\sum_{k=1}^n (X_k - \mu_k)}{s_n} \xrightarrow{d} Z \sim N(0, 1) , \quad (16.3.4)$$

де  $s_n^2 = \sum_{k=1}^n \sigma_k^2$ .

Справедливість цієї теореми можна довести, показавши, що умову Ліндеберга можна обмежити зверху умовою Ляпунова. Ми цього робити не будемо.

**Приклад 16.3.3.** Нехай маємо послідовність незалежних рівномірно обмежених випадкових величин  $X_1, X_2, \dots$  з нульовими сподіваннями. Якщо  $s_n^2 \rightarrow \infty$ , то матимемо, що  $\frac{\sum_{k=1}^n (X_k - \mu_k)}{s_n} \xrightarrow{d} N(0, 1)$ . Справді, нехай  $|X_n| \leq K$  для деякого  $K \in \mathbb{R}$  і для всіх  $n = 1, 2, \dots$ . Тоді

$$\sum_{k=1}^n \frac{1}{s_n^3} \mathbb{E} [|X_k|^3] \leq \sum_{k=1}^n \frac{K \mathbb{E} [X_k^2]}{s_n^3} = \frac{K}{s_n} \rightarrow 0 ,$$

<sup>10</sup>Названа так на честь російського математика Олександра Ляпунова (1857–1918).

а отже виконується умова Ляпунова (16.3.3) для  $\delta = 1$ .  $\square$

Існують і версії ЦГТ, які працюють для випадкових величини, які не є незалежними, але вони виходять за рамки нашого курсу.

Наочанок розгляньмо хіба що варіант ЦГТ для випадкових векторів.

**Теорема 16.3.4.** Нехай маємо послідовність незалежних випадкових векторів  $\mathbf{X}_n = (X_{n1}, \dots, X_{nk})$  з однаковим розподілом,  $n = 1, 2, \dots$ . Нехай  $\mathbb{E}[X_{nj}^2] < \infty$  для всіх  $n, j$ . Нехай  $\mathbb{E}[\mathbf{X}_n] = \boldsymbol{\mu} = (\mu_1, \dots, \mu_k)^\top$  і нехай  $\mathbf{Cov}(\mathbf{X})$  містить елементи  $\sigma_{ij} = \text{Cov}(X_{ni}, X_{nj})$ ,  $i, j = 1, \dots, k$ . Тоді

$$\frac{\sum_{i=1}^n \mathbf{X}_i - n\boldsymbol{\mu}}{\sqrt{n}} \xrightarrow{d} N(0, \Sigma). \quad (16.3.5)$$

## Література

- [1] J. K. Blitzstein and J. Hwang, *Introduction to Probability*, 2nd ed. Boca Raton: Taylor & Francis Group, LLC, 2019, 636 pp.
- [2] G. G. Roussas, *Introduction to Probability*, 2nd ed. Amsterdam: Elsevier, 2014, 548 pp.
- [3] P. Billingsley, *Probability and Measure*, 4th ed. New York: John Wiley & Sons, Inc., 2012, 608 pp.
- [4] G. G. Roussas, *An Introduction to Measure-Theoretic Probability*, 2nd ed. Amsterdam: Elsevier, 2014, 557 pp.
- [5] R. Durrett, *Probability: Theory and Examples*, 4th ed. Cambridge: Cambridge University Press, 2010, 490 pp.
- [6] R. Leadbetter, S. Cambanis, and V. Pipiras, *A Basic Course in Measure and Probability: Theory for Applications*. Cambridge: Cambridge University Press, 2014, 376 pp.
- [7] S. S. Venkatesh, *The Theory of Probability: Explorations and Applications*. Cambridge: Cambridge University Press, 2013, 832 pp.
- [8] S. Ross, *A First Course in Probability*, 10th ed. Harlow: Pearson, 2020, 530 pp.
- [9] J. Pitman, *Probability*. New York: Springer-Verlag, 1993, 564 pp.
- [10] J. S. Rosenthal, *A First Look at Rigorous Probability Theory*, 2nd ed. Singapore: World Scientific Publishing Co., 2006, 237 pp.
- [11] I. Floreescu, *Probability and Stochastic Processes*. Hoboken: John Wiley & Sons, Inc., 2015, 579 pp.
- [12] G. Casella and R. L. Berger, *Statistical inference*, 2nd ed. Pacific Grove, CA: Duxbury, 2002, 686 pp.

# Предметний покажчик

$\chi^2$  distribution, 133, 223

## А

абсолютна неперервність відносно міри, 109  
 абсолютні моменти, 259  
 абсолютні центральні моменти, 261  
 абсолютно неперервна випадкова величина, 110  
 $\sigma$ -адитивна міра, 23  
 алгебра, 13  
 $\sigma$ -алгебра, 16  
 $\sigma$ -алгебра, яку породжує випадковий вектор, 274  
 апостеріорна ймовірність, 40  
 апріорна ймовірність, 40  
 аргумент комплексного числа, 340  
 асимптотична теорія, 290  
 асимптотичний розподіл, 351  
 асоціативність, 6

## Б

багатовимірний нормальний розподіл, 229  
 Беєсівське виведення, 153  
 Беєсівський підхід, 12  
 Бета-розподіл, 150  
 біномний розподіл, 68  
 Борелева  $\sigma$ -алгебра, 19  
 Борелева дійсна вісь, 21  
 Борелева множина, 19  
 Борелева функція, 128  
 Борелів простір, 21

## В

версія умовної ймовірності, 271  
 верхня границя послідовності подій, 290  
 вимірна множина, 16  
 вимірна функція, 55  
 вимірний простір, 16  
 вимірний прямокутник, 166  
 випадкова величина, 56

випадковий вектор, 128

випадковий експеримент, 3

випробування, 11

випробування Бернуллі, 67

від'ємний біномний розподіл, 84

внутрішньогрупова варіація, 283

## Г

Гамма-розподіл, 219  
 геометрична ймовірність, 10  
 геометричний розподіл, 81  
 гіпергеометричний розподіл, 88  
 гістограма, 322  
 гранична теорія, 290

## Д

де Моргана закони, 6  
 дельта-метод, 358  
 дискретна випадкова величина, 57  
 дискретна ймовірнісна міра, 32  
 дискретний імовірнісний простір, 32  
 дисперсія, 64  
 дистрибутивність, 6  
 діаграми Венна, 5  
 дійсна частина комплексного числа, 339  
 добуток мір, 167  
 добуток просторів, 165  
 друга лема Бореля-Кантеллі, 294

## Е

експоненційний розподіл, 157  
 елементарна подія, 4  
 емпірична функція розподілу, 319

## З

з імовірністю 1, 36  
 закон ітерованих сподіван, 280  
 Закон несвідомого статистика, 63  
 закон несвідомого статистика, 117  
 закон повного сподівання, 277  
 закон повної дисперсії, 282

закон повної ймовірності, 46  
 збіжність з імовірністю 1, 297  
 збіжність за ймовірністю, 298  
 збіжність за розподілом, 346  
 збіжність майже напевно, 297  
 збіжність у середньому порядку  $p$ , 300  
 згортка, 192

**I**

індикаторна випадкова величина, 100, 104  
 інтеграл, 103  
 інтеграл за мірою, 103  
 інтеграл Лебега, 103  
 інтегральна теорема де Муавра-Лапласа, 357  
 інтегральне перетворення ймовірностей, 146  
 інтегровна випадкова величина, 108  
 інтегрування методом Монте-Карло, 324

**K**

квантиль, 264  
 квартиль, 265  
 класична ймовірність, 8  
 класичне визначення ймовірності, 8  
 коваріація, 180  
 коефіцієнт асиметрії, 262  
 коефіцієнт ексцесу, 264  
 коефіцієнт кореляції, 182  
 комплексне число, 339  
 комутативність, 6

**Л**

лічна міра, 23  
 логістичний розподіл, 138  
 логнормальний розподіл, 136

**M**

майже напевно, 36  
 майже скрізь, 36  
 маржинальна функція розподілу, 175  
 маржинальна щільність розподілу, 175  
 маржинальний розподіл, 173  
 матриця коваріацій, 183  
 медіана, 253  
 метод моментів, 318  
 міжгрупова варіація, 283  
 міра, 23  
 міра Лебега, 33  
 міри центральної тенденції, 253  
 множина Кантора, 36

множина міри нуль за Лебегом, 35  
 мода, 253  
 модуль комплексного числа, 340  
 моменти, 259  
 монотонно неспадна міра, 25  
 мультиномій розподіл, 225

**H**

напівнормальний розподіл, 217  
 незалежні події, 51  
 незалежність випадкових величин, 185  
 некорельовані випадкові величини, 190  
 неможлива подія, 4  
 неперервна міра, 29  
 неперервність зверху, 29  
 неперервність знизу, 29  
 неперервність у порожній множині, 29  
 непояснена дисперсія, 283  
 нерівність Була, 25  
 нерівність Гольдера, 245  
 нерівність Єнсена, 241  
 нерівність Коші-Буняковського, 247  
 нерівність Маркова, 249  
 нерівність Чебишова, 250  
 несумісні події, 4  
 нижня границя послідовності подій, 290  
 нормальний розподіл, 139  
 носій випадкової величини, 57, 112  
 нульова подія, 35

**O**

образ міри, 56  
 опукла функція, 241  
 оцінка параметра, 317  
 оцінка Чернова, 334

**P**

парадокс Симпсона, 50  
 певна подія, 4  
 перетворення зсуву та масштабування, 137  
 перетворення Лапласа, 339  
 перетворення Фур'є, 339  
 персентиль, 265  
 перша лема Бореля-Кантеллі, 292  
 повторний інтеграл, 167  
 подвійний інтеграл, 167  
 подія, 4  
 подовження міри, 32

поле, 13  
 породжена алгебра, 18  
 породжена  $\sigma$ -алгебра, 18  
 порядкові статистики, 156  
 посилений закон великих чисел, 310  
 поточкова збіжність, 296  
 потужність, 4  
 похідна Радона-Нікодима, 110  
 пояснена дисперсія, 283  
 прообраз, 55, 98  
 проста випадкова величина, 104  
 простір  $L^1$ , 108  
 простір  $L^p$ , 261  
 простір елементарних подій, 4  
 процес Пуассона, 162  
 Пуассонівська парадигма, 79

## P

рівність за розподілом, 57  
 рівномірний розподіл, 66, 138  
 розбиття, 4  
 розподіл  $\chi^2$ , 133, 223  
 розподіл без пам'яти, 83  
 розподіл Бернуллі, 67  
 розподіл випадкової величини, 56  
 розподіл Пуассона, 74  
 розподіл Радемахера, 230  
 розподіл Ст'юдента, 224  
 розподіл Фішера-Сnedекора, 224  
 розподіл, який визначають його моменти, 336

## C

Санкт-Петербурзький парадокс, 243  
 середнє вибіркове значення, 191  
 середня абсолютна похибка, 257, 282  
 середньоквадратична похибка, 256  
 середньоквадратичне відхилення, 64  
 симетричний розподіл, 262  
 сім'я розподілів відносно  
     зсуву-масштабування, 137  
 скінченна міра, 23  
 $\sigma$ -скінченна міра, 23  
 скінченно-адитивна міра, 24  
 слабка збіжність, 344  
 слабкий закон великих чисел, 309  
 спадної граничної корисності, 244  
 спільна функція ймовірності, 172

спільна функція розподілу, 171  
 спільна щільність розподілу, 172  
 спільний розподіл, 171  
 сподівання, 61, 116, 179  
 спряжене комплексне число, 339  
 спряжений апріорний розподіл, 153  
 стандартний нормальний розподіл, 124  
 стандартний рівномірний розподіл, 122  
 стандартний розподіл Коші, 118  
 $\sigma$ -субадитивна міра, 25  
 схема Бернуллі, 68

## T

таблиця спряженості, 177  
 твірна функція моментів, 327  
 теорема Беєса, 46  
 теорема Боннера-Хінчина, 341  
 Теорема Гайне-Бореля, 38  
 теорема Глівенко-Кантеллі, 319  
 теорема Каратеодорі, 32  
 теорема Крамера-Волда, 349  
 теорема Леві, 349  
 теорема про добуток мір, 167  
 теорема про мажоровану збіжність, 241  
 теорема про монотонну збіжність, 106, 240  
 теорема про неперервне відображення, 299, 348  
 теорема Радона-Нікодима, 110  
 теорема Слуцього, 348  
 теорема Фубіні, 167  
 теорія великих вибірок, 290  
 теорія ймовірностей, 2  
 теорія сподіваної корисності, 243  
 тотожність Волда, 280  
 трикутний розподіл, 193

## У

умова Ліндеберга, 360  
 умовна дисперсія, 282  
 умовна ймовірність, 40, 270, 274  
 умовна медіана, 282  
 умовна функція ймовірності, 93, 266  
 умовна функція розподілу, 266, 287  
 умовна щільність розподілу, 283  
 умовне сподівання, 275, 277  
 уявна одиниця, 339  
 уявна частина комплексного числа, 339

**Ф**

- формула заміни змінних, 134  
 формула заміни змінних для випадкових векторів, 213  
 формула Стирлінга, 356  
 формула обернення для характеристичних функцій, 342  
 функція від множини, 23  
 функція ймовірності, 59  
 функція корисності, 244  
 функція надійності, 160  
 функція розподілу, 98

**Х**

- характеристична функція, 339

**Ц**

- центральна гранична теорема Ліндеберга, 361  
 центральна гранична теорема Ліндеберга-Леві, 350  
 центральні моменти, 262

**Ч**

- частотний підхід, 11

**ІІІ**

- щільність розподілу, 110

**А**

- absolute central moments, 261  
 absolute moments, 259  
 absolutely continuity with respect to measure, 109  
 absolutely continuous random variable, 110  
 $\sigma$ -additive measure, 23  
 algebra, 13  
 $\sigma$ -algebra, 16  
 $\sigma$ -algebra generated by a random vector, 274  
 almost everywhere, 36  
 almost surely, 36  
 argument of a complex number, 340  
 associativity, 6  
 asymptotic distribution, 351  
 asymptotic theory, 290

**В**

- Bayes' Theorem, 46  
 Bayesian approach, 12

- Bayesian inference, 153  
 Bernoulli distribution, 67  
 Bernoulli scheme, 68  
 Bernoulli trial, 67  
 Beta distribution, 150  
 between-group variation, 283  
 Binomial distribution, 68  
 Bochner's theorem, 341  
 Boole's inequality, 25  
 Borel  $\sigma$ -algebra, 19  
 Borel function, 128  
 Borel real line, 21  
 Borel set, 19  
 Borel space, 21

**С**

- Cantor set, 36  
 Carathéodory theorem, 32  
 cardinality, 4  
 Cauchy-Schwarz inequality, 247  
 central moments, 262  
 certain event, 4  
 change of variables formula, 134  
 change of variables formula for random vectors, 213  
 characteristic function, 339  
 Chebyshev's inequality, 250  
 Chernoff bound, 334  
 classical probability, 8  
 commutativity, 6  
 complex conjugate, 339  
 complex number, 339  
 conditional distribution function, 266, 287  
 conditional expectation, 275, 277  
 conditional median, 282  
 conditional probability, 40, 270, 274  
 conditional probability density, 283  
 conditional probability mass function, 93, 266  
 conditional variance, 282  
 conjugate prior, 153  
 contingency table, 177  
 continuous mapping theorem, 299, 348  
 continuity at the empty set, 29  
 continuity from above, 29  
 continuity from below, 29  
 continuous measure, 29  
 convergence almost surely, 297  
 convergence in distribution, 346

convergence in probability, 298  
 convergence in  $p$ th mean, 300  
 convergence with probability 1, 297  
 convex function, 241  
 convolution, 192  
 correlation, 182  
 counting measure, 23  
 covariance, 180  
 covariance matrix, 183  
 Cramér–Wold theorem, 349

**D**

de Morgan's laws, 6  
 delta method, 358  
 DeMoivre-Laplace limit theorem, 357  
 diminishing marginal utility, 244  
 discrete probability measure, 32  
 discrete probability space, 32  
 discrete random variable, 57  
 distributivity, 6  
 distribution determined by its moments, 336  
 distribution function, 98  
 distribution of a random variable, 56  
 dominated convergence theorem, 241  
 double integral, 167

**E**

empirical distribution function, 319  
 equality in distribution, 57  
 estimator, 317  
 expectation, 61, 116, 179  
 expected utility theory, 243  
 explained variance, 283  
 exponential distribution, 157  
 extension of a measure, 32

**F**

$F$ -distribution, 224  
 field, 13  
 finite measure, 23  
 $\sigma$ -finite measure, 23  
 finitely additive measure, 24  
 first Borel-Cantelli lemma, 292  
 Fisher–Snedecor distribution, 224  
 Fourier transform, 339  
 frequentist approach, 11  
 Fubini theorem, 167  
 $F$ -розподiл, 224

**G**

Gamma distribution, 219  
 generated algebra, 18  
 generated  $\sigma$ -algebra, 18  
 geometric distribution, 81  
 geometric probability, 10  
 Glivenko-Cantelli, 319

**H**

half-normal distribution, 217  
 Heine-Borel Theorem, 38  
 histogram, 322  
 Hölder's inequality, 245  
 hypergeometric distribution, 88

**I**

imaginary part of a complex number, 339  
 imaginary unit, 339  
 impossible event, 4  
 independent events, 51  
 independent random variables, 185  
 indicator random variable, 100, 104  
 induced measure, 56  
 integrable random variable, 108  
 integral, 103  
 integral with respect to a measure, 103  
 inverse image, 55, 98  
 inversion formula for characteristic functions, 342  
 iterated integral, 167

**J**

Jensen's inequality, 241  
 joint distribution, 171  
 joint distribution function, 171  
 joint probability density function, 172  
 joint probability mass function, 172

**K**

kurtosis, 264

**L**

$L^1$  space, 108  
 $L^p$  space, 261  
 Laplace transform, 339  
 large sample theory, 290  
 law of iterated expectations, 280  
 law of total expectation, 277  
 Law of total probability, 46

law of total variance, 282  
 Law of unconscious statistician, 63  
 law of unconscious statistician, 117  
 Lebesgue integral, 103  
 Lebesgue measure, 33  
 Lebesgue measure zero set, 35  
 Lévy's theorem, 349  
 limit inferior of a sequence of events, 290  
 limit superior of a sequence of events, 290  
 limit theory, 290  
 Lindeberg central limit theorem, 361  
 Lindeberg condition, 360  
 Lindeberg-Lévy central limit theorem, 350  
 location-scale distribution family, 137  
 location-scale transformation, 137  
 logistic distribution, 138  
 log-normal distribution, 136

**M**

marginal distribution, 173  
 marginal distribution function, 175  
 marginal probability density function, 175  
 Markov's inequality, 249  
 mean absolute error, 257, 282  
 mean squared error, 256  
 measurable function, 55  
 measurable rectangle, 166  
 measurable set, 16  
 measurable space, 16  
 measure, 23  
 measures of central tendency, 253  
 median, 253  
 memoryless distribution, 83  
 method of moments, 318  
 mode, 253  
 modulus of a complex number, 340  
 moment generating function, 327  
 moments, 259  
 monotone convergence theorem, 106, 240  
 monotonically nondecreasing, 25  
 Monte Carlo integration, 324  
 multinomial distribution, 225  
 multivariate normal distribution, 229  
 mutually exclusive events, 4

**N**

naive probability, 8  
 negative Binomial distribution, 84

normal distribution, 139  
 null event, 35  
 null set, 35

**O**

order statistics, 156

**P**

partition, 4  
 percentile, 265  
 pointwise convergence, 296  
 Poisson distribution, 74  
 Poisson paradigm, 79  
 Poisson process, 162  
 posterior probability, 40  
 prior probability, 40  
 probability density function, 110  
 probability integral transform, 146  
 probability mass function, 59  
 probability theory, 2  
 product measure, 167  
 product measure theorem, 167  
 product space, 165

**Q**

quantile, 264  
 quartile, 265

**R**

Rademacher distribution, 230  
 Radon-Nikodym derivative, 110  
 Radon-Nikodym theorem, 110  
 random experiment, 3  
 random variable, 56  
 random vector, 128  
 real part of a complex number, 339

**S**

sample, 4  
 sample mean, 191  
 sample space, 4  
 second Borel-Cantelli lemma, 294  
 set function, 23  
 simple random variable, 104  
 Simpson paradox, 50  
 skewness, 262  
 Slutsky's theorem, 348  
 St. Petersburg paradox, 243  
 standard Cauchy distribution, 118

standard deviation, 64  
standard normal distribution, 124  
standard uniform distribution, 122  
Stirling's formula, 356  
strong law of large numbers, 310  
Student's  $t$ -distribution, 224  
 $\sigma$ -subadditive measure, 25  
support of a random variable, 57, 112  
survival function, 160  
symmetric distribution, 262

**T**

$t$ -distribution, 224  
trial, 11  
triangular distribution, 193  
 $t$ -розподіл, 224

**U**  
uncorrelated random variables, 190

unexplained variance, 283  
uniform distribution, 66, 138  
utility function, 244

**V**  
variance, 64  
Venn diagrams, 5  
version of a conditional probability, 271

**W**

Wald's equation, 280  
weak convergence, 344  
weak law of large numbers, 309  
with probability 1, 36  
within-group variation, 283

**Z**

Z-score, 141  
Z-оцінка, 141