

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

**Д. Ю. Тавров,
О. Л. Темнікова**

ТЕОРІЯ ЙМОВІРНОСТЕЙ ПРАКТИКУМ

Навчальний посібник

Рекомендовано Методичною радою КПІ ім. Ігоря Сікорського
як навчальний посібник для здобувачів ступеня бакалавра
за освітньою програмою «Наука про дані та математичне моделювання»
спеціальності 113 Прикладна математика

Електронне мережне навчальне видання

Київ
КПІ ім. Ігоря Сікорського
2024

Рецензент

Яковлев С. В., канд. техн. наук, доцент, завідувач кафедри
математичних методів захисту інформації НН ФТІ

Відповідальний
редактор

Тарасенко-Клятченко О.В., канд. техн. наук, доцент, доцент кафедри
системного програмування і спеціалізованих комп'ютерних систем
ФПМ

*Гриф надано Методичною радою КПІ ім. Ігоря Сікорського
(протокол № 8 від 20.06.2024 р.)
за поданням Вченої ради факультету прикладної математики
(протокол № 11 від 27.05.2024 р.)*

Навчальний посібник розроблено для оволодіння студентами факультету прикладної математики КПІ ім. Ігоря Сікорського, які навчаються за спеціальністю 113 Прикладна математика, освітньою програмою «Наука про дані та математичне моделювання», навичками розв'язання основних типів задач з освітньої компоненти «Теорія ймовірностей». Практикум складається з 17 розділів, які містять короткі теоретичні відомості з основних розділів теорії ймовірностей та детальний розбір розв'язання задач з таких тем: ймовірнісний простір та міра, умовна ймовірність, дискретні та абсолютно неперервні випадкові величини, випадкові вектори, сподівання випадкових величин, умовні сподівання та умовні розподіли, граничні теореми. Засвоєння матеріалу цього практикуму стане в пригоді для подальшого вивчення математичної статистики, машинного навчання, аналізу даних та інших освітніх компонент відповідної освітньої програми.

Національний технічний університет України
«Київський політехнічний інститут імені Ігоря Сікорського»
проспект Берестейський, 37, м. Київ, 03056
<https://kpi.ua>

Свідоцтво про внесення до Державного реєстру видавців, виготовлювачів
і розповсюджувачів видавничої продукції ДК № 5354 від 25.05.2017 р.

© Д. Ю. Тавров, О. Л. Темнікова, 2024
© КПІ ім. Ігоря Сікорського, 2024

Зміст

Передмова	6
1 Обчислення класичної ймовірності	8
1.1 Основні теоретичні відомості	8
1.2 Обчислення класичних імовірностей для скінченних просторів	10
1.2.1 Правило множення	10
1.2.2 Вибір	12
1.3 Додаткові задачі	22
2 Комбінаторні доведення. Геометричні ймовірності. Вимірні множини	26
2.1 Основні теоретичні відомості	26
2.2 Комбінаторні доведення	29
2.3 Метод відбиття	32
2.4 Геометричні ймовірності	37
2.5 σ -алгебри подій	40
2.6 Додаткові задачі	42
3 Властивості ймовірнісної міри	46
3.1 Основні теоретичні відомості	46
3.2 Застосування властивостей імовірностей	49
3.3 Теорема включення-виключення	51
3.4 Властивості міри	57
3.5 Додаткові задачі	62
4 Умовна ймовірність	65
4.1 Основні теоретичні відомості	65
4.2 Визначення умовної ймовірності	68
4.3 Теорема Беєса та повна ймовірність	69
4.4 Незалежність подій	74
4.5 Додаткові вправи	81
5 Дискретні випадкові величини. Схема Бернуллі	85
5.1 Основні теоретичні відомості	85
5.2 Загальні дискретні випадкові величини	90
5.3 Схема Бернуллі	96
5.4 Додаткові вправи	99

6 Деякі поширені дискретні розподіли	103
6.1 Основні теоретичні відомості	103
6.2 Розподіл Пуассона	108
6.3 Інші дискретні розподіли	111
6.4 Додаткові вправи	115
7 Неперервні випадкові величини	119
7.1 Основні теоретичні відомості	119
7.2 Функції розподілу	125
7.3 Щільності розподілу	127
7.4 Сподівання та дисперсія	133
7.5 Додаткові вправи	137
8 Функції від випадкових величин. Деякі неперервні розпо- діли	142
8.1 Основні теоретичні відомості	142
8.2 Рівномірний розподіл	147
8.3 Нормальний розподіл	149
8.4 Функції від випадкових величин	153
8.5 Додаткові вправи	157
9 Окремі застосування неперервних випадкових величин	161
9.1 Основні теоретичні відомості	161
9.2 Інтегральне перетворення ймовірностей	166
9.3 Бета-розподіл та елементи Беєсівського виведення	171
9.4 Порядкові статистики	175
9.5 Експоненційний розподіл та процес Пуассона	177
9.6 Додаткові вправи	181
10 Випадкові вектори та незалежні випадкові величини. Ча- стина 1	185
10.1 Основні теоретичні відомості	185
10.2 Спільні функції та щільності розподілів	194
10.3 Сподівання та коваріації, незалежність	200
10.4 Додаткові вправи	206
11 Випадкові вектори та незалежні випадкові величини. Ча- стина 2	211
11.1 Основні теоретичні відомості	211
11.2 Незалежність випадкових величин	212
11.3 Обчислення сподівань через суму індикаторних величин	215
11.4 Суми незалежних випадкових величин	217
11.5 Додаткові вправи	221

12 Функції від випадкових векторів та їхні розподіли	224
12.1 Основні теоретичні відомості	224
12.2 Розподіл функції від випадкового вектора	229
12.3 Гамма-розподіл	235
12.4 Мультиномний розподіл	237
12.5 Багатовимірний нормальний розподіл	240
12.6 Додаткові вправи	242
13 Сподівання та інші характеристики випадкових величин	249
13.1 Основні теоретичні відомості	249
13.2 Нерівності зі сподіваннями	256
13.3 Міри центральної тенденції	261
13.4 Моменти вищих порядків та квантілі	267
13.5 Додаткові вправи	271
14 Умовні розподіли	276
14.1 Основні теоретичні відомості	276
14.2 Умовні функції ймовірності	279
14.3 Умовні щільності	284
14.4 Умовні сподівання й медіана як найліпші предиктори	289
14.5 Додаткові вправи	292
15 Умовні сподівання	295
15.1 Основні теоретичні відомості	295
15.2 Обумовлення подіями з розбиття	299
15.3 Обумовлення випадковими величинами	303
15.4 Додаткові вправи	309
16 Збіжність послідовностей випадкових величин. Закон великих чисел	314
16.1 Основні теоретичні відомості	314
16.2 Послідовності випадкових величин	319
16.3 Закон великих чисел	325
16.4 Додаткові вправи	328
17 Твірні функції моментів	330
17.1 Основні теоретичні відомості	330
17.2 Вправи	332
Література	342
Предметний показник	343

Передмова

Навчальне видання призначено для студентів, які вивчають дисципліну «Теорія ймовірностей» з циклу загальної підготовки освітньо-професійної програми «Наука про дані та математичне моделювання» (спеціальність 113 «Прикладна математика»). Ця дисципліна є ключовою для освітньої програми, адже вона закладає підвалини для вивчення прикладних аспектів застосування теорії ймовірностей у математичній статистиці, машинному навчанні та аналізі даних.

Цей практикум забезпечує дисципліну «Теорію ймовірностей» обсягом 36 академічних годин лекційних занять та 36 академічних годин практичних занять, яку вивчають на факультеті прикладної математики у 3 семестрі. Він повністю відповідає навчальному посібнику «Теорія ймовірностей. Курс лекцій» (<https://ela.kpi.ua/handle/123456789/57153>), який автори цього практикуму підготували для опанування лекційної частини курсу. Зокрема, у цьому практикумі ми посилаємося на твердження, теореми, визначення з зазначеного курсу лекцій, додаючи до відповідних номерів префікс КЛ.

Метою вивчення дисципліни «Теорія ймовірностей» є засвоєння студентами основних принципів, ідей та методів теорії ймовірностей, потрібних для дальшого вивчення загальнотеоретичних та практично орієнтованих дисциплін та формування практичних навичок їх застосування в обраній предметній галузі. Предметом навчальної дисципліни є випадкові події та обчислення їхніх ймовірностей, випадкові величини та їхні розподіли, послідовності випадкових величин та їхні граничні розподіли, гранична поведінка послідовностей випадкових величин, зокрема закони великих чисел та центральні граничні теореми.

Пререквізитами для вивчення дисципліни є дисципліни, предметом яких є математичний аналіз, алгебра та геометрія та дискретна математика. Особливо корисними для опанування дисципліни є детальне освоєння основ матричної алгебри, умов збіжності послідовностей та функцій, безумовної оптимізації функцій багатьох змінних, теоретико-множинних операцій, елементів теорії міри тощо. Вивчення дисципліни є передумовою вивчення математичної статистики, машинного навчання та аналізу даних.

Практикум містить 17 розділів, які приблизно відповідають практичним заняттям. У кожному розділі наведено стислий теоретичний матеріал (ретельний виклад можна знайти в Курсі лекцій), детальні приклади розв'язання типових задач, а також додаткові задачі для самоконтролю. Задачі для практичних занять переважно адаптовано з джерел, наведених у спи-

ску використаної літератури, окремі задачі автори підготували самостійно. Увесь ілюстративний матеріал підготували автори, графіки було підготовлено з використанням мови програмування Python. Наприкінці практикуму наведено перелік посилань, які було використано під час його підготовки.

Практикум має стати в пригоді під час проведення практичних занять, а також для самостійного поглибленого опанування матеріалу та підготовки до контрольних заходів та іспиту з дисципліни.

1 Обчислення класичної ймовірності

1.1 Основні теоретичні відомості

Детальний огляд понять, потрібний для розв’язання задач на цьому занятті, наведено в Курсі лекцій, Розд. КЛ1.2–КЛ1.3.

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 варіант із деякого набору можливих.

Визначення (КЛ1.2.1). *Простором елементарних подій* (sample space) Ω для деякого випадкового експерименту називатимемо множину всіх можливих його результатів. Простір Ω може бути скінченним (finite), зліченим (countable) або незліченим (uncountable). Елементами $\omega \in \Omega$ цього простору є *елементарні події* (samples). Тоді *подія* (event) A — це деяка підмножина (subset) простору Ω ¹. $\square$

Ми кажемо, що *сталася подія* A , якщо результат експерименту належить множині A . Оскільки події є підмножинами простору S , до них застосовні всі теоретико-множинні операції та їхні властивості. Відповідну таблицю з Курсу лекцій ми повторюємо тут у Табл. 1.1.1.

Події $A_1, A_2, \dots$ утворюють *розбиття* (partition) простору Ω , якщо вони є попарно несумісні — $A_i \cap A_j = \emptyset$, $i \neq j$ — а їх об’єднання дає весь простір: $\bigcup_{i=1}^{\infty} A_i = \Omega$. Розбиття може бути скінченним або нескінченним.

Деякі властивості операцій над множинами та пов’язані з цим закони такі:

- комутативність (commutativity) об’єднання й перетину: $A \cup B = B \cup A$, $A \cap B = B \cap A$;
- асоціативність (associativity) об’єднання й перетину: $A \cup (B \cap C) = (A \cup B) \cap C$, $A \cap (B \cup C) = (A \cap B) \cup C$;

¹Варто розрізняти події як множини, наприклад, події $\{\omega_1\}$, $\{\omega_4\}$ тощо, та *елементарні* події як елементи простору — ω_1 , ω_4 тощо. У цьому сенсі англомовна термінологія вдаліша, адже елементи ω називають samples, а відповідні множини $\{\omega\}$ — простими подіями (simple events).

Табл. 1.1.1: 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) Ω	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
неперетинні (disjoint) множини $A \cap B = \emptyset$	події A і B несумісні (mutually exclusive)
доповнення (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

- дистрибутивність (disributivity): $A \cup (B \cap C) = (A \cup B) \cap (A \cup C)$, $A \cap (B \cup C) = (A \cap B) \cup (A \cap C)$;
- закони де Моргана² (de Morgan's laws):

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

Визначення (КЛ1.3.1). Нехай простір елементарних подій є скінченним: $\Omega = \{\omega_1, \dots, \omega_n\}$. У цьому випадку за побудовою:

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

Нехай $A \subseteq \Omega$ — деяка подія. Тоді *класичною ймовірністю* (classical, naive

²Августус де Морган (Augustus De Morgan, 1806–1871) — британський математик і логік.

probability) події вважатимемо відношення

$$\mathbb{P}(A) = \frac{|A|}{|\Omega|}. \quad (\text{КЛ1.3.1})$$

□

Із визначення класичної ймовірності безпосередньо випливають 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 $\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B)$;
- $\mathbb{P}(A^c) = 1 - \mathbb{P}(A)$.

Для зліченного простору поняття класичної ймовірності визначають як границю відносних частот настання певної події:

$$\mathbb{P}(A) = \lim_{N \rightarrow \infty} \frac{N(A)}{N}, \quad (\text{КЛ1.3.2})$$

де $N(A)$ — число настань події A серед N повторень випадкового експерименту. Вищенаведені властивості ймовірностей зберігаються і в цьому випадку.

Для незліченного простору поняття класичної ймовірності перетворюється в поняття *геометричної ймовірності*, яку рахують як відношення деякої міри події A як множини (довжина відрізка, площа пласкої фігури, об'єм тіла тощо) до загальної міри всього простору (довжини, площі, об'єму такого простору).

Виходячи з визначення класичної ймовірності події $A \subseteq \Omega$, її підрахунок зовсім нескладний: достатньо тільки знати потужності множини $|A|$ і всього простору $|\Omega|$. Цьому й буде присвячено перше практичне заняття.

1.2 Обчислення класичних імовірностей для скінченних просторів

1.2.1 Правило множення

Завжди існує можливість обчислити потужність деякої скінченної множини шляхом підрахунку її елементів. Проте такий підхід працює тільки для найпростіших просторів на кшталт простору для підкидань гральної кісточки. У загальному випадку для визначення числа елементів множини

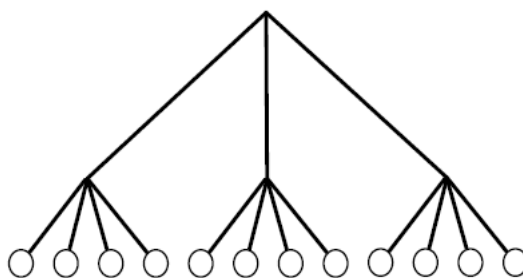


Рис. 1.2.1: Ілюстрація до правила множення: якщо експеримент A може закінчитися 3 варіантами, а експеримент B — чотирма, то загальне число можливих упорядкованих варіантів після виконання обох експериментів дорівнює $3 \cdot 4 = 12$

застосовують декілька прийомів. Найважливіший із них — так зване *правило множення* (multiplication rule).

Твердження 1.2.1 (Правило множення). Нехай експеримент складається з n експериментів, $A_1, \dots, A_n$. Якщо експеримент A_i може завершитися r_i варіантами, $i = 1, \dots, n$, то загальне число впорядкованих варіантів завершення всіх експериментів дорівнює $r_1 \cdot \dots \cdot r_n$.

Ілюстрацію цього правила наведено на Рис. 1.2.1.

Іноколи для спрощення підрахунків можна уявляти, що складові експерименту відбуваються послідовного в часі, хоча насправді вони можуть виконуватися паралельно або в якомусь іншому порядку.

Вправа 1.2.2. Нехай у забігу беруть участь 10 спортсменів. Покладімо, що нічия неможлива, тобто перше, друге і третє місця повинні посісти різні учасники. Скільки існує можливих варіантів посідання (перших трьох) призових місць учасниками забігу? $\square$

Розв'язання. Уявімо, що мають місце три різні експерименти — посідання першого місця A , посідання другого місця B і посідання третього місця C .

Кожен учасник може посісти перше місце, тому всього існує 10 варіантів завершення експерименту A . Після його завершення залишилось 9 можливих кандидатів на друге місце, отже експеримент B має 9 можливих варіантів. Аналогічно C має 8 можливих варіантів.

Отже загальне число варіантів посідання трьох призових місць дорівнює $10 \cdot 9 \cdot 8 = 720$. $\square$

Зауваження 1.2.3. Розв'язуючи Вправу 1.2.2, ми почали аналіз із першого місця, хоча могли це зробити з будь-якого іншого. Наприклад, ми могли сказати, що експеримент C має 10 можливих варіантів, потім що A має

9 можливих варіантів, і нарешті, що B має 8 можливих варіантів. Хронологічний порядок виконання експериментів значення не має, важливий тільки той факт, що результати виконання трьох експериментів впорядковані, тобто що дві особи не можуть посісти одне місце. $\square$

Приклад 1.2.4. Нехай студент планує придбати морозиво. Морозивник торгує ванільним, полуничним та шоколадним морозивом, яке можна замовити у вафельному ріжку або в пластиковому цеберку. Цілком очевидно, що існує $3 \cdot 2 = 6$ усіх можливих варіантів придбання морозива. При цьому обчислення можна було почати з типу фасування і сказати, що існує $2 \cdot 3 = 6$ варіантів.

Також зовсім неприпринципово, *які саме* смаки морозива доступні для кожного типу фасування, головне, що їх є однакова кількість — 3. Якби для морозива в ріжку та морозива в цеберку були доступні смаки в різній кількості, то правило множення не було б застосовно. $\square$

1.2.2 Вибір

Значна кількість експериментів у теорії ймовірностей та статистиці полягає у здійсненні *вибору* (sampling) об'єктів із деякого набору. Вибирати об'єкти можна одним із чотирьох способів.

Вибір можна класифікувати за принципом наявності повторень:

- *із повтореннями* (with replacement): якщо після вибору об'єкта його повертають назад до початкового набору і можливий його повторний вибір у майбутньому;
- *без повторень* (without replacement): якщо об'єкт після вибору вилучають із початкового набору.

Вибір можна класифікувати за принципом збереження порядку:

- *зі збереженням порядку* (order matters): вибір об'єктів у різному порядку веде до різних результатів;
- *без збереження порядку* (order doesn't matter): головне, які об'єкти вибрано, а не в якому порядку.

Поки що зосередьмося на експериментах, у яких порядок здійснення вибору має значення.

Вибір зі збереженням порядку

Твердження 1.2.5 (Вибір із повтореннями). Нехай набір містить n об'єктів. Загальне число вибору k об'єктів із такого набору з повтореннями, зі збереженням порядку, дорівнює n^k .

Доведення. Нехай у деякій вазі міститься n куль, пронумерованих від 1 до n . Вибір із повтореннями передбачає послідовне витягування однієї кулі, фіксації її номеру та повернення її назад до вази.

Витягування однієї кулі можна розглядати як окремий експеримент, який має n можливих варіантів. Загальне число таких експериментів — k штук. Оскільки важливий порядок результатів експериментів, можна застосувати правило множення, згідно з яким існує $n \cdot n \cdot \dots \cdot n = n^k$ варіантів завершення всього експерименту. $\square$

Твердження 1.2.6 (Вибір без повторень). Нехай набір містить n об'єктів. Загальне число вибору k об'єктів із такого набору без повторень, зі збереженням порядку, дорівнює $\frac{n!}{(n-k)!}$.

Доведення. Нехай у деякій вазі міститься n куль, пронумерованих від 1 до n . Вибір без повторень передбачає послідовне витягування однієї кулі, фіксації її номеру та відкидання її набік.

Витягування першої кулі можна розглядати як окремий експеримент, який має n можливих варіантів. Після витягу першої кулі у вазі залишається $(n - 1)$ куль, і тому витягування другої кулі має $(n - 1)$ можливих варіантів завершення, і т.д.

Загальне число таких експериментів — k штук. Оскільки важливий порядок результатів експериментів, можна застосувати правило множення, згідно з яким існує

$$n \cdot (n - 1) \cdot \dots \cdot (n - (k - 1)) = \frac{n!}{(n - k)!}$$

варіантів завершення всього експерименту. $\square$

Зауваження 1.2.7. Очевидно, що якщо стоїть задача вибору n об'єктів із посеред n штук (тобто всіх) без повторень, то фактично потрібно здійснити їх *перестановку* (permutation). Згідно з Твердженням 1.2.6, існує $n!$ таких перестановок. $\square$

Вправа 1.2.8 (Задача про дні народження (Birthday problem)). У кімнаті присутні k осіб. Для спрощення покладімо, що день народження кожного з них міг припасти на будь-який із 365 днів року (без 29 лютого) з

однаковою ймовірністю, а самі дні народження незалежні один від одного (виключаємо ситуації типу близнюків). Чому дорівнює ймовірність події A , що принаймні в одній парі людей день народження припадає на один день? $\square$

Розв’язання. Простір елементарних подій у цій задачі складається з усіх можливих варіантів розподілу днів народження між людьми в кімнаті. Розглядаючи дні народження як об’єкти, які ми витягуємо з повтореннями і присвоюємо людям у кімнаті, доходимо висновку, що існує 365^k можливих варіантів розподілу днів народження у кімнаті, тобто що $|\Omega| = 365^k$.

Тепер нас цікавить число варіантів присвоєння днів народження, за яких принаймні в одній парі осіб дні народження будуть однакові. Такий підрахунок є задачею непростою, адже потрібно врахувати не тільки можливість того, що Іван і Петро мають однаковий день народження, а всі інші — різні, а й можливість, що Іван і Петро мають один спільний день народження, а Марія й Анастасія — також спільний, але якийсь інший. Або що Іван, Петро і Марія мають спільний день народження на трьох, і т.п.

У таких ситуаціях часто буває простіше підрахувати потужність множини A^c , яка є *доповненням* шуканої множини A . У нашому випадку потрібно підрахувати кількість варіантів розподілу днів народження таких, що вони *всі різні*. Це зробити значно простіше, адже маємо справу з вибором k об’єктів із 365 можливих *без повторень*. Відповідно,

$$|A| = |\Omega| - |A^c| = 365^k - \frac{365!}{(365 - k)!},$$

а тому, використовуючи (КЛ1.3.1), маємо:

$$\mathbb{P}(A) = \frac{|A|}{|\Omega|} = 1 - \frac{365 \cdot 364 \cdot \dots \cdot (365 - (k - 1))}{365^k}.$$

Ймовірність $\mathbb{P}(A)$ як функцію від k зображено на Рис. 1.2.2. Як можна бачити, можливо до певної міри несподівано, але вже для $k = 23$ осіб ймовірність збігу днів народження перевищує 0.5. А в групі з $k = 57$ осіб ймовірність збігу днів народження перевищує 0.99! $\square$

Вправа 1.2.9 (Помилка Ляйбніця). Якщо викинути правильну гральну кісточку двічі, то яка подія ймовірніша — A = «сума двох чисел дорівнює 11» чи B = «сума двох чисел дорівнює 12»? $\square$

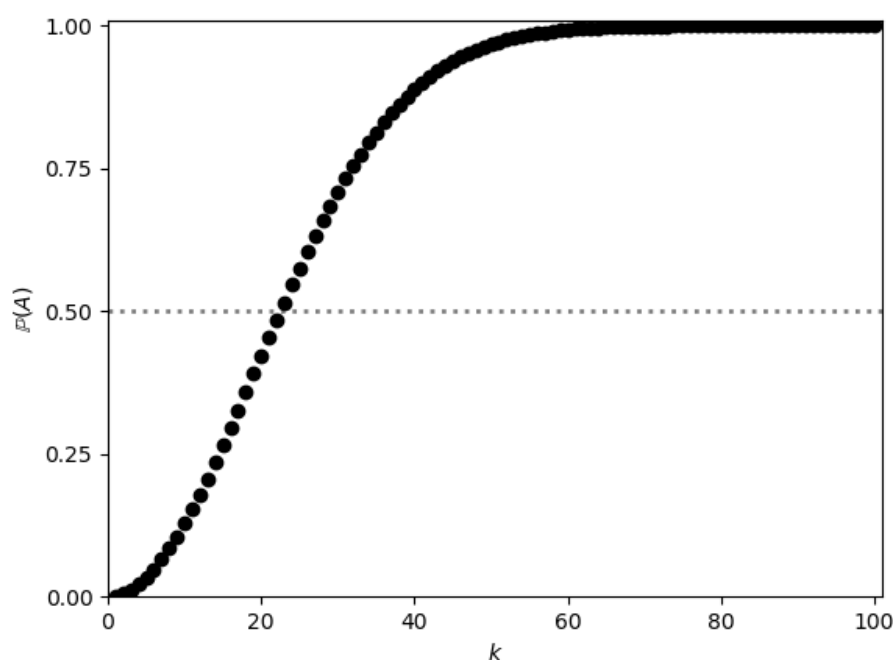


Рис. 1.2.2: Імовірність $\mathbb{P}(A)$ як функція від числа осіб у кімнаті k для Вправи 1.2.8

Розв'язання. Підкидання кожної кісточки можна вважати окремим експериментом, тому існує 36 можливих варіантів випадів чисел на двох кісточках — упорядкованих пар $\Omega = (x, y)^\top \in \mathbb{N}^2$, $1 \leq x, y \leq 6$. Усі вони рівноймовірні за симетричністю задачі. Шукані події можна записати так:

$$A = \{(x, y)^\top \in \mathbb{N}^2 : x + y = 11, 1 \leq x, y \leq 6\},$$

$$B = \{(x, y)^\top \in \mathbb{N}^2 : x + y = 12, 1 \leq x, y \leq 6\}.$$

Число 11 можна утворити сумою 5 і 6, тобто $A = \{(5, 6)^\top, (6, 5)^\top\}$, а число 12 — тільки сумою двох шісток, тобто $B = \{(6, 6)^\top\}$. Отже $\mathbb{P}(A) = 2\mathbb{P}(B)$.

Проте Ляйбніць стверджував, що $\mathbb{P}(A) = \mathbb{P}(B)$, оскільки, на його думку, і 11, і 12 можна дістати тільки в один спосіб. Міркуючи в такий спосіб, він припускався тієї помилки, що кісточки є *різними* об'єктами, а тому *порядок* випадку чисел має значення. $\square$

Вибір без збереження порядку

Якщо потрібно вибрати k об'єктів *без збереження порядку*, то застосування формул із попереднього розділу буде систематично завищувати справжнє число можливих варіантів, оскільки варіанти, які еквівалентні між собою з точністю до перестановки, будуть рахуватися кратно. У загальному випадку, оскільки k об'єктів можна переставляти місцями в $k!$ способів, загальне число варіантів вибору k об'єктів за формулами з попереднього розділу в

умовах без збереження порядку потрібно ділити на $k!$.

Приклад 1.2.10. Із членів комітету в n осіб, відповідно до правила множення, можна обрати $n(n-1)$ пар керівників (голова, заступник голови), але $\frac{n(n-1)}{2}$ пар співголів, оскільки співголови рівноцінні. $\square$

При цьому потрібно завжди пам'ятати, про що саме нас питають.

Приклад 1.2.11. Продовжмо розгляд Прикладу 1.2.4. Нехай студент планує придбати два морозива — одне в обід, а інше увечері. Згідно з правилом множення, усього існує $6 \cdot 6 = 36$ можливих таких пар (x, y) , де x і y позначають певний варіант морозива (наприклад, шоколадне в ріжку). Нехай нас тепер цікавить число варіантів морозива, які можна придбати деякого дня, *незважаючи на їхній порядок*. На перший погляд може здатися, що відповідь дорівнює $36/2 = 18$ варіантів, оскільки пари (x, y) і (y, x) еквівалентні. Проте це неправильно, адже студент міг придбати однаковий варіант морозива (наприклад, шоколадне в ріжку) і в обід, і на вечір, і таких пар виду (x, x) є 6 штук. Вони не мають своїх «дзеркальних відповідників», тому просто вполювинити їх кількість було б неправильно. Отже, згідно з правилом множення, може бути тільки $6 \cdot 5 = 30$ можливих пар (x, y) із *різними* варіантами морозива: $x \neq y$. Відтак загальне число різних варіантів морозива, які можна придбати в один день, становить $30/2 + 6 = 21$. $\square$

У цій та аналогічних ситуаціях порядок результатів окремих експериментів не має значення.

Для спрощення підрахунків числа варіантів у схожих ситуаціях використовують поняття *біномного коефіцієнту* (binomial coefficient).

Твердження 1.2.12. Нехай набір містить n об'єктів. Загальне число вибору k об'єктів із такого набору без повторень, *без збереження порядку*, дорівнює

$$\binom{n}{k} \equiv \begin{cases} \frac{n!}{k!(n-k)!}, & k \leq n \\ 0, & k > n \end{cases}, \quad (1.2.1)$$

де $\binom{n}{k}$ називають біномним коефіцієнтом.

Доведення. Згідно з Твердженням 1.2.6, існує $\frac{n!}{(n-k)!}$ варіантів вибору k елементів з-посеред n можливих. Але порядок вибраних елементів нас не цікавить. Усього існує $k!$ варіантів перестановок цих елементів, тому ми повинні поділити загальне число варіантів на $k!$. $\square$

Вправа 1.2.13. Скільки існує варіантів перестановок літер у слові «математика»? $\square$

Розв’язання. Незважаючи на те, що в питанні стоїть слово «перестановка», нас, очевидно, питають не про ті перестановки, яких буває $n!$ штук, а про різні варіанти формування рядків із символів, що складають слово «математика», з урахуванням того, що деякі літери повторюються.

Існують щонайменше два різні способи відповісти на це питання. З одного боку, ми справді можемо сказати, що існує $10!$ перестановок літер слова «математика», але оскільки літера «а» повторюється тричі, а літери «м» і «т» — двічі, то загальне число формальних перестановок потрібно поділити на відповідні кількості перестановок «локальних», щоб дістати

$$\frac{10!}{3!2!2!} = 151\,200.$$

З іншого боку, ми можемо розглянути експерименти, для кожної літери, у рамках яких вибирають k позицій, які може зайняти відповідна літера, де k — кількість її повторів. Наприклад, спочатку ми хочемо визначити позиції для літери «а». Існує $\binom{10}{3}$ таких варіантів, оскільки ми вибираємо позиції без повторень і без збереження порядку. Після розміщення літер «а» для літери «м» залишається 7 вільних позицій. Тому існує $\binom{7}{2}$ варіантів розташування літери «м», оскільки вона повторюється двічі. Аналогічні міркування зрештою дають

$$\binom{10}{3} \binom{7}{2} \binom{5}{2} \binom{3}{1} \binom{2}{1} \binom{1}{1} = 151\,200.$$

□

Суть біномного коефіцієнту як числа можливих варіантів вибору без повторень і без збереження порядку дає можливість довести біномну теорему.

Приклад 1.2.14 (Біномна теорема). Згідно з біномною теоремою,

$$(x + y)^n = \sum_{k=0}^n \binom{n}{k} x^k y^{n-k}, \quad n \in \mathbb{N}^+.$$

Розкриймо дужки у виразі $(x + y)(x + y) \dots (x + y)$. Кожний доданок матиме вид $x^k y^{n-k}$, $k = 0, 1, \dots, n$, оскільки щоразу перемножуємо точно n множників, по одному з кожної дужки. Потрібно з’ясувати, який коефіцієнт стоїть біля члена $x^k y^{n-k}$.

Оскільки в кожного такого члена сума ступенів завжди дорівнює n , достатньо зосередити увагу на x -ах. У процесі перемноження дужок із кожної дужки вибирають або x , або y . Фактично, нас цікавить вибір дужок з x -ами. Оскільки в цьому випадку мова про вибір елементів без повторень і без збереження порядку, біля $x^k y^{n-k}$ стоїть біномний коефіцієнт $\binom{n}{k}$. □

Вправа 1.2.15. У грі в покер гравець отримує 5 гральних карт, у випадковий спосіб вибраних із колоди в 52 карти³. Кажуть, що має місце *фул-хаус* (full house), якщо з 5 карт три карти — деякого одного звання, а інші дві — деякого іншого (наприклад, три сімки і дві кралі, незалежно від мастей). Чому дорівнює ймовірність події A — дістати фул-хаус? $\square$

Розв’язання. Простором елементарних подій є всі можливі набори по 5 різних карт із колоди в 52 карту. Його потужність дорівнює $\binom{52}{5}$, оскільки ми вибираємо карти без повторень, а порядок карт роли не грає.

Тепер потрібно з’ясувати, скільки існує варіантів дістати фул-хаус. Для цього застосуємо правило множення, розбивши наш експеримент на такі складові частини:

- вибір звання, якому належать три карти. Таких варіантів є 13;
- вибір трьох карт відповідного звання. Тут мова про вибір трьох карт без повторень і без збереження порядку з-посеред чотирьох можливих, тобто маємо $\binom{4}{3}$ варіантів;
- вибір звання, якому належать дві карти. Оскільки одне звання ми вже вибрали, маємо 12 варіантів;
- вибір двох карт відповідного звання, маємо $\binom{4}{2}$ варіантів.

Отже ймовірність дістати фул-хаус за формулою (КЛ1.3.1) дорівнює

$$\mathbb{P}(A) = \frac{13 \binom{4}{3} 12 \binom{4}{2}}{\binom{52}{5}} = \frac{3\,744}{2\,598\,960} \approx 0.001.$$

Ця ймовірність достатньо мала. $\square$

Вправа 1.2.16 (Задача Ньютона-Піпса (Newton-Perus problem)). У 1693 р. до Айзека Ньютона (Isaac Newton) звернувся британський діяч Семюел Піпс (Samuel Perus) із таким питанням. Ймовірність якої з подій більша:

- A = «на 6 гральних кісточках випала принаймні одна шістка»;
- B = «на 12 гральних кісточках випало принаймні дві шістки»;
- C = «на 18 гральних кісточках випало принаймні три шістки»?

$\square$

³Розглядатимемо стандартну колоду карт із 13 званнями та 4 мастями.

Розв'язання. У кожній ситуації просторами елементарних подій є, відповідно, $\{1, \dots, 6\}^6$, $\{1, \dots, 6\}^{12}$ і $\{1, \dots, 6\}^{18}$.

Розгляньмо перший експеримент. Як і у Вправі 1.2.8, простіше підрахувати число випадів кісточок, де немає жодної шістки. У цьому випадку фактично маємо ситуацію з вибором чисел від 1 до 5 шість разів поспіль, із повтореннями. Тобто застосовне Твердження 1.2.5:

$$\mathbb{P}(A) = 1 - \frac{5^6}{6^6} \approx 0.67.$$

Для другого експерименту також підрахуймо число випадів від зворотного, тобто число випадів, коли випала або одна шістка, або жодної. Варіант із жодною шісткою аналогічний попередньому випадку і дає 5^{12} випадів. Випад точно однієї шістки можна інтерпретувати як експеримент, у якому спочатку вибираємо кісточку, на якій випаде 6 (таких варіантів є $\binom{12}{1}$), а потім розглядаємо вибір чисел від 1 до 5, як у попередніх випадках. Отже

$$\mathbb{P}(B) = 1 - \frac{5^{12} + \binom{12}{1}5^{11}}{6^{12}} \approx 0.62.$$

Нарешті, для третього експерименту маємо:

- жодного разу 6 не випаде для 5^{18} випадків;
- точно один раз 6 випаде для $\binom{18}{1}5^{17}$ випадків;
- точно два рази 6 випаде для $\binom{18}{2}5^{16}$, оскільки ми спочатку повинні вибрати дві кісточки, на яких випаде шістка.

Отже

$$\mathbb{P}(C) = 1 - \frac{5^{18} + \binom{18}{1}5^{17} + \binom{18}{2}5^{16}}{6^{18}} \approx 0.60.$$

Тому A має найвищу ймовірність.

Саме так Ньютон і відповів Піпсу. □

Твердження 1.2.17. Нехай набір містить n об'єктів. Загальне число варіантів вибору k об'єктів із такого набору з повтореннями, але без збереження порядку, дорівнює $\binom{n+k-1}{k}$.

Доведення. Розгляньмо ізоморфну задачу: розміщення k нерозрізненних куль в n пронумерованих вазах⁴.

Кожне таке розміщення можна еквівалентно подати як послідовність вертикальних рисок і точок, де риска відповідає стінці між вазами, а точка — кулі (Рис. 1.2.3). Коректна послідовність повинна, звісно, починатися

⁴У такому формулюванні маємо справу з задачею, пов'язаною з так званим *конденсатом Боуза-Айнштейна*, існування якого фізики Сат'єндра Нат Боуз (Satyendra Nath Bose, 1894–1974) та Альберт Айнштайн (Albert Einstein, 1879–1955) передбачили на основі аналізу нерозрізненних частинок і їх розташування на дискретних енергетичних рівнях.

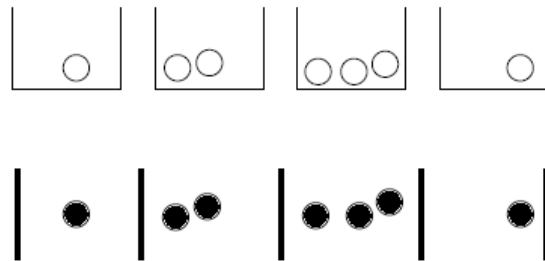


Рис. 1.2.3: Ілюстрація до доведення Твердження 1.2.17: розміщення 7 нерозрізненних об'єктів у 4 вазах можна подати як послідовність вертикальних рисок і точок

і закінчуватися рисою. Якщо не брати їх до уваги, до маємо $n-1$ риску і k точок, які можна в довільний спосіб перетасовувати. Уявімо собі, що існує $n+k-1$ позицій, на яких потрібно розмістити k точок *без повторень* (якщо розмістити k точок, то решту позицій автоматично буде зайнято рисками). Ця задача відома і має розв'язок — $\binom{n+k-1}{k}$ ⁵.

Те, що розв'язок цієї задачі відповідає числу вибору k об'єктів із набору з n об'єктів *з повтореннями* та без збереження порядку, можна побачити, якщо кожен вазу розглядати як об'єкт, а кількість куль у вазі — числу разів, скільки було вибрано цю конкретну вазу. Оскільки кулі нерозрізненні, маємо вибір без збереження порядку, а оскільки куля може потрапити в будь-яку вазу, маємо вибір із повтореннями.

□

Зауваження 1.2.18. Ще однією задачею, яка є ізоморфною для Твердження 1.2.17, є задача підрахунку розв'язків рівняння $x_1 + x_2 + \dots + x_n = k$, де $x_i \in \mathbb{Z}^+$, $i = 1, \dots, n$. Ми можемо розглядати кожний x_i як кількість куль в i -ій вазі.

□

Приклад 1.2.19. Хоча Твердження 1.2.17 дає можливість підрахувати число варіантів вибору k об'єктів із-поміж n можливих із повтореннями, але *без збереження порядку*, його застосування до обчислення класичної ймовірності є дуже обмеженим. Класичне визначення ймовірності працює тільки у випадках, коли результати експерименту рівноймовірні.

Уявімо, наприклад, що для деякого опитування з групи в n осіб у випадковий спосіб по черзі вибирають респондентів, здійснюючи вибір із повтореннями. Якщо розглядати ці вибірки як *упорядковані*, тобто що вибірка (Іван, Петро, Марія, Іван) і (Іван, Марія, Іван, Петро) — це різні вибірки, то кожна з n^k таких можливих вибірок має однакову ймовірність. Але якщо розглядати вибірки як *неупорядковані*, то дві наведені вище вибірки

⁵В англійській літературі такий метод розв'язання цієї задачі називають «stars and bars».

Табл. 1.2.1: Способи підрахунку кількостей можливих варіантів для виборів різного типу

	Зі збереженням порядку	Без збереження порядку
Із повтореннями	n^k	$\binom{n+k-1}{k}$
Без повторень	$\frac{n!}{(n-k)!}$	$\binom{n}{k}$

будуть вважати однаковими, і тоді ймовірність кожної з $\binom{n+k-1}{k}$ вибірок *не буде однакою*. $\square$

Підсумувати результати розгляду різних методів підрахунку можна в Табл. 1.2.1.

Вправа 1.2.20. Для захисту своїх телефонів користувачі можуть задавати пароль із чотирьох цифр. Дамо відповіді на такі питання:

- (i) скільки існує можливих паролів?
- (ii) якщо за «унікальний пароль» вважати пароль із чотирьох *різних* цифр (тобто пароль 6713 — унікальний, а 8824 — ні), то скільки існує таких паролів?
- (iii) якщо за «зростаючий» пароль вважати пароль, де кожна цифра не менша від попередньої (тобто 4789 і 4488 — зростаючі, а 8239 — ні), то скільки існує паролів, які одночасно унікальні і зростаючі?
- (iv) скільки існує паролів, які зростаючі, але необов'язково унікальні?

$\square$

Розв'язання. (i) У цьому випадку маємо вибір 4 цифр із 10 можливих із повтореннями зі збереженням порядку, тому всього існує 10^4 паролів;

(ii) у цьому випадку маємо вибір 4 цифр із 10 можливих *без повторень* зі збереженням порядку, тому всього існує $\frac{10!}{6!} = 10 \cdot 9 \cdot 8 \cdot 7$ паролів;

(iii) потрібно помітити, що будь-яку послідовність із 4 різних цифр можна в *єдиний спосіб* упорядкувати за зростанням. Наприклад, цифри 6, 2, 5, 0 можна впорядкувати і дістати 0, 2, 5, 6. Отже з усіх $4!$ можливих перестановок нас цікавить тільки одна, а значить загальне число зростаючих і унікальних паролів можна дістати з попереднього випадку шляхом ділення на $4!$: $\frac{10!}{6!4!} = \binom{10}{4}$. Інше пояснення — ми вибираємо 4

цифри з-поміж 10 без повторень і *без збереження порядку*, оскільки вибравши 4 цифри, ми самі можемо задати той порядок, який нас цікавить;

- (iv) у цьому випадку ми вибираємо 4 цифри *з повтореннями*, але також *без збереження порядку* (оскільки самі потім їх можемо впорядкувати). Отже загальне число зростаючих паролів (у тому числі необов'язково унікальних) дорівнює $\binom{10+4-1}{4}$.

□

1.3 Додаткові задачі

Вправа 1.3.1. Область поділено на 5 районів, у кожному з яких є 6 міст, які турист планує відвідати. Турист не хоче заїжджати в один район більше одного разу: заїхавши один раз, він хоче відвідати всі міста району перед переїздом в інший район. Скільки існує можливих варіантів сформувати план подорожі, який відповідає цим вимогам? □

Розв'язання. Заїхавши в район, турист має відвідати всі 6 міст, оскільки повторно заїжджати в район турист не планує. Для кожного району міста можна впорядкувати у $6!$ способів. Самі ж райони можна впорядкувати у $5!$ способів. Отже, згідно з правилом множення, ми спочатку вибираємо послідовність районів, а потім для кожного району — послідовність міст у районі, що дає $5!(6!)^5$ можливих варіантів. □

Вправа 1.3.2. П'ятеро друзів — Андрій, Богдана, Володимир, Ганна та Дмитро — у випадковий спосіб займають місця за круглим обіднім столом. Чому дорівнює ймовірність події A , що Богдана сидітиме поруч з Андрієм (неважливо, справа чи зліва)? □

Розв'язання. Спочатку потрібно визначити, який у нас простір елементарних подій Ω . Це — множина всіх можливих перестановок друзів за столом, тому $|\Omega| = 5! = 120$.

Подія A містить такі варіанти розміщення друзів, що Андрій сидить поруч із Богданою. Для підрахунку загального числа таких розміщень розбиймо весь експеримент на такі складові:

- спочатку вибираємо місце, за яким сидітиме Андрій. Усього таких варіантів розміщень є 5;

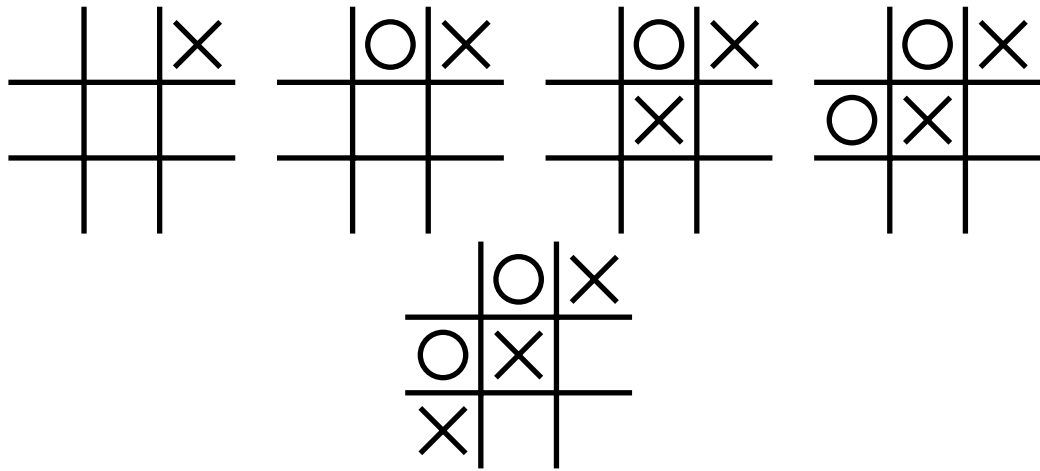


Рис. 1.3.1: Ілюстрація до Вправи 1.3.3

- потім вибираємо місце, за яким сидітиме Богдана. Оскільки вона повинна сидіти поруч з Андрієм, то для неї є 2 можливі варіанти, куди сісти — справа або зліва від Андрія;
- розсаджуємо трьох друзів, що залишилися. Таких варіантів існує $3 \cdot 2 \cdot 1 = 6$.

За правило множення, $|A| = 5 \cdot 2 \cdot 6 = 60$. Отже

$$\mathbb{P}(A) = \frac{|A|}{|\Omega|} = \frac{60}{120} = \frac{1}{2}.$$

□

Вправа 1.3.3. Розгляньмо гру в хрестики-нулики. Якщо гравці будуть ставити хрестики та нулики в абсолютно випадковому порядку, чому дорівнює ймовірність події A , що перший гравець виграє на третьому кроці (один із можливих варіантів розвитку гри, який веде до такого результату, зображено на Рис. 1.3.1)?

□

Розв'язання. Спочатку потрібно визначити, який у нас простір елементарних подій Ω . Якщо перший гравець виграє на третьому кроці, то обидва гравці встигли поставити 3 хрестики та 2 нулики. Усього на дошці 9 клітинок, і тому можна розглянути ізоморфну задачу — формування «слова» з символів X , O та пробілів. При цьому X та O розрізненні, оскільки різна послідовність хрестиків і нуликів визначає різні варіанти розвитку гри. Отже маємо задачу, аналогічну Вправі 1.2.13, де маємо «слово» з 6 літер, де «літера пробіл» повторюється 4 рази. Отже,

$$|\Omega| = \frac{9!}{4!}.$$

Розгляньмо тепер усі можливі варіанти, за яких перший гравець перемагає. Усього існує 8 можливостей виставити три хрестики в ряд (три горизонталі, три вертикалі і дві діагоналі). Як було встановлено вище, порядок слідування хрестиків має значення, тому для кожної можливості існує $3!$ варіантів розміщення трьох хрестиків.

Для нуликів існує $\binom{6}{2}$ варіантів вибору клітинок для заповнення та $2!$ варіантів розміщення двох нуликів по двох клітинках.

Згідно з правилом множення, загальне число варіантів розташування хрестиків і нуликів на дошці, які ведуть до перемоги першого гравця на третьому кроці, дорівнює $8 \cdot 3! \cdot \binom{6}{2} \cdot 2!$, а відтак ймовірність такої події A дорівнює

$$\mathbb{P}(A) = \frac{8 \cdot 3! \cdot \binom{6}{2} \cdot 2!}{\frac{9!}{4!}} \approx 0.095 .$$

□

Вправа 1.3.4. Розгляньмо неорієнтований граф із 5 вершинами. Яка ймовірність дістати цикл із 5 вершин, якщо у випадковий спосіб з'єднати між собою 5 пар вершин? На Рис. 1.3.2а–1.3.2б зображено ситуації, які відповідають цій умові, а на Рис. 1.3.2в — ситуацію, що не відповідає. □

Розв'язання. Простір елементарних подій Ω у цьому випадку містить усі можливі п'ятірки з можливих ребер графа. Усього в такому графі можна побудувати $\binom{5}{2}$ ребер (саме стільки існує *пар* вершин, які можна утворити з 5 наявних). Відповідно, загальне число п'ятірок, які можна вибрати з $\binom{5}{2}$ ребер, дорівнює

$$|\Omega| = \binom{\binom{5}{2}}{5} = \binom{\frac{5!}{2!3!}}{5} = \binom{10}{5} = 252 .$$

Загальне число тих п'ятірок ребер, які відповідають замкненим циклом із 5 вершин, можна дістати з розгляду перестановок чисел від 1 до 5, оскільки в будь-якій такій перестановці сусідні числа можна розглядати як вершини, з'єднані ребром. Наприклад, перестановка $(1, 2, 5, 3, 4)$ говорить, що спочатку потрібно з'єднати першу вершину з другою, потім другою з п'ятою і т.д, і замкнути цикл, з'єднавши четверту вершину з першою. Усього таких перестановок існує $5!$. Проте, *зсув* перестановки не змінює суті справи: $(1, 2, 5, 3, 4)$ і $(5, 3, 4, 1, 2)$ описують один і той самий граф. Отже один і той самий граф описує 5 варіантів однієї й тієї ж перестановки, наприклад, $(1, 2, 5, 3, 4)$, $(2, 5, 3, 4, 1)$, $(5, 3, 4, 1, 2)$, $(3, 4, 1, 2, 5)$, $(4, 1, 2, 5, 3)$. Відповідно, загальне число перестановок потрібно поділити на 5.

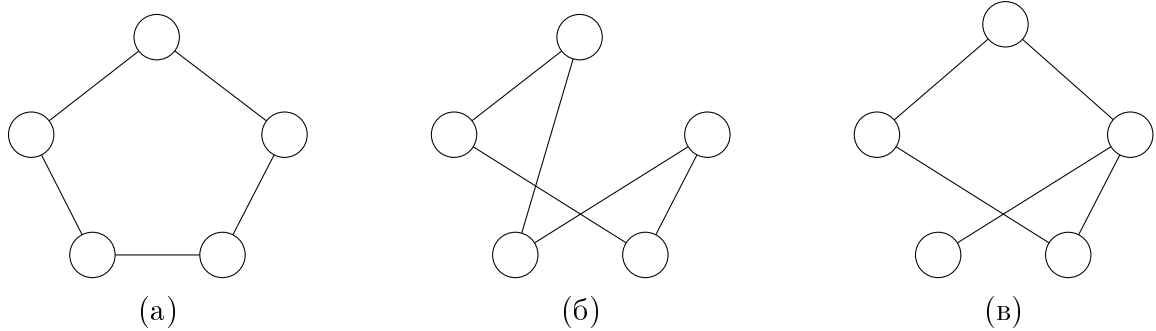


Рис. 1.3.2: Ілюстрація до Вправи 1.3.4

Понад те, перестановку можна записати ззаду наперед, не змінивши при цьому графа: $(1, 2, 3, 4, 5)$ і $(5, 4, 3, 2, 1)$ описують один і той самий граф. Отже додатково загальне число потрібно поділити ще й на 2.

Отже шукана ймовірність дорівнює

$$\mathbb{P}(A) = \frac{5!/(5 \cdot 2)}{252} = \frac{12}{252} \approx 0.048. \quad \square$$

Вправа 1.3.5. Нехай маємо $2n$ людей у групі. Скільки існує варіантів розбити їх на пари?

Наприклад, якщо група складається з 6 людей ($n = 3$), пронумерованих від 1 до 6, то можливим розбиттям на пари є $\{\{1, 5\}, \{2, 3\}, \{4, 6\}\}$. $\square$

Розв'язання. Якщо записати числа від 1 до $2n$ у деякому порядку, то таку послідовність чисел можна інтерпретувати як розбиття чисел на пари. Наприклад, розбиттю $\{\{1, 5\}, \{2, 3\}, \{4, 6\}\}$ відповідає послідовність $(1, 5, 2, 3, 4, 6)$. Усіх можливих перестановок із $2n$ чисел є $(2n)!$.

Серед цих перестановок є такі, що відповідають одному й тому ж розбиттю, наприклад, $(1, 5, 2, 3, 4, 6)$ і $(5, 1, 3, 2, 4, 6)$, оскільки можна у довільний спосіб переставляти числа в рамках кожної окремої пари. Для кожної пари є 2 варіанти перестановок чисел у ній, а всього таких пар є n , тому існує 2^n варіантів перестановок чисел у рамках кожної пари, на які потрібно поділити, щоб позбутися подвійного врахування однакових варіантів.

Еквівалентними також є перестановки виду $(1, 5, 2, 3, 4, 6)$ і $(2, 3, 4, 6, 1, 5)$, тобто такі, у яких самі пари змінено місцями. Усього існує $n!$ варіантів довільних перестановок пар між собою. Отже і на $n!$ потрібно поділити, щоб позбутися дублювань.

Остаточнo маємо

$$\frac{(2n)!}{2^n n!}$$

варіантів розбиття $2n$ людей на пари. $\square$

2 Комбінаторні доведення. Геометричні ймовірності. Вимірні множини

2.1 Основні теоретичні відомості

Задачі на цьому занятті покривають доволі різні теми. Частково поняття та твердження, потрібні для розв'язання задач, наведено в Розд. 1.1. Матеріал, потрібний для розв'язання задач, пов'язаних із вимірними множинами, наведено в Курсі лекцій, Розд. КЛ1.5.

Нехай маємо абстрактний простір елементарних подій Ω . Позначмо через $\mathcal{C}$ деякий клас (сім'ю) подій — тобто множину, яка містить інші множини. Природним є бажання мати можливість обчислювати ймовірності для якомога більшого класу подій. Зasadнича вимога замкненості класу $\mathcal{C}$ відносно основних теоретико-множинних операцій веде до поняття алгебри.

Визначення (КЛ1.5.1). Клас $\mathcal{F}$ підмножин деякого простору Ω 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 $A \cup B \in \mathcal{F}$.

□

Твердження (КЛ1.5.2). Із визначення алгебри випливають такі факти:

- (i) $\emptyset \in \mathcal{F}$;
- (ii) $\mathcal{F}$ замкнена відносно *скінченного* 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 *скінченного* числа перетинів:

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

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

Твердження (КЛ1.5.4). 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.

Ця алгебра доволі обмежена, адже вона:

- не містить множин із окремих чисел $\{x\}$, $x \in \mathbb{R}$, оскільки вони є результатом *зліченного* числа перетинів (Зауваження КЛ1.5.5);
- не є замкнутою відносно *зліченного* числа об'єднань і перетинів (Приклад КЛ1.5.7). Скажімо,

$$\bigcup_{n=2}^{\infty} \left(0; 1 - \frac{1}{n}\right] = (0; 1) \notin \mathcal{B}_0,$$

оскільки цей інтервал не містить своєї правої межі.

Потреба в використанні *зліченного* числа об'єднань і перетинів веде до поняття σ -алгебри.

Визначення (КЛ1.5.8). Клас $\mathcal{A}$ підмножин деякого простору Ω є σ -алгеброю¹ (σ -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 $\bigcup_{n=1}^{\infty} A_n \in \mathcal{A}$.

□

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

Визначення (КЛ1.5.9). Пару $(\Omega, \mathcal{A})$ називають *вимірним простором* (measurable space), а множини в σ -алгебрі $\mathcal{A}$ — *вимірними множинами* (measurable sets). □

Якщо простір Ω скінченний ($|\Omega| < \infty$) або злічений ($|\Omega| = \aleph_0$), то можна покласти $\mathcal{A} = 2^\Omega$, тобто розглянути як σ -алгебру множину всіх підмножин простору Ω . Якщо ж простір Ω незлічений ($|\Omega| = \mathfrak{c}$), то Приклад КЛ1.5.10

¹Літера σ походить від першої літери німецького слова *Summe*.

ілюструє, що існують такі підмножини, для яких неможливо визначити ймовірність з інтуїтивно зрозумілими властивостями.

Отже для коректного визначення ймовірності потрібно зосередитися на деякій σ -алгебрі подій, яка не дорівнює множині всіх підмножин простору елементарних подій. При цьому ми прагнемо, щоб у такій σ -алгебрі містилися всі події, які можуть становити практичний інтерес. Відтак ми розглянемо найменшу σ -алгебру, яка містить усі цікаві для нас множини.

Визначення (КЛ1.5.12). 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 $\mathcal{C}$ σ -алгеброю, яку він породжує (σ -algebra generated by $\mathcal{C}$), називають найменшу σ -алгебру, яка містить $\mathcal{C}$. Таку σ -алгебру позначають через $\mathcal{A}(\mathcal{C})$. $\square$

Виявляється, (σ) -алгебра, яку породжує деякий клас $\mathcal{C}$, не просто завжди існує, а й є єдиною (Лема КЛ1.5.13).

Також справедливі такі властивості.

Лема (КЛ1.5.14). 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 $\mathcal{A}(\mathcal{C}_2) = \mathcal{A}(\mathcal{C}_1)$.

За визначенням можна бачити, що $\mathcal{B}_0 = \mathcal{F}(\mathcal{I})$. Отже, нам потрібно дістати σ -алгебру, яку породжує $\mathcal{B}_0$. Виявляється, такою σ -алгеброю є так звана Борелева σ -алгебра.

Визначення (КЛ1.5.15). σ -алгебра $\mathcal{B} = \mathcal{A}(\mathcal{O})$, яку породжує клас $\mathcal{O}$ усіх відкритих підмножин $\mathbb{R}$, має назву *Борелевої σ -алгебри* (Borel σ -algebra)², а множини з неї називають *Борелевими множинами* (Borel sets). $\square$

Які множини є Борелевими?

Твердження (КЛ1.5.16). (i) Усі замкнені множини є Борелевими;

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

(iii) усі злічені множини є Борелевими;

²Еміль Борель (Félix Édouard Justin Émile Borel, 1871–1956) — французький математик і політик.

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

Понад те, виявляється, що Борелевими є навіть множини, які неможливо утворити зліченими теоретико-множинними операціями над інтервалами. Клас Борелевих множин є настільки широким, що в рамках теорії ймовірностей більшість із них ніколи не виринає на практиці.

Твердження (КЛ1.5.19). Борелеву σ -алгебру породжує алгебра $\mathcal{B}_0$: $\mathcal{B} = \mathcal{A}(\mathcal{B}_0)$.

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 (КЛ1.5.21). σ -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*, а множини, що їй належать, називають *Борелевими множинами*. $\square$

Аналогічно одновимірному випадку, можна показати, що k -вимірну Борелеву σ -алгебру породжують також:

- k -вимірні прямокутники $I_1 \times I_2 \times \dots \times I_k \in \mathbb{R}^k$, де кожний $I_j \subset \mathcal{I}$, $j = 1, \dots, k$;
- аналоги променів $(-\infty; x]$, а саме множини виду

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

Визначення (КЛ1.5.22). Вимірний простір $(\mathbb{R}^k, \mathcal{B}^k)$ називають *Борелевим простором* (Borel space). $\square$

2.2 Комбінаторні доведення

Існують математичні тотожності, доведення яких можна зробити або за допомогою методу математичної індукції, або виконанням алгебричних операцій. Проте в низці випадків такі тотожності можна довести, розглянувши кожний бік від знаку рівності як спосіб підрахунку числа варіантів завершення деякого експерименту.

Приклад 2.2.1. Наприклад, тотожність

$$\binom{n}{k} = \binom{n}{n-k}$$

можна довести, безпосередньо використавши визначення біномного коефіцієнту (1.2.1):

$$\binom{n}{k} = \frac{n!}{k!(n-k)!} = \frac{n!}{(n-k)!(n-(n-k))!} = \binom{n}{n-k}.$$

Проте альтернативно можна було б сказати, що вирази і зліва, і справа від знаку рівності відповідають задачі формування комітету в k осіб із-поміж n наявних. Лівий вираз описує кількість можливих способів вибрати членів комітету (без повторень і без збереження порядку), а правий вираз описує кількість можливих способів вибрати осіб, які залишаться за межами комітету. Очевидно, результатом цих двох виборів буде одна й та сама ситуація: k осіб будуть членами комітету, а решта $n - k$ залишаться поза його межами. $\square$

Приклад 2.2.2. У Вправі 1.3.5 ми показали, чому дорівнює загальне число варіантів розбиття групи з $2n$ осіб на n пар. Виявляється, існує така тотожність:

$$\frac{(2n)!}{2^n n!} = (2n-1)(2n-3) \dots 3 \cdot 1. \quad (2.2.1)$$

Довести її можна за методом математичної індукції:

- *база індукції*: якщо $n = 2$, то

$$\frac{4!}{2^2 \cdot 2!} = 3 = 3 \cdot 1;$$

- *індукційний перехід*: нехай (2.2.1) виконується для $n - 1$. Тоді

$$\begin{aligned} \frac{(2n)!}{2^n n!} &= \frac{(2(n-1))!}{2^{n-1}(n-1)!} \cdot \frac{(2n)(2n-1)}{2n} \\ &= (2n-3) \dots 3 \cdot 1 \cdot \frac{(2n)(2n-1)}{2n} \\ &= (2n-1)(2n-3) \dots 3 \cdot 1. \end{aligned}$$

З іншого боку, ми могли б сказати, що справа також стоїть кількість варіантів утворити n пар осіб. Справді, для першої особи існує $2n - 1$ варіантів вибрати пару. Оскільки дві особи вже мають пару, їх можна вилучити з загального розгляду, тому залишається $2n - 2$ осіб. Для чергової особи

існує $2n - 3$ варіантів вибрати пару. Нову пару також можна вилучити з розгляду, тому залишається $2n - 4$ осіб. Цей процес можна продовжувати і дістати в результаті вираз $(2n - 1)(2n - 3) \dots 3 \cdot 1$. $\square$

Розгляньмо декілька вправ на комбінаторні доведення.

Вправа 2.2.3. Доведімо комбінаторно тотожність

$$n \binom{n-1}{k-1} = k \binom{n}{k} = (n-k+1) \binom{n}{k-1}, \quad k \leq n. \quad (2.2.2)$$

$\square$

Розв'язання. Розгляньмо експеримент із вибору команди в k осіб із-поміж n наявних та призначення її капітана. Цей експеримент можна розбити на складові у три різні способи:

- спочатку визначити капітана (n варіантів), і помістити його в команду, а потім добрати решту $k - 1$ членів команди з-поміж $n - 1$ осіб ($\binom{n-1}{k-1}$ варіантів);
- вибрати членів команди ($\binom{n}{k}$ варіантів), а потім серед них вибрати капітана (k варіантів);
- вибрати членів команди, *які не є капітаном* ($\binom{n}{k-1}$ варіантів), а потім вибрати капітана серед $n - (k - 1) = n - k + 1$ осіб, що залишилися.

Правило множення (Твердження 1.2.1) для всіх трьох випадків дає вирази в (2.2.2). $\square$

Вправа 2.2.4 (Тотожність Вандермонда³). Доведімо комбінаторно тотожність

$$\binom{m+n}{k} = \sum_{j=0}^k \binom{m}{j} \binom{n}{k-j}. \quad (2.2.3)$$

$\square$

³Названа так на честь французького математика Александра-Теофіля Вандермонда (Alexandre-Theophile Vandermonde, 1735–1796), який довів її у 1792 р., проте вперше про цю тотожність згадка є в працях китайського математика Чжу Шицзе за 1303 р.

Розв'язання. Доведення цього твердження методом математичної індукції або алгебрично видається задачею нетривіальною.

Розгляньмо експеримент із формування команди на олімпіаду з математики серед студентів двох різних факультетів, де m студентів представляють факультет 1, а n — факультет 2. Тоді $\binom{m+n}{k}$ дає загальне число варіантів формування команди в k осіб.

Розгляньмо тепер ситуацію з іншого боку. Якщо в команді є k осіб, то серед них повинно бути j представників факультету 1 і $k-j$ представників факультету 2, $j = 0, 1, \dots, k$. Існує $\binom{m}{j}$ способів вибрати j представників факультету 1 і $\binom{n}{k-j}$ способів вибрати $k-j$ представників факультету 2. За правилом множення (Твердження 1.2.1) загальне число варіантів сформування команду, щоб у ній було j представників факультету 1, дорівнює $\binom{m}{j} \cdot \binom{n}{k-j}$.

Але цей результат дає тільки кількість варіантів для деякого конкретного $0 \leq j \leq k$. Сума відповідних варіантів для кожного j дає вираз справа від знака рівності в (2.2.3). $\square$

2.3 Метод відбиття

Окремий клас комбінаторних задач можна розв'язати не тільки шляхом застосування методів обчислення класичної ймовірності, розглянутих у Розділі 1, а й за допомогою так званого *методу відбиття* (reflection method)⁴.

Вправа 2.3.1 (Задача Бертрана про вибори (Bertrand's ballot problem)⁵). У виборах беруть участь двоє кандидатів, A і B . Якщо за кандидата A подадуть n голосів, а за B — m голосів, $m < n$ (тобто кандидат A переміг), а голоси подаватимуть *послідовно і у випадковому порядку*, то чому дорівнює ймовірність, що кандидат A буде лідерувати протягом усього часу? $\square$

Для розв'язання цієї задачі розгляньмо модель, відому як *випадкове блукання* (random walk). Починаючи з точки 0, деяка частинка зміщується на одиницю вправо або на одиницю вліво з однаковою ймовірністю. Ця модель надзвичайно поширена в теорії випадкових процесів і часових рядів. Наразі ми не будемо аналізувати властивості випадкового блукання як такого, а натомість просто розглянемо можливі його траєкторії.

Зображувати траєкторії руху частинки зручно в двовимірному просторі, де перша координата відповідає поточному числу виконаних кроків n , а друга — поточній позиції частинки після n кроків, $S(n)$ (Рис. 2.3.1a). Для всіх натуральних чисел $n, k \in \mathbb{N}$ визначмо $N_n(k)$ як число можливих *різних*

⁴Цей метод у 1887 р. запропонував французький математик Дезіре Андре (Desiré André, 1840–1917).

⁵Названа так на честь французького математика Жозефа Бертрана (Joseph Louis Francois Bertrand, 1822–1900), який у 1887 р. *повторно довів* її, хоча вперше її опублікував у 1878 р. англійський математик Вільям Витворт (William Allen Whitworth, 1840–1905).

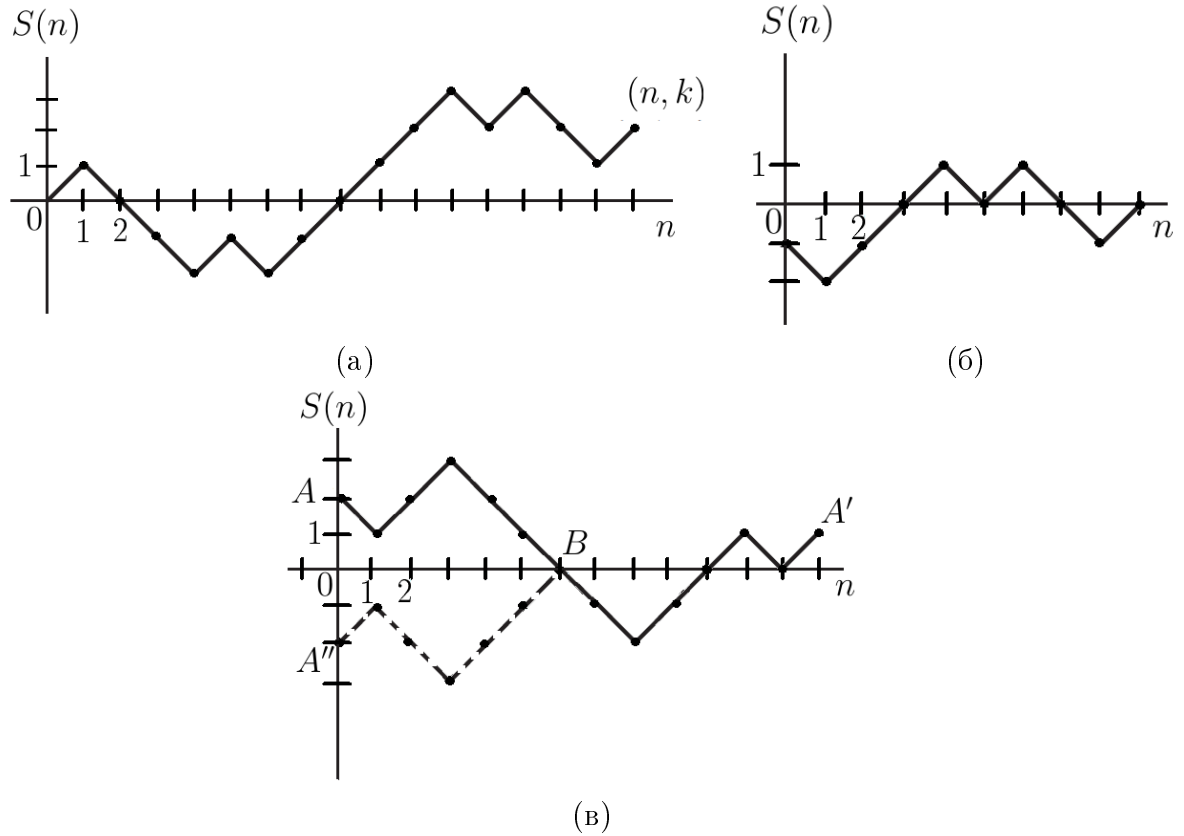


Рис. 2.3.1: Ілюстрація до Вправи 2.3.1: (а) можлива траєкторія частинки під час випадкового блукання; (б) конкретна ілюстрація траєкторії для Лема 2.3.2; (в) ілюстрація до Твердження 2.3.3

траєкторій від точки $(0, 0)$ до точки (n, k) , тобто від початку руху до позиції $k = S(n)$ на n -ому кроці. Аналогічно можна визначити $N_n^+(k)$ як число таких траєкторій, які постійно лежать над віссю абсцис, тобто для яких $S(n) > 0$ для всіх $n > 0$.

Частинка може рухатися в обидва боки з однаковою ймовірністю, усі траєкторії є рівноймовірними, а отже застосовне Визначення КЛ1.3.1. Загальне число всіх можливих траєкторій можна обчислити комбінаторно.

Лема 2.3.2. Нехай $a, a', b, b' \in \mathbb{Z}$, $0 \leq a < a'$. Тоді число можливих траєкторій від точки (a, b) до точки (a', b') залежить тільки від різниць у координатах $n = a' - a$, $k = b' - b$ і дорівнює $N_n(k) = \binom{n}{\frac{n+k}{2}}$.

Доведення. Нехай траєкторія від точки (a, b) до точки (a', b') передбачає r кроків у додатному напрямку («угору») та s кроків у від'ємному напрямку («униз»). Очевидно, що загальне число кроків є $n = r + s$, а кінцева позиція точки буде $S(n) = b + k$, де $k = r - s$. Звідси випливає

$$r = \frac{n + k}{2}, \quad s = \frac{n - k}{2}.$$

Наприклад, на Рис. 2.3.16 $(a, b) = (0, -1)$, $(a', b') = (9, 0)$, $r = 5$, $s = 4$, $n = 9$, $k = 5 - 4 = 1$, $S(9) = -1 + 1 = 0$.

Кроки «вгору» і «вниз» можна впорядкувати в будь-який спосіб, іншими словами, існує $\binom{n}{r}$ варіантів вибрати r додатних кроків із n можливих, решта $n - r = s$ автоматично будуть від'ємними.

Якщо покласти $(a, b) = (0, 0)$, то $(a', b') = (n, k)$, а тому

$$N_n(k) = \binom{n}{r} = \binom{n}{\frac{n+k}{2}}. \quad \square$$

Метод відбиття можна сформулювати так.

Твердження 2.3.3 (Метод відбиття). Нехай $a, a', b, b' \in \mathbb{Z}$, $0 \leq a < a'$, $b, b' > 0$ (тобто розглядаємо тільки траєкторії, які починаються і закінчуються вище нуля). Тоді число можливих траєкторій від точки (a, b) до точки (a', b') , які *перетинають* (або *торкають*) вісь абсцис, дорівнює числу всіх траєкторій від точки $(a, -b)$ до точки (a', b') .

Це число залежить тільки від $n = a' - a$ і $m = b' + b$ і дорівнює $N_n(m)$.

Доведення. Розгляньмо будь-яку траєкторію від точки (a, b) до точки (a', b') , яка перетинає (або торкає) вісь абсцис. Повинен існувати *найменший* індекс $1 \leq j \leq n - 1$, для якого $S(j) = 0$, тобто $S(0) \equiv b > 0$, $S(1) > 0$, $\dots$, $S(j - 1) > 0$, а $S(j) = 0$. Позначмо такі точки:

- $A = (a, b)$ — початкова точка траєкторії;
- $B = (a + j, 0)$ — точка першого дотику до осі абсцис;
- $A' = (a', b')$ — кінцева точка траєкторії;
- $A'' = (a, -b)$ — дзеркальне відображення точки A відносно осі абсцис.

Наприклад, для Рис. 2.3.1в маємо $A = (0, 2)$, $A' = (13, 1)$, $B = (6, 0)$, $A'' = (0, -2)$.

Для будь-якої траєкторії, яка торкає вісь абсцис у точці $(j, 0)$, можна збудувати траєкторію, частина якої до j -го кроку буде дзеркальним відображенням початкової траєкторії (штрихова лінія на Рис. 2.3.1в). Згідно з Лемою 2.3.2, існує $N_n(b' - (-b)) = N_n(m)$ траєкторій від точки $(a, -b)$ до точки (a', b') , а отже існує така сама кількість траєкторій від точки (a, b) до точки (a', b') , що перетинають (або торкають) вісь абсцис. $\square$

Використовуючи ці результати, можемо довести основне твердження.

Теорема 2.3.4 (Теорема про вибори (The Ballot Theorem)). Число можливих траєкторій від точки $(0, 0)$ до точки (n, k) , для яких $S(j) > 0$ для всіх $j = 1, \dots, n$, дорівнює

$$N_n^+(k) = \frac{k}{n} N_n(k). \quad (2.3.1)$$

Доведення. Оскільки траєкторія починається з точки $(0, 0)$, і на кожному кроці вона додатна, перший крок був «угору», і тому $S(1) = 1$. Згідно з Лемою 2.3.2, загальне число таких траєкторій дорівнює $N_{n-1}(k-1)$. А з-посеред цих траєкторій число таких траєкторій, які перетинають (або торкаються) вісь абсцис, згідно з Твердженням 2.3.3, дорівнює $N_{n-1}(k+1)$. Отже число траєкторій, які *не перетинають (і не торкаються)* осі абсцис, дорівнює

$$N_n^+(k) = N_{n-1}(k-1) - N_{n-1}(k+1) = \binom{n-1}{\frac{n+k}{2}-1} - \binom{n-1}{\frac{n+k}{2}}.$$

Згідно з (2.2.2),

$$\binom{n-1}{k-1} = \frac{k}{n} \binom{n}{k}, \quad \binom{n}{k} = \frac{n-k+1}{k} \binom{n}{k-1},$$

а тому

$$\begin{aligned} \binom{n-1}{\frac{n+k}{2}-1} &= \frac{n+k}{2n} \binom{n}{\frac{n+k}{2}}, \\ \binom{n-1}{\frac{n+k}{2}} &= \frac{(n-1) - \frac{n+k}{2} + 1}{\frac{n+k}{2}} \binom{n-1}{\frac{n+k}{2}-1} = \frac{(n-1) - \frac{n+k}{2} + 1}{\frac{n+k}{2}} \cdot \frac{n+k}{2n} \binom{n}{\frac{n+k}{2}}, \end{aligned}$$

отже

$$N_n^+(k) = \left(\frac{n+k}{2n} - 2 \frac{n - \frac{n+k}{2}}{n+k} \cdot \frac{n+k}{2n} \right) \binom{n}{\frac{n+k}{2}} = \frac{k}{n} \binom{n}{\frac{n+k}{2}}.$$

□

Використовуючи результат Теорема 2.3.4, нарешті можемо визначити відповідь для Вправи 2.3.1.

Розв'язання. Можемо розглянути випадкове блукання частинки, де k позначає поточну кількість поданих голосів, $S(k)$ — різницю між кількістю голосів, поданих за кандидата A та кандидата B на k -ому кроці. Наприклад, якщо після подання 7 голосів кандидат A має 5 голосів, а кандидат B — 2 голоси, маємо $S(7) = 5 - 2 = 3$.

Подію «кандидат A лідує на кожному кроці голосування» складають усі можливі траєкторії частинки, що лежать над віссю абсцис. Таких траєкторій, згідно з Лемою 2.3.4, $N_{n+m}^+(n-m) = \frac{n-m}{n+m} N_{n+m}(n-m)$ штук. Загальне число всіх можливих траєкторій дорівнює, згідно з Лемою 2.3.2, $N_{n+m}(n-m)$ штук, тобто $|\Omega| = N_{n+m}(n-m)$. Згідно з класичним визна-

ченням імовірності,

$$\begin{aligned}\mathbb{P}(\text{«кандидат } A \text{ постійно лідерує»}) &= \frac{n-m}{n+m} N_{n+m}(n-m) \cdot \frac{1}{N_{n+m}(n-m)} \\ &= \frac{n-m}{n+m}.\end{aligned}$$

□

Розгляньмо додаткову задачу, розв'язання якої передбачає використання цього ж методу.

Вправа 2.3.5. У маршрутне таксі протягом виконання маршруту по черзі у випадковому порядку сідають 100 пасажирів. У 75 з них є монети по 5 грн, у 25 — по 10 грн. Проїзд коштує 5 грн. На початку маршруту у водія немає вільних коштів.

Чому дорівнює ймовірність події $A = \text{«у водія завжди буде принаймні 5 грн для видачі решти»}$?

Чому дорівнює ймовірність події $B = \text{«у водія завжди будуть гроші для решти»}$? □

Розв'язання. Цю задачу можна розглянути в термінах Вправи 2.3.1, розглянувши блукання частинки. Поточну позицію частинки зіставмо числу наявних у водія монет по 5 грн після заходу пасажира k . Кожний новий пасажир або дає 5 грн, тобто $S(k) = S(k-1) + 1$, або дає 10 грн, і тоді водій повинен видати решту в 5 грн, тобто $S(k) = S(k-1) - 1$.

Кінцева точка частинки дорівнює $(100, 75 - 25) = (100, 50)$, а отже число можливих траєкторій від точки $(0, 0)$ до точки $(100, 50)$, згідно з Лемою 2.3.2, дорівнює

$$|\Omega| = N_{100}(50) = \binom{100}{\frac{100+50}{2}} = \binom{100}{75}.$$

Подія A передбачає, що у водія завжди залишається принаймні 5 грн для видачі решти, і вона відповідає траєкторіям, для яких $S(k) \geq 1$, $k = 1, \dots, 100$, іншими словами, які ніколи не перетинають (і не торкаються) осі абсцис. Число таких траєкторій від точки $(0, 0)$ до точки $(100, 50)$, згідно з Теоремою 2.3.4, дорівнює

$$|A| = N_{100}^+(50) = \frac{50}{100} \binom{100}{75} = \frac{1}{2} \binom{100}{75}.$$

Отже

$$\mathbb{P}(A) = \frac{|A|}{|\Omega|} = \frac{1}{2}.$$

Подія B передбачає, що водій завжди матиме можливість видати решту, хоча в нього необов'язково завжди повинні бути на руках гроші (скажімо, він віддав останню монету в 5 грн, а наступний пасажир відразу сплатив йому чергові 5 грн). Така подія відповідає траєкторіям, для яких $S(1) = 1$, $S(k) \geq 0$, $k = 2, \dots, 100$.

Щоб звести цю задачу до попередньої, ми можемо *зсунути* систему координат на одиницю вліво і вниз так, щоб рух відбувався від точки $(0, 0)$ до точки $(101, 51)$, і щоб вісь абсцис не перетинали (і не торкали). Число таких траєкторій, згідно з Теоремою 2.3.4, дорівнює, використовуючи (2.2.2),

$$|B| = N_{101}^+(51) = \frac{51}{101} \binom{101}{\frac{101+51}{2}} = \frac{51}{101} \binom{101}{76} = \frac{51}{101} \frac{101}{76} \binom{100}{75} = \frac{51}{76} \binom{100}{75}.$$

Отже

$$\mathbb{P}(B) = \frac{|B|}{|\Omega|} = \frac{51}{76} \binom{100}{75} \cdot \frac{1}{\binom{100}{75}} \approx 0.671.$$

□

2.4 Геометричні ймовірності

Як зазначалося в Розд. 1.1, визначення класичної ймовірності можна *спробувати* узагальнити на випадок незліченного простору Ω як відношення міри події A до міри всього простору:

$$\mathbb{P}(A) = \frac{m(A)}{m(\Omega)}, \quad (2.4.1)$$

де $m(\cdot)$ — міра множини в k -вимірному просторі $\mathbb{R}^k$ (довжина прямої, площа пласкої фігури, об'єм тіла і т.д.).

Розгляньмо декілька задач на застосування такого спрощеного *геометричного* поняття ймовірності⁶. Почнімо розгляд із задачі, яка історично була однією з перших у цьому класі.

Вправа 2.4.1 (Задача про голку Бюффона (Buffon's needle problem⁷)). Нехай паркетна підлога складається з паралельних дошок однакової ширини d . На підлогу падає голка довжиною $l \leq d$. Чому дорівнює ймовірність події A = «голка перетинатиме межу двох дошок»? □

⁶Як ми далі розглядатимемо в цьому курсі, ці задачі фактично використовують поняття рівномірного розподілу випадкового вектора, але суто з міркувань шани традиції ми коротко розглянемо ці приклади саме як приклади на геометричні ймовірності.

⁷Названа на честь французького натураліста та математика Жоржа-Луї Леклерка, графа де Бюффона (Georges-Louis Leclerc, Comte de Buffon, 1707–1788), який уперше розв'язав її в 1777 р.

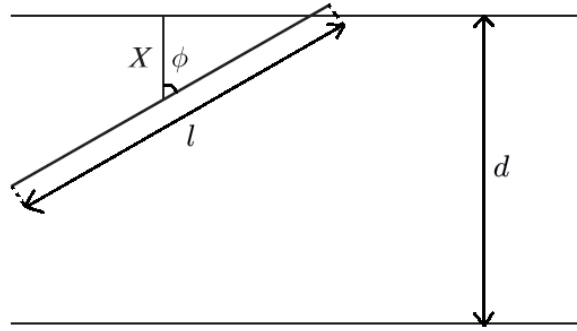


Рис. 2.4.1: Ілюстрація до Вправи 2.4.1: одне з можливих розташувань випадково пущеної голки на паркетній підлозі

Розв'язання. Голка довжиною l може приземлитися або в межах деякої дошки, або перетнути межу двох дошок. Останній варіант зображено на Рис. 2.4.1. Позицію голки можна описати двома змінними:

- відстань X від середини голки до найближчої межі дошок;
- кут ϕ між голкою та перпендикуляром, опущеним із середини голки на межу між двома дошками.

Подія A станеться в тому випадку, якщо гіпотенуза утвореного прямокутного трикутника буде меншою від $l/2$, тобто якщо

$$\frac{X}{\cos \phi} < \frac{l}{2} \Rightarrow X < \frac{l}{2} \cos \phi. \quad (2.4.2)$$

У такій моделі цієї ситуації випадковий експеримент із кидання голки завершується парою (X, ϕ) . Отже простором елементарних подій буде квадрат $\Omega = [0; \frac{d}{2}] \times [0; \frac{\pi}{2}]$, оскільки відстань до найближчої дошки не може перевищувати $\frac{d}{2}$, а кут не може перевищувати $90^\circ = \frac{\pi}{2}$. Звідси випливає

$$m(\Omega) = \frac{\pi d}{4}.$$

Подія A — це множина точок у відповідному квадраті, які задовольняють (2.4.2). Її зображено на Рис. 2.4.2. Площу зафарбованої області можна порахувати за допомогою інтегралу

$$m(A) = \int_0^{\frac{\pi}{2}} \frac{l}{2} \cos \phi \, d\phi = \frac{l}{2} (\sin \phi) \Big|_0^{\frac{\pi}{2}} = \frac{l}{2}.$$

Отже, відповідно до (2.4.1) маємо:

$$\mathbb{P}(A) = \frac{m(A)}{m(\Omega)} = \frac{l}{2} \cdot \frac{4}{\pi d} = \frac{2l}{\pi d}.$$

□

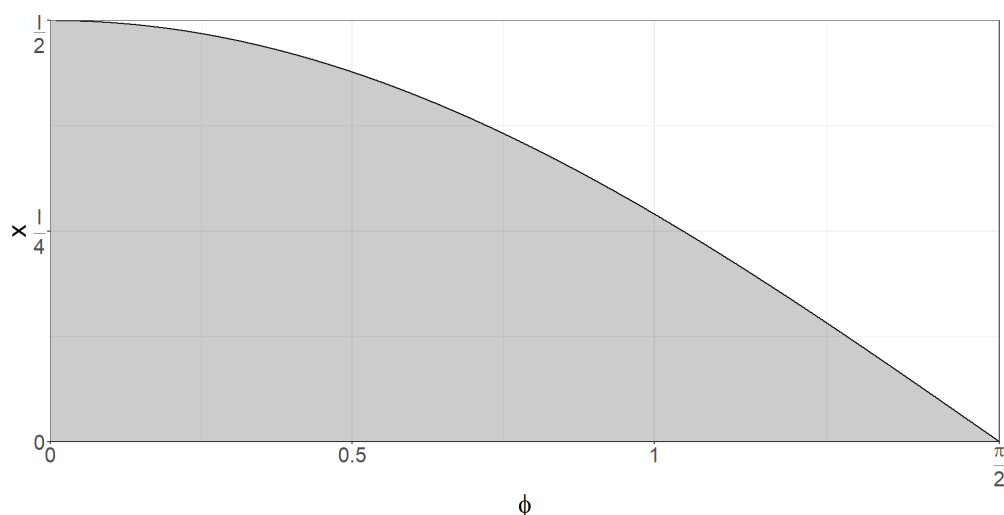


Рис. 2.4.2: Ілюстрація до Вправи 2.4.1: множина A (зафарбована область)

Вправа 2.4.2. Чоловік і дружина здійснюють собі покупки, використовуючи одну й ту саму картку, тільки на різних вебсайтах. На картці лежать 10 000 грн. Чому дорівнює ймовірність події A = «в одного з них не пройде оплата через нестачу коштів на картці»? $\square$

Розв'язання. Позначмо через x та y кількість коштів, які в підсумку витратять чоловік і дружина відповідно. Оскільки кожний із подружжя може розраховувати на використання до 10 000 грн, простором Ω буде квадрат стороною 10 000: $\Omega = [0; 10\,000] \times [0; 10\,000]$.

Подія A відповідає ситуації, коли $x + y > 10\,000$, тобто це — половина квадрата, отже

$$\mathbb{P}(A) = \frac{1}{2}.$$

$\square$

Вправа 2.4.3 (Задача про зустріч). Двоє друзів умовилися зустрітися випити кави, але в силу невизначеності своїх робочих графіків вони вмовилися зустрітися «десь» між дванадцятою й першою годиною. Обидва друга не мають багато часу на очікування, тому ухвалено рішення, що той, хто прийде першим, очікуватиме іншого протягом 15 хвилин, але не пізніше 1-ої години дня. Якщо за цей час ніхто не з'явиться, то зустріч не відбудеться. Чому дорівнює ймовірність події A , що зустріч таки відбудеться? $\square$

Розв'язання. Для цього спочатку визначмо простір елементарних подій. «Момент часу», у який один із друзів може з'явитися на каву, визначмо як

точку на відрізку довжиною 1⁸ (що відповідає проміжку в одну годину). Тоді простір елементарних подій є $\Omega = [0; 1] \times [0; 1]$, де дві координати, x та y , відповідають моментам появи двох друзів на каву відповідно.

Тоді подія $A = \text{«зустріч відбулася»}$ являє собою смугу, зображену на Рис. 2.4.3. Справді, зустріч станеться, якщо один друг прийде в той час, коли інший друг обідає, тобто протягом проміжку в 15 хвилин, або $1/4$ години, від приходу іншого друга:

$$x \leq y \leq x + \frac{1}{4} \Rightarrow \begin{cases} y - x \leq \frac{1}{4} \\ y \geq x \end{cases}$$

або

$$y \leq x \leq y + \frac{1}{4} \Rightarrow \begin{cases} x - y \leq \frac{1}{4} \\ x \geq y \end{cases}.$$

Компактніше це можна записати як те, що *відстань* (у часі) між двома моментами прибуття друзів не перевищує $1/4$ години:

$$A = \left\{ (x, y)^T \in [0; 1] \times [0; 1] : |x - y| \leq \frac{1}{4} \right\}.$$

Згідно з (2.4.1),

$$\mathbb{P}(A) = \frac{m(A)}{m(\Omega)} = \frac{m(A)}{1} = m(A).$$

Із геометричних міркувань, шляхом вирізання двох прямокутних трикутників із катетами в $3/4$ із цілого квадрата, маємо:

$$\mathbb{P}(A) = m(A) = 1 - 2 \cdot \frac{1}{2} \cdot \left(\frac{3}{4}\right)^2 = \frac{7}{16}.$$

□

2.5 σ -алгебри подій

Вправа 2.5.1. Нехай $\Omega = \{1, 2, 3, 4\}$. Нехай $\mathcal{C} = \{\{1\}, \{2\}\}$. Яку алгебру $\mathcal{F}$ (вона ж буде σ -алгебра) породжує цей клас множин? □

⁸Звісно, годину часу можна поділити на скінченну, але велику кількість дискретних моментів часу, наприклад, наносекунд. Проте тоді обчислення стають занадто громіздкими, і тому, як це часто роблять у природничих і соціальних науках, використовують незліченні множини як апроксимацію дійсності, із якою працювати простіше.

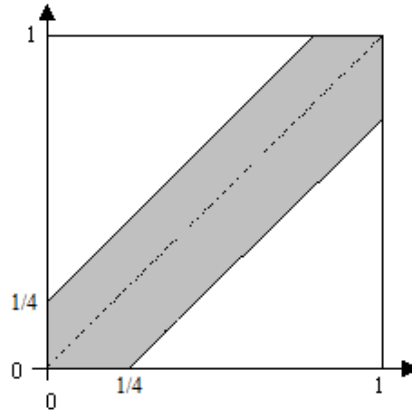


Рис. 2.4.3: Ілюстрація до Вправи 2.4.3: імовірнісний простір і подія A , якій відповідає зафарбована область

Розв'язання. За визначенням (σ) -алгебри, вона повинна містити як самі множини $\{1\}$ та $\{2\}$, так і їхні доповнення — $\{2, 3, 4\}$ та $\{1, 3, 4\}$ відповідно. Також алгебрі повинні належати всі можливі об'єднання цих множин. У результаті доходимо висновку, що

$$\mathcal{F} = \{\emptyset, \{1\}, \{2\}, \{1, 2\}, \{3, 4\}, \{1, 3, 4\}, \{2, 3, 4\}, \Omega\}.$$

□

Вправа 2.5.2. Доведіть, що такі множини належать Борелевій σ -алгебрі у двовимірному просторі $\mathcal{B}^2$:

- (i) $A = \{(x, y)^\top \in \mathbb{R}^2 : x^2 + y^2 \leq 1\}$;
- (ii) $B = \mathbb{Q} \times \mathbb{Q}$;
- (iii) $C = (0; 1) \times [2; 3]$;
- (iv) $D = \{(x, y)^\top \in [0; 1] \times [0; 1] : x = y\}$ (діагональ одиничного квадрата).

□

Розв'язання. (i) Борелева σ -алгебра містить усі замкнені множини, тому $A \in \mathcal{B}^2$;

(ii) Борелева σ -алгебра містить усі зліченні множини, тому $B \in \mathcal{B}^2$;

(iii) квадрат C не є ні відкритою, ні замкненою множиною, оскільки він містить тільки нижню свою межу, та й то не всю (не містить точок

$(0, 2)$ і $(1, 2)$, оскільки 0 і 1 не належать інтервалу $(0; 1)$). Тому не можна відразу сказати, що він є Борелевим.

За аналогією з тим, як кожному одноелементну множину $\{x\}$ у просторі $\mathbb{R}$ можна подати як злічений перетин

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

ми можемо розглянути злічений перетин, наприклад,

$$\bigcap_{n=1}^{\infty} \left(1 - \frac{1}{n}; 1 \right] \times (2; 3],$$

результатом якого є інтервал $\{(x, y) : x = 1, 2 < y \leq 3\}$.

Оскільки Борелевими в $\mathbb{R}$ є також відкриті інтервали та інтервали, замкнені зліва і відкриті справа, у схожий спосіб можна утворити інтервал $\{(x, y) : 0 < x < 1, y = 2\}$. Остаточню шуканий прямокутник можна подати як об'єднання такого інтервалу та відкритої множини $(0; 1) \times (2; 3)$;

- (iv) довести відповідний факт можна, якщо довести, що *доповнення* діагоналі є Борелевим. Справді, розгляньмо два числа $x \neq y$. Тоді можна підібрати два неперетинні пів інтервали I_x та I_y такі, які містять відповідні точки і мають *раціональні* межі.

Оскільки множина декартових добутків пів інтервалів породжує Борелеву σ -алгебру на площині, $I_x \times I_y \in \mathcal{B}^2$. Але цей прямокутник не перетинається з діагоналлю. Оскільки множина $\mathbb{Q}$ є зліченною, існує зліченне число таких прямокутників, (зліченне) об'єднання яких є доповненням діагоналі. Оскільки зліченне об'єднання Борелевих множин є Борелевим, Борелевим також є його доповнення, тобто шукана діагональ.

□

2.6 Додаткові задачі

Вправа 2.6.1. Доведіть комбінаторну тотожність Ферма (Fermat's combinatorial identity)⁹:

$$\binom{n}{k} = \sum_{i=k}^n \binom{i-1}{k-1}, \quad n \geq k. \quad (2.6.1)$$

⁹Названа так на честь французького математика П'єра де Ферма (Pierre de Fermat, 1607–1665).

□

Розв'язання. Вираз зліва визначає число підмножин розміром k , які можна вибрати з-посеред натуральних чисел від 1 до n . Скільки існує підмножин таких, що найбільший їхній елемент дорівнює деякому i ?

Найменше значення, якого може набувати i , дорівнює k . Існує всього одна така множина — $\{1, \dots, k\}$. Можна сказати, що існує $\binom{k-1}{k-1}$ варіантів вибору $k - 1$ чисел, менших від k , із-посеред $k - 1$ можливих варіантів.

Для $i = k + 1$ існує k різних варіантів утворення множин. Оскільки $k + 1$ обов'язково належить множині, нам потрібно вибрати $k - 1$ чисел із-поміж k наявних, тобто $k = \binom{k}{k-1}$.

За цією ж логікою, якщо $i = k + 2$, існує $\binom{k+1}{k-1}$ варіантів утворення відповідних множин.

Якщо додати всі можливі варіанти утворення множин розміром k , де $i = k, \dots, n$ є найбільшим числом, то ми і дістанемо вираз (2.6.1). □

Вправа 2.6.2. Машини їдуть автомагістраллю, зберігаючи відстань 50 м одна від одної. Кожна машина має довжину 4 м. На найближчому надземному переході висить велика бурулька. Якщо вона впаде на якусь машину або на відстані до 10 метрів попереду неї, станеться ДТП. Чому дорівнює ймовірність події A , що ДТП таки станеться? □

Розв'язання. У цьому випадку простір елементарних подій — це відрізок в $\mathbb{R}$, який відповідає відстані між переднім бампером однієї машини до переднього бампера іншої, тобто відрізок довжиною $50 + 4 = 54$ метри, отже $|\Omega| = |[0; 54]| = 54$. Подія A полягає в тому, що бурулька впаде на відрізок від 0 до $4 + 10 = 14$ метрів, а отже

$$\mathbb{P}(A) = \frac{14}{54} \approx 0.259.$$

□

Вправа 2.6.3. Пілот компанії з вантажних авіап перевезень щодня здійснює вильоти в міста, розташовані не далі, як 200 км від аеродрому. Щоранку пілот заправляє літак паливом, достатнім для виконання польоту загальною довжиною 450 км.

Сьогодні вранці пілот забув заправити літак. Чому дорівнює ймовірність, що для сьогоднішнього польоту йому вистачить палива? □

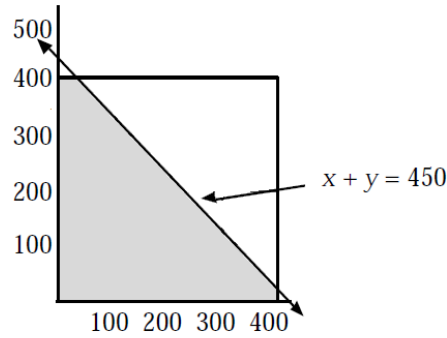


Рис. 2.6.1: Ілюстрація до Вправи 2.6.3: імовірнісний простір і подія A , якій відповідає зафарбована область

Розв'язання. Очевидно, паливо закінчиться в тому випадку, якщо сумарна відстань польотів за вчора і за сьогодні перевищить 450 км. Позначмо через x відстань нальоту за вчора, а через y — відстань, яку потрібно подолати сьогодні.

Простором Ω буде квадрат, кожна сторона якого дорівнює 400 км, оскільки міста розташовано не далі як на 200 км, тому максимально можливий політ може бути на відстань 200 км в обидва боки. Відповідно, $|\Omega| = 400^2$ км².

Множина A містить усі точки (x, y) такі, що $x + y \leq 450$ (Рис. 2.6.1), отже $|A| = 400^2 - \frac{1}{2}350^2$ км², а

$$\mathbb{P}(A) = 1 - \frac{350^2}{2 \cdot 400^2} = \frac{79}{128} \approx 0.617. \quad \square$$

Вправа 2.6.4. Чому у Вправі 2.3.5 дорівнює ймовірність події C , що у водія завжди знаходитиметься решта, якщо в нього з самого початку було 5 грн? $\square$

Розв'язання. $\mathbb{P}(C)$ можна дістати, якщо розглянути траєкторії, які рухаються з точки $(0, 1)$ до точки $(100, 51)$. Потужність простору $|\Omega|$ у цьому випадку не зміниться, оскільки згідно з Лемою 2.3.2 принциповою є тільки різниця між початковою і кінцевою точкою.

Зсувом системи координат на одиницю вліво знову зводимо задачу до підрахунку числа невід'ємних траєкторій від точки $(0, 0)$ до точки $(101, 51)$. Понад те, у цьому випадку перший крок обов'язково повинен бути додатний $S(1) = 1$, тобто не може бути $S(1) = 0$. Отже маємо ситуацію, аналогічну випадку з подією B у Вправі 2.3.5 — водій починає маршрут без власних коштів, але тепер пасажирів 101:

$$\mathbb{P}(C) = \frac{|C|}{|\Omega|} = \frac{N_{102}^+(52)}{\binom{100}{75}} = \frac{52}{102} \cdot \frac{102}{77} \cdot \frac{101}{76} \approx 0.897.$$

Отже завжди непогано мати з собою зайві гроші. $\square$

Вправа 2.6.5. Чи можна казати, що клас множин виду $(-b; b)$, $b \in \mathbb{R}$, породжує Борелеву σ -алгебру? $\square$

Розв'язання. Позначмо клас таких множин через $\mathcal{I}_b$. Для доведення рівності двох множин (σ -алгебр) потрібно показати вкладення цих множин в обидва боки: $\mathcal{A}(\mathcal{I}_b) \subseteq \mathcal{A}(\mathcal{I})$ та $\mathcal{A}(\mathcal{I}) \subseteq \mathcal{A}(\mathcal{I}_b)$.

Той факт, що $\mathcal{I}_b \subseteq \mathcal{A}(\mathcal{I})$, є очевидним, адже можна завжди показати, як ми це вже неодноразово робили раніше:

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

Проте показати, що будь-який пів інтервал $(-a; c]$, $a \neq c$, можна утворити шляхом злічених доповнень та об'єднань симетричних інтервалів $(-b; b)$ — неможливо. Справді, об'єднання будь-яких симетричних інтервалів також повинно бути симетричним. Доповнення $(-b; b)^c = (-\infty; -b] \cup [b; \infty)$ також симетричне. Відповідно, будь-які теоретико-множинні операції над такими симетричними множинами будуть давати множини, які матимуть властивість симетричності, наприклад

$$((-\infty; -3] \cup [3; \infty)) \cap (-4; 4) = (-4; -3] \cup [3; 4) .$$

Отже утворити інтервал $(-a; c]$, $a \neq c$, тільки за допомогою симетричних інтервалів неможливо, а тому клас симетричних інтервалів не породжує Борелевої σ -алгебри. $\square$

3 Властивості ймовірнісної міри

3.1 Основні теоретичні відомості

Детальний огляд понять, потрібний для розв’язання задач на цьому занятті, наведено в Курсі лекцій, Розд. КЛ2.1–КЛ2.4.

Визначення (КЛ2.1.1). *Функцією від множини* (set function) називають функцію $f : \mathcal{C} \rightarrow \mathbb{R}$, де $\mathcal{C}$ — деякий клас множин елементів простору Ω : $\mathcal{C} = \{A : A \subseteq \Omega\}$. $\square$

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 перерахування аксіом, які будь-яка корисна на практиці міра повинна задовольняти.

Визначення (КЛ2.1.2). Функція $\mu : \mathcal{A} \rightarrow \mathbb{R}$ має назву *міри* (measure), якщо:

- (i) $\mu(A) \geq 0$ для всіх $A \in \mathcal{A}$;
- (ii) μ є σ -адитивною (σ -additive): для деякої послідовності неперетинних множин справедливо

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

(КЛ2.1.1)

$\square$

У загальному випадку мірою простору $\mu(\Omega)$ може бути будь-яке дійсне число (і навіть нескінченність). Міру називають *скінченною* (finite), якщо $\mu(\Omega) < \infty$. Також корисним для нас буде поняття σ -скінченної міри.

Визначення (КЛ2.1.3). Міра μ є σ -скінченною (σ -finite), якщо $\mu(\Omega) = \infty$, але існує розбиття $\{A_i, i = 1, 2, \dots\}$ простору Ω таке, що $\mu(A_i) < \infty, i = 1, 2, \dots$ $\square$

Імовірність від міри відрізняє наявність специфічної нормалізації.

Визначення (КЛ2.1.5). Міру $\mathbb{P}$, яка задовольняє всі аксіоматичні властивості з Визначення КЛ2.1.2, і для якої $\mathbb{P}(\Omega) = 1$, називають *імовірнісною мірою* (probability measure). $\square$

Визначення (КЛ2.1.6). Трійку $(\Omega, \mathcal{A}, \mu)$ називають *мірним простором* (measure space). Зокрема, $(\Omega, \mathcal{A}, \mathbb{P})$ називають *імовірнісним простором* (probability space). $\square$

Кількість аксіом, які задають міру і які наведені у Визначенні КЛ2.1.2, є дуже малою, що дає змогу (відносно) легко перевіряти, чи є деяка функція мірою. Проте з цих двох аксіом можна вивести низку корисних властивостей, які задовольняє будь-яка міра.

Теорема (КЛ2.2.1). Будь-яка міра μ , задана на мірному просторі $(\Omega, \mathcal{A}, \mu)$, має такі властивості:

$$(i) \quad \mu(\emptyset) = 0;$$

(ii) міра є скінченно-адитивною (адитивною, finitely additive):

$$\mu\left(\bigcup_{j=1}^n A_j\right) = \sum_{j=1}^n \mu(A_j), \quad A_j \in \mathcal{A}, \quad j = 1, 2, \dots, \quad A_i \cap A_j = \emptyset, \quad i \neq j; \quad (3.1.1)$$

(iii) μ 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) μ є σ -субадитивною (σ -subadditive)

$$\mu\left(\bigcup_{j=1}^{\infty} A_j\right) \leq \sum_{j=1}^{\infty} \mu(A_j), \quad A_j \in \mathcal{A}, \quad j = 1, 2, \dots, \quad (3.1.2)$$

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

Ця нерівність також відома як *нерівність Була* (Boole's inequality)¹;

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

Для ймовірнісної міри додатково також справедливо, що $\mathbb{P}(A^c) = 1 - \mathbb{P}(A)$ і $\mathbb{P}(A) \leq 1$.

Також можна показати, що будь-яка міра є неперервною в такому розумінні.

Визначення (КЛ2.2.7). Міра μ , 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 μ *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 μ є просто *неперервною*, якщо вона одночасно неперервна як зверху, так і знизу.

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 $\lim_{i \rightarrow \infty} \mu(A_i) = 0$. $\square$

Теорема (КЛ2.2.8). (i) Будь-яка міра μ на мірному просторі $(\Omega, \mathcal{A}, \mu)$ є неперервна;

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

Теорема (Теорема Каратеодорі (Carathéodory theorem)², КЛ2.4.1). Нехай на деякій алгебрі $\mathcal{F}$ задано міру μ . Тоді цю міру можна подовжити на σ -алгебру $\mathcal{A}(\mathcal{F})$, породжену цією алгеброю. Причому якщо міра скінченна (або σ -скінченна), то таке подовження єдине і скінченне (або σ -скінченне).

¹Джордж Бул (George Boole, 1815–1864) — видатний англійський математик.

²Константин Каратеодорі (Constantin Carathéodory, 1873–1950) — грецько-німецький математик.

3.2 Застосування властивостей імовірностей

У нижченаведених задачах вважатимемо, якщо явно не зазначено інше, що ми працюємо з деяким імовірнісним простором $(\Omega, \mathcal{A}, \mathbb{P})$, а всі множини, що фігурують у вправах, належать $\mathcal{A}$.

Вправа 3.2.1. Якщо $\mathbb{P}(A) = \frac{1}{3}$, а $\mathbb{P}(B^c) = \frac{1}{4}$, чи можуть A та B бути несумісними? $\square$

Розв'язання. За умовами задачі маємо $\mathbb{P}(B) = 1 - \mathbb{P}(B^c) = \frac{3}{4}$. Для несумісних подій справедливо, що

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

У нашому випадку це неможливо, адже

$$\mathbb{P}(A \cup B) = \frac{1}{3} + \frac{3}{4} > 1 ,$$

хоча ймовірнісна міра не може перевищувати 1. $\square$

Вправа 3.2.2. Доведіть такі нерівності:

$$\mathbb{P}(A \cap B) \leq \mathbb{P}(A \cup B) \leq \mathbb{P}(A) + \mathbb{P}(B) . \quad (3.2.1)$$

Вкажіть умови, за яких ці нерівності обертаються в рівності. $\square$

Розв'язання. Із того факту, що $A \cap B \subseteq A \cup B$, та властивості (iii) з Теореми КЛ2.2.1 випливає

$$\mathbb{P}(A \cap B) \leq \mathbb{P}(A \cup B) .$$

Ця нерівність обернеться в рівність, якщо $A \cap B = A \cup B$, тобто $A = B$ ³.

Із властивості (v) із Теореми КЛ2.2.1 та невід'ємності міри випливає

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

Ця нерівність обернеться в рівність, якщо $A \cap B = \emptyset$. $\square$

³Строго кажучи, достатньо було б виконання $A = B$ майже напевно. Справді, нехай $B = A \cup Z$, де $\mathbb{P}(Z) = 0$. Тоді $\mathbb{P}(B) = \mathbb{P}(A)$, хоча дві множини не є абсолютно ідентичними. Ми будемо часто стикатися з поняттям «майже напевно» далі в цьому курсі.

Вправа 3.2.3. Нехай S і L позначають події, що після візиту до сімейного лікаря пацієнту дадуть скерування до спеціаліста та скерування на аналізи відповідно. Нехай $\mathbb{P}(S) = 0.25$, $\mathbb{P}(L) = 0.35$, а ймовірність, що візит не завершиться жодною з цих подій, дорівнює 0.45. Чому дорівнює:

- (i) ймовірність скерування одночасно і до спеціаліста, і на аналізи;
- (ii) ймовірність скерування щонайменше на одну з цих опцій? □

Розв'язання. (i) У першому питанні нас цікавить ймовірність перетину двох подій, $\mathbb{P}(S \cap L)$. У нас (поки що) немає жодної властивості ймовірностей, яка б давала змогу обчислювати ймовірності перетинів, проте ми вміємо обчислювати ймовірності об'єднань, наприклад, за рахунок властивості (v) із Теорема КЛ2.2.1. Тому перейдімо від перетину до об'єднання через доповнення і закони де Моргана:

$$\begin{aligned}\mathbb{P}(S \cap L) &= 1 - \mathbb{P}((S \cap L)^c) = 1 - \mathbb{P}(S^c \cup L^c) \\ &= 1 - (\mathbb{P}(S^c) + \mathbb{P}(L^c) - \mathbb{P}(S^c \cap L^c)) \\ &= 1 - (1 - \mathbb{P}(S) + 1 - \mathbb{P}(L) - \mathbb{P}(S^c \cap L^c)) .\end{aligned}$$

Ймовірність перетину доповнень нам дано: $\mathbb{P}(S^c \cap L^c) = 0.45$. Отже

$$\mathbb{P}(S \cap L) = 1 - (1 - 0.25 + 1 - 0.35 - 0.45) = 0.05 ;$$

- (ii) у другому питанні нас цікавить ймовірність об'єднання, для обчислення якої ми застосуємо властивість (v) із Теорема КЛ2.2.1 та щойно здобутий результат:

$$\mathbb{P}(S \cup L) = \mathbb{P}(S) + \mathbb{P}(L) - \mathbb{P}(S \cap L) = 0.25 + 0.35 - 0.05 = 0.55 .$$

□

Вправа 3.2.4. Нехай A_n , $n = 1, 2, \dots$ — деякі події в ймовірнісному просторі $(\Omega, \mathcal{A}, \mathbb{P})$ такі, що $\mathbb{P}(A_n) = 1$ для всіх n . Доведіть, що

$$\mathbb{P}\left(\bigcap_{n=1}^{\infty} A_n\right) = 1 . \quad \square$$

Розв'язання. Із властивостей ймовірності безпосередньо випливає, що ймовірність зліченного перетину дорівнюватиме 1 тоді й тільки тоді, коли

$$\mathbb{P}\left(\left(\bigcap_{n=1}^{\infty} A_n\right)^c\right) = \mathbb{P}\left(\bigcup_{n=1}^{\infty} A_n^c\right) = 0 .$$

Оскільки $\mathbb{P}(A_n^c) = 0$ для всіх n , згідно з нерівністю Була маємо

$$\mathbb{P}\left(\bigcup_{n=1}^{\infty} A_n^c\right) \leq \sum_{n=1}^{\infty} \mathbb{P}(A_n^c) = 0 \quad \Rightarrow \quad \mathbb{P}\left(\bigcup_{n=1}^{\infty} A_n^c\right) = 0. \quad \square$$

Вправа 3.2.5. Нехай на вимірному просторі $(\Omega, \mathcal{A})$ задано дві ймовірнісні міри, $\mathbb{P}$ і $\mathbb{Q}$. Нехай $\mathbb{P}(A) = \mathbb{Q}(A)$ для всіх $A \in \mathcal{A}$ таких, що $\mathbb{P}(A) \leq \frac{1}{2}$. Доведіть, що $\mathbb{P}(A) = \mathbb{Q}(A)$ для всіх $A \in \mathcal{A}$. $\square$

Розв'язання. Нехай $A \in \mathcal{A}$ така, що $\mathbb{P}(A) \leq \frac{1}{2}$. Тоді за припущенням $\mathbb{Q}(A) = \mathbb{P}(A)$.

Якщо ж $A \in \mathcal{A}$ така, що $\mathbb{P}(A) > \frac{1}{2}$, то $\mathbb{P}(A^c) = 1 - \mathbb{P}(A) \leq \frac{1}{2}$, а тому $1 - \mathbb{P}(A) = \mathbb{P}(A^c) = \mathbb{Q}(A^c) = 1 - \mathbb{Q}(A)$, звідки знову-таки випливає, що $\mathbb{Q}(A) = \mathbb{P}(A)$.

Отже $\mathbb{P}(A) = \mathbb{Q}(A)$ для всіх $A \in \mathcal{A}$. $\square$

Варто зазначити, що у Вправі 3.2.5 умова $\mathbb{P}(A) \leq \frac{1}{2}$ є посутньою, адже якщо $\mathbb{P}(A) < \frac{1}{2}$ (строго), то міри можуть бути різні. Наприклад, нехай $\mathcal{A} = \{\emptyset, \{1\}, \{2\}, \{1, 2\}\}$. Тоді нехай

$$\begin{aligned} \mathbb{P}(\emptyset) &= 0, \quad \mathbb{P}(\Omega) = 1, \quad \mathbb{P}(\{1\}) = \mathbb{P}(\{2\}) = \frac{1}{2}, \\ \mathbb{Q}(\emptyset) &= 0, \quad \mathbb{Q}(\Omega) = 1, \quad \mathbb{Q}(\{1\}) = \frac{1}{3}, \quad \mathbb{Q}(\{2\}) = \frac{2}{3}. \end{aligned}$$

Суто формально можна стверджувати, що $\mathbb{P}(A) = \mathbb{Q}(A)$ для всіх $A \in \mathcal{A}$ таких, що $\mathbb{P}(A) < \frac{1}{2}$, оскільки така множина єдина — це порожня множина. Але очевидно, що цієї вимоги недостатньо для рівності двох мір.

3.3 Теорема включення-виключення

Властивість із Теорема КЛ2.2.1(v) можна узагальнити на довільну кількість множин. Сформулюймо відповідну теорему в застосуванні до ймовірнісних мір, хоча вона справедлива і для довільних мір також.

Теорема (Теорема включення-виключення (Inclusion-exclusion theorem), КЛ2.2.6). Нехай $(\Omega, \mathcal{A}, \mathbb{P})$ — деякий імовірнісний простір. Тоді

$$\begin{aligned} \mathbb{P}\left(\bigcup_{i=1}^n A_i\right) &= \sum_i \mathbb{P}(A_i) - \sum_{i < j} \mathbb{P}(A_i \cap A_j) + \sum_{i < j < k} \mathbb{P}(A_i \cap A_j \cap A_k) - \dots \\ &\quad + (-1)^{n+1} \mathbb{P}(A_1 \cap \dots \cap A_n). \end{aligned} \quad (\text{КЛ2.2.3})$$

Приклад 3.3.1 (Задача зіставлень де Монморта (de Montmort's matching problem)⁴). Розгляньмо добре перетасовану колоду з n карт, пронумерованих від 1 до n та викладених на стіл у випадковому порядку. Гравець по черзі перегортає карти і виграє в тому випадку, якщо i -та перегорнута карта матиме нанесений на ній номер i . Чому дорівнює ймовірність виграшу?

Альтернативна інтерпретація цієї задачі така. Нехай у перинатальному центрі перебувають n матерів, у кожної з яких по одному новонародженому. Уявімо, що всі матері покинули центр і пішли додому переночувати, а повернувшись наступного ранку, вони виявили, що всіх дітей переплутано. Чому дорівнює ймовірність того, що у випадку випадкової роздачі дітей хоча б одна мати дістане назад свою власну дитину?

Позначмо через A_i подію, що i -та карта, витягнута з колоди, має нанесений на ній номер i . Тоді нас цікавить ймовірність $\mathbb{P}(A_1 \cup \dots \cup A_n)$. Підрахунок вестимемо за Теоремою КЛ2.2.6.

Подія $A_{i_1} \cap A_{i_2} \cap \dots \cap A_{i_k}$, де $i_1, \dots, i_k \in \{1, \dots, n\}$ — деякі *різні* індекси, описує ситуацію, коли точно k карт стоять на своїх позиціях, а саме карти з номерами $i_1, \dots, i_k$ стоять на позиціях $i_1, \dots, i_k$ відповідно. Оскільки карти можуть опинитися на будь-якій позиції з однаковою ймовірністю, тут застосовне класичне визначення ймовірності.

Простором елементарних подій Ω буде множина всіх можливих перестановок із n карт, яких існує $n!$ штук. Кожна подія $A_{i_1} \cap A_{i_2} \cap \dots \cap A_{i_k}$ відповідає таким перестановкам, де позиції $i_1, \dots, i_k$ *зафіксовані*, і переставляти можна тільки $n - k$ позицій, що залишилися. Отже загальне число перестановок, які відповідають події $A_{i_1} \cap A_{i_2} \cap \dots \cap A_{i_k} = (n - k)!$.

Отже, наприклад, маємо

$$\mathbb{P}(A_1) = \frac{(n-1)!}{n!} = \frac{1}{n}, \quad \mathbb{P}(A_2 \cap A_5 \cap A_7) = \frac{(n-3)!}{n!} = \frac{1}{n(n-1)(n-2)}.$$

Застосування (КЛ2.2.3) дає

$$\mathbb{P}\left(\bigcup_{i=1}^n A_i\right) = \sum_{i=1}^n \frac{1}{n} - \sum_{i < j} \frac{1}{n(n-1)} + \sum_{i < j < k} \frac{1}{n(n-1)(n-2)} - \dots + (-1)^{n+1} \frac{1}{n!}.$$

Залишилось тільки підрахувати, скільки існує доданків у кожній сумі. Це впливає з визначення біномних коефіцієнтів: усіх можливих перетинів

⁴Уперше цю задачу запропонував французький математик П'єр де Монморт (Pierre Remond de Montmort, 1678–1719) у 1708 р.

по дві множини $\in \binom{n}{2}$, по три множини $\in \binom{n}{3}$ тощо. Остаточню маємо:

$$\begin{aligned}\mathbb{P}\left(\bigcup_{i=1}^n A_i\right) &= n \cdot \frac{1}{n} - \frac{\binom{n}{2}}{n(n-1)} + \frac{\binom{n}{3}}{n(n-1)(n-2)} - \dots + (-1)^{n+1} \cdot \frac{1}{n!} \\ &= 1 - \frac{1}{2!} + \frac{1}{3!} - \dots + (-1)^{n+1} \cdot \frac{1}{n!}.\end{aligned}\quad (3.3.1)$$

Використовуючи розвинення експоненти в ряд Тейлора

$$e^{-1} = 1 - 1 + \frac{1}{2!} - \frac{1}{3!} + \dots,$$

маємо наближену оцінку шуканої ймовірності:

$$\mathbb{P}\left(\bigcup_{i=1}^n A_i\right) \approx 1 - \frac{1}{e}.$$

Код на Лістингу 3.3.1 виконує симуляцію за *методом Монте-Карло* (Monte Carlo simulation)⁵, суть якого полягає у відтворенні деякого випадкового експерименту велику кількість разів і використання одержаних результатів для дослідження властивостей процесу, який їх породжує.

Так, відповідна програма симулює проведення незалежних випробувань із витягуванням карт і фіксує успішність кожного з них у розумінні появи принаймні одного збігу номерів. Відносна частота успішних випробувань відіграє роль *емпіричної* ймовірності настання відповідної події. Також програма рахує теоретичну ймовірність як суму з (3.3.1).

Симуляції виконуємо для різного числа карт, $n = 1, \dots, 10$. Для кожного n виконуємо 10 000 повторень експерименту з перетасування карт. Як можна бачити з Рис. 3.3.1, емпіричні ймовірності доволі близькі до ймовірностей теоретичних, до того ж обидві послідовності ймовірностей дуже швидко прямують до $1 - 1/e \approx 0.632$. Із цього можна зробити цікавий практичний висновок, що ймовірність буде приблизно однаковою для різних n , хоча «інтуїція» підказує, що зі збільшенням n повинно бути більше «шансів», щоб хоча б одна карта стала на свою позицію. $\square$

Лістинг 3.3.1: Код для Прикладу 3.3.1

```
import matplotlib.pyplot as plt
from matplotlib import rc
import numpy as np
from numpy.random import seed
5 from scipy.special import factorial
```

⁵Відповідну 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.

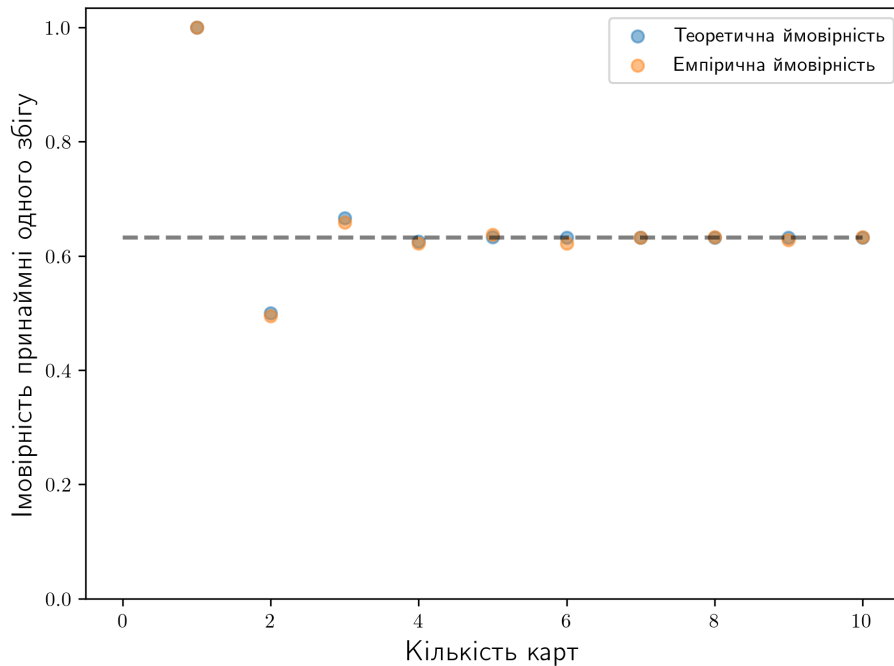


Рис. 3.3.1: Теоретичні й емпіричні ймовірності для Прикладу 3.3.1

```

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

def prob_theoretical(n):
    k = np.arange(1, n + 1)
15     return np.sum((1 / factorial(k)) * (-1)**(k + 1))

def prob_empirical(n, trials):
    results = np.zeros(trials)
    ordered = np.arange(1, n + 1)
20     for i in range(trials):
        shuffled = np.random.permutation(ordered)
        results[i] = np.sum(shuffled == ordered)

25     return np.sum(results >= 1) / trials

seed(100)
trials = 10000
30 number_of_cards_max = 10

probs_theoretical = \
    np.array([prob_theoretical(n) for n
              in range(1, number_of_cards_max + 1)])
35 probs_empirical = \
    np.array([prob_empirical(n, trials) for
              n in range(1, number_of_cards_max + 1)])

40 plt.scatter(
    np.arange(1, number_of_cards_max + 1), probs_theoretical,
    label="Теоретична ймовірність", alpha=0.5

```

```

)
plt.scatter(
    np.arange(1, number_of_cards_max + 1), probs_empirical,
45     label="Емпірична ймовірність", alpha=0.5
)
plt.hlines(1 - 1/np.e, 0, number_of_cards_max,
           linestyle='dashed', lw=2, alpha=0.5)
plt.xticks(fontsize=10)
50 plt.yticks([0.0, 0.2, 0.4, 0.6, 0.8, 1.0], fontsize=10)
plt.xlabel("Кількість карт", fontsize=15)
plt.ylabel("Ймовірність принаймні одного збігу", fontsize=15)
plt.legend()

55 plt.tight_layout()
plt.savefig("../images/Chapter 3/montmort.png", dpi=300)

```

Розгляньмо інші ситуації, де застосування Теорема КЛ2.2.6 цілком виправдано.

Вправа 3.3.2. Нехай маємо групу з 7 людей. Чому дорівнює ймовірність, що серед них є щонайменше по одній людині з днем народження, що припадає на кожну пору року? Вважатимемо, що ймовірності народитися в кожну пору року однакові. $\square$

Розв'язання. Як і в попередній задачі, і як і в усіх схожих вправах, спочатку потрібно визначити по події для кожної пори року, а потім обчислити ймовірність їх об'єднання за Теоремою КЛ2.2.6.

У даній задачі нас цікавить ймовірність одночасного виконання, тобто перетину, подій A_i , $i = 1, 2, 3, 4$, що в групі є принаймні одна людина з днем народження, що припадає на пору року i . Згідно з законом де Моргана,

$$\mathbb{P}(A_1 \cap A_2 \cap A_3 \cap A_4) = 1 - \mathbb{P}(A_1^c \cup A_2^c \cup A_3^c \cup A_4^c),$$

де кожна з подій A_i^c , $i = 1, 2, 3, 4$, має інтерпретацію, що в групі немає жодної людини з днем народження в пору року i .

Отже

$$\begin{aligned} \mathbb{P}(A_1^c \cup A_2^c \cup A_3^c \cup A_4^c) &= \sum_{i=1}^4 \mathbb{P}(A_i^c) - \sum_{i < j} \mathbb{P}(A_i^c \cap A_j^c) \\ &\quad + \sum_{i < j < k} \mathbb{P}(A_i^c \cap A_j^c \cap A_k^c) - \mathbb{P}(A_1^c \cap A_2^c \cap A_3^c \cap A_4^c). \end{aligned}$$

Оскільки всі варіанти народження рівноймовірні, застосуємо правила обчислення класичної ймовірності. Зокрема, ймовірність кожної A_i^c , $i = 1, 2, 3, 4$, дорівнює $(3/4)^7$, ймовірність кожного перетину $A_i^c \cap A_j^c$, $j = 1, 2, 3, 4$,

$0 < i < j$, дорівнює $(2/4)^7 = (1/2)^7$, а ймовірність усіх потрібних перетинів дорівнює $(1/4)^7$. Ймовірність $\mathbb{P}(A_1^c \cap A_2^c \cap A_3^c \cap A_4^c)$, очевидно, дорівнює 0.

Отже,

$$\mathbb{P}(A_1^c \cup A_2^c \cup A_3^c \cup A_4^c) = 4 \cdot \frac{3^7}{4^7} - \binom{4}{2} \frac{1}{2^7} + \binom{4}{3} \frac{1}{4^7} \approx 0.487 ,$$

а шукана ймовірність дорівнює

$$\mathbb{P}(A_1 \cap A_2 \cap A_3 \cap A_4) = 1 - \mathbb{P}(A_1^c \cup A_2^c \cup A_3^c \cup A_4^c) \approx 0.513 .$$

□

Вправа 3.3.3. За святковий стіл у випадковому порядку сідають 4 подружні пари. Чому дорівнює ймовірність того, що жодна жінка не сидітиме поруч зі своїм чоловіком? □

Розв'язання. Позначмо через E_i подію, що чоловік і дружина з подружньої пари номер i сидять поруч. Нас цікавить ймовірність

$$\mathbb{P}\left(\left(\bigcup_{i=1}^4 E_i\right)^c\right) = 1 - \mathbb{P}\left(\bigcup_{i=1}^4 E_i\right) ,$$

де ймовірність об'єднання можна обчислити за (КЛ2.2.3):

$$\begin{aligned} \mathbb{P}\left(\bigcup_{i=1}^4 E_i\right) &= \sum_{i=1}^4 \mathbb{P}(E_i) - \sum_{i < j} \mathbb{P}(E_i \cap E_j) + \sum_{i < j < k} \mathbb{P}(E_i \cap E_j \cap E_k) \\ &\quad - \mathbb{P}(E_1 \cap E_2 \cap E_3 \cap E_4) . \end{aligned}$$

Як і в попередніх випадках, маємо ситуацію з класичною ймовірністю, де простір елементарних подій Ω складають усі можливі розсадки за столом. Будь-яку групу з 8 осіб можна всадити за стіл у $8!$ способів: $|\Omega| = 8!$.

Подія-перетин $E_{i_1} \cap \dots \cap E_{i_k}$, де $i_1, \dots, i_k \in \{1, \dots, n\}$ — деякі *різні* індекси, описує ситуацію, коли точно k подружніх пар сидять поруч. Загальне число варіантів розсаджування, які ведуть до такого результату, можна підрахувати так:

- розгляньмо k подружніх пар як нерозривні. Тоді загальне число варіантів розсаджування k пар і $8-2k$ окремих осіб, тобто $8-k$ «об'єктів», дорівнює $(8-k)!$;
- у рамках кожної пари чоловіка й дружину можна переставляти місцями, тому є 2^k варіантів перестановок *усередині* відповідних пар;

- згідно з правилом множення, загальне число можливих варіантів розсаджування, за яких k пар сидять поруч, становить $2^k(8-k)!$.

Через симетрію всіх розсадок остаточно маємо

$$\mathbb{P}\left(\bigcup_{i=1}^4 E_i\right) = \binom{4}{1} \frac{2^1 \cdot 7!}{8!} - \binom{4}{2} \frac{2^2 \cdot 6!}{8!} + \binom{4}{3} \frac{2^3 \cdot 5!}{8!} - \binom{4}{4} \frac{2^4 \cdot 4!}{8!} \approx 0.657,$$

а тому ймовірність, що жодне подружжя не сидітиме поруч, дорівнює приблизно 0.343. $\square$

3.4 Властивості міри

Розгляньмо декілька задач, пов'язаних із застосуванням властивостей мір.

Вправа 3.4.1. Нехай маємо вимірний простір $(\Omega, \mathcal{A}, \mathbb{P})$, де Ω — деякий незліченний простір елементарних подій, а σ -алгебра $\mathcal{A}$ містить події, що або самі є не більш ніж зліченими, або чії доповнення є не більш ніж зліченими.

Розгляньмо функцію $\mathbb{P}$ таку, що $\mathbb{P}(A) = 0$, якщо A не більш ніж зліченна, та $\mathbb{P}(A) = 1$, якщо A^c не більш ніж зліченна.

Покажіть, що $\mathbb{P}$ є ймовірнісною мірою. $\square$

Розв'язання. Оскільки $\mathbb{P}(A) \geq 0$ для будь-якої $A \in \mathcal{A}$, а $\mathbb{P}(\Omega) = 1$, бо доповнення простору $\Omega^c = \emptyset$ не більш ніж зліченне, достатньо перевірити σ -адитивність. Розгляньмо послідовність несумісних подій $A_n \in \mathcal{A}$, $n = 1, 2, \dots$, $A_i \cap A_j = \emptyset$, $i \neq j$, та їх об'єднання $A = \bigcup_{n=1}^{\infty} A_n$.

Якщо A не більш ніж зліченна, то і всі A_n , $n = 1, 2, \dots$, повинні бути не більш ніж зліченні, і тому

$$\mathbb{P}(A) = 0 = 0 + 0 + \dots = \sum_{n=1}^{\infty} \mathbb{P}(A_n).$$

Якщо ж A^c не більш ніж зліченна, то A повинна бути незліченною, а тому щонайменше одна з A_n також повинна бути незліченною. Насправді, *тільки одна* з A_n є незліченною. Доведемо це від супротивного. Якщо A_i і A_j , $i \neq j$, одночасно є незліченими, то, відповідно, A_i^c і A_j^c одночасно є не більш ніж зліченими. Оскільки $A_i \cap A_j = \emptyset$, $A_i \subset A_j^c$. Але якщо A_j^c не більш ніж зліченна, то її підмножина A_i також повинна бути не більш ніж зліченна, що суперечить нашому початковому припущенню. Отже, існує

єдина множина з-посеред A_n , яка є незліченною, наприклад, A_{n_0} . Тоді

$$\mathbb{P}(A) = 1 = 0 + 0 + \dots + 0 + 1 + 0 + \dots = \sum_{n=1}^{n_0-1} \mathbb{P}(A_n) + \mathbb{P}(A_{n_0}) + \sum_{n=n_0+1}^{\infty} \mathbb{P}(A_n) .$$

□

Розгляньмо цікаву ілюстрацію неперервності ймовірнісної міри.

Приклад 3.4.2. Уявімо, що маємо нескінченно велику вазу і нескінченно великий набір куль, пронумерованих натуральними числами. Розгляньмо такий експеримент:

- за 1 хвилину до півночі у вазу кладуть кулі з номерами від 1 до 10 і відразу витягають кулю з номером 10 (усе відбувається «миттєво»);
- за 0.5 хвилини до півночі у вазу кладуть кулі з номерами від 11 до 20 і відразу витягають кулю з номером 20;
- за 0.25 хвилини до півночі у вазу кладуть кулі з номерами від 21 до 30 і відразу витягають кулю з номером 30 і т.д.

Іншими словами, на n -ому кроці, тобто за $2^{-(n-1)}$ хвилин до півночі, $n = 1, 2, \dots$, у вазу кладуть кулі з номерами від $10(n-1) + 1$ до $10n$ і відразу витягають кулю з номером $10n$. Скільки у вазі буде куль станом на північ?

Оскільки на кожному кроці витягають кулю з номером, кратним 10, у сухому залишку у вазі залишиться нескінченна кількість куль, номери яких *не будуть* кратні 10.

Розгляньмо тепер модифікацію цього експерименту, у якому на кожному кроці n витягають не кулю з номером $10n$, $n = 1, 2, \dots$, а кулю з номером n , тобто за хвилину до півночі витягають кулю з номером 1, за пів хвилини — кулю з номером 2 тощо. Отже, як і в попередньому експерименті, на кожному кроці у вазу кладуть 10 куль, а витягають усього одну. Проте в цьому випадку, як би це не було на перший погляд парадоксально, станом на північ ваза буде порожня! Справді, за $2^{-(n-1)}$ хвилин до півночі з вази буде витягнуто кулю з номером n . Отже для кожного $n = 1, 2, \dots$ справедливо, що кулі з номером n у вазі не буде.

Жодного «парадоксу» тут немає, адже ми маємо справу з різними ситуаціями, які у випадку нескінченно великих чисел дають на перший погляд контрінтуїтивні результати.

Розгляньмо ще одну варіацію цього експерименту, у якому на кожному кроці витягають деяку кулю з номером від 1 до $10n$ *у випадковий спосіб*. Щоб відповісти на питання про кількість куль, які залишаться станом на північ, ми застосуємо властивість неперервності ймовірнісної міри з Визначення КЛ2.2.7.

Позначмо через A_n подію, що після n витягів у вазі залишатиметься куля з номером 1. Очевидно, що

$$\mathbb{P}(A_1) = \frac{9}{10}.$$

На другому кроці ймовірність, що куля з номером 1 залишається у вазі, дорівнює ймовірності, що її не вибрали ні на першому кроці, ні на другому:

$$\mathbb{P}(A_2) = \frac{9}{10} \cdot \frac{18}{19}.$$

У схожий спосіб можна показати, що

$$\mathbb{P}(A_n) = \prod_{k=1}^n \frac{9k}{9k+1}.$$

Усі події A_n , $n = 1, 2, \dots$, є вкладеними одна в одну: $A_1 \supset A_2 \supset \dots$

Відповідно до властивості неперервності, ймовірність того, що куля з номером 1 залишається у вазі станом на північ, дорівнює

$$\mathbb{P}\left(\bigcap_{n=1}^{\infty} A_n\right) = \mathbb{P}\left(\lim_{n \rightarrow \infty} A_n\right) = \lim_{n \rightarrow \infty} \mathbb{P}(A_n) = \lim_{n \rightarrow \infty} \prod_{k=1}^n \frac{9k}{9k+1}.$$

Розгляньмо границю

$$\frac{1}{\lim_{n \rightarrow \infty} \prod_{k=1}^n \frac{9k}{9k+1}} = \lim_{n \rightarrow \infty} \prod_{k=1}^n \left(1 + \frac{1}{9k}\right).$$

Можна помітити, що

$$\begin{aligned} \prod_{k=1}^n \left(1 + \frac{1}{9k}\right) &= \left(1 + \frac{1}{9}\right) \left(1 + \frac{1}{18}\right) \dots \left(1 + \frac{1}{9n}\right) \\ &> \frac{1}{9} + \frac{1}{18} + \dots + \frac{1}{9n} = \frac{1}{9} \sum_{k=1}^n \frac{1}{k}. \end{aligned}$$

Із теорії рядів нам відомо, що

$$\sum_{k=1}^{\infty} \frac{1}{k} = \infty,$$

а отже

$$\lim_{n \rightarrow \infty} \prod_{k=1}^n \left(1 + \frac{1}{9k}\right) = \infty .$$

У підсумку ми маємо:

$$\mathbb{P} \left(\bigcap_{n=1}^{\infty} A_n \right) = 0 ,$$

іншими словами, куля з номером 1 не залишиться у вазі станом на північ із ймовірністю 1.

В аналогічний спосіб можна показати, що куля з *будь-яким* номером не залишиться у вазі станом на північ.

Позначмо через F_n подію, що куля з номером n залишиться у вазі станом на північ. Ймовірність того, що ваза *не буде* порожньою станом на північ, дорівнює ймовірності об'єднання F_n , яку можна оцінити за допомогою нерівності Була:

$$\mathbb{P} \left(\bigcup_{n=1}^{\infty} F_n \right) \leq \sum_{n=1}^{\infty} \mathbb{P}(F_n) = 0 .$$

Отже ваза буде порожньою станом на північ з ймовірністю 1. $\square$

Згідно з умовами Теорема Каратеодорі КЛ2.4.1, подовження міри з алгебри на σ -алгебру буде єдиним, якщо міра скінченна або σ -скінченна. Розгляньмо простий контрприклад.

Приклад 3.4.3. Нехай $\Omega = \{\omega_1, \omega_2, \omega_3, \omega_4\}$ і нехай маємо клас множин $\mathcal{C} = \{\emptyset, \{\omega_1, \omega_2\}, \{\omega_1, \omega_3\}, \{\omega_2, \omega_4\}, \{\omega_3, \omega_4\}, \Omega\}$. Цей клас множин не є алгеброю, що видно, наприклад, із того факту, що

$$\{\omega_1, \omega_2\} \cup \{\omega_1, \omega_3\} = \{\omega_1, \omega_2, \omega_3\} \notin \mathcal{C} .$$

Розгляньмо на цьому класі множин міру μ :

$$\begin{aligned} \mu(\{\omega_1, \omega_2\}) &= \mu(\{\omega_1, \omega_3\}) = \mu(\{\omega_2, \omega_4\}) = \mu(\{\omega_3, \omega_4\}) = 3 , \\ \mu(\Omega) &= 6 , \\ \mu(\emptyset) &= 0 . \end{aligned}$$

Це справді міра, адже вона невід'ємна, а адитивність виконується, оскільки вона виконується для єдиних двох пар неперетинних множин, об'єднання яких також лежить у $\mathcal{C}$:

$$\begin{aligned} \mu(\Omega) &= \mu(\{\omega_1, \omega_2\}) + \mu(\{\omega_3, \omega_4\}) = 3 + 3 = 6 , \\ \mu(\Omega) &= \mu(\{\omega_1, \omega_3\}) + \mu(\{\omega_2, \omega_4\}) = 3 + 3 = 6 . \end{aligned}$$

Розгляньмо два різні подовження міри μ на σ -алгебру 2^Ω (множину всіх

підмножин Ω):

$$\begin{aligned}\mu_1(\{\omega_1\}) &= \mu_1(\{\omega_4\}) = \mu_2(\{\omega_2\}) = \mu_2(\{\omega_3\}) = 1, \\ \mu_1(\{\omega_2\}) &= \mu_1(\{\omega_3\}) = \mu_2(\{\omega_1\}) = \mu_2(\{\omega_4\}) = 2.\end{aligned}$$

Обидві ці функції є подовженнями μ , оскільки

$$\begin{aligned}\mu_1(\emptyset) &= \mu_2(\emptyset) = 0, \\ \mu_1(\Omega) &= \mu_2(\Omega) = 6, \\ \mu_1(\{\omega_1, \omega_2\}) &= \mu_2(\{\omega_1, \omega_2\}) = 3, \\ \mu_1(\{\omega_1, \omega_3\}) &= \mu_2(\{\omega_1, \omega_3\}) = 3, \\ \mu_1(\{\omega_2, \omega_4\}) &= \mu_2(\{\omega_2, \omega_4\}) = 3, \\ \mu_1(\{\omega_3, \omega_4\}) &= \mu_2(\{\omega_3, \omega_4\}) = 3,\end{aligned}$$

хоча це зовсім різні міри.

Отже, якщо клас $\mathcal{C}$ не є алгеброю, подовження деякої міри, визначеної на ньому, необов'язково буде єдиним. $\square$

Вправа 3.4.4. Нехай маємо мірний простір $(\Omega, \mathcal{A}, \lambda) = ((0; 1], \mathcal{B}(0; 1], \lambda)$, де λ — міра Лебега. Визначмо новий простір так:

$$\Omega' = \{(x, y) \in \mathbb{R}^2 : 0 < x \leq 1, 0 < y \leq 1\},$$

тобто цей простір є декартовим добутком $(0; 1] \times (0; 1]$.

У цьому просторі розгляньмо клас множин $\mathcal{A}'$, до якого входять множини

$$\{(x, y) \in \mathbb{R}^2 : x \in A, 0 < y \leq 1\}, \quad A \in \mathcal{A}.$$

Нарешті, нехай

$$\mathbb{P}(A \times (0; 1]) = \lambda(A), \quad A \in \mathcal{A}.$$

Доведіть, що $(\Omega', \mathcal{A}', \mathbb{P})$ є ймовірнісним простором. $\square$

Розв'язання. Нам потрібно перевірити, що $\mathcal{A}'$ є σ -алгеброю, а $\mathbb{P}$ — імовірнісною мірою.

Для перевірки $\mathcal{A}'$ потрібно перевірити три властивості σ -алгебр:

- $\Omega' \in \mathcal{A}'$, оскільки $\Omega' = \Omega \times (0; 1]$;
- нехай $B \in \mathcal{A}'$, тобто B є декартовим добутком деякої множини $A \in \mathcal{A}$ з пів інтервалом $(0; 1]$. Оскільки $A \in \mathcal{A}$, то через те, що $\mathcal{A}$ сама є σ -алгеброю, $A^c \in \mathcal{A}$, а тому і $A^c \times (0; 1]$ належить $\mathcal{A}'$;

- нарешті, якщо

$$B = \bigcup_{i=1}^{\infty} (A_i \times (0; 1]) \text{ , } A_i \in \mathcal{A} \text{ , } i = 1, 2, \dots \text{ ,}$$

то через те, що $\mathcal{A}$ сама є σ -алгеброю, $\bigcup_{i=1}^{\infty} A_i \in \mathcal{A}$, а отже і

$$\bigcup_{i=1}^{\infty} (A_i \times (0; 1]) = \left(\bigcup_{i=1}^{\infty} A_i \right) \times (0; 1] \in \mathcal{A}' \text{ .}$$

Тепер потрібно перевірити три аксіоми міри для функції $\mathbb{P}$. По-перше, вона невід'ємна через невід'ємність міри Лебега. По-друге,

$$\mathbb{P}(\Omega') = \mathbb{P}(\Omega \times (0; 1]) = \lambda(\Omega) = 1 \text{ .}$$

Що стосується σ -адитивності, то нехай $B_i = A_i \times (0; 1]$ для деяких неперетинних $A_i \in \mathcal{A}$, $i = 1, 2, \dots$. Тоді

$$\begin{aligned} \mathbb{P}\left(\bigcup_{i=1}^{\infty} B_i\right) &= \mathbb{P}\left(\left(\bigcup_{i=1}^{\infty} A_i\right) \times (0; 1]\right) = \lambda\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} \lambda(A_i) \\ &= \sum_{i=1}^{\infty} \mathbb{P}(A_i \times (0; 1]) = \sum_{i=1}^{\infty} \mathbb{P}(B_i) \text{ .} \end{aligned}$$

□

3.5 Додаткові задачі

Вправа 3.5.1. Доведіть таку властивість імовірності симетричної різниці множин:

$$\mathbb{P}(A \Delta B) = \mathbb{P}(A) + \mathbb{P}(B) - 2\mathbb{P}(A \cap B) \text{ .} \quad (3.5.1)$$

□

Розв'язання. Згідно з Табл. КЛ1.2.1, симетрична різниця дорівнює

$$A \Delta B = (A \cap B^c) \cup (A^c \cap B) \text{ .}$$

Альтернативним формулюванням симетричної різниці є різниця об'єднання й перетину:

$$A \Delta B = (A \cup B) \setminus (A \cap B) = (A \cup B) \cap ((A \cap B)^c) = (A \cup B) \cap (A^c \cup B^c) \text{ .}$$

Нескладно переконатися, що

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

наприклад, застосувавши властивість дистрибутивності до виразу справа («розкрити дужки»).

Для обчислення ймовірності різниці використаємо властивість (iii) з Теорема КЛ2.2.1:

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

Застосовуючи до цього виразу властивість (v) із Теорема КЛ2.2.1, остаточно маємо:

$$\begin{aligned} \mathbb{P}(A \Delta B) &= \mathbb{P}(A \cup B) - \mathbb{P}(A \cap B) = \mathbb{P}(A) + \mathbb{P}(B) - \mathbb{P}(A \cap B) - \mathbb{P}(A \cap B) \\ &= \mathbb{P}(A) + \mathbb{P}(B) - 2\mathbb{P}(A \cap B) . \end{aligned}$$

□

Вправа 3.5.2. Абітурієнт планує вступити до одного з двох ЗВО, A і B . Нехай імовірність вступити до ЗВО A дорівнює p_1 , імовірність *не вступити* до ЗВО B дорівнює p_2 , а ймовірність не вступити принаймні до одного з цих ЗВО дорівнює p_3 .

Чому дорівнює ймовірність вступити хоча б до одного з цих ЗВО? □

Розв'язання. Позначмо через A та B події вступу до відповідних ЗВО. Згідно з умовами вправи, ми знаємо, що

$$\mathbb{P}(A) = p_1 , \quad \mathbb{P}(B^c) = p_2 , \quad \mathbb{P}(A^c \cup B^c) = p_3 .$$

Нас цікавить імовірність об'єднання:

$$\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B) - \mathbb{P}(A \cap B) = p_1 + (1 - p_2) - (1 - p_3) = p_1 - p_2 + p_3 .$$

□

Вправа 3.5.3. За результатами роздавання карт у гравця на руках опиняються 13 гральних карт зі стандартної колоди з 4 мастями і 13 званнями. Чому дорівнює ймовірність події, що в нього взагалі не буде карт принаймні однієї масти? □

Розв'язання. Позначмо через S , H , D , C події, що у гравця немає карт кожної з чотирьох мастей відповідно — вино (spades), чирва (hearts), дзвінка (diamonds) та жир (clubs). Тоді нас цікавить ймовірність об'єднання цих подій. Разом із тим, через повну симетрію маємо, що всі перетини мають однакову ймовірність, наприклад $\mathbb{P}(S \cap H) = \mathbb{P}(D \cap C)$. Відповідно до (КЛ2.2.3), шукану ймовірність можна обчислити так:

$$\begin{aligned} \mathbb{P}(S \cup D \cup H \cup C) &= 4 \cdot \mathbb{P}(S) - \binom{4}{2} \mathbb{P}(S \cap H) + \binom{4}{3} \mathbb{P}(S \cap H \cap D) \\ &\quad - \mathbb{P}(S \cap D \cap H \cap C) . \end{aligned}$$

Підрахуймо кожен ймовірність із цього виразу. Оскільки всі можливі варіанти роздачі карт є рівноймовірні, можемо застосувати класичне визначення ймовірності. Простором елементарних подій є множина всіх можливих роздач карт, і його потужність є $|\Omega| = \binom{52}{13}$. Загальне число варіантів не дістати жодної карти деякої масті дорівнює $\binom{52-13}{13} = \binom{39}{13}$, деяких двох мастей — $\binom{52-13-13}{13} = \binom{26}{13}$, деяких трьох мастей — $\binom{52-13-13-13}{13} = 1$. Нарешті, ймовірність не дістати карти жодної з чотирьох мастей дорівнює 0.

Отже, маємо

$$\mathbb{P}(S \cup D \cup H \cup C) = 4 \cdot \frac{\binom{39}{13}}{\binom{52}{13}} - 6 \cdot \frac{\binom{26}{13}}{\binom{52}{13}} + 4 \frac{1}{\binom{52}{13}} \approx 0.051 .$$

□

Вправа 3.5.4. Нехай μ_1, μ_2 — деякі σ -скінченні міри на деякому вимірному просторі $(\Omega, \mathcal{A})$, і нехай μ — деяка міра така, що

$$\mu(A) = \mu_1(A) + \mu_2(A) , \quad A \in \mathcal{A} .$$

Покажіть, що μ також σ -скінченна.

□

Розв'язання. Нехай маємо дві σ -скінченні міри μ_1 і μ_2 . Розгляньмо розбиття простору $\{A_1^1, A_2^1, \dots\}$ та $\{A_1^2, A_2^2, \dots\}$ такі, що

$$\mu_1(A_i^1) < \infty , \quad \mu_2(A_i^2) < \infty , \quad i = 1, 2, \dots .$$

Можна бачити, що множина всіх можливих перетинів $\{A_i^1 \cap A_j^2, i, j \geq 1\}$ також є розбиттям простору Ω . Тоді

$$\mu(A_i^1 \cap A_j^2) = \mu_1(A_i^1 \cap A_j^2) + \mu_2(A_i^1 \cap A_j^2) < \infty , \quad i, j \geq 1 ,$$

а тому μ також є σ -скінченною мірою.

□

4 Умовна ймовірність

4.1 Основні теоретичні відомості

Детальний огляд понять, потрібний для розв'язання задач на цьому занятті, наведено в Курсі лекцій, Розд. КЛЗ.1, КЛЗ.3–КЛЗ.4.

Визначення (КЛЗ.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

$$\mathbb{P}(A | B) = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)}. \quad (\text{КЛЗ.1.1})$$

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

Зауваження (КЛЗ.1.3). Варто зазначити, що запис $\mathbb{P}(A | B)$ не означає, що існує подія $A | B$, ймовірність 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 .

Формально, 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. $\square$

Уведена в такий спосіб умовна ймовірність справді є мірою.

Твердження (КЛЗ.1.4). Нехай маємо вимірний простір $(\Omega, \mathcal{A})$, $B \in \mathcal{A}$, і до того ж $\mathbb{P}(B) > 0$. Тоді умовна ймовірність $\mathbb{P}(\cdot | B)$ є ймовірнісною мірою в розумінні Визначення КЛ2.1.5.

Обумовлення події A подією B означає, що ми *перенормалізовуємо* ймовірнісну міру. 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.

Обумовлення однієї події іншою не є комутативним: $\mathbb{P}(A | B) \neq \mathbb{P}(B | A)$ у загальному випадку.

Із визначення умовної ймовірності (КЛЗ.1.1) випливає, що ймовірність перетину двох подій можна подати так:

$$\mathbb{P}(A \cap B) = \mathbb{P}(B) \mathbb{P}(A | B) = \mathbb{P}(A) \mathbb{P}(B | A), \quad \mathbb{P}(A) > 0, \mathbb{P}(B) > 0. \quad (\text{КЛЗ.1.2})$$

Відповідну властивість можна узагальнити.

Твердження (КЛЗ.1.7). Нехай $A_1, \dots, A_n$ — події такі, що для них виконується $\mathbb{P}(A_1 \cap \dots \cap A_n) > 0$. Тоді

$$\begin{aligned} \mathbb{P}(A_1 \cap \dots \cap A_n) &= \mathbb{P}(A_1) \mathbb{P}(A_2 | A_1) \mathbb{P}(A_3 | A_1, A_2) \dots \mathbb{P}(A_n | A_1, \dots, A_{n-1}) \\ &= \mathbb{P}(A_2) \mathbb{P}(A_1 | A_2) \mathbb{P}(A_3 | A_1, A_2) \dots \mathbb{P}(A_n | A_1, \dots, A_{n-1}) \\ &= \dots \end{aligned} \quad (\text{КЛЗ.1.3})$$

і т.д. для всіх можливих комбінацій подій.

Із визначення (КЛЗ.1.2) безпосередньо випливають такі твердження.

Теорема (Теорема Беєса (Bayes' Theorem)¹, КЛЗ.3.1). Нехай маємо вимірний простір $(\Omega, \mathcal{A})$. Для довільних подій $A, B \in \mathcal{A}$, $\mathbb{P}(B) > 0$, справедливо:

$$\mathbb{P}(A | B) = \frac{\mathbb{P}(B | A) \mathbb{P}(A)}{\mathbb{P}(B)}. \quad (\text{КЛЗ.3.1})$$

Теорема (Закон повної ймовірності (Law of total probability), КЛЗ.3.2). Нехай $A_1, \dots, A_n \in \mathcal{A}$ утворюють розбиття вимірного простору $(\Omega, \mathcal{A})$, і для всіх i $\mathbb{P}(A_i) > 0$, тобто $\Omega = \bigcup_{i=1}^n A_i$, $A_i \cap A_j = \emptyset$, $i \neq j$. Тоді

$$\mathbb{P}(B) = \sum_{i=1}^n \mathbb{P}(B | A_i) \mathbb{P}(A_i). \quad (\text{КЛЗ.3.2})$$

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

Твердження (КЛЗ.3.7). Нехай маємо вимірний простір $(\Omega, \mathcal{A})$. Для довільних подій $A, B, C \in \mathcal{A}$, $\mathbb{P}(B \cap C) > 0$, справедливо:

$$\mathbb{P}(A | B, C) = \frac{\mathbb{P}(B | A, C) \mathbb{P}(A | C)}{\mathbb{P}(B | C)}. \quad (\text{КЛЗ.3.3})$$

¹Томас Беєс (Thomas Bayes, 1701–1761) — англійський статистик, філософ і священник.

Твердження (КЛЗ.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

$$\mathbb{P}(B | C) = \sum_{i=1}^n \mathbb{P}(B | A_i, C) \mathbb{P}(A_i | C) . \quad (\text{КЛЗ.3.4})$$

Визначення (КЛЗ.4.1). Події A і B називають *незалежними* (independent) і позначають $A \perp B$, якщо

$$\mathbb{P}(A \cap B) = \mathbb{P}(A) \mathbb{P}(B) . \quad (\text{КЛЗ.4.1})$$

Якщо $\mathbb{P}(A) > 0$ і $\mathbb{P}(B) > 0$, то з (КЛЗ.1.1) випливає, що події незалежні, якщо

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

□

Іншими словами, події незалежні, коли обумовлення однієї події іншою в жодний спосіб не впливає на її ймовірність.

Твердження (КЛЗ.4.3). Якщо події $A \perp B$, то $A \perp B^c$, $A^c \perp B$, $A^c \perp B^c$.

Для декількох подій незалежність визначають так.

Визначення (КЛЗ.4.4). Події $A_1, \dots, A_n$ незалежні, що позначають $A_1 \perp \dots \perp A_n$, 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 події з будь-якої скінченної множини членів послідовності. □

Вимога незалежності для всіх пар, трійок і т.п. є принциповою, адже з попарної незалежності не випливає незалежність, а виконання тільки умови $\mathbb{P}(A_i \cap A_j \cap A_k) = \mathbb{P}(A_i) \mathbb{P}(A_j) \mathbb{P}(A_k)$ для всіх i, j, k різних не гарантує попарної незалежності.

Події також можуть бути незалежні умовно.

Визначення (КЛЗ.4.7). Події A і B називають *умовно незалежними* (conditionally independent) за умови настання події C , що позначають через $A \perp B | C$, якщо

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

□

Умовна й безумовна незалежності в загальному випадку між собою ніяк не пов'язані. Так, одні й ті ж події можуть бути умовно незалежні за умови однієї події і умовно залежні за умови деякої іншої. Події можуть бути незалежні умовно, але залежні безумовно, і навпаки.

4.2 Визначення умовної ймовірності

Вправа 4.2.1. Я впевнений на 80%, що мої ключі лежать в одній із двох кишень куртки — по 40% на кожну з двох кишень. Якщо я пориюся в одній кишені і не знайду там ключів, чому дорівнює ймовірність, що я знайду їх у другій кишені? $\square$

Розв'язання. Нехай L = «ключі лежать у лівій кишені», R = «ключі лежать у правій кишені». Тоді за (КЛЗ.1.1) маємо:

$$\mathbb{P}(R \mid L^c) = \frac{\mathbb{P}(R \cap L^c)}{\mathbb{P}(L^c)} = \frac{\mathbb{P}(R)}{1 - \mathbb{P}(L)} = \frac{2}{3}.$$

$\square$

Вправа 4.2.2. У вазі містяться 8 червоних куль і 4 білі кулі. Ми витягаємо 2 кулі без повторень. Якщо будь-яку кулю витягнути можна з однаковою ймовірністю, чому дорівнює ймовірність, що обидві кулі будуть червоні?

Нехай тепер червоні кулі мають вагу r кожна, а білі — w кожна. Нехай ймовірність витягнути кулю дорівнює відношенню її ваги до загальної ваги всієї вази. Чому в цьому випадку дорівнює ймовірність, що обидві кулі будуть червоні? $\square$

Розв'язання. Позначмо R_i = « i -та витягнута куля червона», $i = 1, 2$.

Для першого випадку, очевидно, застосовне класичне визначення ймовірності, і тому

$$\mathbb{P}(R_1 \cap R_2) = \frac{\binom{8}{2}}{\binom{12}{2}} = \frac{14}{33}.$$

Проте для другого випадку такий розв'язок буде зовсім незастосовний, а отже розгляньмо цю задачу через призму умовної ймовірності і (КЛЗ.1.1):

$$\mathbb{P}(R_1 \cap R_2) = \mathbb{P}(R_1) \mathbb{P}(R_2 \mid R_1) = \frac{2}{3} \cdot \frac{7}{11} = \frac{14}{33}.$$

Розгляньмо тепер другий випадок. Спочатку у вазі містяться кулі загальною вагою $8r + 4w$, тому ймовірність витягти деяку червону кулю дорівнює

$$\mathbb{P}(R_1) = 8 \cdot \frac{r}{8r + 4w} ,$$

оскільки вибрати можна будь-яку з 8 наявних червоних куль.

Після першого витягу у вазі залишаються кулі вагою $7r + 4w$, і за тією ж логікою остаточно маємо

$$\mathbb{P}(R_1 \cap R_2) = \mathbb{P}(R_1) \mathbb{P}(R_2 \mid R_1) = \frac{8r}{8r + 4w} \cdot \frac{7r}{7r + 4w} .$$

□

Вправа 4.2.3. У групі з 8 футбольних команд 4 команди вважають сильними, а інші 4 — слабкими. Пари суперників визначають у випадковий спосіб. Чому дорівнює ймовірність, що жодна з сильних команд не зіграє з сильним суперником у першому раунді? □

Розв'язання. Нехай W_i = «сильна команда i зіграє проти слабого суперника», $i = 1, 2, 3, 4$. Тоді за (КЛЗ.1.3) маємо:

$$\begin{aligned} \mathbb{P}(W_1 \cap W_2 \cap W_3 \cap W_4) &= \mathbb{P}(W_1) \mathbb{P}(W_2 \mid W_1) \mathbb{P}(W_3 \mid W_1, W_2) \\ &\quad \times \mathbb{P}(W_4 \mid W_1, W_2, W_3) . \end{aligned}$$

Команда 1 може зіграти з будь-якою командою з однаковою ймовірністю, тобто в неї 7 потенційних суперників, 4 з яких є слабкі, а отже $\mathbb{P}(W_1) = \frac{4}{7}$. Якщо ми знаємо, що команда 1 зіграла зі слабким суперником (W_1 сталася), то для команди 3 залишилося 2 сильні і 3 слабкі суперники, звідки $\mathbb{P}(W_2 \mid W_1) = \frac{3}{5}$. Аналогічні міркування для решти ситуацій дають

$$\mathbb{P}(W_1 \cap W_2 \cap W_3 \cap W_4) = \frac{4}{7} \cdot \frac{3}{5} \cdot \frac{2}{3} \cdot 1 = \frac{8}{35} .$$

□

4.3 Теорема Беєса та повна ймовірність

Вправа 4.3.1. У страховій компанії вважають, що клієнтів можна розділити на дві групи — з високим ризиком і з низьким ризиком. Історичні дані свідчать про те, що клієнт із високим ризиком потрапить у страховий випадок протягом 1 року з ймовірністю 0.4, тоді як для клієнта з низьким ризиком така ймовірність падає до 0.2.

Нехай серед усього населення 30% мають високий ризик настання страхового випадку. Чому дорівнює ймовірність того, що новий клієнт протягом 1 року потрапить у страховий випадок?

Нехай клієнт придбав поліс, і протягом 1 року стався страховий випадок. Чому дорівнює ймовірність, що він мав високий ризик?

Чому дорівнює ймовірність, що новий клієнт матиме страховий випадок після 1 року після купівлі полісу, *за умови*, що протягом 1 року вже мав місце страховий випадок? Можна вважати, що ймовірність настання страхового випадку після 1 року *для деякої конкретної особи* точно така ж, як і протягом 1 року. $\square$

Розв'язання. Нехай A_1 = «станеться страховий випадок протягом 1 року», R = «клієнт має високий ризик». Цілком очевидно, що R і R^c утворюють розбиття простору («усіх людей»). Тоді можна застосувати (КЛЗ.3.2):

$$\mathbb{P}(A_1) = \mathbb{P}(A_1 | R) \mathbb{P}(R) + \mathbb{P}(A_1 | R^c) \mathbb{P}(R^c) = 0.4 \cdot 0.3 + 0.2 \cdot 0.7 = 0.26.$$

Для другого питання застосуємо (КЛЗ.3.1):

$$\mathbb{P}(R | A_1) = \frac{\mathbb{P}(R) \mathbb{P}(A_1 | R)}{\mathbb{P}(A_1)} = \frac{0.3 \cdot 0.4}{0.26} = \frac{6}{13} \approx 0.462.$$

Для третього питання розгляньмо подію A_2 = «станеться страховий випадок після 1 року». Тоді шукана ймовірність, згідно з (КЛЗ.3.4), дорівнює

$$\begin{aligned} \mathbb{P}(A_2 | A_1) &= \mathbb{P}(A_2 | R, A_1) \mathbb{P}(R | A_1) + \mathbb{P}(A_2 | R^c, A_1) \mathbb{P}(R^c | A_1) \\ &= \mathbb{P}(A_2 | R) \mathbb{P}(R | A_1) + \mathbb{P}(A_2 | R^c) (1 - \mathbb{P}(R | A_1)) \\ &= 0.4 \cdot \frac{6}{13} + 0.2 \cdot \frac{7}{13} \approx 0.29. \end{aligned}$$

Варто звернути увагу, що $\mathbb{P}(A_2 | R, A_1) = \mathbb{P}(A_2 | R)$, тобто події $A_1 \perp\!\!\!\perp A_2 | R$ (*умовно* незалежні). Але $\mathbb{P}(A_2 | A_1) = 0.29 \neq 0.26 = \mathbb{P}(A_2)$, тобто вони не є *безумовно* незалежні. $\square$

Вправа 4.3.2. Стався злочин. Його скоїв один із n чоловіків, кожен із яких міг стати злочинцем з однаковою ймовірністю. Свідок стверджує, що злочин скоїв чоловік із 6 пальцями на правій руці.

Нехай p_0 — ймовірність того, що невинуватий чоловік має 6 пальців на правій руці, а p_1 — ймовірність того, що зловмисник має 6 пальців на правій руці, до того ж $p_0 < p_1$. Ймовірність p_1 необов'язково дорівнює 1, оскільки свідок міг помилятися або брехати. Також вважатимемо, що чоловіки можуть мати 6 пальців на правій руці незалежно один від одного.

Заарештовано підозрюваного. У нього 6 пальців на правій руці. Чому дорівнює ймовірність, що саме він є злочинець?

Уявімо собі, що поліція перевірила кожного з n можливих чоловіків, і виявилось, що затриманий — *єдиний* чоловік із 6 пальцями на правій руці. Чому дорівнює ймовірність, що *він* є злочинцем? $\square$

Розв'язання. Нехай A = «заарештований є злочинець» і B = «заарештований має 6 пальців на правій руці». Тоді ймовірності, викладені в умові вправи, можна записати як $\mathbb{P}(B | A) = p_1$, $\mathbb{P}(B | A^c) = p_0$, $\mathbb{P}(A) = \frac{1}{n}$.

За Теоремою КЛЗ.3.1 і формулою (КЛЗ.3.2),

$$\begin{aligned}\mathbb{P}(A | B) &= \frac{\mathbb{P}(B | A) \mathbb{P}(A)}{\mathbb{P}(B | A) \mathbb{P}(A) + \mathbb{P}(B | A^c) \mathbb{P}(A^c)} \\ &= \frac{p_1 \cdot \frac{1}{n}}{p_1 \cdot \frac{1}{n} + p_0 \left(1 - \frac{1}{n}\right)} = \frac{p_1}{p_1 + p_0(n - 1)}.\end{aligned}$$

Тепер нехай додатково N = «ніхто, окрім заарештованого, не має 6 пальців на правій руці». Тоді можемо повторно застосувати Теорему КЛЗ.3.1 і формулу (КЛЗ.3.2):

$$\mathbb{P}(A | B, N) = \frac{\mathbb{P}(B \cap N | A) \mathbb{P}(A)}{\mathbb{P}(B \cap N | A) \mathbb{P}(A) + \mathbb{P}(B \cap N | A^c) \mathbb{P}(A^c)}.$$

Ймовірність $\mathbb{P}(B \cap N | A)$ відповідає ймовірності того, що заарештований є єдиний, хто має 6 пальців, враховуючи, що він справді винуватий. Як і в попередньому випадку, заарештований має 6 пальців з ймовірністю p_1 , проте в цьому випадку ми повинні вимагати, щоб усі інші (невинуваті) чоловіки *не мали* 6 пальців, тобто $\mathbb{P}(B \cap N | A) = p_1(1 - p_0)^{n-1}$.

У свою чергу ймовірність $\mathbb{P}(B \cap N | A^c)$ відповідає ймовірності того, що заарештований є єдиний, хто має 6 пальців, враховуючи, що він справді невинуватий. Тоді заарештований має 6 пальців з ймовірністю p_0 (оскільки він невинуватий), а всі інші чоловіки не повинні мати 6 пальців. Серед $n - 1$ чоловіків один винуватий, тому для нього ймовірність не мати 6 пальців дорівнює $1 - p_1$, а для всіх інших — $1 - p_0$. Отже, $\mathbb{P}(B \cap N | A^c) = p_0(1 - p_1)(1 - p_0)^{n-2}$.

Остаточного маємо

$$\begin{aligned}\mathbb{P}(A | B, N) &= \frac{p_1(1 - p_0)^{n-1} \cdot \frac{1}{n}}{p_1(1 - p_0)^{n-1} \cdot \frac{1}{n} + p_0(1 - p_1)(1 - p_0)^{n-2} \cdot \left(1 - \frac{1}{n}\right)} \\ &= \frac{p_1(1 - p_0)}{p_1(1 - p_0) + p_0(1 - p_1)(n - 1)}.\end{aligned}$$

Приклад 4.3.3 (Задача Монті Голла (Monty Hall problem)). Розгляньмо класичну задачу, базовану на телевізійному шоу *Let's Make a Deal*, що багато років ішло на американському телеканалі NBC. У найпростішій версії цієї гри за одними з трьох зачинених дверей розташовано машину, а за двома іншими — козу. Учасники гри знають, що машина ховається тільки за одними з дверей, але не знають, за якими саме.

Спочатку учасник вибирає одні з дверей, але не відчиняє їх. Ведучий шоу Монті Голл, на честь якого і названо відповідну задачу, відчиняє одні з двох дверей, що залишилися, і показує, що за ними ховається коза. Оскільки незалежно від вибору перших дверей за одними з невибраних завжди буде перебувати коза, Монті Голл завжди може відчинити двері саме з козою.

Учаснику пропонують ухвалити рішення: зберегти свій початковий вибір або змінити його на користь тих із двох невибраних дверей, які Монті Холл поки ще не відчиняв. Після остаточного рішення гравця відчиняють усі двері і стає зрозумілим його остаточний виграш.

Питання, яке вирінає в цьому контексті: чи повинен учасник міняти своє початкове рішення і вибирати інші двері?

Пронумеруймо двері від 1 до 3 і, не зменшуючи загалу, вважатимемо, що учасник завжди вибирає двері номер 1 (в іншому випадку їх можна просто перенумерувати). Уведімо події C_i = «машина ховається за дверима номер i », $i = 1, 2, 3$. Імовірність події A = «виграно машину, якщо початковий вибір не змінено» можна обчислити за допомогою формули (КЛЗ.3.2):

$$\mathbb{P}(A) = \mathbb{P}(A | C_1) \cdot \frac{1}{3} + \mathbb{P}(A | C_2) \cdot \frac{1}{3} + \mathbb{P}(A | C_3) \cdot \frac{1}{3} = 1 \cdot \frac{1}{3} + 0 \cdot \frac{1}{3} + 0 \cdot \frac{1}{3} = \frac{1}{3},$$

оскільки ймовірність, що машина ховається за кожними з дверей, є однаковою, а за умови збереження початкового вибору виграти можна тільки тоді, коли машина справді ховається за дверима 1.

Якщо ж гравець змінить свій початковий вибір, то ймовірність події B = «виграно машину, якщо початковий вибір змінено» буде

$$\mathbb{P}(B) = \mathbb{P}(B | C_1) \cdot \frac{1}{3} + \mathbb{P}(B | C_2) \cdot \frac{1}{3} + \mathbb{P}(B | C_3) \cdot \frac{1}{3} = 0 \cdot \frac{1}{3} + 1 \cdot \frac{1}{3} + 1 \cdot \frac{1}{3} = \frac{2}{3},$$

оскільки Монті Голл відчиняє або двері 2, або двері 3, і тому коли учасник змінює свій початковий вибір, він точно знає, які з двох дверей — 2 чи 3 — вибрати. Ілюстративно відповідні міркування зображено на Рис. 4.3.1.

Цей результат можна дістати, використавши (КЛЗ.3.1). Нехай M_j = «Монті відчинив двері j », $j = 2, 3$. Тоді

$$\mathbb{P}(C_1 | M_2) = \frac{\mathbb{P}(M_2 | C_1) \mathbb{P}(C_1)}{\mathbb{P}(M_2)} = \frac{\frac{1}{2} \cdot \frac{1}{3}}{\frac{1}{2}} = \frac{1}{3},$$

тобто ймовірність, що виграш ховається за дверима 1, за умови, що Монті

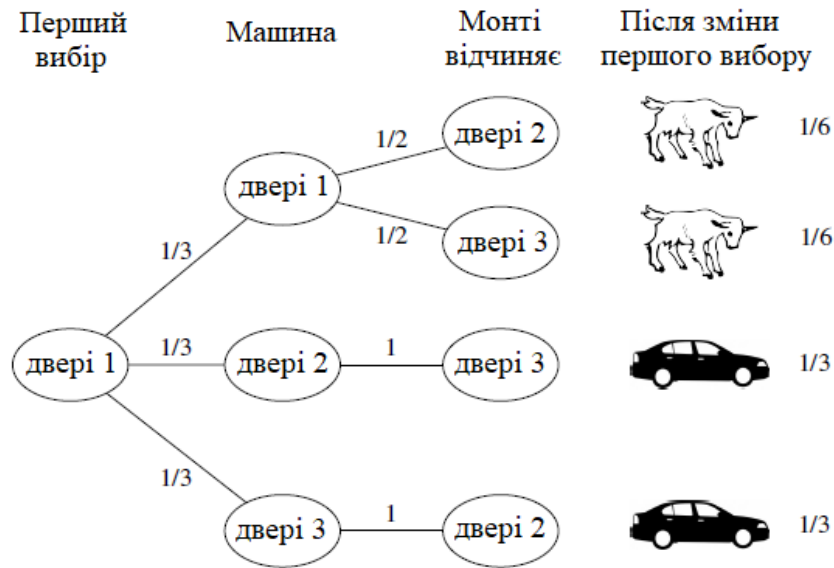


Рис. 4.3.1: Дерево ухвалення рішень для Прикладу 4.3.3 (адаптовано з [1], Рис. 2.5)

відчинив двері 2, дорівнює $\frac{1}{3}$. Оскільки Монті може відчинити двері 2 або 3 з однаковою ймовірністю, також маємо $\mathbb{P}(C_1 | M_3) = \frac{1}{3}$. Нарешті,

$$\begin{aligned} \mathbb{P}(M_2) &= \mathbb{P}(M_2 | C_1) \mathbb{P}(C_1) + \mathbb{P}(M_2 | C_2) \mathbb{P}(C_2) + \mathbb{P}(M_2 | C_3) \mathbb{P}(C_3) \\ &= \frac{1}{2} \cdot \frac{1}{3} + 0 \cdot \frac{1}{3} + 1 \cdot \frac{1}{3} = \frac{1}{2}. \end{aligned}$$

Тобто, які б двері Монті не відчинив, доцільно змінити початковий вибір і збільшити ймовірність виграшу до $\frac{2}{3}$.

Багато людей вважають, що початковий вибір змінювати немає сенсу, адже після того, як Монті відчинив двері без машини, залишається ймовірність $\frac{1}{2}$, що машина ховається за одними з невідчинених дверей. Такі міркування неправильні, адже вони застосовують класичне визначення ймовірностей там, де це недоречно. Після того, як американська колумністка Мерилін vos Савант (Marilyn vos Savant) у 1990 р. у своїй колонці в журналі «Parade» розглянула розв'язок цієї задачі, їй почали надходити тисячі листів із гнівними доповненнями її розв'язку.

Для підсилення розуміння цієї задачі, уявіть, що існує не 3, а 1 000 000 дверей, і після початкового вибору Монті відчиняє 999 998 дверей і показує, що там ховаються кози. Чи продовжували б Ви далі триматися свого початкового вибору?.. $\square$

Вправа 4.3.4. У коробці містяться ліхтарики трьох типів: 20% типу 1, 30% типу 2 і 50% типу 3. На основі наявних даних відомо, що ліхтарик типу 1 пропрацює понад 100 годин з ймовірністю 0.7. Для ліхтариків типів 2 і 3 відповідні ймовірності дорівнюють 0.4 і 0.3.

Чому дорівнює ймовірність, що вибраний навімання ліхтарик пропрацює понад 100 годин? Якщо вибраний ліхтарик пропрацював понад 100 годин, чому дорівнює ймовірність, що він був типу 1? 2? 3? $\square$

Розв'язання. Нехай A = «ліхтарик пропрацює понад 100 годин», F_j = «вибрано ліхтарик типу j », $j = 1, 2, 3$. Оскільки три типи ліхтариків утворюють розбиття простору, для обчислення $\mathbb{P}(A)$ можна застосувати (КЛЗ.3.2):

$$\mathbb{P}(A) = \sum_{j=1}^3 \mathbb{P}(A | F_j) \mathbb{P}(F_j) = 0.7 \cdot 0.2 + 0.4 \cdot 0.3 + 0.3 \cdot 0.5 = 0.41 .$$

Для обчислення ймовірності $\mathbb{P}(F_j | A)$, $j = 1, 2, 3$, застосуємо (КЛЗ.3.1):

$$\begin{aligned} \mathbb{P}(F_1 | A) &= \frac{\mathbb{P}(A | F_1) \mathbb{P}(F_1)}{\mathbb{P}(A)} = \frac{0.7 \cdot 0.2}{0.41} = \frac{14}{41} , \\ \mathbb{P}(F_2 | A) &= \frac{\mathbb{P}(A | F_2) \mathbb{P}(F_2)}{\mathbb{P}(A)} = \frac{0.4 \cdot 0.3}{0.41} = \frac{12}{41} , \\ \mathbb{P}(F_3 | A) &= \frac{\mathbb{P}(A | F_3) \mathbb{P}(F_3)}{\mathbb{P}(A)} = \frac{0.3 \cdot 0.5}{0.41} = \frac{15}{41} . \end{aligned} \quad \square$$

4.4 Незалежність подій

Вправа 4.4.1. Покажіть, що якщо події A , B і C незалежні, то

$$\mathbb{P}(A \cup B \cup C) = 1 - (1 - \mathbb{P}(A))(1 - \mathbb{P}(B))(1 - \mathbb{P}(C)) . \quad (4.4.1)$$

$\square$

Розв'язання. Для того, щоб використати умову незалежності подій, потрібно перейти від їх об'єднання до їх перетину. Це можна зробити за допомогою законів де Моргана:

$$\mathbb{P}(A \cup B \cup C) = \mathbb{P}((A^c \cap B^c \cap C^c)^c) = 1 - \mathbb{P}(A^c \cap B^c \cap C^c) .$$

Згідно з Твердженням КЛЗ.4.3, якщо події незалежні, то незалежні також і їх доповнення, отже можемо застосувати Визначення КЛЗ.4.4:

$$\begin{aligned} 1 - \mathbb{P}(A^c \cap B^c \cap C^c) &= 1 - \mathbb{P}(A^c) \mathbb{P}(B^c) \mathbb{P}(C^c) \\ &= 1 - (1 - \mathbb{P}(A))(1 - \mathbb{P}(B))(1 - \mathbb{P}(C)) . \end{aligned}$$

$\square$

Вправа 4.4.2. Немовля плаче тоді й тільки тоді, коли воно голодне, втомлене, або одночасно голодне і втомлене. Нехай C = «дитина плаче», H = «дитина голодна», T = «дитина втомлена», і нехай $\mathbb{P}(C) = c$, $\mathbb{P}(H) = h$, $\mathbb{P}(T) = t$, $0 < c, h, t < 1$. Нехай $H \perp T$.

Чому дорівнюють імовірності $\mathbb{P}(H | C)$, $\mathbb{P}(T | C)$ і $\mathbb{P}(H \cap T | C)$?

Чи можна казати, що $H \perp T | C$? □

Розв'язання. Застосуємо Теорему КЛЗ.3.1 для обчислення умовних імовірностей:

$$\begin{aligned}\mathbb{P}(H | C) &= \frac{\mathbb{P}(C | H) \mathbb{P}(H)}{\mathbb{P}(C)} = \frac{1 \cdot h}{c} = \frac{h}{c}, \\ \mathbb{P}(T | C) &= \frac{\mathbb{P}(C | T) \mathbb{P}(T)}{\mathbb{P}(C)} = \frac{1 \cdot t}{c} = \frac{t}{c}, \\ \mathbb{P}(H \cap T | C) &= \frac{\mathbb{P}(C | H, T) \mathbb{P}(H \cap T)}{\mathbb{P}(C)} = \frac{1 \cdot ht}{c} = \frac{ht}{c}.\end{aligned}$$

Для перевірки умовної незалежності перевірмо виконання рівності

$$\mathbb{P}(H \cap T | C) = \mathbb{P}(H | C) \mathbb{P}(T | C).$$

Вона очевидно *не виконується*, адже зліва стоїть щойно підрахований вираз $\frac{ht}{c}$, а справа стоїть $\frac{h}{c} \cdot \frac{t}{c} = \frac{ht}{c^2}$.

Інтуїтивне пояснення, чому незалежність не виконується для дитини, що плаче, дуже просте: якщо дитина плаче і вона не голодна, то вона повинна бути втомлена, тобто голод і втома для заплаканої дитини не можуть бути незалежними. □

Розгляньмо класичну задачу, у якій розв'язання шукають за допомогою ітеративного застосування закону повної ймовірності.

Приклад 4.4.3 (Задача про розорення гравця (Gambler's ruin problem)). Двоє осіб, A і B , беруть участь у такій грі: вони підкидають монетку, і якщо випадає герб, то гравець B віддає гравцю A 1 гривню, а якщо випадає число, то 1 гривню віддає гравець A гравцю B . Гра триває, допоки хтось із учасників не втратить усіх грошей. Підкидання монетки незалежні одне від одного, а ймовірність випадку герба дорівнює $\mathbb{P}(H) = p$.

Чому дорівнює ймовірність події E , що A виграє всі гроші, якщо на початку гри він має i гривень, а B має $(N - i)$ гривень?

Позначмо ймовірність $\mathbb{P}(E)$ через P_i , підкреслюючи, що це ймовірність виграти, якщо на початку гри учасник A має i гривень. Застосуємо фор-

мулу повної ймовірності (КЛ3.3.2), обумовивши першим випадом монетки:

$$\begin{aligned} P_i &= \mathbb{P}(E) = \mathbb{P}(E \mid H) \mathbb{P}(H) + \mathbb{P}(E \mid H^c) \mathbb{P}(H^c) \\ &= p \mathbb{P}(E \mid H) + (1 - p) \mathbb{P}(E \mid H^c) . \end{aligned}$$

Якщо перший випад був герб, то в учасника A буде $i + 1$ гривень, а в учасника B буде $N - (i + 1)$ гривень. Оскільки підкидання монетки незалежні одне від одного, можемо розглянути *частину гри*, яка починається після випаду герба, як *окрему гру*. Імовірність виграшу учасника A в такій грі, $\mathbb{P}(E \mid H)$, можна позначити через P_{i+1} , підкресливши, що це ймовірність виграти, якщо на початку гри учасник A має $i + 1$ гривень.

У схожий спосіб впливає, що $\mathbb{P}(E \mid H^c) = P_{i-1}$. Отже

$$P_i = p P_{i+1} + (1 - p) P_{i-1} , \quad i = 1, 2, \dots, N - 1 . \quad (4.4.2)$$

Ми записали $N - 1$ рівнянь з $N + 1$ невідомими (а саме $P_0, P_1, \dots, P_N$). Можна додати два додаткові рівняння, а саме $P_0 = 0$ і $P_N = 1$, оскільки в цих випадках учасники B і A відповідно виграють на самому початку гри.

Розв'язати цю систему можна, переписавши (4.4.2) для $i = 1, 2, \dots, N - 1$:

$$p P_i + (1 - p) P_i = p P_{i+1} + (1 - p) P_{i-1} \quad \Rightarrow \quad P_{i+1} - P_i = \frac{1 - p}{p} (P_i - P_{i-1}) .$$

Оскільки $P_0 = 0$, можемо «розплутати клубок» так:

$$\begin{aligned} P_2 - P_1 &= \frac{1 - p}{p} P_1 , \\ P_3 - P_2 &= \frac{1 - p}{p} (P_2 - P_1) = \left(\frac{1 - p}{p} \right)^2 P_1 , \\ &\vdots \\ P_i - P_{i-1} &= \frac{1 - p}{p} (P_{i-1} - P_{i-2}) = \left(\frac{1 - p}{p} \right)^{i-1} P_1 , \\ &\vdots \\ P_N - P_{N-1} &= \frac{1 - p}{p} (P_{N-1} - P_{N-2}) = \left(\frac{1 - p}{p} \right)^{N-1} P_1 . \end{aligned}$$

Додавши перші $i - 1$ рівнянь, дістаємо

$$P_i - P_1 = P_1 \cdot \sum_{k=1}^{i-1} \left(\frac{1 - p}{p} \right)^k ,$$

звідки, з урахуванням формули суми геометричної прогресії,

$$P_i = \begin{cases} \frac{1 - ((1-p)/p)^i}{1 - (1-p)/p} P_1, & \frac{1-p}{p} \neq 1 \\ iP_1, & \frac{1-p}{p} = 1 \end{cases}.$$

Оскільки $P_N = 1$, маємо

$$1 = \begin{cases} \frac{1 - ((1-p)/p)^N}{1 - (1-p)/p} P_1, & p \neq \frac{1}{2} \\ NP_1, & p = \frac{1}{2} \end{cases},$$

звідки

$$P_1 = \begin{cases} \frac{1 - (1-p)/p}{1 - ((1-p)/p)^N}, & p \neq \frac{1}{2} \\ \frac{1}{N}, & p = \frac{1}{2} \end{cases}.$$

Остаточно маємо

$$P_i = \begin{cases} \frac{1 - ((1-p)/p)^i}{1 - ((1-p)/p)^N}, & p \neq \frac{1}{2} \\ \frac{i}{N}, & p = \frac{1}{2} \end{cases}.$$

Позначивши через Q_i імовірність виграшу учасника B , якщо він починає з $N - i$ гривнями, дістаємо, за абсолютною симетрією,

$$Q_i = \begin{cases} \frac{1 - (p/(1-p))^{N-i}}{1 - (p/(1-p))^N}, & 1-p \neq \frac{1}{2} \\ \frac{N-i}{N}, & 1-p = \frac{1}{2} \end{cases}.$$

Якщо додати P_i і Q_i , то у випадку $p = \frac{1}{2}$ ми маємо $\frac{i}{N} + \frac{N-i}{N} = 1$, але цей же результат також справедливий і для випадку $p \neq \frac{1}{2}$:

$$\begin{aligned} P_i + Q_i &= \frac{1 - ((1-p)/p)^i}{1 - ((1-p)/p)^N} + \frac{1 - (p/(1-p))^{N-i}}{1 - (p/(1-p))^N} \\ &= \frac{p^N - p^N \left(\frac{1-p}{p}\right)^i}{p^N - (1-p)^N} + \frac{(1-p)^N - (1-p)^N \left(\frac{p}{1-p}\right)^{N-i}}{(1-p)^N - p^N} \\ &= \frac{p^N - p^{N-i}(1-p)^i - (1-p)^N + (1-p)^i p^{N-i}}{p^N - (1-p)^N} \\ &= 1. \end{aligned}$$

Що ми показали? Оскільки сума ймовірностей перемог кожного з учасників дорівнює 1, це означає, що ймовірність того, що ніхто з них не виграє, дорівнює 0. Іншими словами, гра *обов'язково* повинна закінчитися за скінченну кількість кроків.

Наприклад, нехай A починає з 5 гривнями, а B — з 10 гривнями. Тоді, якщо $p = 0.5$,

$$P_5 = \frac{5}{15} = \frac{1}{3},$$

а якщо $p = 0.6$,

$$P_5 = \frac{1 - ((1 - 0.6)/0.6)^5}{1 - ((1 - 0.6)/0.6)^{15}} \approx 0.87.$$

Імовірність p має значний вплив на результат гри. Нехай обидва гравці починають грати зі 100 гривнями кожен. Тоді, якщо $p = 0.5$, очевидно, що $P_{100} = 0.5$, проте якщо $p = 0.49$, маємо

$$P_{100} = \frac{1 - ((1 - 0.49)/0.49)^{100}}{1 - ((1 - 0.49)/0.49)^{200}} \approx 0.018,$$

а якщо $p = 0.51$, маємо

$$P_{100} = \frac{1 - ((1 - 0.51)/0.51)^{100}}{1 - ((1 - 0.51)/0.51)^{200}} \approx 0.982.$$

□

Вправа 4.4.4. У турнірі планують зіграти 64 команди, які розбито на 4 групи по 16 команд². У кожній групі команди пронумеровано від 1 до 16 залежно від їхнього рейтингу. У кожній групі команди грають одна з одною на виліт. Графік ігор наведено на Рис. 4.4.1.

Позначмо через $r(i, j) = r(j, i)$, $i \neq j$, номер раунду, у якому можуть зіграти між собою команди i і j . Наприклад, $r(1, 16) = 1$, оскільки команди 1 і 16 грають між собою в першому ж раунді, а $r(1, 5) = 3$, оскільки команди 1 і 5 потенційно можуть зіграти між собою у півфіналі.

Покладімо, що якщо команди i і j грають одна з одною, команда i виграє з імовірністю $p_{i,j} = 1 - p_{j,i}$. За такого підходу до моделювання ми не беремо до уваги історію турніру: ми вважаємо, що ймовірність перемоги деякої команди не залежить від її попередніх успіхів чи невдач. Це дуже велике припущення, яке на практиці може виконуватися або не виконуватися.

Чому дорівнює ймовірність P_i перемоги команди i у групі, $i = 1, \dots, 16$?

□

²Опис відповідного турніру максимально наближено до турніру між університетськими командами з баскетболу в США під егідою Національної асоціації університетського спорту (National Collegiate Athletic Association, NCAA).

Розв'язання. Щоб обчислити P_i , ми розглянемо чотири ймовірності $P_i(k)$, $k = 1, 2, 3, 4$, того, що команда i виграє в k -ому раунді.

Нехай $O_i(k) = \{j : r(i, j) = k\}$ — множина можливих суперників команди i в раунді k . Для обчислення $P_i(k)$ ми використаємо підхід, схожий до Прикладу 4.4.3, і розглянемо ітеративне застосування закону повної ймовірності. Так, цілком очевидно, що для перемоги в раунді k потрібно перемогти всіх суперників у раунді $k - 1$:

$$P_i(k) = \sum_{j \in O_i(k)} \mathbb{P}(\text{ком. } i \text{ виграла } k \text{ ігор} \mid \text{ком. } j \text{ дійшла до раунду } k) \\ \times P_j(k - 1) .$$

За побудовою турніру команда i не може зустрітися з жодною командою з $O_i(k)$ у раундах $1, \dots, k - 1$, а тому

$$\begin{aligned} & \mathbb{P}(i \text{ виграла } k \text{ ігор} \mid j \text{ дійшла до раунду } k) \\ &= \mathbb{P}((i \text{ виграла } k - 1 \text{ ігор}) \cap (i \text{ перемогла } j) \mid j \text{ дійшла до раунду } k) \\ &= \mathbb{P}((i \text{ виграла } k - 1 \text{ ігор}) \cap (i \text{ перемогла } j)) \cdot \mathbb{P}(j \text{ дійшла до раунду } k) \\ &= P_i(k - 1)p_{i,j} , \end{aligned}$$

де ми використали незалежність двох подій, оскільки перемоги команд i і j до раунду k ніяк між собою не пов'язані.

Тоді маємо

$$P_i(k) = \sum_{j \in O_i(k)} P_i(k - 1)p_{i,j}P_j(k - 1) = P_i(k - 1) \sum_{j \in O_i(k)} p_{i,j}P_j(k - 1) .$$

Очевидно, $P_i(0) = 1$, $i = 1, \dots, 16$, тому можна підрахувати всі ймовірності в рекурсивний спосіб.

Наприклад, нехай $p_{i,j} = \frac{j}{i+j}$. Тоді підрахунок $P_i \equiv P_i(4)$ вестиметься так. Ймовірності перемог у першому раунді будуть

$$\begin{aligned} P_1(1) &= p_{1,16} = \frac{16}{17} = 1 - P_{16}(1) , \\ P_2(1) &= p_{2,15} = \frac{15}{17} = 1 - P_{15}(1) , \\ P_3(1) &= p_{3,14} = \frac{14}{17} = 1 - P_{14}(1) , \\ P_4(1) &= p_{4,13} = \frac{13}{17} = 1 - P_{13}(1) , \\ P_5(1) &= p_{5,12} = \frac{12}{17} = 1 - P_{12}(1) , \\ P_6(1) &= p_{6,11} = \frac{11}{17} = 1 - P_{11}(1) , \end{aligned}$$

$$P_7(1) = p_{7,10} = \frac{10}{17} = 1 - P_{10}(1) ,$$

$$P_8(1) = p_{8,9} = \frac{9}{17} = 1 - P_9(1) .$$

Ймовірності $P_i(2)$ дістанемо рекурсивно. Наприклад, $O_1(2) = \{8, 9\}$, тому

$$P_1(2) = P_1(1)(P_8(1)p_{1,8} + P_9(1)p_{1,9}) = \frac{16}{17} \left(\frac{9}{17} \cdot \frac{8}{9} + \frac{8}{17} \cdot \frac{9}{10} \right) \approx 0.842 .$$

Так само рекурсивно можна дістати й усі інші ймовірності. $\square$

4.5 Додаткові вправи

Вправа 4.5.1. Нехай $\mathbb{P}(C) \mathbb{P}(C^c) > 0$. Доведіть, що одночасне виконання нерівностей $\mathbb{P}(A | C) > \mathbb{P}(B | C)$ і $\mathbb{P}(A | C^c) > \mathbb{P}(B | C^c)$ має своїм наслідком $\mathbb{P}(A) > \mathbb{P}(B)$.

Інтуїтивна інтерпретація цілком зрозуміла: якщо одна подія ймовірніша від іншої як за умови C , так і за умови C^c , то вона ймовірніша й безумовно. $\square$

Розв'язання. Із (КЛЗ.1.1) випливає

$$\begin{aligned} \mathbb{P}(A | C) > \mathbb{P}(B | C) &\Rightarrow \mathbb{P}(A \cap C) > \mathbb{P}(B \cap C) , \\ \mathbb{P}(A | C^c) > \mathbb{P}(B | C^c) &\Rightarrow \mathbb{P}(A \cap C^c) > \mathbb{P}(B \cap C^c) , \end{aligned}$$

де ми неявно використали той факт, що $\mathbb{P}(C) > 0$, $\mathbb{P}(C^c) > 0$. Тоді

$$\mathbb{P}(A \cap C) + \mathbb{P}(A \cap C^c) > \mathbb{P}(B \cap C) + \mathbb{P}(B \cap C^c) ,$$

звідки $\mathbb{P}(A) > \mathbb{P}(B)$, адже $A = (A \cap C) \cup (A \cap C^c)$, $B = (B \cap C) \cup (B \cap C^c)$. $\square$

Вправа 4.5.2. Нехай правильну монетку підкидують двічі. Чому дорівнює ймовірність, що обидва випадки будуть герби, якщо перший випадок був герб? Щонайменше один випадок був герб? $\square$

Розв'язання. Нехай B = «випало два герби», F = «перший випадок був герб», A = «принаймні один випадок був герб». Тоді за (КЛЗ.1.1) маємо

$$\begin{aligned} \mathbb{P}(B | F) &= \frac{\mathbb{P}(B \cap F)}{\mathbb{P}(F)} = \frac{1/4}{2/4} = \frac{1}{2} , \\ \mathbb{P}(B | A) &= \frac{\mathbb{P}(B \cap A)}{\mathbb{P}(A)} = \frac{1/4}{3/4} = \frac{1}{3} . \end{aligned}$$

$\square$

Вправа 4.5.3. Якщо у випадковий спосіб роздати 52 гральні карти 4 гравцям, по 13 карт кожному, чому дорівнює ймовірність, що в кожного буде по одному тузу? $\square$

Розв'язання. Розгляньмо події $E_i = \text{«у гравця } i \text{» міститься точно 1 туз, } i = 1, 2, 3, 4$. Тоді за (КЛЗ.1.3) маємо:

$$\mathbb{P}(E_1 \cap E_2 \cap E_3 \cap E_4) = \mathbb{P}(E_1) \mathbb{P}(E_2 | E_1) \mathbb{P}(E_3 | E_1, E_2) \mathbb{P}(E_4 | E_1, E_2, E_3) .$$

За класичним визначенням ймовірності маємо, що

$$\mathbb{P}(E_1) = \frac{\binom{4}{1} \binom{48}{12}}{\binom{52}{13}} .$$

За умови, що E_1 сталася, у колоді залишається 39 карт і 3 тузи, тому

$$\mathbb{P}(E_2 | E_1) = \frac{\binom{3}{1} \binom{36}{12}}{\binom{39}{13}} .$$

Аналогічні міркування дають

$$\mathbb{P}(E_3 | E_1, E_2) = \frac{\binom{2}{1} \binom{24}{12}}{\binom{26}{13}} , \quad \mathbb{P}(E_4 | E_1, E_2, E_3) = \frac{\binom{1}{1} \binom{12}{12}}{\binom{13}{13}} = 1 ,$$

отже

$$\mathbb{P}(E_1 \cap E_2 \cap E_3 \cap E_4) \approx 0.106 .$$

$\square$

Вправа 4.5.4. Нехай є дві монетки, правильна та така, що в неї ймовірність випадку герба дорівнює p . Якщо у випадковий спосіб вибрана монетка тричі поспіль випадає гербом, яка ймовірність, що вона правильна?

Якщо підкинути монетку четвертий раз, яка ймовірність випадку герба? $\square$

Розв'язання. Нехай $A = \text{«на монетці тричі випав герб»}$ і $F = \text{«вибрано правильну монетку»}$. Нас цікавить ймовірність $\mathbb{P}(F | A)$, але значно простіше обчислити ймовірності $\mathbb{P}(A | F)$ і $\mathbb{P}(A | F^c)$, тому ми використаємо Теорему Беєса КЛЗ.3.1 і формулу повної ймовірності (КЛЗ.3.2):

$$\begin{aligned} \mathbb{P}(F | A) &= \frac{\mathbb{P}(A | F) \mathbb{P}(F)}{\mathbb{P}(A)} \\ &= \frac{\mathbb{P}(A | F) \mathbb{P}(F)}{\mathbb{P}(A | F) \mathbb{P}(F) + \mathbb{P}(A | F^c) \mathbb{P}(F^c)} \end{aligned}$$

$$\begin{aligned}
&= \frac{\frac{1}{2^3} \cdot \frac{1}{2}}{\frac{1}{2^3} \cdot \frac{1}{2} + p^3 \cdot \frac{1}{2}} \\
&= \frac{1}{1 + 8p^3}.
\end{aligned}$$

Так, наприклад, якщо ймовірність випадку герба на другій монетці дорівнює $p = 0.75$, то $\mathbb{P}(F | A) \approx 0.23$.

Нехай тепер H = «монетка випала гербом четвертий раз». Нас цікавить ймовірність $\mathbb{P}(H | A)$. Можемо застосувати формулу (КЛЗ.3.4):

$$\mathbb{P}(H | A) = \mathbb{P}(H | A, F) \mathbb{P}(F | A) + \mathbb{P}(H | A, F^c) \mathbb{P}(F^c | A).$$

Ми вже знаємо, що $\mathbb{P}(F | A) = \frac{1}{1+8p^3}$, а отже $\mathbb{P}(F^c | A) = 1 - \frac{1}{1+8p^3}$. Що ж стосується ймовірностей $\mathbb{P}(H | A, F)$ і $\mathbb{P}(H | A, F^c)$, то обумовлення подією A ні на що не впливає, адже випадки монетки незалежні один від одного:

$$\begin{aligned}
\mathbb{P}(H | A, F) &= \mathbb{P}(H | F) = \frac{1}{2}, \\
\mathbb{P}(H | A, F^c) &= \mathbb{P}(H | F^c) = p.
\end{aligned}$$

Остаточнo маємо

$$\begin{aligned}
\mathbb{P}(H | A) &= \mathbb{P}(H | A, F) \mathbb{P}(F | A) + \mathbb{P}(H | A, F^c) \mathbb{P}(F^c | A) \\
&= \frac{1}{2} \cdot \frac{1}{1 + 8p^3} + p \left(1 - \frac{1}{1 + 8p^3} \right) = \frac{8p^4}{1 + 8p^3}.
\end{aligned}$$

Так, наприклад, якщо ймовірність випадку герба на другій монетці дорівнює $p = 0.75$, то $\mathbb{P}(H | A) \approx 0.69$. □

Вправа 4.5.5. Відповідаючи на питання тесту з декількома варіантами відповідей, студент на якісь питання знає відповідь, а на якісь відповідає наважання. Нехай p — ймовірність того, що студент знає відповідь, а $(1-p)$ — що він угадує відповідь. Додатково вважатимемо, що студент може вгадати відповідь з ймовірністю $\frac{1}{m}$, де m — кількість варіантів відповідей.

Чому дорівнює ймовірність, що студент знає відповідь на питання, якщо він відповів на нього правильно? □

Розв'язання. Нехай C = «студент відповідає правильно», K = «студент

знає відповідь на питання». Використаймо (КЛЗ.3.1) і (КЛЗ.3.2):

$$\begin{aligned}\mathbb{P}(K | C) &= \frac{\mathbb{P}(C | K) \mathbb{P}(K)}{\mathbb{P}(C | K) \mathbb{P}(K) + \mathbb{P}(C | K^c) \mathbb{P}(K^c)} \\ &= \frac{1 \cdot p}{1 \cdot p + \frac{1}{m} \cdot (1 - p)} = \frac{mp}{1 + (m - 1)p}.\end{aligned}$$

Наприклад, якщо $m = 5$, $p = 0.5$, студент справді знає відповідь на питання, на яке він дав правильну відповідь, з ймовірністю $\frac{5}{6}$, тобто важко стверджувати, що студент, який правильно відповів, «усього лише вгадав».

□

Вправа 4.5.6. Нехай $\mathbb{P}(B) \mathbb{P}(B^c) > 0$. Доведіть, що події A і B незалежні тоді й тільки тоді, коли $\mathbb{P}(A | B) = \mathbb{P}(A | B^c)$.

Це твердження має очевидну інтуїтивну інтерпретацію: якщо A незалежна від B , то додаткова інформація про B в жодний спосіб не повинна впливати на ймовірність A . Наведене твердження додатково каже, що інформація про B^c також у жодний спосіб не повинна впливати. Понад те, якщо і інформація про B , і інформація про B^c не впливає на ймовірність A , то $A \perp B$.

□

Розв'язання. Доведення в один бік очевидне, оскільки для незалежних подій $A \perp B$, відповідно до Визначення КЛЗ.4.1, моментально виконується $\mathbb{P}(A | B) = \mathbb{P}(A)$. Згідно з Твердженням КЛЗ.4.3, $A \perp B^c$, а отже і $\mathbb{P}(A | B^c) = \mathbb{P}(A)$, звідки $\mathbb{P}(A | B) = \mathbb{P}(A | B^c)$.

Покладімо тепер, що $\mathbb{P}(A | B) = \mathbb{P}(A | B^c)$. За формулою (КЛЗ.1.1),

$$\frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)} = \frac{\mathbb{P}(A \cap B^c)}{\mathbb{P}(B^c)} \Rightarrow \mathbb{P}(A \cap B) \mathbb{P}(B^c) = \mathbb{P}(A \cap B^c) \mathbb{P}(B).$$

Ми можемо собі дозволити виконати це перемноження, оскільки умова вправи $\mathbb{P}(B) \mathbb{P}(B^c) > 0$ означає, що обидві ймовірності, $\mathbb{P}(B)$ і $\mathbb{P}(B^c)$, повинні бути ненульові.

Звідси випливає

$$\begin{aligned}\mathbb{P}(A \cap B) (1 - \mathbb{P}(B)) &= \mathbb{P}(A \cap B^c) \mathbb{P}(B) \Rightarrow \\ \mathbb{P}(A \cap B) &= (\mathbb{P}(A \cap B) + \mathbb{P}(A \cap B^c)) \mathbb{P}(B) = \mathbb{P}(A) \mathbb{P}(B),\end{aligned}$$

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

□

5 Дискретні випадкові величини. Схема Бернуллі

5.1 Основні теоретичні відомості

Детальний огляд понять, потрібний для розв'язання задач на цьому занятті, наведено в Курсі лекцій, Розд. КЛ4.

Для кожного експерименту можна визначити деякий простір елементарних подій Ω , задати на ньому σ -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 (елемент $\omega \in \Omega$), а деяка числова функція від нього (наприклад, зріст людини, тривалість роботи пристрою, дохід страхової компанії тощо). Такі функції називають випадковими величинами. Формально випадкова величина — це деяка вимірна функція, що діє на дійсну вісь.

Визначення (КЛ4.1.2). 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 σ -алгебри функція є вимірною, то просто кажуть, що вона вимірна. $\square$

Зауваження (КЛ4.1.4). Будь-яка функція $X : \Omega \rightarrow \mathbb{R}$, де $(\Omega, 2^\Omega)$ — дискретний імовірнісний простір, є вимірною, адже 2^Ω включає всі можливі підмножини Ω . $\square$

Визначення (КЛ4.1.5). Дійснозначну вимірну функцію $X : \Omega \rightarrow \mathbb{R}$ називають *випадковою величиною* (random variable).

За замовчуванням кладуть, що σ -алгеброю для простору $\mathbb{R}$ є Борелева σ -алгебра $\mathcal{B}$. $\square$

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 зафіксований, і випадкова величина набуває конкретного значення $X(\omega)$.

Отже 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 (КЛ4.1.6). Функція (випадкова величина) X задає на вимірному просторі $(\mathbb{R}, \mathcal{B})$ ймовірнісну міру, яка для кожної $A \in \mathcal{B}$ дорівнює

$$\mathbb{P}_X(A) \equiv \mathbb{P}_X(X \in A) = \mathbb{P}(\{\omega \in \Omega : X(\omega) \in A\}) = \mathbb{P}(X^{-1}(A)) . \quad (\text{КЛ4.1.1})$$

Цю міру називають *розподілом 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 (КЛ4.1.7). Із Визначення КЛ4.1.6 стає зрозумілим, навіщо ми вимагали, щоб випадкова величина була вимірною функцією: обчислення ймовірнісних тверджень відносно випадкових величин зводиться до обчислення ймовірностей підмножин Ω . Вимагаючи, щоб випадкова величина була вимірною, ми гарантуємо, що такі ймовірності можна буде обчислити. $\square$

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

Твердження (КЛ4.1.8). 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 КЛ2.1.5.

Визначення 5.1.1 (КЛ4.1.10). 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): $X \stackrel{d}{=} Y$. $\square$

У Розд. КЛ2.3 ми розглядали спосіб побудови ймовірнісної міри на дискретному просторі. Розподіли, які є образами дискретних ймовірнісних мір, також називають дискретними, як і відповідні випадкові величини.

Визначення (КЛ4.2.1). Випадкову величину X називають *дискретною*, як і відповідний їй розподіл, якщо її образ є не більш ніж зліченим. $\square$

Отже, хоча формально ми пишемо $X : \Omega \rightarrow \mathbb{R}$, потрібно пам'ятати, що насправді образом 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 її образ: $\text{supp}(X) = \{x_1, x_2, \dots\}$.

Було б корисно мати можливість оперувати деякими функціями, які давали б змогу обчислювати ймовірності різних подій, не використовуючи щоразу формальне Визначення КЛ4.1.5.

Визначення (КЛ4.2.3). Для опису розподілу $\mathbb{P}_X$ дискретної випадкової величини $X : \Omega \rightarrow \mathbb{R}$ із 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:

$$p_X(x) \begin{cases} \neq 0, & x \in \{x_1, x_2, \dots\} \\ = 0, & x \notin \{x_1, x_2, \dots\} \end{cases}, \quad \sum_i p_X(x_i) = 1. \quad (\text{КЛ4.2.1})$$

Тоді, за Теоремою КЛ2.3.1, функція ймовірності визначатиме міру $\mathbb{P}_X$ — розподіл випадкової величини. Зокрема,

$$\mathbb{P}_X(A) = \sum_{x_i \in A} p_X(x_i). \quad (\text{КЛ4.2.2})$$

$\square$

Чи не найважливішою числовою характеристикою випадкової величини є її сподівання, яке можна розглядати як її «середнє значення».

Визначення (КЛ4.3.1). *Сподіванням* (expectation) дискретної випадкової величини X є зважене середнє

$$\mathbb{E}[X] = \sum_{x \in \text{supp}(X)} x \cdot \mathbb{P}_X(X = x) = \sum_{x \in \text{supp}(X)} x p_X(x). \quad (\text{КЛ4.3.1})$$

Сподівання може бути нескінченно великим, а якщо одночасно

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

то кажуть, що сподівання *не існує* (does not exist)¹. $\square$

¹Інколи також кажуть, що сподівання не існує, якщо воно не є скінченним. Це нюанси термінології, на яких ми не будемо зупинятися.

Зауваження (КЛ4.3.2). Надзвичайно важливо усвідомлювати, що $\mathbb{E}[X]$, якщо воно існує, є сталим (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, сподівання не несе в собі жодної випадковості: це є фіксоване значення, яке заздалегідь відоме і може бути обчислено. $\square$

Можна показати, що сподівання має властивість лінійності.

Твердження (КЛ4.3.5). Для будь-яких випадкових величин² 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:

$$\mathbb{E}[aX + bY] = a\mathbb{E}[X] + b\mathbb{E}[Y] . \quad (\text{КЛ4.3.2})$$

Сподівання також можна обчислити для деякої функції від випадкової величини.

Теорема (Закон несвідомого статистика для дискретних величин (Law of unconscious statistician for discrete variables), КЛ4.3.6). Нехай X — деяка дискретна випадкова величина з носієм $\text{supp}(X) = \{x_1, x_2, \dots\}$ і функцією ймовірності p_X , і нехай $g : \mathbb{R} \rightarrow \mathbb{R}$ — деяка функція. Тоді

$$\mathbb{E}[g(X)] = \sum_{x \in \text{supp}(X)} g(x)p_X(x) . \quad (\text{КЛ4.3.3})$$

Визначення (КЛ4.3.8). *Дисперсією* (variance) дискретної випадкової величини X є сподівання квадратів відхилень її значень від її сподівання:

$$\text{Var}(X) = \mathbb{E}[(X - \mathbb{E}[X])^2] = \sum_{x \in \text{supp}(X)} (x - \mathbb{E}[X])^2 \cdot \mathbb{P}_X(X = x) . \quad (\text{КЛ4.3.4})$$

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

$$\text{sd}(X) = \sqrt{\text{Var}(X)} = \sqrt{\sum_{x \in \text{supp}(X)} (x - \mathbb{E}[X])^2 \cdot \mathbb{P}_X(X = x)} . \quad (\text{КЛ4.3.5})$$

$\square$

Часто дисперсію позначають через σ_X^2 , і тоді середньоквадратичне відхилення має позначення σ_X .

²Будь-яких, не тільки дискретних, але тут ми доведемо для дискретних.

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

На практиці корисна альтернативна формула дисперсії.

Твердження (КЛ4.3.10). Для будь-якої випадкової величини X ³

$$\text{Var}(X) = \mathbb{E}[X^2] - (\mathbb{E}[X])^2. \quad (\text{КЛ4.3.6})$$

Варто відзначити, що дисперсія *не* є лінійною: для довільних констант $a, b \in \mathbb{R}$ та довільної випадкової величини X

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

Розгляньмо декілька відомих дискретних розподілів.

Визначення (КЛ4.4.1). Випадкова величина X має *рівномірний розподіл* (uniform distribution), якщо її функція ймовірності має вид:

$$p_X(x) = \frac{1}{|\text{supp}(X)|}. \quad (\text{КЛ4.4.1})$$

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

Нехай деякий експеримент може завершитися успіхом із деякою ймовірністю $p \in [0; 1]$, або ж невдачею з ймовірністю $q \equiv 1 - p$. Такий експеримент називають *випробуванням Бернуллі*⁴ (Bernoulli trial).

Визначення (КЛ4.5.1). Випадкова величина X має *розподіл Бернуллі* (Bernoulli distribution) із параметром p , що позначають через $X \sim \text{Bern}(p)$, якщо її функція ймовірності має вид $p_X(1) = p$, $p_X(0) = q = 1 - p$, або, компактніше,

$$p_X(x) = p^x(1 - p)^{1-x} \cdot \mathbb{1}\{x \in \{0, 1\}\}. \quad (\text{КЛ4.5.1})$$

□

³Будь-якої, не тільки дискретної.

⁴Яків Бернуллі (Jacob Bernoulli, 1654–1705) — швейцарський математик.

Твердження (КЛ4.5.3). Якщо $X \sim \text{Bern}(p)$, то сподівання $\mathbb{E}[X] = p$, дисперсія $\text{Var}(X) = p(1 - p)$.

Зверніть увагу, що в цьому випадку $\mathbb{E}[X] = \mathbb{P}_X(X = 1)$.

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

Визначення (КЛ4.5.4). 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 $X \sim \text{Binom}(n, p)$. $\square$

Можна вивести функцію ймовірності такої випадкової величини.

Твердження (КЛ4.5.5). Якщо випадкова величина $X \sim \text{Binom}(n, p)$, то її функція ймовірності дорівнює

$$p_X(k) = \binom{n}{k} p^k (1 - p)^{n-k} \cdot \mathbb{1}\{k \in \{0, 1, \dots, n\}\} . \quad (\text{КЛ4.5.2})$$

Твердження (КЛ4.5.9). Нехай $X \sim \text{Binom}(n, p)$. Тоді випадкова величина $Y = n - X \sim \text{Binom}(n, 1 - p)$.

Відповідний результат очевидний, адже величина Y просто рахує кількість невдач, що є повністю дзеркальною задачею.

Твердження (КЛ4.5.10). Якщо $X \sim \text{Binom}(n, p)$, то сподівання $\mathbb{E}[X] = np$, а дисперсія $\text{Var}(X) = np(1 - p)$.

5.2 Загальні дискретні випадкові величини

Вправа 5.2.1. Колесо для гри в рулетку містить 18 чорних комірок, 18 червоних комірок та 2 зелені комірки. Якщо гравець поставить 10 грн на червоне, чому дорівнюватиме його сподіваний виграш? $\square$

Розв'язання. Спочатку потрібно з'ясувати, яка функція ймовірності у випадкової величини X = «виграш гравця в рулетку». Учасник може виграти 10 грн з імовірністю $18/38$ і програти 10 грн з імовірністю $20/38$, тому

$$p_X(x) = \begin{cases} \frac{20}{38}, & x = -10 \\ \frac{18}{38}, & x = 10 \\ 0, & x \notin \{10, -10\} \end{cases} .$$

Отже

$$\mathbb{E}[X] = 10 \cdot \frac{18}{38} - 10 \cdot \frac{20}{38} = -\frac{20}{38} \approx -0.526.$$

Інтерпретація цього результату така, що якщо гравець повторюватиме гру нескінченно (або принаймні «дуже велику») кількість разів, у середньому його виграш буде -0.526 (грн). $\square$

Вправа 5.2.2. У страхової компанії двоє клієнтів поважного віку, кожен із яких застрахований на випадок смерті на суму 1 000 000 грн. Нехай Y = «молодший із клієнтів помре протягом року», O = «старший із клієнтів помре протягом року». Нехай $Y \perp O$, а $\mathbb{P}(Y) = 0.05$, $\mathbb{P}(O) = 0.10$.

Розгляньмо випадкову величину X = «загальна сума, яку виплатить страхова компанія протягом року родинам цих двох клієнтів». Чому дорівнює її функція ймовірності p_X ? Чому дорівнюють сподівання, дисперсія та середньоквадратичне відхилення? $\square$

Розв'язання. Носієм випадкової величини X є множина з трьох елементів $\text{supp}(X) = \{0, 1\,000\,000, 2\,000\,000\}$, оскільки можливі три варіанти розвитку подій: ніхто не помре, помре тільки один клієнт, помруть обидва клієнти. Функція ймовірності p_X повинна кожному елементу $x \in \text{supp}(X)$ зіставити імовірність $\mathbb{P}_X(X = x)$.

Випадкова величина набуде значення 0, якщо ніхто не помре, тобто

$$\begin{aligned} p_X(0) &= \mathbb{P}_X(X = 0) = \mathbb{P}(Y^c \cap O^c) = \mathbb{P}(Y^c) \cdot \mathbb{P}(O^c) \\ &= (1 - 0.05) \cdot (1 - 0.10) = 0.855, \end{aligned}$$

де ми використали незалежність відповідних подій для обчислення їх ймовірності через добуток.

Випадкова величина набуде значення 1 000 000, якщо помре або старший клієнт, або молодший, але не обидва, тобто

$$\begin{aligned} p_X(1\,000\,000) &= \mathbb{P}_X(X = 1\,000\,000) = \mathbb{P}((Y \cap O^c) \cup (Y^c \cap O)) \\ &= \mathbb{P}(Y) \cdot \mathbb{P}(O^c) + \mathbb{P}(Y^c) \cdot \mathbb{P}(O) \\ &= 0.05(1 - 0.10) + (1 - 0.05) \cdot 0.10 = 0.140, \end{aligned}$$

де ми використали незалежність та несумісність відповідних подій.

Нарешті, випадкова величина набуде значення 2 000 000, якщо помруть обидва клієнти, тобто

$$\begin{aligned} p_X(2\,000\,000) &= \mathbb{P}_X(X = 2\,000\,000) = \mathbb{P}(Y \cap O) \\ &= \mathbb{P}(Y) \cdot \mathbb{P}(O) = 0.05 \cdot 0.10 = 0.005, \end{aligned}$$

де ми використали незалежність відповідних подій.

Оскільки $0.855 + 0.140 + 0.005 = 1$, це справді є функція ймовірності. Для підрахунку сподівання застосуємо (КЛ4.3.1):

$$\begin{aligned}\mathbb{E}[X] &= 0.855 \cdot 0 + 0.140 \cdot 1\,000\,000 + 0.005 \cdot 2\,000\,000 \\ &= 140\,000 + 10\,000 = 150\,000 ,\end{aligned}$$

тобто страхова компанія може сподіватися витратити 150 000 грн.

Для підрахунку дисперсії застосуємо (КЛ4.3.6), для чого спочатку підрахуємо сподівання X^2 :

$$\mathbb{E}[X^2] = 0.855 \cdot 0^2 + 0.140 \cdot (1\,000\,000)^2 + 0.005 \cdot (2\,000\,000)^2 = 1.6 \cdot 10^{11} ,$$

звідки

$$\text{Var}(X) = \mathbb{E}[X^2] - (\mathbb{E}[X])^2 = 1.6 \cdot 10^{11} - 2.25 \cdot 10^{10} = 1.375 \cdot 10^{11} ,$$

а отже $\text{sd}(X) = \sqrt{\text{Var}(X)} = 370\,809.9$, тобто сподіваний розкид значень навколо сподівання буде 370 809.9 грн.

На відміну від Вправи 5.2.1, тут сподівання не має інтерпретації середнього значення, адже ми не можемо повторити експеримент не те що нескінченну кількість разів, а навіть двічі. Сподівання має беєсівську інтерпретацію і відіграє роль прогнозного значення видатків. $\square$

Вправа 5.2.3. Деякий продукт, який має сезонний попит, приносить прибуток $b > 0$ за кожную продану одиницю і втрати $l > 0$ за кожную одиницю, яка залишилася непроданою по завершенні сезону. Кількість одиниць товару X , на які може бути попит протягом сезону, є випадковою величиною з функцією ймовірності p_X та носієм $\text{supp}(X) = \{0, 1, 2, \dots\}$.

Якщо потрібно закупити продукт наперед, скільки одиниць варто закупити для максимізації сподіваного прибутку? $\square$

Розв'язання. Позначмо прибуток у випадку закупівлі s одиниць продукту через P_s . Нам потрібно знайти значення s , яке максимізуватиме $\mathbb{E}[P(s)]$.

Прибуток дорівнює

$$P_s = \begin{cases} bX - (s - X)l , & X \leq s \\ sb , & X > s \end{cases} ,$$

оскільки якщо весь товар буде розкуплено, то прибуток буде sb (по b за кожную продану одиницю), а якщо буде залишок $(s - X)$ одиниць, то кожна тягтиме за собою по l втрат.

Фактично, P_s є функцією від X . Тому сподіваний прибуток можна обчислити за допомогою Закону несвідомого статистика (КЛ4.3.3):

$$\begin{aligned}
 \mathbb{E}[P_s(X)] &= \sum_{i=0}^s (bi - (s-i)l)p_X(i) + \sum_{i=s+1}^{\infty} sbp_X(i) \\
 &= (b+l) \sum_{i=0}^s ip_X(i) - sl \sum_{i=0}^s p_X(i) + sb \left(1 - \sum_{i=0}^s p_X(i)\right) \\
 &= (b+l) \sum_{i=0}^s ip_X(i) - (b+l)s \sum_{i=0}^s p_X(i) + sb \\
 &= (b+l) \sum_{i=0}^s (i-s)p_X(i) + sb.
 \end{aligned}$$

Суто теоретично можна взяти першу похідну і прирівняти її до 0. Проте в цьому випадку маємо суму зі змінною верхньою межею, тому взяття похідної ускладнене. Розгляньмо натомість вираз

$$\begin{aligned}
 \mathbb{E}[P(s+1)] - \mathbb{E}[P(s)] &= (b+l) \sum_{i=0}^{s+1} (i - (s+1))p_X(i) + (s+1)b \\
 &\quad - (b+l) \sum_{i=0}^s (i-s)p_X(i) + sb \\
 &= b - (b+l) \sum_{i=0}^s p_X(i).
 \end{aligned}$$

Можна помітити, що збільшення кількості одиниць товару з s до $s+1$ вестиме до збільшення сподіваного прибутку, якщо

$$b - (b+l) \sum_{i=0}^s p_X(i) > 0 \quad \Leftrightarrow \quad \sum_{i=0}^s p_X(i) < \frac{b}{b+l}.$$

Нехай s^* — найбільше s таке, що вищевказана нерівність виконується. Тоді для всіх $s = 1, \dots, s^* + 1$ сподіваний прибуток зростає, а після $s^* + 1$ починає спадати. Отже $s^* + 1$ є оптимальною кількістю товарів, яку потрібно закупити для максимізації сподіваного прибутку. $\square$

Приклад 5.2.4. Розгляньмо відомий *парадокс дружби* (friendship paradox), який коротко полягає в тому, що Ваші друзі в середньому мають більше друзів, ніж Ви.

Нехай у популяції є n осіб, пронумеровані від 1 до n . Тоді мережу друзів можна подати як неорієнтований граф, де кожна вершина відповідає особі,

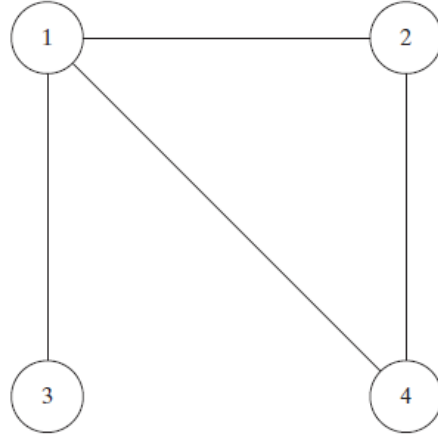


Рис. 5.2.1: Можлива мережа друзів для Прикладу 5.2.4

а ребро з'єднує двох друзів. Приклад такої мережі наведено на Рис. 5.2.1: особа 1 дружить з усіма, а, скажімо, особа 3 дружить тільки з особою 1.

Позначмо через $f(i)$ число друзів особи i , $i = 1, \dots, n$. Позначмо $f = \sum_{i=1}^n f(i)$. Наприклад, для Рис. 5.2.1 маємо

$$f(1) = 3, \quad f(2) = 2, \quad f(3) = 1, \quad f(4) = 2, \quad f = 8.$$

Розгляньмо випадкову величину X = «номер випадково вибраної особи». Очевидно, що ця величина має рівномірний розподіл $X \sim U(\{1, \dots, n\})$, тобто $\mathbb{P}_X(X = i) = \frac{1}{n}$. Можемо обчислити сподіване число друзів, використовуючи Закон несвідомого статистика (КЛ4.3.3):

$$\mathbb{E}[f(X)] = \sum_{i=1}^n f(i) \mathbb{P}_X(X = i) = \sum_{i=1}^n \frac{f(i)}{n} = \frac{f}{n}.$$

Також можемо обчислити сподівання $f^2(X)$:

$$\mathbb{E}[f^2(X)] = \sum_{i=1}^n f^2(i) \mathbb{P}_X(X = i) = \sum_{i=1}^n \frac{f^2(i)}{n}.$$

Тоді

$$\frac{\mathbb{E}[f^2(X)]}{\mathbb{E}[f(X)]} = \frac{\sum_{i=1}^n f^2(i)}{f}.$$

Уявімо тепер, що кожна з n осіб напише на окремих аркушах паперу імена своїх друзів. Оскільки кожна особа має $f(i)$ друзів, усього буде $f = \sum_{i=1}^n f(i)$ аркушів. Нехай Y = «номер особи, чие ім'я написано на випадково вибраному аркуші». Тоді

$$\mathbb{E}[f(Y)] = \sum_{i=1}^n f(i) \mathbb{P}_Y(Y = i) = \sum_{i=1}^n \frac{f^2(i)}{f},$$

оскільки ймовірність витягнути аркуш з іменем i -ої особи, очевидно, дорівнює $\mathbb{P}_Y(Y = i) = \frac{f(i)}{f}$.

Отже, ми маємо

$$\mathbb{E}[f(Y)] = \frac{\mathbb{E}[f^2(X)]}{\mathbb{E}[f(X)]} \geq \mathbb{E}[f(X)] ,$$

де нерівність випливає з того факту, що $\mathbb{E}[X^2] \geq (\mathbb{E}[X])^2$ для будь-якої випадкової величини X (адже $\text{Var}(X) \geq 0$).

Тому маємо, що середнє число друзів у деякої особи X не перевищує середнього числа друзів у деякого у випадковий спосіб вибраного її друга. $\square$

Вправа 5.2.5. Нехай у Вас є 10 000 грн і Ви плануєте придбати деякий товар, ціна якого наразі 20 грн за штуку. За тиждень його ціна може бути або 10, або 40 грн за штуку (з однаковою ймовірністю).

Якщо Вашою ціллю є максимізувати сподіваний достаток (готівка та ринкова вартість товару станом на момент, коли Ви будете визначати свій достаток) по завершенні одного тижня, яку стратегію варто застосувати?

Якщо Вашою ціллю є максимізувати сподівану кількість товарів після зміни ціни (уже придбаних та які можна буде придбати на залишок коштів після зміни ціни товару), яку стратегію варто застосувати? $\square$

Розв'язання. Нехай на початку тижня Ви купляете x штук товару. У Вас лишиться $10\,000 - 20x$ готівки, і по завершенні одного тижня достаток буде

$$V = 10\,000 - 20x + Yx ,$$

де Y — випадкова величина, що відповідає ціні за одиницю товару по завершенні одного тижня. Отже,

$$\begin{aligned} \mathbb{E}[V] &= 10\,000 - 20x + \mathbb{E}[Y]x \\ &= 10\,000 - 20x + x \cdot \left(10 \cdot \frac{1}{2} + 40 \cdot \frac{1}{2}\right) \\ &= 10\,000 + 5x . \end{aligned}$$

Вочевидь, максимального значення цей вираз досягне, якщо x буде максимальним, а саме $x = 500$, оскільки на 10 000 грн можна придбати максимум 500 одиниць товару за ціною 20 грн за штуку. Тобто, іншими словами, потрібно витратити всю наявну готівку. Незважаючи на те, що з ймовірністю 0.5 ціна може впасти удвічі, саме така стратегія максимізуватиме сподіваний

достаток. Це зрозуміло, адже сподівана ціна на другому тижні дорівнює $0.5 \cdot 10 + 0.5 \cdot 40 = 25$ грн, що більше від 20 грн на першому тижні.

У другому випадку ми воліємо максимізувати сподівану кількість товару A після зміни ціни товару. Якщо ми придбаємо x одиниць товару на першому тижні, у нас лишиться $10\,000 - 20x$ готівки, на які можна буде придбати додаткову кількість товару за новою ціною. Отже

$$A = x + \frac{10\,000 - 20x}{Y},$$

і нам потрібно відшукати сподівання $\mathbb{E}[A]$. Це можна зробити з застосуванням (КЛ4.3.3):

$$\begin{aligned}\mathbb{E}[A] &= x + \mathbb{E}\left[\frac{10\,000 - 20x}{Y}\right] \\ &= x + \frac{10\,000 - 20x}{10} \cdot \frac{1}{2} + \frac{10\,000 - 20x}{40} \cdot \frac{1}{2} \\ &= 625 - \frac{x}{4}.\end{aligned}$$

Максимального значення цей вираз досягне для $x = 0$. Тобто, іншими словами, потрібно відкласти купівлю товарів до зміни ціни, а вже потім придбавати її на всю готівку. Хоча з імовірністю 0.5 ціна може підскочити удвічі, саме така стратегія максимізуватиме сподівану кількість товару на руках. Сподівана ціна після одного тижня буде 25 грн, проте в цьому випадку ми не можемо застосувати її безпосередньо, як для обчислення сподіваного достатку. Ми маємо справу з нелінійною функцією $\frac{10\,000 - 20x}{Y}$, а тому не можна казати, що сподівана кількість товару буде $\frac{10\,000 - 20x}{25}$. $\square$

5.3 Схема Бернуллі

Вправа 5.3.1. Відомо, що компанія виробляє гайки, які можуть, незалежно одна від одної, мати дефект з імовірністю 0.1. Компанія продає гайки в пакетах по 10 штук і обіцяє повернути кошти, якщо в пакеті буде більше від 1 дефективної гайки.

За яку частку пакетів компанія вимушена буде сплатити компенсацію? $\square$

Розв'язання. Нехай X = «число дефективних гайок у пакеті». За описом ситуації, $X \sim \text{Binom}(10, 0.1)$, оскільки кожну гайку можна розглядати як окреме випробування Бернуллі, «успіхом» у якому є наявність дефекту.

Нам потрібно знайти ймовірність $\mathbb{P}_X(X > 1)$, що можна зробити за допомогою (КЛ4.5.2):

$$\begin{aligned}\mathbb{P}_X(X > 1) &= 1 - \mathbb{P}_X(X = 0) - \mathbb{P}_X(X = 1) \\ &= 1 - \binom{10}{0}(0.01)^0(1 - 0.01)^{10} - \binom{10}{1}(0.01)^1(1 - 0.01)^9 \approx 0.004 ,\end{aligned}$$

тобто компанії потрібно буде компенсувати вартість 0.4% проданих пакетів. $\square$

Вправа 5.3.2. Нехай для ухвалення рішення комітетом із 12 осіб потрібно зібрати щонайменше дві третини голосів, тобто 8 і більше. Нехай кожен член комітету діє незалежно від інших і голосує на користь ухвалення рішення з ймовірністю θ .

Чому дорівнює ймовірність, що комітет ухвалить рішення? $\square$

Розв'язання. Голосування членів комітету можна розглядати як схему Бернуллі, де кожне випробування може завершитися «успіхом» (голосуванням на користь рішення) з ймовірністю θ . Тоді X = «загальне число поданих голосів» має біномний розподіл $X \sim \text{Binom}(12, \theta)$. Отже ймовірність ухвалення рішення дорівнює

$$\mathbb{P}_X(X \geq 8) = \sum_{i=8}^{12} \binom{12}{i} \theta^i (1 - \theta)^{12-i} .$$

$\square$

Розгляньмо трішки ускладнену версію цієї задачі.

Вправа 5.3.3. Нехай для ухвалення осудного рішення судом присяжних із 12 осіб потрібно зібрати щонайменше 9 голосів. Нехай кожний присяжний може виправдати злочинця з ймовірністю 0.2 та засудити невинуватого з ймовірністю 0.1. Кожний присяжний ухвалює рішення незалежно від інших. Усього серед обвинувачених 65% справді винуватих.

Чому дорівнює ймовірність, що суд присяжних ухвалить правильне рішення? Яку частку підсудних буде засуджено? $\square$

Розв'язання. Позначмо події E = «суд присяжних ухвалив правильне рішення», G = «підсудний винуватий», C = «присяжний проголосував «за» засудження підсудного». Тоді нас цікавить ймовірність $\mathbb{P}(E)$, яку можна обчислити за формулою повної ймовірності (КЛ3.3.2):

$$\mathbb{P}(E) = \mathbb{P}(E | G) \mathbb{P}(G) + \mathbb{P}(E | G^c) \mathbb{P}(G^c) .$$

Апріорну ймовірність винуватості дано в умові задачі, тому $\mathbb{P}(G) = 0.65$, $\mathbb{P}(G^c) = 1 - \mathbb{P}(G) = 0.35$.

Для ухвалення осудного рішення судом присяжних потрібно набрати щонайменше 9 голосів. Позначмо $X = \text{«кількість голосів, поданих «за» засудження підсудного»}$. Тоді $X \sim \text{Binom}(12, p)$, де p — ймовірність, що присяжний проголосує «за» засудження.

На цьому етапі можна припуститися помилки, і сказати, що

$$p = \mathbb{P}(C | G) \mathbb{P}(G) + \mathbb{P}(C | G^c) \mathbb{P}(G^c) = (1 - 0.2) \cdot 0.65 + 0.1 \cdot 0.35 = 0.555 ,$$

проте це буде неправильно, адже якщо перед судом присяжних стоїть насправді винувата людина (тобто ми обчислюємо ймовірність $\mathbb{P}(E | G)$), то голос присяжного можна спрогнозувати значно точніше. Так, якщо підсудний справді винуватий, то «правильним» буде рішення засудити його, тому $p = \mathbb{P}(C | G) = 1 - 0.2 = 0.8$, а тому

$$\mathbb{P}(E | G) = \sum_{i=9}^{12} \binom{12}{i} (0.8)^i (0.2)^{12-i} \approx 0.795 .$$

Аналогічно можна підрахувати $\mathbb{P}(E | G^c)$. Щоправда для спрощення підрахунків ми розглянемо ймовірність $\mathbb{P}(E^c | G^c)$, тобто ймовірність засудження невинуватої людини:

$$\mathbb{P}(E | G^c) = 1 - \mathbb{P}(E^c | G^c) = 1 - \sum_{i=9}^{12} \binom{12}{i} (0.1)^i (0.9)^{12-i} \approx 1.000 .$$

Остаточного маємо

$$\mathbb{P}(E) = \mathbb{P}(E | G) \mathbb{P}(G) + \mathbb{P}(E | G^c) \mathbb{P}(G^c) = 0.795 \cdot 0.65 + 1.000 \cdot 0.35 \approx 0.867 .$$

Щоб підрахувати частку засуджених підсудних, потрібно обчислити

$$\mathbb{P}(E | G) \mathbb{P}(G) + \mathbb{P}(E^c | G^c) \mathbb{P}(G^c) = 0.795 \cdot 0.65 + 0.000 \cdot 0.35 \approx 0.517 .$$

□

Вправа 5.3.4. Викладачі перевіряють екзамен. Одне з питань складається з 10 тверджень, щодо кожного з яких потрібно вказати, істинне воно чи хибне. Перший викладач помічає, що в одного студента правильні всього 2 відповіді і коментує: «Цей студент був настільки недалекий, що навіть не додумався просто підкидати монетку. Так би він відповів правильно принаймні на 5 питань». Його колега відповідає: «Я би не поспішав із висновками. Якби всі 340 студентів підкидали монетку, то обов'язково серед

них був би принаймні один студент з 2 правильними відповідями або й менше». А третій додає: «Я б навіть очікував знайти принаймні один екзамен взагалі без правильних відповідей».

Хто з викладачів має більше рації? $\square$

Розв'язання. У цій задачі маємо справу з двома біномними величинами одночасно. Перша з них — X = «кількість правильно вгаданих відповідей» — має біномний розподіл $X \sim \text{Binom}(10, 0.5)$. Друга — Y = «кількість студентів із двома і менше відповідями» — має біномний розподіл $Y \sim \text{Binom}(340, \mathbb{P}_X(X \leq 2))$.

Ми маємо

$$\mathbb{P}_X(X \leq 2) = \sum_{i=0}^2 \binom{10}{i} (0.5)^i (0.5)^{10-i} = \frac{1 + 10 + 45}{2^{10}} = \frac{56}{1024}.$$

Чи був другий професор правий? Маємо

$$\begin{aligned} \mathbb{P}_Y(Y \geq 1) &= 1 - \mathbb{P}_Y(Y = 0) = 1 - \binom{340}{0} \left(\frac{56}{1024}\right)^0 \left(\frac{968}{1024}\right)^{340} \\ &\approx 1 - 4.962 \cdot 10^{-9} \approx 1.000. \end{aligned}$$

Іншими словами, ймовірність того, що принаймні один студент матиме 2 правильні відповіді або менше, надзвичайно велика, тому той факт, що знайшовся один такий студент, зовсім не є несподіваним.

З іншого боку, ймовірність того, що студент не відповів правильно на жодне питання, дорівнює $\mathbb{P}_X(X = 0) = \frac{1}{1024}$. Розгляньмо випадкову величину Z = «кількість студентів з жодною правильною відповіддю», яка має біномний розподіл $Z \sim \text{Binom}(340, \mathbb{P}_X(X = 0))$. Ймовірність того, що знайдеться хоча б один такий студент, дорівнює $\mathbb{P}_Z(Z \geq 1)$, або ж

$$\mathbb{P}_Z(Z \geq 1) = 1 - \mathbb{P}_Z(Z = 0) = 1 - \binom{340}{0} \left(\frac{1}{1024}\right)^0 \left(\frac{1023}{1024}\right)^{340} \approx 0.717.$$

Ця ймовірність не є аж настільки великою, щоб бути впевненим, що принаймні один екзамен буде без жодної правильної відповіді. $\square$

5.4 Додаткові вправи

Вправа 5.4.1. Число вмикань деякого приладу до першого виходу з ладу X має функцію ймовірності

$$p_X(x) = c \left(\frac{9}{10}\right)^{x-1} \cdot \mathbb{1}\{x \in \mathbb{N}\}.$$

Чому дорівнює стала c ? Чому дорівнює ймовірність, що прилад не вийде з ладу щонайменше до 10 вмикання включно? $\square$

Розв'язання. Сума значень функції ймовірності повинна дорівнювати 1:

$$\sum_{x=1}^{\infty} c \left(\frac{9}{10} \right)^{x-1} = 1 .$$

За сумою нескінченно спадної геометричної прогресії маємо

$$\frac{c}{1 - \frac{9}{10}} = 1 \quad \Rightarrow \quad c = \frac{1}{10} .$$

Обчислити ймовірність того, що прилад не вийде з ладу щонайменше до 10 вмикання включно, можна так:

$$\mathbb{P}_X (X > 10) = \frac{1}{10} \cdot \sum_{x=11}^{\infty} \left(\frac{9}{10} \right)^{x-1} = \frac{1}{10} \cdot \frac{\left(\frac{9}{10} \right)^{10}}{1 - \frac{9}{10}} \approx 0.349 . \quad \square$$

Вправа 5.4.2 (Закон Бенфорда). Випадкова величина X має *розподіл Бенфорда*⁵, якщо її функція ймовірності має вигляд

$$p_X(i) = \mathbb{P}_X (X = i) = \log_{10} \left(\frac{i+1}{i} \right) , \quad i = 1, 2, \dots, 9 .$$

Відповідний розподіл на практиці непогано описує розподіл перших цифр чисел у багатьох реальних наборах даних.

Перевірте, що це є функція ймовірності. Обчисліть $\mathbb{P}_X (X \leq j)$ та $\mathbb{E} [X]$. $\square$

Розв'язання. Очевидно, що функція невід'ємна, оскільки $\frac{i+1}{i} > 1$. Тому потрібно перевірити, що сума її значень дорівнює одиниці:

$$\sum_{i=1}^9 \mathbb{P}_X (X = i) = \sum_{i=1}^9 \log_{10} \left(\frac{i+1}{i} \right) = \log_{10} 10 - \log_{10} 1 = 1 .$$

Відповідно,

$$\mathbb{P}_X (X \leq j) = \sum_{i=1}^j \log_{10} \left(\frac{i+1}{i} \right) = \log_{10}(j+1) - \log_{10} 1 = \log_{10}(j+1) .$$

⁵Френк Бенфорд-молодший (Frank Albert Benford Jr., 1883–1948) — американський фізик.

Сподівання можна обчислити чисельно

$$\mathbb{E}[X] = \sum_{i=1}^9 i \log_{10} \left(\frac{i+1}{i} \right) \approx 3.44 . \quad \square$$

Вправа 5.4.3. Деяка комунікаційна система складається з n компонент, кожна з яких може незалежно від інших вийти з ладу з імовірністю $0 < p < 1$. Система не припинить працювати, якщо з ладу вийдуть менше половини компонент.

Якими повинні бути значення p , щоб система з 5 компонент мала більшу ймовірність не вийти з ладу, аніж система з 3 компонент? $\square$

Розв'язання. Очевидно, випадкова величина X = «число компонент, що вийшли з ладу» має біномний розподіл: $X \sim \text{Binom}(n, p)$. Імовірність, що система з 5 компонент буде працювати, дорівнює

$$\begin{aligned} \mathbb{P}_X(X \leq 2) &= \mathbb{P}_X(X = 0) + \mathbb{P}_X(X = 1) + \mathbb{P}_X(X = 2) \\ &= \binom{5}{0} p^0 (1-p)^5 + \binom{5}{1} p^1 (1-p)^4 + \binom{5}{2} p^2 (1-p)^3 \\ &= (1-p)^5 + 5p(1-p)^4 + 10p^2(1-p)^3 . \end{aligned}$$

Для системи з 3 компонент аналогічна ймовірність дорівнюватиме

$$\begin{aligned} \mathbb{P}_X(X \leq 1) &= \mathbb{P}_X(X = 0) + \mathbb{P}_X(X = 1) \\ &= \binom{3}{0} p^0 (1-p)^3 + \binom{3}{1} p^1 (1-p)^2 \\ &= (1-p)^3 + 3p(1-p)^2 . \end{aligned}$$

Система з 5 компонент буде ліпшою, якщо

$$(1-p)^5 + 5p(1-p)^4 + 10p^2(1-p)^3 > (1-p)^3 + 3p(1-p)^2 ,$$

тобто, після всіх скорочень, якщо

$$p < \frac{1}{2} .$$

Тобто система з 5 компонентами надійніша, якщо ймовірність виходу компоненти з ладу менша від 50%. Іншими словами, збільшення кількості компонент не означає, що система буде надійнішою. $\square$

Вправа 5.4.4. Нехай $X \sim \text{Binom}(n, p)$. Покажіть, що

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

□

Розв'язання. Можна безпосередньо застосувати закон несвідомого статистика (КЛ4.3.3):

$$\begin{aligned} \mathbb{E}[(1-p)^X] &= \sum_{i=0}^n (1-p)^i \cdot \binom{n}{i} p^i (1-p)^{n-i} \\ &= \sum_{i=0}^n \binom{n}{i} p^i (1-p)^n \\ &= (1-p)^n \sum_{i=0}^n \binom{n}{i} p^i \\ &= (1-p)^n \cdot (1+p)^n \\ &= ((1-p)(1+p))^n \\ &= (1-p^2)^n, \end{aligned}$$

де ми використали біномну теорему.

□

6 Деякі поширені дискретні розподіли

6.1 Основні теоретичні відомості

Детальний огляд понять, потрібний для розв’язання задач на цьому занятті, наведено в Курсі лекцій, Розд. КЛ5.

Визначення (КЛ5.1.1). Випадкова величина X має *розподіл Пуассона* (Poisson distribution)¹ із параметром $\lambda > 0$, що позначають через $X \sim \text{Pois}(\lambda)$, якщо її функція ймовірності має вид

$$p_X(k) = \mathbb{P}_X(X = k) = \frac{e^{-\lambda} \lambda^k}{k!} \cdot \mathbb{1}\{k \in \mathbb{N} \cup \{0\}\} . \quad (\text{КЛ5.1.1})$$

□

Параметр λ можна інтерпретувати як частоту появи деяких подій («успіхів»), які стаються з малою ймовірністю.

Розподіл Пуассона доречно застосовувати для моделювання явищ, які стаються нечасто. Проте це не означає, що такі явища мають розподіл Пуассона. Розподіл Пуассона є всього лише апроксимацією, яку зручно використовувати на практиці.

Теорема (КЛ5.1.4). Нехай $X \sim \text{Binom}(n, p_n)$, $n \rightarrow \infty$, $p_n \rightarrow 0$, але $np_n \rightarrow \lambda$, тобто залишається фіксованим. Тоді функція ймовірності p_X прямує до функції ймовірності величини $Y \sim \text{Pois}(\lambda)$.

Теорема КЛ5.1.4 застосовна, коли число випробувань n велике, ймовірність 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;

¹Уперше запропонував французький математик Симеон Дені Пуассон (Siméon Denis Poisson, 1781–1840) у 1837 р. у роботі *Recherches sur la probabilité des jugements en matière criminelle et en matière civile* («Дослідження ймовірностей судових рішень у кримінальних та цивільних справах»).

- 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 α -частинок, випромінюваних протягом фіксованого періоду.

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.

Дуже часто розподіл Пуассона використовують для опису явищ, які стаються в деякі випадкові моменти часу. Тоді загальне число подій, які стаються в часовому інтервалі довжиною h , має розподіл Пуассона з параметром λh , 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 із нагрітого катода за певний період часу;
- число смертей серед клієнтів страхової компанії за певний період часу.

Розподіл Пуассона також можна використовувати як апроксимацію розподілу суми випадкових величин, навіть якщо ймовірності подій різні.

Твердження (Пуассонівська парадигма (Poisson paradigm), КЛ5.1.8). 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 $\lambda = \sum_{j=1}^n p_j$.

Підрахуймо сподівання та дисперсію розподілу Пуассона.

Твердження (КЛ5.1.9). Якщо $X \sim \text{Pois}(\lambda)$, то $\mathbb{E}[X] = \text{Var}(X) = \lambda$.

Розгляньмо декілька розподілів, пов'язаних зі схемою Бернуллі.

Визначення (КЛ5.2.1). Нехай маємо послідовні *незалежні* випробування Бернуллі, у кожному з яких імовірність успіху дорівнює p . Тоді, випадкова величина $X = \text{«кількість випробувань до появи першого успіху»}$ має *геометричний розподіл* (geometric distribution) із параметром p , що позначають через $X \sim \text{Geom}(p)$. □

Ця випадкова величина має таку функцію ймовірності.

Твердження (КЛ5.2.2). Якщо випадкова величина $X \sim \text{Geom}(p)$, то її функція ймовірності має вид:

$$p_X(k) = \mathbb{P}_X(X = k) = (1 - p)^k p \cdot \mathbb{1}\{k \in \mathbb{Z}^+\} . \quad (\text{КЛ5.2.1})$$

Зауваження (КЛ5.2.3). У літературі не існує єдиного бачення, який розподіл визначати як геометричний. Деякі автори подають геометричний розподіл як розподіл $Y = \text{«кількість випробувань до появи першого успіху, включно з самим успіхом»}$. Ми вважатимемо, що така величина має окремий розподіл: $Y \sim \text{FS}(p)$ ². Очевидно, що якщо $X \sim \text{Geom}(p)$, то $X + 1 \sim \text{FS}(p)$. $\square$

Для величини $X \sim \text{Geom}(p)$ справедлива властивість

$$\mathbb{P}_X(X \geq k) = (1 - p)^k .$$

Понад те, умовна ймовірність того, що $X \geq k + j$ за умови, що всі перші j випробувань завершилися невдачами, дорівнює

$$\begin{aligned} \mathbb{P}_X(X \geq k + j \mid X \geq j) &= \frac{\mathbb{P}_X((X \geq k + j) \cap (X \geq j))}{\mathbb{P}_X(X \geq j)} \\ &= \frac{\mathbb{P}_X(X \geq k + j)}{\mathbb{P}_X(X \geq j)} \\ &= \frac{(1 - p)^{k+j}}{(1 - p)^j} = (1 - p)^k . \end{aligned}$$

Отже геометричний розподіл є розподілом *без пам'яті* (memoryless), тобто тривалість очікування першого успіху, за умови, що він до цього ще не встиг настати, не залежить від попередньої історії.

Твердження (КЛ5.2.5). Якщо $X \sim \text{Geom}(p)$, то сподівання $\mathbb{E}[X] = \frac{1-p}{p}$, а дисперсія $\text{Var}(X) = \frac{1-p}{p^2}$.

Зауваження (КЛ5.2.6). Якщо $Y \sim \text{FS}(p)$, а, відповідно, $X = Y - 1 \sim \text{Geom}(p)$, то

$$\mathbb{E}[Y] = \mathbb{E}[X + 1] = \mathbb{E}[X] + 1 = \frac{1 - p}{p} + 1 = \frac{1}{p} .$$

$\square$

²За першими літерами словосполучення «перший успіх» англійською — «first success».

У повній аналогії з біномним розподілом як узагальненням розподілу Бернуллі можна увести від'ємний біномний розподіл як узагальнення розподілу геометричного.

Визначення (КЛ5.3.1). 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 $X \sim \text{NegBinom}(r, p)$. $\square$

Ця випадкова величина має таку функцію ймовірності.

Твердження (КЛ5.3.2). Якщо випадкова величина $X \sim \text{NegBinom}(r, p)$, то її функція ймовірності має вид:

$$p_X(j) = \mathbb{P}_X(X = j) = \binom{j+r-1}{r-1} p^r (1-p)^j, \quad j = 0, 1, \dots \quad (\text{КЛ5.3.1})$$

Зауваження (КЛ5.3.3). 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 = «загальна кількість випробувань, потрібна для появи r успіхів». Тоді функція ймовірності має вид

$$\mathbb{P}_Y(Y = k) = \binom{k-1}{r-1} p^r (1-p)^{k-r}, \quad k = r, r+1, \dots$$

$\square$

Твердження (КЛ5.3.5). Якщо $X \sim \text{NegBinom}(r, p)$, то сподівання $\mathbb{E}[X] = r \frac{1-p}{p}$, а дисперсія $\text{Var}(X) = r \frac{1-p}{p^2}$.

Наостанок розгляньмо ще один корисний на практиці розподіл.

Визначення (КЛ5.4.1). Нехай маємо вазу, у якій міститься w білих та b чорних кульок. Випадкова величина X = «кількість білих кульок серед n 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. $\square$

Функція ймовірності цієї випадкової величини така.

Твердження (КЛ5.4.2). Якщо випадкова величина $X \sim \text{HGeom}(w, b, n)$, то функція її розподілу має вид³:

$$\mathbb{P}_X(X = k) = \frac{\binom{w}{k} \binom{b}{n-k}}{\binom{w+b}{n}}, \quad k = 0, 1, \dots, n. \quad (\text{КЛ5.4.1})$$

У загальному випадку гіпергеометричний розподіл застосовний до ситуацій, коли об'єкти мають два типи міток, і мітки принаймні одного типу присвоюють у випадковий спосіб. Тоді випадкова величина X відповідає числу об'єктів із мітками обох типів. У прикладі з кульками такими мітками є колір кульки та той факт, відібрано її чи ні.

Твердження (КЛ5.4.5). Випадкова величина $X \sim \text{HGeom}(w, b, n)$ рівна за розподілом випадковій величині $Y \sim \text{HGeom}(n, w+b-n, w)$.

Результат Твердження КЛ5.4.5 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 (КЛ5.4.6). Якщо $X \sim \text{HGeom}(w, b, n)$, то $\mathbb{E}[X] = \frac{nw}{w+b}$, а $\text{Var}(X) = \frac{nwb(w+b-n)}{(w+b)^2(w+b-1)}$.

Як розподіл Пуассона є граничним для біномного, так і біномний є граничним — для гіпергеометричного.

Твердження (КЛ5.4.7). Якщо $X \sim \text{HGeom}(w, b, n)$, і $N = w+b \rightarrow \infty$, а $\frac{w}{w+b} \rightarrow p$ залишається сталою, то функція ймовірності $p_X(x)$ прямує до функції ймовірності величини $Y \sim \text{Binom}(n, p)$.

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 .

³Варто 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 $n-b$.

На практиці часто виникають ситуації, коли потрібно обчислювати умовні ймовірності за участю випадкових величин.

Визначення (КЛ5.5.1). 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 ймовірності можна виконати, використовуючи Визначення КЛ4.1.6:

$$\begin{aligned} \mathbb{P}_{X|Z}(X = x | Z = z) &= \mathbb{P}(\{\omega : X(\omega) = x\} | \{\omega : Z(\omega) = z\}) \\ &= \frac{\mathbb{P}(\{\omega : (X(\omega) = x) \cap (Z(\omega) = z)\})}{\mathbb{P}(\{\omega : Z(\omega) = z\})}. \end{aligned} \quad (\text{КЛ5.5.1})$$

Зокрема, справедливе правило Беєса з Теорема КЛ3.3.1:

$$\mathbb{P}_{X|Z}(X = x | Z = z) = \frac{\mathbb{P}_{Z|X}(Z = z | X = x) \mathbb{P}_X(X = x)}{\mathbb{P}_Z(Z = z)}. \quad (\text{КЛ5.5.2})$$

6.2 Розподіл Пуассона

Вправа 6.2.1. Нехай експеримент полягає в підрахунку числа α -частинок, які протягом однієї секунди випромінює 1 грам радіоактивної речовини. Із попереднього досвіду відомо, що в середньому за секунду випромінюються 3.2 таких частинок. Чому (приблизно) дорівнює ймовірність, що протягом найближчої секунди буде випромінено не більше від 2 частинок? $\square$

Розв'язання. Для моделювання випадкової величини X = «загальне число випромінених частинок протягом однієї секунди» доречно застосувати розподіл Пуассона, адже грам радіоактивної речовини можна розглядати як сукупність значної кількості (n) атомів, кожен із яких має незначну ймовірність $(3.2/n)$ розпастися й випромінити α -частинку. Оскільки за одну секунду в середньому випромінюються 3.2 частинки, маємо $X \sim \text{Pois}(3.2)$.

Відповідно,

$$\begin{aligned} \mathbb{P}_X(X \leq 2) &= \mathbb{P}_X(X = 0) + \mathbb{P}_X(X = 1) + \mathbb{P}_X(X = 2) \\ &= \frac{e^{-3.2}(3.2)^0}{0!} + \frac{e^{-3.2}(3.2)^1}{1!} + \frac{e^{-3.2}(3.2)^2}{2!} \approx 0.3799. \end{aligned}$$

$\square$

Вправа 6.2.2. Нехай дощ випадає з інтенсивністю 20 крапель на квадратний сантиметр за хвилину. Який розподіл можна використати для опису випадкової величини $X_t =$ «число крапель, що падають на ділянку площею 5 квадратних сантиметрів за t хвилин»? Чому дорівнює ймовірність, що на ділянку площею 5 квадратних сантиметрів протягом 3 секунд не впаде жодної краплі дощу? $\square$

Розв'язання. Доброю апроксимацією розподілу величини X_t можна вважати розподіл Пуассона з параметром λt , де $\lambda = 20 \cdot 5 = 100$. Це впливає з того, що 20 — це інтенсивність опадів на 1 квадратний сантиметр, а нас цікавить ділянка площею в 5 квадратних сантиметрів.

Оскільки 3 секунди становлять 0.05 від однієї хвилини, ймовірність жодної краплі дощу становить

$$\mathbb{P}_{X_{0.05}}(X_{0.05} = 0) = \frac{e^{-\lambda h}(\lambda h)^0}{0!} = \frac{e^{-100 \cdot 0.05}(100 \cdot 0.05)^0}{0!} = e^{-5} \approx 0.007.$$

 $\square$

Вправа 6.2.3. Група з n друзів вирішила зіграти в гру «таємний Санта», у рамках якої кожен пише своє ім'я на аркуші паперу, після чого всі аркуші кладуть у мішок. Далі кожен учасник у випадковий спосіб витягає аркуш із мішка. За умовами потрібно зробити подарунок особі, чиє ім'я ти витягнеш.

Нехай $X =$ «число осіб, що витягнули своє ім'я». Чому дорівнює $\mathbb{E}[X]$?

Нехай $Y =$ «число пар осіб таких, що вони витягнули ім'я одне одного». Чому дорівнює $\mathbb{E}[Y]$?

Нарешті, який розподіл *приблизно* має величина X , якщо n велике? Чому дорівнює $\mathbb{P}_X(X = 0)$, коли $n \rightarrow \infty$? $\square$

Розв'язання. Нехай $S_i = j$, якщо i -а особа витягнула ім'я j -ої особи, $i, j = 1, \dots, n$. Тоді нехай $X_i = 1$, якщо $S_i = i$, і 0 у протилежному випадку, $i = 1, \dots, n$. Це, звісно, величина Бернуллі з параметром $p = 1/n$. Проте не можна казати, що їх сума $X = \sum_{i=1}^n X_i$ є величиною з біномним розподілом, адже величини X_i не є незалежними. Справді, якщо відомо, що всі особи витягнули імена, окрім n -го, то n -а особа витягне своє ім'я з одиничною ймовірністю.

Утім, властивість лінійності сподівання залишається справедливою:

$$\mathbb{E}[X] = \sum_{i=1}^n \mathbb{E}[X_i] = n \cdot \frac{1}{n} = 1.$$

Нехай $Y_{ij} = 1$, якщо $S_i = j$ і $S_j = i$ одночасно, і 0 у протилежному випадку. Очевидно, що це може статися з імовірністю

$$\mathbb{P}((S_i = j) \cap (S_j = i)) = \mathbb{P}(S_i = j \mid S_j = i) \cdot \mathbb{P}(S_j = i) = \frac{1}{n-1} \cdot \frac{1}{n}.$$

Варто помітити, що всього існує $\binom{n}{2}$ пар, які ми розглядаємо. Отже

$$\mathbb{E}[Y] = \mathbb{E}\left[\sum_{i=1}^n \sum_{j=i+1}^n Y_{ij}\right] = \binom{n}{2} \frac{1}{n(n-1)} = \frac{1}{2}.$$

Виконуються всі умови Твердження КЛ5.1.8⁴, тому X має приблизний розподіл $\text{Pois}(n \cdot \frac{1}{n}) = \text{Pois}(1)$. Отже $\mathbb{P}_X(X = 0) \rightarrow e^{-1} \approx 0.368$. $\square$

Приклад 6.2.4. Нехай число B хромосом, які пошкоджуються протягом деякого періоду часу, має розподіл Пуассона з параметром λ . Пошкоджена хромосома може відновитися з імовірністю p . Відновлення однієї хромосоми не залежить від відновлення інших.

Нехай R = «загальне число відновлених хромосом». Тоді R має умовний біномний розподіл за умови, що $B = n$: $R \mid B = n \sim \text{Binom}(n, p)$. Щоб дістати *безумовний* розподіл R , використаємо формулу повної ймовірності:

$$\begin{aligned} \mathbb{P}_R(R = k) &= \sum_{n=1}^{\infty} \mathbb{P}_{R|B}(R = k \mid B = n) \cdot \mathbb{P}_B(B = n) \\ &= \sum_{n=k}^{\infty} \binom{n}{k} p^k (1-p)^{n-k} \cdot \frac{e^{-\lambda} \lambda^n}{n!} \\ &= \frac{e^{-\lambda} (\lambda p)^k}{k!} \sum_{n=k}^{\infty} \frac{(\lambda(1-p))^{n-k}}{(n-k)!} \\ &= \frac{e^{-\lambda} (\lambda p)^k}{k!} \sum_{m=0}^{\infty} \frac{(\lambda(1-p))^m}{m!} \\ &= \frac{e^{-\lambda} (\lambda p)^k}{k!} \cdot e^{\lambda(1-p)} \\ &= \frac{e^{-\lambda p} (\lambda p)^k}{k!}. \end{aligned}$$

Отже $R \sim \text{Pois}(\lambda p)$, звідки сподіване число відновлених хромосом дорівнює $\mathbb{E}[R] = \lambda p$. $\square$

⁴Ми це доводити не будемо, але інтуїтивно зрозуміло, що величини X_i залишаються залежними, проте зі збільшенням n «сила» цієї залежності падає.

6.3 Інші дискретні розподіли

Вправа 6.3.1. Подружня пара планує народжувати дітей, допоки в них не народяться і хлопчик, і дівчинка. Нехай дитина може народитися будь-якої з двох статей з однаковою ймовірністю. Яке сподіване число дітей у такої подружньої пари? $\square$

Розв'язання. Очевидно, перша дитина буде деякої з двох статей. Тому фактично потрібно порахувати сподівання величини $X =$ «загальне число дітей, починаючи з другого, до появи дитини статі, відмінної від статі першої дитини». Ця величина матиме розподіл FS(0.5), тому сподіване число дітей у родині дорівнює

$$1 + \mathbb{E}[X] = 1 + 2 = 3 .$$

 $\square$

Вправа 6.3.2. Чому дорівнює сподівання та дисперсія числа разів, які потрібно підкинути гральну кісточку, допоки цифра 1 не випаде 4 рази? $\square$

Розв'язання. Нехай $X =$ «загальне число підкидань для чотирьох випадів одиниці». Тоді $X - 4$ має від'ємний біномний розподіл із параметрами $r = 4$, $p = 1/6$. За Твердженням КЛ5.3.5 маємо:

$$\begin{aligned}\mathbb{E}[X - 4] &= r \frac{1 - p}{p} = 4 \cdot 5 = 20 , \\ \text{Var}(X - 4) &= r \frac{1 - p}{p^2} = 4 \cdot 30 = 120 .\end{aligned}$$

Оскільки $\mathbb{E}[X - 4] = \mathbb{E}[X] - 4 = 20$, маємо, що $\mathbb{E}[X] = 24$.

За властивостями дисперсії маємо, що $\text{Var}(X - 4) = \text{Var}(X) = 120$. Це впливає з того факту, що

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

 $\square$

Вправа 6.3.3. Покупець придбаває електронні компоненти партіями по 10 штук. Як правило, він перевіряє 3 довільні компоненти з 10 і купляє партію, тільки якщо всі 3 не мають дефектів.

Якщо 30% усіх партій мають по 4 дефективні компоненти, а 70% — тільки по одній, яку частку партій купить клієнт? $\square$

Розв'язання. Нехай X = «число дефективних компонент партії серед 3 витягнутих на перевірку». Нас цікавить $\mathbb{P}_X(X = 0)$, проте ми не знаємо, який розподіл має X . Проблема полягає в тому, що для кожної *конкретної* партії число дефективних компонент серед 3 витягнутих матиме гіпергеометричний розподіл, але з різними параметрами. Тобто ми можемо розписати за формулою повної ймовірності

$$\begin{aligned}\mathbb{P}_X(X = 0) &= \mathbb{P}_X(X = 0 \mid \text{«у партії 4 дефекти»}) \cdot 0.3 \\ &+ \mathbb{P}_X(X = 0 \mid \text{«у партії 1 дефект»}) \cdot 0.7 \\ &= 0.3\mathbb{P}_{X_4}(X_4 = 0) + 0.7\mathbb{P}_{X_1}(X_1 = 0),\end{aligned}$$

де X_4 та X_1 мають гіпергеометричний розподіл, але з різними параметрами. Зокрема, $X_4 \sim \text{HGeom}(4, 6, 3)$ та $X_1 \sim \text{HGeom}(1, 9, 3)$.

Остаточню

$$\mathbb{P}_X(X = 0) = \frac{\binom{4}{0}\binom{6}{3}}{\binom{10}{3}} \cdot 0.3 + \frac{\binom{1}{0}\binom{9}{3}}{\binom{10}{3}} \cdot 0.7 = 0.54.$$

Отже покупець придбає лише 54% всіх партій. $\square$

Вправа 6.3.4. Іван і Марія щойно познайомилися, і їм цікаво, скільки в них є спільних друзів серед працівників їхньої компанії. Нехай усього в компанії працюють 1000 осіб, а в Івана і Марії є 40 і 60 друзів відповідно. Для спрощення покладемо, що друзі кожного з них становлять випадкову вибірку з-посеред 1000 працівників компанії. Також будемо вважати, що інформація про друзів Івана не проливає світло на друзів Марії і навпаки.

Чому дорівнює сподіване число їхніх спільних друзів? $\square$

Розв'язання. Можна помітити, що число спільних друзів X має гіпергеометричний розподіл, оскільки маємо справу з мітками двох типів (друг Івана і друг Марії), які присвоюють незалежно одну від одної. Отже $X \sim \text{HGeom}(40, 960, 60)$ або ж, що повністю еквівалентно з Твердження КЛ5.4.5, $X \sim \text{HGeom}(60, 940, 40)$, і тому з Твердження КЛ5.4.6

$$\mathbb{E}[X] = \frac{nw}{w+b} = \frac{40 \cdot 60}{40 + 960} = \frac{60 \cdot 40}{60 + 940} = 2.4.$$

$\square$

Приклад 6.3.5. У галузі цифрових комунікацій повідомлення, як правило, розбивають на окремі пакети, які послідовно передають мережею.

Один із протоколів передавання інформації полягає в тому, що відправник жде підтвердження від приймача, що пакет доставлено успішно. Якщо приймач одержує пакет без пошкоджень, він посилає назад код підтвердження АСК (acknowledgement), а якщо пакет надійшов з невиправними помилками, приймач посилає назад код негативного підтвердження НАК (negative acknowledgement).

Якщо відправник одержує код АСК, він надсилає наступний пакет, в іншому випадку (одержання коду НАК або дуже тривале очікування без відповіді) він повторює поточний пакет.

Нехай кожний пакет може бути пошкоджено або втрачено, незалежно один від одного, з імовірністю q . Також нехай коди АСК та НАК надходять без помилок. Тоді загальне число X повторень одного й того ж пакету є випадковою величиною з геометричним розподілом із параметром $1 - q$. Тому сподіване число таких надлишкових пакетів дорівнює, згідно з Твердженням КЛ5.2.5, $\mathbb{E}[X] = \frac{q}{1-q}$. $\square$

Вправа 6.3.6. Нехай у щоденній лотереї потрібно вгадати три цифри від 0 до 9. Правильна комбінація абсолютно випадкова і генерується незалежно кожного дня.

Поява трьох однакових цифр завжди привертає увагу. Який розподіл має випадкова величина X = «число днів до появи першої комбінації з однакових цифр»?

Нехай гравець у лотерею прождав 120 днів, і жодного разу не випала комбінація з однакових цифр. Гравець міркує так, що наступного дня «вже точно» повинні випасти однакові цифри. Чи має він рацію? $\square$

Розв'язання. Нескладно бачити, що $X \sim \text{Geom}(p)$, оскільки будь-який день без трьох однакових цифр можна розглядати як невдачу. Імовірність успіху p можна обчислити з класичного визначення ймовірності: $p = 0.01$, оскільки існує всього 10 комбінацій з однакових цифр серед 1000 можливих варіантів (від 000 до 999). Зокрема, потрібно в середньому зачекати $\mathbb{E}[X] = \frac{1-p}{p} = 99$ днів до появи комбінації з однакових цифр.

Геометричний розподіл не має пам'яті, тобто $\mathbb{P}_X(X \geq k + j \mid X \geq j) = \mathbb{P}_X(X \geq k) = (1 - p)^k$. Тому немає жодних підстав сподіватися, що комбінація випаде «ось уже». Навпаки, через відсутність пам'яті сподівана кількість днів, які потрібно ждати, усе одно залишається 99.

Відповідна похибка мислення має назву *помилки гравця* (gambler's fallacy). Вона виникає тоді, коли людина не усвідомлює, що якщо події з деякої послідовності стаються незалежно одна від одної, то немає жодного сенсу прогнозувати майбутні події на основі попередніх. $\square$

Вправа 6.3.7. Нехай двоє гравців намагаються влучити баскетбольним м'ячем у кошик. Імовірність, що гравець А влучить, дорівнює p , незалежно від номеру кидку, і він кидатиме доти, доки не влучить r разів. Імовірність, що гравець В влучить, дорівнює, незалежно від номеру кидку, mp , $mp < 1$, $m \in \{2, 3, \dots\}$. Він кидатиме доти, доки не влучить mr разів.

Нехай X_A, X_B — кількості кидків, які повинен зробити кожний гравець для досягнення свого цільового показника. У кого з гравців вища сподівана кількість кидків $\mathbb{E}[X_i]$, $i = A, B$? Дисперсія $\text{Var}(X_i)$, $i = A, B$? $\square$

Розв'язання. Цілком зрозуміло, що $Y_A = X_A - r \sim \text{NegBinom}(r, p)$. Згідно з лінійністю сподівання й результатом Твердження КЛ5.3.5, маємо

$$\mathbb{E}[X_A] = \mathbb{E}[Y_A + r] = \mathbb{E}[Y_A] + r = r \frac{1-p}{p} + r = \frac{r}{p}.$$

Для другого гравця маємо $Y_B = X_B - mr \sim \text{NegBinom}(mr, mp)$, і тому

$$\mathbb{E}[X_B] = \mathbb{E}[Y_B + mr] = \mathbb{E}[Y_B] + mr = mr \frac{1-mp}{mp} + mr = \frac{r}{p},$$

тобто $\mathbb{E}[X_A] = \mathbb{E}[X_B]$.

Що стосується дисперсії, то

$$\begin{aligned} \text{Var}(X_A) &= \text{Var}(Y_A + r) = \text{Var}(Y_A) = \frac{r(1-p)}{p^2} = \frac{r}{p^2}(1-p), \\ \text{Var}(X_B) &= \text{Var}(Y_B + mr) = \text{Var}(Y_B) = \frac{mr(1-mp)}{(mp)^2} = \frac{r}{p^2} \left(\frac{1}{m} - p \right). \end{aligned}$$

Отже, $\text{Var}(X_B) < \text{Var}(X_A)$. $\square$

Приклад 6.3.8 (Задача про колекціонера купонів (Coupon collector's problem)). Нехай існують купони n видів, і задача полягає в зібранні повної колекції купонів. Купони збираються у послідовному порядку, і кожний наступний купон може бути будь-якого з n видів з однаковою ймовірністю.

Нехай N = «число купонів, які потрібно перебрати, щоб зібрати по одному купону кожного з n видів». Нас цікавить $\mathbb{E}[N]$.

Позначмо через N_i число купонів, які розділяють перші появи купонів раніше не бачених видів. Наприклад, $N_1 = 1$ за побудовою. Якщо другий купон був такий самий, як перший, а третій купон був іншого виду, то $N_1 = 1$, $N_2 = 2$. Відповідну ілюстрацію зображено на Рис. 6.3.1.

Очевидно, $N = \sum_{i=1}^n N_i$. Також можемо з'ясувати, які розподіли мають N_i , $i = 2, \dots, n$. Оскільки в кожному випадку ми рахуємо загальне число невдач перед новим успіхом (появою нового купону), *включно з самим*

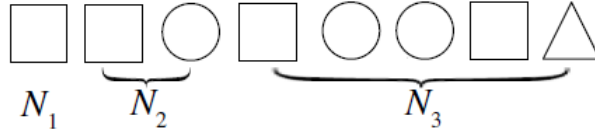


Рис. 6.3.1: Ілюстрація до Прикладу 6.3.8

успіхом, маємо $N_2 \sim \text{FS} \left(\frac{n-1}{n} \right)$, оскільки після появи купону першого виду існує ймовірність $\frac{n-1}{n}$ дістати купон нового виду.

З аналогічних міркувань можна бачити, що $N_i \sim \text{FS} \left(\frac{n-(i-1)}{n} \right)$. Згідно з лінійністю сподівання та з урахуванням Зауваження КЛ5.2.6, маємо

$$\mathbb{E}[N] = \sum_{i=1}^n \mathbb{E}[N_i] = \sum_{i=1}^n \frac{n}{n-(i-1)} = n \sum_{i=1}^n \frac{1}{i}.$$

Між іншим, у схожий спосіб можна обчислити дисперсію. Ми цього поки що не доводили, але для *незалежних* випадкових величин дисперсія суми дорівнює сумі дисперсій:

$$\text{Var}(N) = \sum_{i=1}^n \text{Var}(N_i) = \sum_{i=1}^n \frac{1 - \frac{n-(i-1)}{n}}{\left(\frac{n-(i-1)}{n} \right)^2} = n \sum_{i=1}^n \frac{i-1}{(n-(i-1))^2}.$$

Наприклад, якщо маємо купони 10 видів, $\mathbb{E}[N] \approx 29.290$, $\text{Var}(N) \approx 125.687$, $\text{sd}(N) \approx 11.211$, а якщо кількість видів зростає до 20, то $\mathbb{E}[N] = 71.955$, $\text{Var}(N) = 566.511$, $\text{sd}(N) = 23.801$. Можна показати, що для великих n сподівання дорівнює $\mathbb{E}[N] \approx n(\ln n + 0.577)$, що впливає з властивостей гармонійних чисел. $\square$

6.4 Додаткові вправи

Вправа 6.4.1. У покері роял-флеш (royal flush) випадає, коли комбінація карт на руках містить туза, короля, кралою, валета та десятку однієї масти. Чому дорівнює ймовірність p дістати роял-флеш? Скільки разів потрібно роздавати карти, щоб роял-флеш випав принаймні один раз з імовірністю, не меншою від $1 - e^{-1} \approx 2/3$? $\square$

Розв'язання. Усього існує $\binom{52}{5}$ варіантів витягти 5 карт із 52 наявних, а роял-флеш може бути одного з 4 видів (за кількістю мастей), тому

$$p = \frac{4}{\binom{52}{5}} = \frac{1}{649\,740} \approx 1.539 \cdot 10^{-6}.$$

Цю ймовірність можна розглядати як ймовірність успіху в послідовності з n випробувань Бернуллі. Оскільки вона доволі мала, якщо n буде доволі велике, загальне число X роял-флешів буде мати приблизний розподіл Пуассона з параметром $\lambda = np$. Отже нас цікавить n , за якого

$$\mathbb{P}_X(X \geq 1) = 1 - \mathbb{P}_X(X = 0) = 1 - e^{-np} \geq 1 - e^{-1} \Leftrightarrow n \geq \frac{1}{p} = 649\,740.$$

□

Вправа 6.4.2. Нехай маємо k пронумерованих куль і n пронумерованих коробок. Кулі у випадковий спосіб розміщують по коробках. Звісно, будь-який варіант із n^k розміщень має однакову ймовірність. Задачі такого класу мають назву *задач про розміщення* (occupancy problems), вони часто виринають у теорії алгоритмів.

Чому дорівнює сподіване число порожніх коробок? Чому дорівнює ймовірність, що принаймні одна коробка порожня? Якщо $n = 1000$, $k = 5806$, як можна обчислити відповідну ймовірність наближено? □

Розв'язання. Нехай $X_i = \langle i\text{-та коробка порожня} \rangle$. Це випадкова величина Бернуллі з параметром

$$p = \left(\frac{n-1}{n} \right)^k,$$

оскільки це ймовірність того, що всі коробки буде заповнено, окрім i -ої.

Отже загальне число порожніх коробок є $X = \sum_{i=1}^n X_i$, а тому

$$\mathbb{E}[X] = \sum_{i=1}^n \left(\frac{n-1}{n} \right)^k = n \cdot \left(\frac{n-1}{n} \right)^k.$$

При цьому варто зазначити, що X не має біномного розподілу, адже величини X_i не є незалежними: якщо ми знаємо, що всі кульки розміщено по $n-1$ коробках, то коробка n буде порожньою з одиничною ймовірністю.

Ймовірність того, що принаймні одна коробка порожня, дорівнює 1, якщо $k < n$. Нехай $A_i = (X_i = 1)$. Якщо $k \geq n$, згідно з формулою включення-виключення,

$$\begin{aligned} \mathbb{P}_X(X \geq 1) &= \mathbb{P}(A_1 \cup \dots \cup A_n) = \sum_{j=1}^n (-1)^{j+1} \binom{n}{j} \mathbb{P}(A_1 \cap \dots \cap A_j) \\ &= \sum_{j=1}^{n-1} (-1)^{j+1} \binom{n}{j} \left(\frac{n-j}{n} \right)^k. \end{aligned}$$

Якщо $n = 1000$, $k = 5806$, маємо $p \approx 0.003$, а тому можна застосувати Твердження КЛ5.1.8 і використати апроксимацію за допомогою розподілу Пуассона, де $\lambda = np \approx 3$:

$$\mathbb{P}_X(X \geq 1) \approx 1 - \frac{e^{-3}3^0}{0!} \approx 0.950.$$

□

Вправа 6.4.3. Нехай гра полягає в тому, що підкидають дві гральні кісточки, і гравець перемагає, якщо випало дві шістки. Якщо плата за кожний підкид дорівнює 1 грн, яким повинен бути чесний виграш у цій грі (у тому розумінні, що в середньому гравець нічого не виграє і нічого не програє)?

□

Розв'язання. Нехай X = «загальне число грошей, які гравець витратить до першого виграшу». Очевидно, ця величина має розподіл $X \sim \text{FS}(1/36)$. Тому її сподівання дорівнює $\mathbb{E}[X] = 36$. Отже виграш у 36 грн буде чесним виграшем.

□

Вправа 6.4.4. Нехай біля святкового столу стоять n стільців, по числу гостей свята. Гості по чергово танцюють і спілкуються за столом. Після кожного танцю кожний окремий стілець може виявитися, незалежно від інших, зайнятим або звільненим з однаковою ймовірністю.

Нехай X_t = «число зайнятих стільців після t -ого танцю». Чому дорівнює $\mathbb{E}[X_t]$? Чому дорівнює сподівання $\mathbb{E}[M]$ випадкової величини M = «найменший номер танцю, після якого всі стільці зайнято»?

□

Розв'язання. Нескладно бачити, що $X_t \sim \text{Binom}(n, 0.5)$, отже $\mathbb{E}[X_t] = 0.5n$.

Величина M має розподіл $\text{FS}(1/2^n)$, оскільки $1/2^n$ — ймовірність того, що всі стільці будуть зайняті. Отже $\mathbb{E}[M] = 2^n$.

□

Вправа 6.4.5. Нехай $x \sim \text{Geom}(p)$. Чому дорівнює ймовірність того, що X набула парного значення?

□

Розв'язання. Носієм величини з геометричним розподілом є множина невід'ємних цілих $\text{supp}(X) = \{0, 1, 2, \dots\}$. Отже нас цікавить ймовірність $\mathbb{P}_X(X \in \{0, 2, 4, \dots\})$.

Оскільки $\mathbb{P}_X(X = x) = p(1 - p)^x$, шукана ймовірність дорівнює

$$\mathbb{P}_X(X \in \{0, 2, 4, \dots\}) = \sum_{i=0}^{\infty} p(1 - p)^{2i} = p \sum_{i=0}^{\infty} ((1 - p)^2)^i = \frac{p}{1 - (1 - p)^2},$$

де ми використали суму нескінченно спадної геометричної прогресії. $\square$

7 Неперервні випадкові величини

7.1 Основні теоретичні відомості

Детальний огляд понять, потрібний для розв'язання задач на цьому занятті, наведено в Курсі лекцій, Розд. КЛ6.

Визначення (КЛ6.1.1). *Функцією розподілу* (distribution function, інколи cumulative distribution function, CDF) випадкової величини X є функція

$$F_X(x) \equiv \mathbb{P}_X(X \leq x) = \mathbb{P}_X((-\infty; x]) , \quad x \in \mathbb{R} . \quad (\text{КЛ6.1.1})$$

□

Розгляньмо декілька властивостей функції розподілу, які впливають безпосередньо з її визначення.

Твердження (КЛ6.1.2). 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 справа¹: $\lim_{x \downarrow c} F_X(x) = F_X(c)$;
- (iii) є нормованою:

$$\begin{aligned} \lim_{x \rightarrow -\infty} F_X(x) &= 0 , \\ \lim_{x \rightarrow \infty} F_X(x) &= 1 . \end{aligned}$$

Зауваження (КЛ6.1.3). Функція розподілу F_X є неперервною справа, але вона не є неперервною зліва.

Повторюючи кроки з доведення властивості (ii), маємо таке. За Теоремою КЛ2.2.8 будь-яка міра є неперервною, у тому числі неперервною

¹Функція *неперервна справа* (continuous from the right) у точці c , якщо $\lim_{\substack{x \rightarrow c \\ x > c}} f(x) \equiv \lim_{x \downarrow c} f(x)$ існує.

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

$$\lim_{n \rightarrow \infty} F_X\left(c - \frac{1}{n}\right) = \lim_{x \uparrow c} F_X(x) \equiv \lim_{\substack{x \rightarrow c \\ x < c}} F_X(x) \equiv F_X(c-).$$

Проте в цьому випадку

$$\lim_{n \rightarrow \infty} I_n = (-\infty; c),$$

оскільки $c \notin (-\infty; c - 1/n]$ для кожного n .

Отже,

$$F_X(c-) = \mathbb{P}_X\left(\lim_{n \rightarrow \infty} I_n\right) = \mathbb{P}_X((-\infty; c)) = \mathbb{P}_X(X < c).$$

Тому $F_X(c-) \neq F_X(c)$ у загальному випадку. У кожній такій точці c функція F_X має стрибок (розрив I роду). $\square$

Наведемо деякі правила обчислення ймовірностей.

Твердження (КЛ6.1.5). Нехай $F_X : \mathbb{R} \rightarrow \mathbb{R}$ — функція розподілу деякої випадкової величини X . Тоді:

- (i) $\mathbb{P}_X(X \geq x) = 1 - F_X(x-);$
- (ii) $\mathbb{P}_X((a; b]) = F_X(b) - F_X(a), a, b \in \mathbb{R};$
- (iii) $\mathbb{P}_X(X = x) = F_X(x) - F_X(x-).$

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 (КЛ6.1.6). Нехай $F_X : \mathbb{R} \rightarrow \mathbb{R}$ — деяка функція, яка задовольняє умови (i)–(ii) з Твердження КЛ6.1.2. Тоді в Борелевому просторі $(\mathbb{R}, \mathcal{B})$ існує єдина міра ν така, що $\nu((a; b]) = F_X(b) - F_X(a), a, b \in \mathbb{R}$.

Якщо ця функція ще задовольняє умову (iii) з Твердження КЛ6.1.2, то міра ν є ймовірнісною, тобто існує випадкова величина $X : (\Omega, \mathcal{A}, \mathbb{P}) \rightarrow (\mathbb{R}, \mathcal{B})$ на деякому ймовірнісному просторі така, що $F_X(x) = \mathbb{P}_X(X \leq x)$.

У будь-якої функції розподілу $F(x)$ є не більш ніж зліченна кількість стрибків (Твердження КЛ6.1.8). Тому існує проста процедура побудови функції розподілу будь-якої дискретної випадкової величини.

Приклад (КЛ6.1.9). 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 КЛ6.1.5 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): $F_X(x) = F_X(x-)$, $x \neq x_i$, $i = 1, 2, \dots$.

Отже 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 :

$$F_X(x) = \sum_{\substack{y \in \text{supp}(X) \\ y \leq x}} p_X(y) . \quad (\text{КЛ6.1.2})$$

□

На практиці винятково корисними є такі випадкові величини, розподіли яких є абсолютно неперервними відносно міри Лебега.

Визначення (КЛ6.3.1). Нехай на вимірному просторі $(\Omega, \mathcal{A})$ визначено дві міри, μ і ν . Тоді міра ν є *абсолютно неперервною відносно μ* (absolutely continuous with respect to μ), 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 ν , абсолютно неперервну відносно μ , можна подати як *інтеграл* від деякої функції за мірою μ .

Теорема (Теорема Радона-Никодима (Radon–Nikodym theorem), КЛ6.3.2). Якщо на вимірному просторі $(\Omega, \mathcal{A})$ визначено дві σ -скінченні міри μ і ν , $\nu \ll \mu$, тоді існує невід’ємна вимірна функція $f : \Omega \rightarrow [0; \infty)$ така, що

$$\nu(A) = \int_A f d\mu , \quad \forall A \in \mathcal{A} . \quad (\text{КЛ6.3.1})$$

Функцію f із Теорема КЛ6.3.2 називають *похідною Радона-Никодима* (Radon–Nikodym derivative).

Безпосереднє застосування цієї теореми до випадкових величин таке.

Визначення (КЛ6.3.3). Випадкова величина X є *абсолютно неперервною* (absolutely continuous), якщо її розподіл абсолютно неперервний відносно

міри Лебега, $\mathbb{P}_X \ll \lambda$:

$$\mathbb{P}_X(A) = \int_A f_X d\lambda, \quad \forall A \in \mathcal{B}. \quad \square$$

З урахуванням (КЛ6.1.1), маємо:

$$F_X(x) = \mathbb{P}_X((-\infty; x]) = \int_{(-\infty; x]} f_X d\lambda. \quad (\text{КЛ6.3.2})$$

Визначення (КЛ6.3.4). Функція f_X із (КЛ6.3.2) має назву *щільності розподілу* (probability density function, PDF). $\square$

Як відомо з курсу «Математичний аналіз», якщо функція інтегровна за Ріманом² (Riemann integrable), то вона інтегровна за Лебегом. Усі корисні на практиці щільності 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 інтегралом Рімана. Іншими словами,

$$f_X(x) = F'_X(x), \quad (\text{КЛ6.3.3})$$

для всіх точок x , у яких функція F_X диференційовна. 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 КЛ6.1.5.

Твердження (КЛ6.3.6). Нехай $F_X : \mathbb{R} \rightarrow \mathbb{R}$ — функція розподілу деякої абсолютно неперервної випадкової величини X . Тоді:

- (i) $\mathbb{P}_X(X \geq x) = 1 - F_X(x)$;
- (ii) $\mathbb{P}_X((a; b]) = \mathbb{P}_X([a; b)) = \mathbb{P}_X((a; b)) = \mathbb{P}_X([a; b]) = F_X(b) - F_X(a) = \int_a^b f_X(x) dx, \quad a, b \in \mathbb{R}$;
- (iii) $\mathbb{P}_X(X = x) = 0$.

²Георг Фридрих Бернгард Ріман (Georg Friedrich Bernhard Riemann, 1826–1866) — німецький математик.

Усі ці результати очевидні через те, що $F(x) = F(x-)$ для всіх $x \in \mathbb{R}$.

Для абсолютно неперервних випадкових величин можна уточнити поняття носія.

Визначення (КЛ6.3.7). *Носієм* (support) абсолютно неперервної величини X зі щільністю f_X називають множину $\text{supp}(X) = \{x \in \mathbb{R} : f_X(x) > 0\}$.

□

Будь-яка щільність має такі ж властивості, як і функція ймовірності: вона невід'ємна, а сума (інтеграл) усіх значень дорівнює 1.

Твердження (КЛ6.3.9). Будь-яка функція $f : \mathbb{R} \rightarrow [0; \infty)$ є щільністю для деякого розподілу тоді й тільки тоді, коли

$$\int_{-\infty}^{\infty} f(t) dt = F(\infty) = 1 .$$

Будь-яка функція ймовірності є щільністю, оскільки розподіл будь-якої дискретної випадкової величини абсолютно неперервний відносно лічної міри. Конкретніше, 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 лічну міру $\#$ визначмо як

$$\#(B) = |X_0 \cap B| , \quad B \in \mathcal{B} .$$

Можна показати, що розподіл $\mathbb{P}_X$ є абсолютно неперервний відносно $\#$ (Зазначення КЛ6.3.12), а отже, за Теоремою КЛ6.3.2,

$$\mathbb{P}_X(X \in B) = \int_B f d\# .$$

Функція f у цьому виразі є функцією ймовірності p_X .

Отже, теорія дискретних величин є складовою частиною загальнішої теорії абсолютно неперервних величин. Понад те, загальна теорія також дає змогу працювати з випадковими величинами, які не є ні дискретними, ні абсолютно неперервними.

На практиці найчастіше використовують абсолютно неперервні величини, тому для спрощення їх часто називають просто неперервними.

Можемо дати визначення сподівання будь-якої випадкової величини.

Визначення (КЛ6.4.1). *Сподіванням* (expectation) будь-якої випадкової величини X є її інтеграл:

$$\mathbb{E}[X] = \int X d\mathbb{P} . \quad (\text{КЛ6.4.1})$$

□

Визначення дисперсії з (КЛ4.3.4) залишається справедливим.

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.

Теорема (Закон несвідомого статистика (Law of unconscious statistician), КЛ6.4.2). 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 μ 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

$$\int g(X(\omega)) d\mathbb{P} = \int g(x) d\mathbb{P}_X = \int g(x) \cdot f_X(x) d\mu, \quad (\text{КЛ6.4.2})$$

у тому розумінні, що якщо існує один інтеграл, то існують усі інші, і вони дорівнюють одне одному.

Зауваження (КЛ6.4.3). Теорема КЛ6.4.2 дає змогу звести обчислення інтегралу в абстрактному просторі до обчислення рядів і звичних інтегралів Рімана.

Для дискретних величин розподіл завжди абсолютно неперервний відносно лічної міри, а тому

$$\mathbb{E}[g(X)] = \int g(x) \cdot p_X(x) d\# = \sum_{x \in \text{supp}(X)} g(x) \cdot p_X(x), \quad (\text{КЛ6.4.3})$$

де p_X — функція ймовірності. Це відповідає формулі (КЛ4.3.3).

Для (абсолютно) неперервних величин розподіл абсолютно неперервний відносно міри Лебега, а тому³

$$\mathbb{E}[g(X)] = \int g(x) \cdot f_X(x) d\lambda = \int_{-\infty}^{\infty} g(x) \cdot f_X(x) dx, \quad (\text{КЛ6.4.4})$$

де f_X — щільність розподілу. □

Із властивостей лінійності інтегралу впливають властивості сподівання та дисперсії, які ми раніше доводили для дискретних випадкових величин.

Твердження (КЛ6.4.6). Для сподівання $\mathbb{E}[X]$ та дисперсії $\text{Var}(X)$ деякої випадкової величини X , якщо вони існують, справедливі такі властивості:

- $\mathbb{E}[aX + bY] = a\mathbb{E}[X] + b\mathbb{E}[Y]$;

³Ми розглядаємо тільки щільності, інтегровані за Ріманом, хоча, як було зазначено раніше, існують і (малоцікаві) неінтегровані за Ріманом щільності.

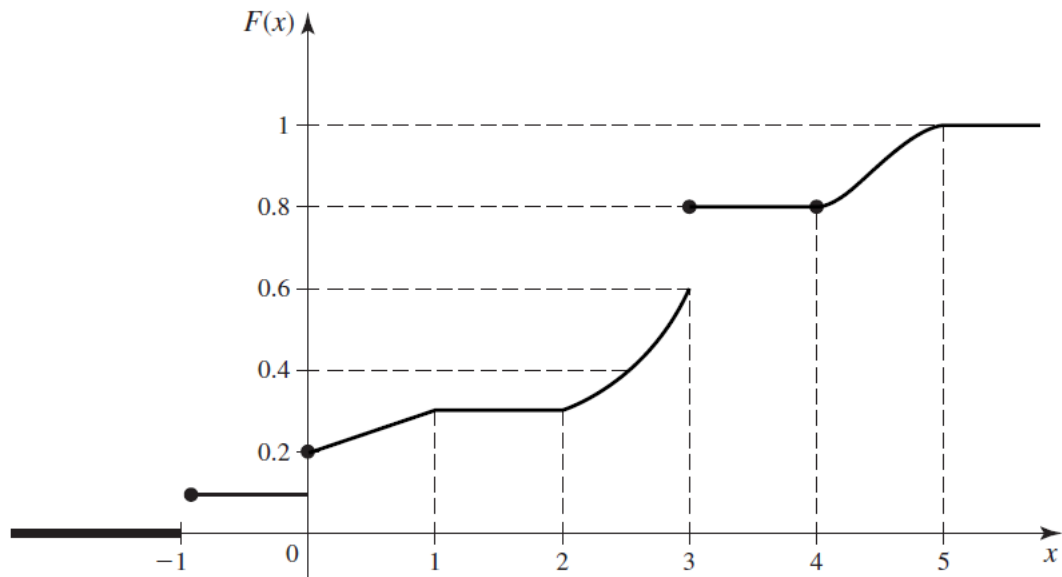


Рис. 7.2.1: Ілюстрація до Вправи 7.2.1 ([7], Рис. 3.9)

- $\text{Var}(aX + b) = a^2 \text{Var}(X)$;
- якщо $\text{Var}(X) = 0$, то $X = c$, $c \in \mathbb{R}$, майже напевно.

Існують такі розподіли, сподівання яких є нескінченним, а є й такі, сподівання яких узагалі не існує.

7.2 Функції розподілу

Вправа 7.2.1. Функцію розподілу випадкової величини зображено на Рис. 7.2.1. Чому дорівнюють такі ймовірності:

- $\mathbb{P}_X(X = -1)$, $\mathbb{P}_X(X \leq 0)$, $\mathbb{P}_X(X < 0)$, $\mathbb{P}_X(X = 1)$;
- $\mathbb{P}_X(0 < X \leq 3)$, $\mathbb{P}_X(0 < X < 3)$, $\mathbb{P}_X(0 \leq X \leq 3)$;
- $\mathbb{P}_X(X > 5)$, $\mathbb{P}_X(X \geq 5)$?

□

Розв'язання. Усі обчислення виконуватимемо за Твердженням КЛ6.1.5. Так, $\mathbb{P}_X(X = -1) = F(-1) - F(-1-) = 0.1 - 0 = 0.1$.

За визначенням, $\mathbb{P}_X(X \leq 0) = F(0) = 0.2$.

Згідно з Зауваженням КЛ6.1.3, $\mathbb{P}_X(X < 0) = F(0-) = 0.1$.

У точці $x = 1$ немає стрибка, тому $\mathbb{P}_X(X = 1) = 0$.

За Твердженням КЛ6.1.5, $\mathbb{P}_X((a; b]) = F(b) - F(a)$, тому $\mathbb{P}_X(0 < X \leq 3) = F(3) - F(0) = 0.8 - 0.2 = 0.6$.

З іншого боку,

$$\begin{aligned}\mathbb{P}_X(0 < X < 3) &= \mathbb{P}_X(0 < X \leq 3) - \mathbb{P}_X(X = 3) \\ &= F(3) - F(0) - (F(3) - F(3-)) \\ &= F(3-) - F(0) = 0.6 - 0.2 = 0.4.\end{aligned}$$

За цією ж логікою, $\mathbb{P}_X(0 \leq X \leq 3) = F(3) - F(0-) = 0.8 - 0.1 = 0.7$.

За Твердженням КЛ6.1.5, $\mathbb{P}_X(X \geq 5) = 1 - F(5-) = 1 - 1 = 0$.

З іншого боку, $\mathbb{P}_X(X > 5) = \mathbb{P}_X(X \geq 5) - \mathbb{P}_X(X = 5) = 1 - F(5-) - (F(5) - F(5-)) = 1 - F(5) = 1 - 1 = 0$. $\square$

Вправа 7.2.2. Нехай підкидають гральну кісточку. Розгляньмо випадкову величину X = «число, що випало на кісточці». Чому дорівнює функція розподілу F_X ? $\square$

Розв'язання. Очевидно, що функція ймовірності дорівнює $p_X(x_i) = \frac{1}{6}$, $x_i \in \{1, \dots, 6\}$. За визначенням функції розподілу маємо

$$F_X(x) = \mathbb{P}_X(X \leq x) .$$

Для $x < 1$ функція $F_X(x) = 0$, оскільки всі елементи носія більше або дорівнюють 1.

Якщо $x \in [1; 2)$, то подія $X \leq x$ еквівалентна події $X = 1$, а тому $F_X(x) = \frac{1}{6}$.

Якщо $x \in [2; 3)$, то подія $X \leq x$ еквівалентна події $(X = 1) \cup (X = 2)$, а тому $F_X(x) = \frac{1}{6} + \frac{1}{6} = \frac{1}{3}$.

У загальному випадку бачимо, що в кожній точці з носія маємо стрибок величиною $\frac{1}{6}$.

Нарешті, якщо $x \geq 6$, функція $F_X(x) = 1$, оскільки подія $X \leq x$ еквівалентна події $X \in \text{supp}(X)$.

Отже

$$F_X(x) = \begin{cases} 0, & x < 1 \\ \frac{i}{6}, & x \in [i; i+1), i = 1, \dots, 5 \\ 1, & x \geq 6 \end{cases} .$$

$\square$

Вправа 7.2.3. Нехай X має функцію розподілу

$$F(x) = a + b \operatorname{arctg} x .$$

Чому дорівнюють константи a і b ? $\square$

Розв'язання. Із властивостей функцій розподілу нам буде корисна властивість нормованості:

$$\begin{aligned}\lim_{x \rightarrow -\infty} F_X(x) &= 0, \\ \lim_{x \rightarrow \infty} F_X(x) &= 1.\end{aligned}$$

Зокрема,

$$\begin{aligned}0 &= \lim_{x \rightarrow -\infty} F_X(x) = \lim_{x \rightarrow -\infty} (a + b \operatorname{arctg} x) = a - b \frac{\pi}{2}, \\ 1 &= \lim_{x \rightarrow \infty} F_X(x) = \lim_{x \rightarrow \infty} (a + b \operatorname{arctg} x) = a + b \frac{\pi}{2}.\end{aligned}$$

Маємо систему лінійних рівнянь, розв'язком якої є

$$a = \frac{1}{2}, \quad b = \frac{1}{\pi},$$

звідки

$$F(x) = \frac{1}{2} + \frac{1}{\pi} \operatorname{arctg} x.$$

Ця функція розподілу відповідає *стандартному розподілу Коші* (standard Cauchy distribution)⁴, який розглянуто в Прикладі КЛ6.4.8. $\square$

7.3 Щільності розподілу

Вправа 7.3.1. Температура у двигуні X має щільність розподілу

$$f(x) = n(1-x)^{n-1} \cdot \mathbb{1}\{0 \leq x \leq 1\},$$

де $n \in \mathbb{N}$.

Для яких n це справді щільність розподілу? Чому дорівнює відповідна функція розподілу? $\square$

Розв'язання. Згідно з Твердженням КЛ6.3.9, f буде щільністю, якщо вона невід'ємна (що очевидно) і якщо

$$\int_{-\infty}^{\infty} f dx = 1.$$

⁴Огюстен-Луї Коші (Augustin-Louis Cauchy, 1789–1857) — французький математик.

У нашому випадку

$$\begin{aligned} \int_{-\infty}^{\infty} n(1-x)^{n-1} \cdot \mathbb{1}\{0 \leq x \leq 1\} dx &= \int_0^1 n(1-x)^{n-1} dx \\ &= -(1-x)^n \Big|_0^1 \\ &= 1, \end{aligned}$$

незалежно від n . Отже f є щільністю для будь-якого $n \in \mathbb{N}$.

Функцію розподілу знайдемо з виразу

$$\begin{aligned} F(x) &= \int_{-\infty}^x f(t) dt \\ &= \int_{-\infty}^x n(1-t)^{n-1} \cdot \mathbb{1}\{0 \leq t \leq 1\} dt. \end{aligned}$$

Для $x < 0$ маємо $F(x) = 0$, оскільки

$$F(x) = \int_{-\infty}^x n(1-t)^{n-1} \cdot \mathbb{1}\{0 \leq t \leq 1\} dt = \int_{-\infty}^x 0 dt = 0.$$

Для $x \in [0; 1)$ маємо

$$\begin{aligned} F(x) &= \int_{-\infty}^x n(1-t)^{n-1} \cdot \mathbb{1}\{0 \leq t \leq 1\} dt \\ &= \int_0^x n(1-t)^{n-1} dt \\ &= -(1-t)^n \Big|_0^x \\ &= 1 - (1-x)^n. \end{aligned}$$

Нарешті, для $x \geq 1$ маємо

$$F(x) = \int_{-\infty}^x n(1-t)^{n-1} \cdot \mathbb{1}\{0 \leq t \leq 1\} dt = \int_0^1 n(1-t)^{n-1} dt + \int_1^x 0 dt = 1.$$

Остаточно маємо

$$F(x) = \begin{cases} 0, & x < 0 \\ 1 - (1-x)^n, & 0 \leq x < 1 \\ 1, & x \geq 1 \end{cases}.$$

□

Вправа 7.3.2. Розгляньмо функцію *логістичного розподілу* (logistic distribution) :

$$F(x) = \frac{e^x}{1 + e^x}, \quad x \in \mathbb{R}. \quad (7.3.1)$$

Чому дорівнює щільність такого розподілу? Чому дорівнює ймовірність, що випадкова величина X , яка має логістичний розподіл, лежатиме на проміжку $[-2; 2]$? $\square$

Розв'язання. Очевидно, відповідна функція розподілу неперервна, і можна додатково перевірити, що вона неперервна абсолютно. Згідно з (КЛ6.3.3), щільність розподілу є похідною від функції розподілу:

$$f(x) = F'(x) = \left(\frac{e^x}{1 + e^x} \right)' = \frac{e^x}{(1 + e^x)^2}, \quad x \in \mathbb{R}.$$

Величина X має носій $\mathbb{R}$, оскільки $f(x) > 0$ скрізь.

Перевірмо, що інтеграл щільності дорівнює 1:

$$\begin{aligned} \int_{-\infty}^{\infty} f(x) dx &= \int_{-\infty}^{\infty} \frac{e^x}{(1 + e^x)^2} dx \\ &= \int_0^{\infty} \frac{du}{(1 + u)^2} = -\frac{1}{1 + u} \Big|_0^{\infty} = 1. \end{aligned}$$

Ймовірність потрапляння X у деяку множину A можна обчислити або за Теоремою Радона-Никодима як

$$\mathbb{P}_X(-2 \leq X \leq 2) = \int_{-2}^2 f(x) dx,$$

або за Твердженням КЛ6.3.6 як

$$\mathbb{P}_X(-2 \leq X \leq 2) = F(2) - F(0).$$

Ці два способи є абсолютно ідентичними, оскільки згідно з фундаментальною теоремою аналізу (fundamental theorem of calculus, вона ж теорема Ньютона-Ляйбніця),

$$\int_{-2}^2 f(x) dx = F(2) - F(0) \approx 0.76.$$

На Рис. 7.3.1 наведено графіки щільності та функції логістичного розподілу, а також проілюстровано процес обчислення $\mathbb{P}_X(-2 \leq X \leq 2)$. $\square$

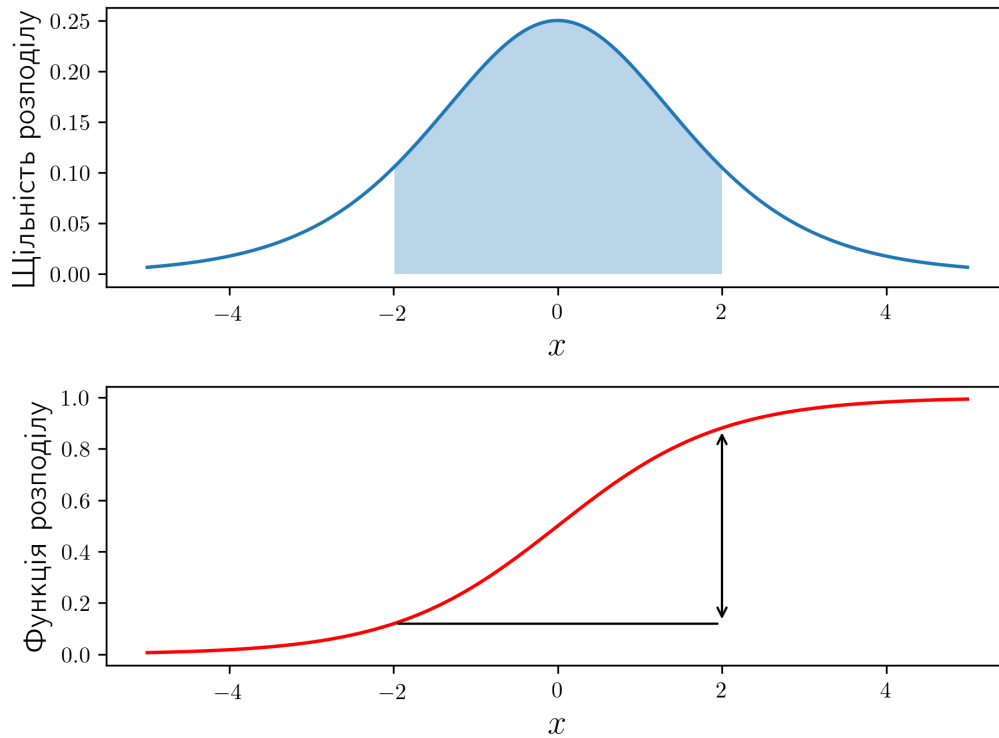


Рис. 7.3.1: Ілюстрація до Вправи 7.3.2: площа зафарбованої області дорівнює ймовірності події $-2 \leq X \leq 2$; таку ж довжину має вертикальний відрізок

Вправа 7.3.3. Функція розподілу *Рейлі*⁵ (Rayleigh distribution) дорівнює

$$F(x) = \left(1 - e^{-\frac{x^2}{2}}\right) \cdot \mathbb{1}\{x > 0\}. \quad (7.3.2)$$

Чому дорівнює щільність такого розподілу? Чому дорівнює ймовірність, що випадкова величина X , яка має розподіл Рейлі, перевищуватиме 2? $\square$

Розв'язання. Згідно з (КЛ6.3.3), щільність розподілу є похідною від функції розподілу в усіх точках її диференційовності. Нескладно бачити, що для всіх $x < 0$ щільність $f(x) = 0$, оскільки $F(x) = 0$. Для $x > 0$ маємо

$$f(x) = F'(x) = \left(1 - e^{-\frac{x^2}{2}}\right)' = x e^{-\frac{x^2}{2}}.$$

У точці $x = 0$ можна коректно порахувати похідні зліва і справа, щоб переконатися, що $F'(0) = 0$. Щоправда, $\{x = 0\}$ є множиною міри нуль, а тому значення щільності в цій точці не впливає на обчислення ймовірностей через відповідні інтеграли. Можна покласти $f(0) = c$ для будь-якого

⁵Названий так на честь лорда Рейлі (John William Strutt, 3rd Baron Rayleigh, 1842–1919) — англійського математика й фізика.

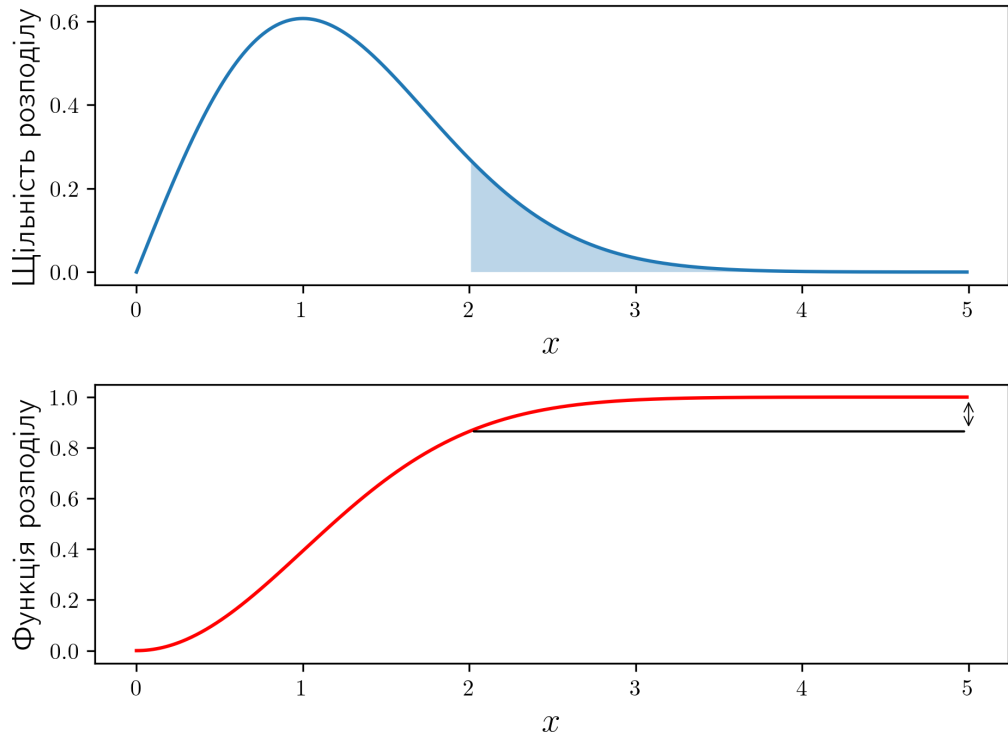


Рис. 7.3.2: Ілюстрація до Вправи 7.3.3: площа зафарбованої області дорівнює ймовірності події $X > 2$; таку ж довжину має вертикальний відрізок

$c \geq 0$. Якщо $f(0) = 0$, матимемо

$$f(x) = xe^{-\frac{x^2}{2}} \cdot \mathbb{1}\{x > 0\}.$$

Перевірмо, що інтеграл щільності дорівнює 1:

$$\begin{aligned} \int_{-\infty}^{\infty} f(x) dx &= \int_{-\infty}^{\infty} xe^{-\frac{x^2}{2}} \cdot \mathbb{1}\{x > 0\} dx \\ &= \int_0^{\infty} xe^{-\frac{x^2}{2}} dx \\ &= \int_0^{\infty} e^{-u} du = -e^{-u} \Big|_0^{\infty} = 1. \end{aligned}$$

Ймовірність, що X перевищує 2, можна обчислити, наприклад, як

$$\mathbb{P}_X(X > 2) = 1 - \mathbb{P}_X(X \leq 2) = 1 - F(2) \approx 0.14.$$

На Рис. 7.3.2 наведено графіки щільності та функції розподілу Рейлі, а також проілюстровано процес обчислення ймовірності $\mathbb{P}_X(X > 2)$.

□

Вправа 7.3.4. Нехай X = «тривалість роботи (у годинах) радіолампи» має щільність розподілу

$$f(x) = \frac{100}{x^2} \cdot \mathbb{1}\{x > 100\}.$$

Чому дорівнює ймовірність, що точно 2 з 5 таких ламп у деякому наборі потрібно буде замінити не пізніше від 150 годин роботи? Вважатимемо, що лампи можуть вийти з ладу незалежно одна від одної. $\square$

Розв'язання. Позначмо E_i = «потрібно замінити i -ту лампу». Тоді ймовірність такої події дорівнює

$$\begin{aligned} \mathbb{P}(E_i) &= \mathbb{P}_X(X \leq 150) = \int_{-\infty}^{150} \frac{100}{x^2} \cdot \mathbb{1}\{x > 100\} dx \\ &= 100 \int_{100}^{150} \frac{1}{x^2} dx \\ &= \frac{1}{3}. \end{aligned}$$

Маємо схему Бернуллі з 5 випробувань, де ймовірність успіху дорівнює $p = \mathbb{P}_X(X \leq 150) = \frac{1}{3}$. Отже випадкова величина «кількість ламп з 5, які доведеться замінити» має розподіл $Y \sim \text{Binom}(5, \frac{1}{3})$. Імовірність того, що $Y = 2$, дорівнює

$$\mathbb{P}_Y(Y = 2) = \binom{5}{2} \left(\frac{1}{3}\right)^2 \left(\frac{2}{3}\right)^3 = \frac{80}{243}.$$

$\square$

Вправа 7.3.5. Нехай X = «напруга в деякій електричній мережі», щільність розподілу якої

$$f_X(x) = \frac{1}{(1+x)^2} \cdot \mathbb{1}\{x > 0\}.$$

Нехай для її вимірювання використовують вольтметр, який коректно міряє напругу до значення 3 включно, але якщо напруга перевищує це значення, вольтметр просто показує 3.

Яку функцію розподілу має випадкова величина Y = «число на дисплеї вольтметра»? $\square$

Розв'язання. Функцію розподілу можна обчислити за визначенням. Так, якщо $x \leq 3$, $F_Y(x) = F_X(x)$. Функція розподілу величини X дорівнює

$$\begin{aligned} F_X(x) &= \begin{cases} 0, & x < 0 \\ \int_0^x \frac{dt}{(1+t)^2}, & x \geq 0 \end{cases} \\ &= \begin{cases} 0, & x < 0 \\ -\frac{1}{1+t} \Big|_0^x, & x \geq 0 \end{cases} \\ &= \begin{cases} 0, & x < 0 \\ 1 - \frac{1}{1+x}, & x \geq 0 \end{cases}. \end{aligned}$$

Для $x \geq 3$ ми маємо $Y = 3$. Отже $F_Y(y) = 1$ для всіх $y \geq 3$. Остаточно:

$$F_Y(y) = \begin{cases} 0, & y < 0 \\ 1 - \frac{1}{1+y}, & 0 \leq y < 3 \\ 1, & y \geq 3 \end{cases}.$$

Оскільки

$$F_y(3-) = \lim_{\substack{y \rightarrow 3 \\ y < 3}} \left(1 - \frac{1}{1+y} \right) = \frac{3}{4},$$

ми робимо висновок, що в точці $y = 3$ функція розподілу має стрибок величиною $1 - \frac{3}{4} = \frac{1}{4}$, що повністю відповідає ймовірності $\mathbb{P}_Y(Y = 3) = \mathbb{P}_X(X \geq 3)$ (Рис. 7.3.3).

Отже, випадкова величина Y не є ні абсолютно неперервною відносно міри Лебега, ні дискретною (тобто абсолютно неперервною відносно лічної міри), тому ні щільність (відносно міри Лебега), ні функція ймовірності не існують. Проте функція розподілу існує завжди.

□

7.4 Сподівання та дисперсія

Вправа 7.4.1. Розгляньмо випадкову величину X зі щільністю $f(x) = 3x^2 \cdot \mathbb{1}_{\{0 < x < 1\}}$.

Чому дорівнюють її сподівання та дисперсія?

Якщо $Y = 3X - 2$, чому дорівнюють її сподівання та дисперсія?

□

Розв'язання. За визначенням, сподівання є інтегралом функції X , який,

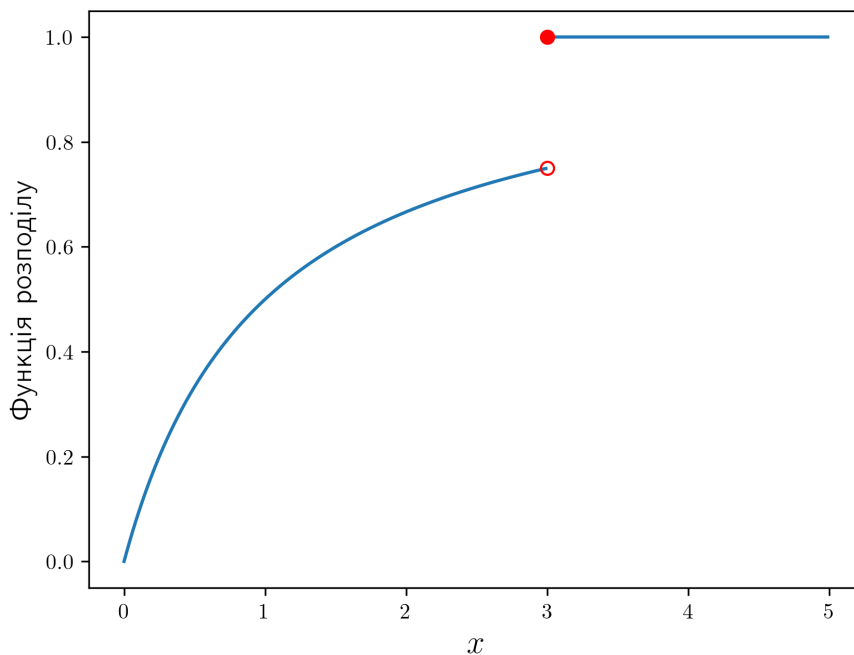


Рис. 7.3.3: Ілюстрація до Вправи 7.3.5: функція розподілу F_Y випадкової величини Y

згідно з Теоремою КЛ6.4.2, дорівнює

$$\mathbb{E}[X] = \int X d\mathbb{P} = \int_{-\infty}^{\infty} x f(x) dx = \int_0^1 x \cdot 3x^2 dx = \frac{3}{4} x^4 \Big|_0^1 = \frac{3}{4}.$$

Для підрахунку дисперсії спочатку знайдемо

$$\mathbb{E}[X^2] = \int_{-\infty}^{\infty} x^2 f(x) dx = \int_0^1 x^2 \cdot 3x^2 dx = \frac{3}{5} x^5 \Big|_0^1 = \frac{3}{5}.$$

Звідси

$$\text{Var}(X) = \mathbb{E}[X^2] - (\mathbb{E}[X])^2 = \frac{3}{5} - \frac{9}{16} = 0.0375.$$

За властивостями сподівання та дисперсії з Твердження КЛ6.4.6 маємо:

$$\begin{aligned} \mathbb{E}[Y] &= \mathbb{E}[3X - 2] = 3\mathbb{E}[X] - 2 = \frac{9}{4} - 2 = \frac{1}{4}, \\ \text{Var}(Y) &= \text{Var}(3X - 2) = 9\text{Var}(X) = 0.3375. \end{aligned}$$

□

Вправа 7.4.2. Ви повинні прийти на деяку зустріч. Якщо Ви прийдете на s хвилин раніше, Ваші втрати будуть cs для деякого $c > 0$. Якщо ж Ви запізнитесь на s хвилин, то Ваші втрати будуть ks для деякого $k > 0$.

Час $X \geq 0$, потрібний, щоб доїхати до місця зустрічі, має щільність f_X .

За скільки хвилин до початку зустрічі потрібно виїхати, що мінімізувати сподівані втрати? $\square$

Розв'язання. Спочатку з'ясуємо, як втрати залежать від часу виїзду X . Якщо виїхати за t хвилин до зустрічі, то втрати будуть дорівнювати

$$C_t(X) = \begin{cases} c(t - X), & X \leq t \\ k(X - t), & X \geq t \end{cases}.$$

Сподівані втрати можна обчислити за законом несвідомого статистика:

$$\begin{aligned} \mathbb{E}[C_t(X)] &= \int_0^\infty C_t(x) f(x) dx \\ &= \int_0^t c(t - x) f(x) dx + \int_t^\infty k(x - t) f(x) dx \\ &= ct \int_0^t f(x) dx - c \int_0^t x f(x) dx + k \int_t^\infty x f(x) dx - kt \int_t^\infty f(x) dx. \end{aligned}$$

Для мінімізації цього виразу потрібно знайти умову першого порядку, тобто взяти похідну за змінною t . Для похідних від інтегралів зі змінними межами використаємо інтегральне правило Ляйбніца (Leibniz integral rule):

$$\frac{d}{dt} \left(\int_{a(t)}^{b(t)} f(t, x) dx \right) = f(t, b(t)) \cdot \frac{d}{dt} b(t) - f(t, a(t)) \cdot \frac{d}{dt} a(t) + \int_{a(t)}^{b(t)} \frac{\partial}{\partial t} f(t, x) dx.$$

Наприклад, для $\int_0^t f(x) dx$ маємо $a(t) = 0$, $b(t) = t$, $f(t, x) = f(x)$, а отже

$$\frac{d}{dt} \int_0^t f(x) dx = f(t) \cdot \frac{d}{dt} t - f(0) \cdot \frac{d}{dt} 0 + \int_0^t \frac{d}{dt} f(x) dx = f(t).$$

Використовуючи аналогічні міркування для інших інтегралів, маємо:

$$\begin{aligned} \frac{\partial}{\partial t} \mathbb{E}[C_t(X)] &= c \int_0^t f(x) dx + ct f(t) - ct f(t) - kt f(t) - k \int_t^\infty f(x) dx + kt f(t) \\ &= c \int_0^t f(x) dx - k \int_t^\infty f(x) dx \\ &= cF(t) - k(1 - F(t)) = (c + k)F(t) - k. \end{aligned}$$

Тоді умова першого порядку має вид

$$(c + k)F(t) - k = 0,$$

а отже оптимальним часом виїзду є t^* , що задовольняє відповідне рівняння.

Друга похідна в цій точці дорівнює $(c + k)f(t^*) \geq 0$, оскільки щільність не може бути від'ємною. Для того, що довести, що в цій точці щільність буде строго додатною (а отже це справді буде точка мінімуму), потрібно довести, що $t^* > 0$.

Справді, оскільки $f(x) = 0$ для $x \leq 0$,

$$F(0) = 0 < \frac{k}{k + c} = F(t^*) ,$$

а отже $0 < t^*$ через неспадність F . □

Приклад 7.4.3. У розділі економіки, пов'язаному з вибором в умовах невизначеності, розглядають *теорію сподіваної корисності* (expected utility theory). Якщо дуже коротко, то кожному з можливих результатів розвитку деякої ситуації зіставляють його *корисність* (utility), і ухвалюють рішення з міркувань максимізації сподіваної корисності.

Наприклад, нехай ми стикаємося з вибором між двома так званими *лотереями* (lotteries) — варіантами розвитку ситуації. У першій лотереї ми можемо дістати $x = 500$ або $x = -350$ з однаковою ймовірністю, а в другій — $y = 40$, $y = 50$ або $y = 60$, також з однаковою ймовірністю. Позначивши виграш через відповідні літери, можемо з легкістю обчислити $\mathbb{E}[X] = 75$ та $\mathbb{E}[Y] = 50$. На перший погляд, може здатися, що перша лотерея вигідніша, проте також і дисперсія в першому випадку значно вища ($\text{Var}(X) = 425^2$, $\text{Var}(Y) \approx 66.67$).

Альтернативний спосіб вибрати лотерею полягає в тому, що кожному результату зіставляється його корисність, яка може набувати різних форм залежно від того, якими критеріями ми послуговуємося. Наприклад, якщо нам не подобаються ситуації, коли ми програємо великі суми, ми можемо використати таку функцію:

$$U(x) = \begin{cases} 100 \ln \frac{x + 100}{100} , & x \geq 0 \\ x , & x < 0 \end{cases} .$$

Тоді сподівані корисності дорівнюватимуть

$$\mathbb{E}[U(X)] = \frac{1}{2} \cdot 100 \ln 6 + \frac{1}{2} \cdot (-350) = -85.412$$

$$\mathbb{E}[U(Y)] = \frac{1}{3} \cdot 100 \ln 1.4 + \frac{1}{3} \cdot 100 \ln 1.5 + \frac{1}{3} \cdot 100 \ln 1.6 = 40.398 ,$$

і тоді друга лотерея стає очевидно привабливішою.

Розгляньмо тепер ситуацію, коли результат лотереї може бути випадковим числом із проміжку $[0; 4]$, тобто $f(x) = \frac{1}{4} \cdot \mathbb{1}\{0 \leq X \leq 4\}$. Нехай

кожний результат має корисність $U(x) = x^\alpha$, $x \geq 0$, $\alpha > 0$.

Сподівана корисність у цьому випадку дорівнює

$$\mathbb{E}[U(X)] = \int_0^4 x^\alpha \cdot \frac{1}{4} dx = \frac{4^\alpha}{\alpha + 1}.$$

Якщо нас попросять продати квиток на таку лотерею за деяку ціну x_0 , то ми повинні погодитися продати його, якщо

$$\mathbb{E}[U(x_0)] > \mathbb{E}[U(X)] \Leftrightarrow x_0^\alpha > \frac{4^\alpha}{\alpha + 1} \Leftrightarrow x_0 > \frac{4}{(\alpha + 1)^{1/\alpha}}.$$

□

7.5 Додаткові вправи

Вправа 7.5.1. Нехай X має функцію розподілу

$$F(x) = \begin{cases} 0, & x < 0 \\ 2c \left(x^2 - \frac{x^3}{3} \right), & 0 \leq x < 2 \\ 1, & x \geq 2 \end{cases}.$$

Чому дорівнює відповідна щільність? Чому дорівнює константа c ? Чому дорівнює $\mathbb{P}_X(X > 0.5)$? □

Розв'язання. Щільність розподілу є похідної функції розподілу. Зокрема, $f(x) = 0$ для $x < 0$ та $x > 2$. Для $0 < x < 2$ маємо

$$f(x) = F'(x) = 4cx - 2cx^2 = 2c(2x - x^2).$$

У точках $x = 0$ та $x = 2$, щоб не возитися з перевіркою диференційовності, можемо дозволити собі покласти $f(0) = f(2) = 0$, оскільки ці дві точки є множинами міри нуль. Тому

$$f(x) = 2c(2x - x^2) \cdot \mathbb{1}_{\{0 < x < 2\}}.$$

Інтеграл щільності повинен дорівнювати 1:

$$\int_{-\infty}^{\infty} f(x) dx = \int_0^2 2c(2x - x^2) dx = 2c \left(x^2 - \frac{x^3}{3} \right) \Big|_0^2 = 2c \left(4 - \frac{8}{3} \right) = 1,$$

звідки $c = \frac{3}{8}$.

Імовірність $\mathbb{P}_X(X > 0.5)$ обчислімо за допомогою функції розподілу:

$$\begin{aligned}\mathbb{P}_X(X > 0.5) &= 1 - \mathbb{P}_X(X \leq 0.5) = 1 - F(0.5) \\ &= 1 - 2 \cdot \frac{3}{8} \left((0.5)^2 - \frac{(0.5)^3}{3} \right) \approx 0.844.\end{aligned}$$

□

Вправа 7.5.2. Нехай випадкова величина X має щільність $f(x) = (ax + b) \cdot \mathbb{1}\{0 \leq x \leq 1\}$ і нехай $\mathbb{P}_X(X \geq \frac{1}{2}) = \frac{1}{3}$.

Чому дорівнюють константи a і b ? Чому дорівнює функція розподілу?

□

Розв'язання. З одного боку, ми знаємо, що інтеграл щільності дорівнює 1:

$$\int_0^1 (ax + b) dx = \left(a \frac{x^2}{2} + bx \right) \Big|_0^1 = \frac{a}{2} + b = 1.$$

З іншого боку,

$$\mathbb{P}_X\left(X \geq \frac{1}{2}\right) = \int_{\frac{1}{2}}^1 (ax + b) dx = \left(a \frac{x^2}{2} + bx \right) \Big|_{\frac{1}{2}}^1 = \frac{a}{2} + b - \frac{a}{8} - \frac{b}{2} = \frac{1}{3}.$$

Маємо систему лінійних рівнянь, розв'язком якої є

$$a = -\frac{4}{3}, \quad b = \frac{5}{3}.$$

Функцію розподілу знайдемо з виразу

$$F(x) = \int_{-\infty}^x \left(-\frac{4}{3}t + \frac{5}{3} \right) \cdot \mathbb{1}\{0 \leq t \leq 1\} dt.$$

Для $x < 0$ маємо $F(x) = 0$, оскільки маємо інтеграл від 0.

Для $x \in [0; 1]$ маємо

$$F(x) = \int_0^x \left(-\frac{4}{3}t + \frac{5}{3} \right) dt = \left(-\frac{2}{3}t^2 + \frac{5}{3}t \right) \Big|_0^x = -\frac{2}{3}x^2 + \frac{5}{3}x.$$

Нарешті, для $x > 1$ маємо

$$F(x) = \int_0^1 f(t) dt + \int_1^x 0 dt = 1.$$

Остаточно маємо

$$F(x) = \begin{cases} 0, & x < 0 \\ -\frac{2}{3}x^2 + \frac{5}{3}x, & 0 \leq x < 1 \\ 1, & x \geq 1 \end{cases}.$$

□

Вправа 7.5.3. Нехай випадкова величина X має щільність

$$f(x) = \begin{cases} c_1 x, & -2 < x < 0 \\ c_2 x, & 0 \leq x < 1 \\ 0, & \text{інакше} \end{cases}.$$

Нехай $\mathbb{E}[X] = \frac{1}{3}$.

Чому дорівнюють константи c_1 та c_2 ?

□

Розв'язання. Для визначення констант c_1 та c_2 нам потрібно дістати два рівняння. Одне випливає з того, що інтеграл щільності дорівнює 1:

$$1 = \int_{-2}^0 c_1 x \, dx + \int_0^1 c_2 x \, dx = -2c_1 + \frac{c_2}{2}.$$

Інше випливає з інформації про сподівання:

$$\frac{1}{3} = \int_{-2}^0 x \cdot c_1 x \, dx + \int_0^1 x \cdot c_2 x \, dx = \frac{8}{3}c_1 + \frac{c_2}{3}.$$

Отже маємо систему рівнянь

$$\begin{cases} -2c_1 + \frac{c_2}{2} = 1 \\ 8c_1 + c_2 = 1 \end{cases},$$

розв'язком якої є

$$c_1 = -\frac{1}{12}, \quad c_2 = \frac{5}{3}.$$

□

Вправа 7.5.4. Нехай на автозаправку бензин завозять один раз на тиждень. Якщо обсяг тижневих продажів автозаправки (у тисячах літрів) є випадковою величиною X зі щільністю $f(x) = 5(1-x)^4 \cdot \mathbb{1}\{0 < x < 1\}$,

якого обсягу повинна бути цистерна, щоб імовірність використання всього тижневого запасу протягом деякого тижня була 0.01? $\square$

Розв'язання. Нехай обсяг цистерни дорівнює V . Нас цікавить імовірність $\mathbb{P}_X(X > V) = 0.01$. Ця ймовірність має такий вираз:

$$\mathbb{P}_X(X > V) = 1 - \int_{-\infty}^V 5(1-x)^4 \cdot \mathbb{1}\{0 < x < 1\} dx.$$

Імовірність може дорівнювати 0.01, тільки якщо $0 < V < 1$, а отже

$$\mathbb{P}_X(X > V) = 1 - \int_0^V 5(1-x)^4 dx = 1 + (1-x)^5 \Big|_0^V = 1 + (1-V)^5 - 1 = (1-V)^5,$$

звідки

$$1 - V = \sqrt[5]{0.01} \Rightarrow V = 1 - \sqrt[5]{0.01} \approx 0.602.$$

$\square$

Вправа 7.5.5. Нехай випадкова величина X має щільність

$$f(x) = \frac{|x|}{c^2} \cdot \mathbb{1}\{-c < x < c\}.$$

Чому дорівнює $\mathbb{E}[X^n]$ для будь-якого $n = 1, 2, \dots$? Чому дорівнює дисперсія X ? $\square$

Розв'язання. За законом несвідомого статистика,

$$\begin{aligned} \mathbb{E}[X^n] &= \int_{-\infty}^{\infty} x^n \cdot \frac{|x|}{c^2} \cdot \mathbb{1}\{-c < x < c\} dx \\ &= \int_{-c}^c x^n \cdot \frac{|x|}{c^2} dx \\ &= \int_{-c}^0 \left(-\frac{x^{n+1}}{c^2} \right) dx + \int_0^c \frac{x^{n+1}}{c^2} dx \\ &= -\frac{1}{c^2(n+2)} x^{n+2} \Big|_{-c}^0 + \frac{1}{c^2(n+2)} x^{n+2} \Big|_0^c \\ &= \frac{1}{c^2(n+2)} (-c)^{n+2} + \frac{1}{c^2(n+2)} c^{n+2} \\ &= \frac{1}{c^2(n+2)} (c^{n+2} + (-c)^{n+2}). \end{aligned}$$

Іншими словами,

$$\mathbb{E}[X^n] = \begin{cases} 0, & n \text{ непарне} \\ \frac{2c^n}{n+2}, & n \text{ парне} \end{cases}.$$

Зокрема,

$$\begin{aligned} \mathbb{E}[X] &= 0, \\ \mathbb{E}[X^2] &= \frac{2c^2}{4} = \frac{c^2}{2}, \end{aligned}$$

а отже

$$\text{Var}(X) = \frac{c^2}{2}.$$

□

8 Функції від випадкових величин. Деякі неперервні розподіли

8.1 Основні теоретичні відомості

Детальний огляд понять, потрібний для розв’язання задач на цьому занятті, наведено в Курсі лекцій, Розд. КЛ7.

У загальному випадку функцію розподілу для будь-якої випадкової величини можна дістати за Визначенням КЛ6.1.1. 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

$$\begin{aligned} F_Y(y) &= \mathbb{P}_Y(Y \leq y) = \mathbb{P}_X(g(X) \leq y) = \mathbb{P}_X(X \in g^{-1}((-\infty; y])) \\ &= \mathbb{P}(\{\omega : X(\omega) \in g^{-1}((-\infty; y])\}) \end{aligned} \quad (\text{КЛ7.4.1})$$

де $g^{-1}((-\infty; y]) = \{x \in \mathbb{R} : g(x) \leq y\}$.

Підхід із використанням (КЛ7.4.1) загальний і працює для будь-яких величин, проте для величин, які мають щільність (у тому числі відносно лічної міри), існують альтернативні способи.

Так, для дискретної випадкової величини X маємо таку формулу для функції ймовірності $Y = g(X)$:

$$\mathbb{P}_Y(Y = y) = \mathbb{P}_X(g(X) = y) = \sum_{x: g(x)=y} \mathbb{P}_X(X = x) \quad (\text{КЛ7.4.2})$$

При цьому варто пам’ятати, що застосування функції g до випадкової величини X не є тотожним до застосування функції g до функції ймовірності p_X відповідної випадкової величини. Зокрема, зовсім некоректно стверджувати, що для здобуття функції ймовірності $Y = 2X$ потрібно помножити $p_X(x)$ на 2 (Зауваження КЛ7.4.3).

Для абсолютно неперервних випадкових величин в окремих випадках справедлива так звана формула заміни змінних.

Теорема (Формула заміни змінних (Change of variables formula), КЛ7.4.4). 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

$$f_Y(y) = f_X(g^{-1}(y)) \cdot \left| \frac{d}{dy} g^{-1}(y) \right| \cdot \mathbb{1}\{y \in g(\text{supp}(X))\} . \quad (\text{КЛ7.4.3})$$

У випадках, коли функція g не є взаємно однозначною, інколи можливо розбити носій величини X на неперетинні підмножини так, що на кожній із них g *буде* взаємно однозначною.

Твердження (КЛ7.4.6). 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 $B = \{y : g_i(x) = y, x \in A_i\}$ однакова для всіх $i = 1, \dots, k$;
- функції $g_i^{-1}(y)$ неперервно диференційовні на B , $i = 1, \dots, k$.

Тоді

$$f_Y(y) = \mathbb{1}\{y \in B\} \cdot \sum_{i=1}^k f_X(g_i^{-1}(y)) \cdot \left| \frac{d}{dy} g_i^{-1}(y) \right| . \quad (\text{КЛ7.4.4})$$

На практиці поширеними є випадкові величини, які можна утворювати одну з одної шляхом *перетворення зсуву та масштабування* (location-scale transformation):

$$Y = aX + b, \quad a, b \in \mathbb{R}, \quad a > 0 . \quad (\text{КЛ7.5.1})$$

Параметр b відповідає за зсув, а параметр a — за масштабування.

Сукупність розподілів, які пов'язано у відповідний спосіб, становить *сім'ю відносно зсуву-масштабування* (location-scale family).

Оскільки для такої сім'ї виконуються всі умови Теорема КЛ7.4.4, можемо записати формулу щільності величини Y :

$$f_Y(y) = f_X\left(\frac{y-b}{a}\right) \cdot \frac{1}{a} \cdot \mathbb{1}\{y \in a \cdot \text{supp}(X) + b\} . \quad (\text{КЛ7.5.2})$$

На практиці поширеними є *рівномірні розподіли* (uniform distributions), які утворюють власну сім'ю відносно зсуву-масштабування. Розгляд їхніх властивостей доречно почати зі стандартизованої версії.

Визначення (КЛ7.1.1). Випадкова величина X має *стандартний* (standard) рівномірний розподіл, $X \sim U((0; 1])$, якщо її щільність розподілу дорівнює

$$\begin{aligned} f_X(x) &= \begin{cases} 1, & x \in (0; 1] \\ 0, & x \notin (0; 1] \end{cases} \\ &= \mathbb{1}\{x \in (0; 1]\} . \end{aligned} \quad (\text{КЛ7.1.1})$$

□

Функція стандартного рівномірного розподілу дорівнює

$$F_X(x) = \int_{(-\infty; x]} \mathbb{1}\{t \in (0; 1]\} d\lambda = \int_{-\infty}^x \mathbb{1}\{t \in (0; 1]\} dt = \begin{cases} 0, & x < 0 \\ x, & 0 \leq x < 1 \\ 1, & x \geq 1 \end{cases} . \quad (\text{КЛ7.1.2})$$

Сподівання та дисперсія стандартного рівномірного розподілу впливають із такого твердження.

Твердження (КЛ7.1.3). Якщо $X \sim U((0; 1])$, то $\mathbb{E}[X] = \frac{1}{2}$, а $\text{Var}(X) = \frac{1}{12}$.

Це інтуїтивно зрозуміло, адже якщо щільність є сталою на інтервалі $(0; 1]$, то «середнім значенням» повинна бути середина інтервалу.

Рівномірний розподіл на деякому довільному проміжку $(a; b]$ можна утворити шляхом перетворення (КЛ7.5.1). Так, нехай $U \sim U((0; 1])$. Розгляньмо випадкову величину $X = a + (b - a)U$. Згідно з (КЛ7.5.2), щільність такої величини дорівнює

$$\begin{aligned} f_Y(y) &= f_X\left(\frac{y - a}{b - a}\right) \cdot \frac{1}{b - a} \cdot \mathbb{1}\{y \in a + (b - a) \cdot \text{supp}(X)\} \\ &= \frac{1}{b - a} \cdot \mathbb{1}\{y \in (a; b]\} . \end{aligned}$$

Визначення (КЛ7.5.2). Випадкова величина X має *рівномірний розподіл* (uniform distribution) на $(a; b]$, що позначають через $X \sim U((a; b])$, якщо її щільність розподілу дорівнює

$$f(x) = \frac{1}{b - a} \cdot \mathbb{1}\{x \in (a; b]\} . \quad (\text{КЛ7.5.4})$$

□

Функція рівномірного розподілу дорівнює

$$F_X(x) = \int_{(-\infty; x]} \mathbb{1}\{x \in (a; b]\} d\lambda = \int_{-\infty}^x \mathbb{1}\{t \in (a; b]\} dt$$

$$= \begin{cases} 0, & x < a \\ \frac{x-a}{b-a}, & a \leq x < b \\ 1, & x \geq b \end{cases} \quad (\text{КЛ7.5.5})$$

Рівномірний розподіл є *нормованою* мірою Лебега, тобто для $x \in (a; b)$ маємо

$$\mathbb{P}_X((a; x]) = F_X(x) - F_X(a) = \frac{x-a}{b-a} = \frac{\lambda((a; x])}{\lambda((a; b])}.$$

Використовуючи взаємозв'язок між стандартним рівномірним розподілом і довільним рівномірним розподілом, можна швидко обчислити відповідні сподівання та дисперсію.

Твердження (КЛ7.5.4). Якщо $X \sim U((a; b])$, то сподівання $\mathbb{E}[X] = \frac{a+b}{2}$, а дисперсія $\text{Var}(X) = \frac{(b-a)^2}{12}$.

Іншим надзвичайно поширеним на практиці розподілом є *нормальний розподіл* (normal distribution). Зокрема, його стандартизована форма має таке визначення.

Визначення (КЛ7.2.1). Випадкова величина Z^1 має стандартний нормальний розподіл, $Z \sim N(0, 1)$, якщо її щільність розподілу дорівнює

$$\phi(z) = \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}}, \quad z \in \mathbb{R}. \quad (\text{КЛ7.2.1})$$

□

Позначення для цієї щільності через ϕ замість f обумовлено історично. Як можна бачити, щільність є парною функцією: $\phi(-z) = \phi(z)$.

Твердження (КЛ7.2.2). Функція (КЛ7.2.1) відповідає вимогам до щільностей розподілу, викладеним у Твердженні КЛ6.3.9.

Варто зазначити, що функція стандартного нормального розподілу

$$\Phi(z) = \int_{-\infty}^z \phi(t) dt = \int_{-\infty}^z \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}} dt \quad (\text{КЛ7.2.2})$$

не має аналітичного виразу. Її значення можна підрахувати тільки чисельно. Позначення цієї функції через Φ також обумовлено історично.

¹Таке позначення замість X обумовлено історично, і ми його дотримуватимемося.

Для функції розподілу справедливі деякі корисні властивості.

Твердження (КЛ7.2.3). Функція стандартного нормального розподілу Φ має такі властивості:

- (i) $\Phi(z) = 1 - \Phi(-z)$;
- (ii) $\Phi^{-1}(p) = -\Phi^{-1}(1 - p)$, $p \in (0; 1)$.

Сподівання та дисперсія стандартного нормального розподілу такі.

Твердження (КЛ7.2.5). Якщо $Z \sim N(0, 1)$, то $\mathbb{E}[Z] = 0$, а $\text{Var}(Z) = 1$.

Тепер стає зрозуміло, чому в позначенні стандартного нормального розподілу $N(0, 1)$ використано числа 0 і 1.

Безпосереднім узагальненням стандартного нормального розподілу є нормальний розподіл зі сподіванням μ та дисперсією σ^2 . Відповідний розподіл можна дістати шляхом застосування перетворення (КЛ7.5.1).

Визначення (КЛ7.5.6). Випадкова величина X має *нормальний розподіл* (normal distribution) зі сподіванням μ та дисперсією σ^2 ($\sigma > 0$), $X \sim N(\mu, \sigma^2)$, якщо $X = \mu + \sigma Z$, де $Z \sim N(0, 1)$. $\square$

Зв'язок між функціями та щільностями стандартного нормального розподілу і довільного нормального розподілу визначає таке твердження.

Твердження (КЛ7.5.7). Нехай $X \sim N(\mu, \sigma^2)$. Тоді

$$\begin{aligned} F(x) &= \Phi\left(\frac{x - \mu}{\sigma}\right), \\ f(x) &= \frac{1}{\sigma} \cdot \phi\left(\frac{x - \mu}{\sigma}\right) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x - \mu)^2}{2\sigma^2}}. \end{aligned} \quad (\text{КЛ7.5.6})$$

Зауваження (КЛ7.5.9). Застосування перетворення (КЛ7.5.1) 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 впливає з (КЛ7.5.2):

$$f_Y(y) = f_X\left(\frac{y - b}{a}\right) \cdot \frac{1}{a} = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{\left(\frac{y-b}{a} - \mu\right)^2}{2\sigma^2}} \cdot \frac{1}{a} = \frac{1}{\sqrt{2\pi}a\sigma} e^{-\frac{(y - (a\mu + b))^2}{2a^2\sigma^2}}. \quad \square$$

Твердження (КЛ7.5.10). Якщо $X \sim N(\mu, \sigma^2)$, то $\mathbb{E}[X] = \mu$, а $\text{Var}(X) = \sigma^2$.

Зауваження (КЛ7.5.11). У загальному випадку до будь-якої величини X зі сподіванням μ та дисперсією σ^2 можна застосувати процедуру стандартизації:

$$Z = \frac{X - \mu}{\sigma} . \quad (\text{КЛ7.5.7})$$

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). $\square$

Для нормального розподілу справедливі такі підрахунки:

$$\begin{aligned} \mathbb{P}_X(|X - \mu| \leq \sigma) &= \mathbb{P}_X\left(\left|\frac{X - \mu}{\sigma}\right| \leq 1\right) = \mathbb{P}_Z(-1 \leq Z \leq 1) \\ &= \Phi(1) - \Phi(-1) = 2\Phi(1) - 1 \approx 0.68 , \\ \mathbb{P}_X(|X - \mu| \leq 2\sigma) &= \mathbb{P}_X\left(\left|\frac{X - \mu}{\sigma}\right| \leq 2\right) = \mathbb{P}_Z(-2 \leq Z \leq 2) \\ &= \Phi(2) - \Phi(-2) = 2\Phi(2) - 1 \approx 0.95 , \\ \mathbb{P}_X(|X - \mu| \leq 3\sigma) &= \mathbb{P}_X\left(\left|\frac{X - \mu}{\sigma}\right| \leq 3\right) = \mathbb{P}_Z(-3 \leq Z \leq 3) \\ &= \Phi(3) - \Phi(-3) = 2\Phi(3) - 1 \approx 0.997 . \end{aligned}$$

Ці значення можна використовувати для швидкого приблизного підрахунку відповідних імовірностей. Зокрема, під час великого числа повторень деякого експерименту, результатом якого є нормальна величина, 99.7% усіх результатів лежатимуть на відстані до трьох середньоквадратичних відхилень від сподівання (на проміжку $[\mu - 3\sigma; \mu + 3\sigma]$). Часто відповідне правило називають «правилом трьох сигм».

8.2 Рівномірний розподіл

Вправа 8.2.1. Нехай маємо палицю довжиною $L = 1$. Нехай її ламають на дві частини у випадковій точці $U \sim U((0; 1])$.

Чому дорівнює сподівана довжина того з двох шматків, який містить точку $0 \leq p \leq 1$? Яким повинно бути p , щоб ця довжина була максимальна? $\square$

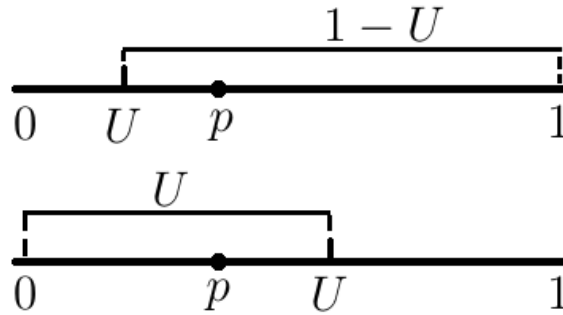


Рис. 8.2.1: Ілюстрація до Вправи 8.2.1: верхній рисунок відображає ситуацію, коли точка p лежить правіше від точки розламу U , нижній — лівіше від цієї точки розламу

Розв'язання. Нехай $L_p(U)$ — довжина шматка, що містить точку p . Тоді, як можна бачити з Рис. 8.2.1,

$$L_p(U) = \begin{cases} 1 - U, & U < p \\ U, & U \geq p \end{cases}.$$

З іншого боку, $f_U(u) = \mathbb{1}\{0 < u < 1\}$. Тому за законом несвідомого статистика маємо:

$$\begin{aligned} \mathbb{E}[L_p(U)] &= \int_{-\infty}^{\infty} L_p(u) \cdot f_U(u) du = \int_0^p (1 - u) du + \int_p^1 u du \\ &= p - \frac{p^2}{2} + \frac{1}{2} - \frac{p^2}{2} = \frac{1}{2} + p(1 - p). \end{aligned}$$

Це сподівання максимізує точка $p = \frac{1}{2}$, тобто сподівана довжина відповідного шматка максимальна, якщо точка p лежить посередині палиці. $\square$

Вправа 8.2.2. Нехай обсяг коштів, які страхова компанія повинна виплатити клієнту, має рівномірний розподіл $X \sim U((0; a])$, $a > 0$. У страховій справі існує поняття *франшизи* (deductible) — певної суми, яку повинен відшкодувати сам клієнт. Наприклад, поліс медичного страхування може передбачати, що франшиза дорівнює d , $0 < d < a$, тобто страхова компанія покриє тільки ту частину витрат пацієнта, які перевищують d .

У цьому випадку сума коштів Y , яку сплатить компанія, дорівнює

$$Y = (X - d) \cdot \mathbb{1}\{d < X \leq a\}.$$

Чому дорівнюють сподівані витрати страхової компанії?

Якщо ми хочемо, щоб $\mathbb{E}[Y] = c\mathbb{E}[X]$ для деякого $0 < c < 1$, яким повинно бути значення d ? $\square$

Розв'язання. Згідно з законом несвідомого статистика, враховуючи, що $f_X(x) = \frac{1}{a} \cdot \mathbb{1}\{0 < x < a\}$,

$$\begin{aligned}\mathbb{E}[Y] &= \int_{-\infty}^{\infty} (x - d) \cdot \mathbb{1}\{d < x \leq a\} \cdot \frac{1}{a} \cdot \mathbb{1}\{0 < x < a\} dx \\ &= \frac{1}{a} \int_d^a (x - d) dx = \frac{1}{2a} (x - d)^2 \Big|_d^a = \frac{(a - d)^2}{2a}.\end{aligned}$$

У нашому випадку, очевидно, $\mathbb{E}[X] = \frac{a}{2}$. Якщо ми хочемо $\mathbb{E}[Y] = c\mathbb{E}[X]$, то потрібно покласти

$$\frac{(a - d)^2}{2a} = c \frac{a}{2},$$

звідки $d = a(1 - \sqrt{c})$. □

Вправа 8.2.3. Нехай між аеропортом та містом що 15 хвилин курсує шатл. Якщо час очікування шатла має рівномірний розподіл, чому дорівнює ймовірність, що пасажиру доведеться ждати шатл понад 6 хвилин?

Якщо пасажир *уже* жде шатл протягом 8 хвилин, чому дорівнює ймовірність, що йому доведеться ждати *додатково* 2 і більше хвилин? □

Розв'язання. За умовами задачі маємо $X \sim U((0; 15])$. Отже

$$\mathbb{P}_X(X \geq 6) = 1 - F_X(6) = 1 - \frac{6}{15} = 0.6.$$

Якщо пасажир уже жде шатл протягом 8 хвилин, то маємо

$$\begin{aligned}\mathbb{P}_X(X \geq 8 + 2 \mid X \geq 8) &= \frac{\mathbb{P}_X(X \geq 10)}{\mathbb{P}_X(X \geq 8)} = \frac{1 - F_X(10)}{1 - F_X(8)} \\ &= \frac{1 - 10/15}{1 - 8/15} = \frac{5}{7} \approx 0.714.\end{aligned}$$

□

8.3 Нормальний розподіл

Вправа 8.3.1. Нехай розподіл IQ серед деякої групи осіб є нормальний зі сподіванням $\mu = 105$ та середньоквадратичним відхиленням $\sigma = 20$. У якій частки осіб IQ щонайменше 50? Щонайбільше 80? Між 95 і 125? □

Розв'язання. Нехай $X = \text{IQ}$ випадково обраної особи». Тоді розподіл буде $X \sim N(105, 20^2)$, і

$$\mathbb{P}_X(X \geq 50) = 1 - \mathbb{P}_X(X < 50) = 1 - F_X(50) .$$

У таких випадках ми обов'язково маємо перейти до стандартної нормальної випадкової величини, оскільки її функція розподілу Φ є загально-відомою і її значення можна чисельно обчислити. Для цього використаємо стандартизацію (КЛ7.5.7):

$$\begin{aligned} \mathbb{P}_X(X \geq 50) &= 1 - \mathbb{P}_X\left(\frac{X - 105}{20} < \frac{50 - 105}{20}\right) \\ &= 1 - \mathbb{P}_Z(Z < -2.75) = \Phi(-2.75) \approx 0.997 . \end{aligned}$$

Аналогічно для інших випадків маємо

$$\begin{aligned} \mathbb{P}_X(X \leq 80) &= \mathbb{P}_X\left(\frac{X - 105}{20} \leq \frac{80 - 105}{20}\right) \\ &= \mathbb{P}_Z(Z \leq -1.25) = \Phi(-1.25) \approx 0.106 , \\ \mathbb{P}_X(95 \leq X \leq 125) &= \mathbb{P}_X\left(\frac{95 - 105}{20} \leq \frac{X - 105}{20} \leq \frac{125 - 105}{20}\right) \\ &= \mathbb{P}_Z(-0.5 \leq Z \leq 1) = \Phi(1) - \Phi(-0.5) \approx 0.533 . \end{aligned}$$

□

Приклад 8.3.2. Побутує така думка, що екзаменаційну контрольну роботу складено вдало, якщо розподіл балів студентів за неї має (приблизно) нормальний розподіл. Тоді екзаменатор може оцінити параметри μ та σ^2 відповідного розподілу на основі наявних балів² і визначити порогові бали для кожної оцінки.

Нехай студент дістав бал X . Тоді можна поставити такі оцінки:

- «відмінно», якщо $X > \mu + \sigma$;
- «дуже добре», якщо $\mu < X < \mu + \sigma$;
- «добре», якщо $\mu - \sigma < X < \mu$;
- «задовільно», якщо $\mu - 2\sigma < X < \mu - \sigma$;
- «незадовільно», якщо $X < \mu - 2\sigma$.

²Цим займаються в рамках статистики.

У цьому випадку матимемо

$$\begin{aligned}
 \mathbb{P}_X(X > \mu + \sigma) &= \mathbb{P}_X\left(\frac{X - \mu}{\sigma} > 1\right) \\
 &= 1 - \Phi(1) \approx 0.1587, \\
 \mathbb{P}_X(\mu < X < \mu + \sigma) &= \mathbb{P}_X\left(0 < \frac{X - \mu}{\sigma} < 1\right) \\
 &= \Phi(1) - \Phi(0) \approx 0.3413, \\
 \mathbb{P}_X(\mu - \sigma < X < \mu) &= \mathbb{P}_X\left(-1 < \frac{X - \mu}{\sigma} < 0\right) \\
 &= \Phi(0) - \Phi(-1) \approx 0.3413, \\
 \mathbb{P}_X(\mu - 2\sigma < X < \mu - \sigma) &= \mathbb{P}_X\left(-2 < \frac{X - \mu}{\sigma} < -1\right) \\
 &= \Phi(-1) - \Phi(-2) \approx 0.1359, \\
 \mathbb{P}_X(X < \mu - 2\sigma) &= \mathbb{P}_X\left(\frac{X - \mu}{\sigma} < -2\right) \\
 &= \Phi(-2) \approx 0.0228.
 \end{aligned}$$

Іншими словами, за такого підходу до виставлення оцінок можна гарантувати, що оцінку «відмінно» здобудуть приблизно 16% студентів, оцінки «дуже добре» і «добре» — приблизно по 34%, оцінку «задовільно» — приблизно 14%, а не складуть екзамену приблизно 2% студентів. $\square$

Вправа 8.3.3. Нехай річний рівень опадів (у см) у деякому регіоні має нормальний розподіл зі сподіванням $\mu = 40$ і середньоквадратичним відхиленням $\sigma = 4$. Вважатимемо, що річний рівень опадів не залежить від рівня опадів за попередній рік. Чому дорівнює ймовірність, що протягом найближчих 10 років щороку випадатиме не більше 50 см опадів? $\square$

Розв'язання. Нехай X = «річний рівень опадів», тоді $X \sim N(40, 4^2)$. Ймовірність того, що випаде не більше 50 см опадів, дорівнює

$$\mathbb{P}_X(X \leq 50) = \mathbb{P}_Z\left(Z \leq \frac{50 - 40}{4}\right) = \Phi(2.5) \approx 0.994.$$

Через незалежність відповідних подій можемо підрахувати, що

$$\mathbb{P}(\text{протягом 10 років не буде більше 50 см}) = (\mathbb{P}_X(X \leq 50))^{10} \approx 0.942.$$

$\square$

Вправа 8.3.4. Нехай річні заробітні платні спеціалістів у певній галузі мають нормальний розподіл. Нехай із даних відомо, що чверть спеціалістів заробляють менше 180 000 грн, а чверть заробляють більше 320 000 грн.

Яка частка спеціалістів заробляє менше 200 000 грн? Заробляє від 280 000 до 320 000 грн? $\square$

Розв'язання. У цій задачі ми знаємо, що дані мають нормальний розподіл (наприклад, нам про це говорить форма відповідної гістограми або інформація про аналогічні вибірки спеціалістів за попередні періоди), проте ми не знаємо сподівання й дисперсії. Інформація, якою ми володіємо, дає нам змогу утворити систему з двох рівнянь:

$$\begin{aligned} 0.25 &= \mathbb{P}_X(X \leq 180) = \mathbb{P}_Z\left(Z \leq \frac{180 - \mu}{\sigma}\right) = \Phi\left(\frac{180 - \mu}{\sigma}\right), \\ 0.25 &= \mathbb{P}_X(X \geq 320) = 1 - \mathbb{P}_Z\left(Z \leq \frac{320 - \mu}{\sigma}\right) = 1 - \Phi\left(\frac{320 - \mu}{\sigma}\right). \end{aligned}$$

Можна порахувати, що $\Phi^{-1}(0.25) \approx -0.674$, а також додатково, згідно з Твердженням КЛ7.2.3, що $\Phi^{-1}(0.75) = -\Phi^{-1}(0.25) = 0.674$.

Отже маємо

$$\begin{aligned} \mu - 0.674\sigma &= 180, \\ \mu + 0.674\sigma &= 320. \end{aligned}$$

Розв'язком цієї системи є $\mu = 250$, $\sigma \approx 103.858$. Отже заробітні платні мають нормальний розподіл $X \sim N(250, 103.858^2)$. Звідси

$$\begin{aligned} \mathbb{P}_X(X \leq 200) &= \mathbb{P}_Z\left(Z \leq \frac{200 - 250}{103.858}\right) = \Phi\left(-\frac{50}{103.858}\right) \approx 0.3151, \\ \mathbb{P}_X(280 \leq X \leq 320) &= \mathbb{P}_Z\left(\frac{30}{103.858} \leq Z \leq \frac{70}{103.858}\right) \\ &= \Phi\left(\frac{70}{103.858}\right) - \Phi\left(\frac{30}{103.858}\right) \approx 0.136. \end{aligned} \quad \square$$

Приклад 8.3.5. У фінансовій математиці для оцінки деякої інвестиції використовують поняття *вартості під ризиком* (value at risk, VaR), яку визначають як значення v таке, за якого існує тільки 1% імовірності, що втрати від інвестиції будуть вищі від v .

Нехай X — дохід від деякої інвестиції, який має нормальний розподіл $N(\mu, \sigma^2)$. Тоді VaR такої інвестиції можна визначити з рівняння

$$0.01 = \mathbb{P}_X(-X > v).$$

Нескладно бачити, що $-X \sim N(-\mu, \sigma^2)$, що безпосередньо впливає, наприклад, із розгляду функції розподілу:

$$\mathbb{P}_{-X}(-X \leq x) = 1 - \mathbb{P}_X(X \leq -x) = 1 - \Phi\left(\frac{-x - \mu}{\sigma}\right) = \Phi\left(\frac{x + \mu}{\sigma}\right).$$

Отже

$$0.01 = \mathbb{P}_X\left(\frac{-X + \mu}{\sigma} > \frac{v + \mu}{\sigma}\right) = 1 - \Phi\left(\frac{v + \mu}{\sigma}\right).$$

Тобто

$$\frac{v + \mu}{\sigma} = \Phi^{-1}(0.99).$$

Можна обчислити, що $\Phi^{-1}(0.99) \approx 2.33$, а тому

$$v = 2.33\sigma - \mu.$$

Отже серед усіх інвестицій із нормально розподіленим доходом найменшу вартість під ризиком має інвестиція з найбільшим значенням $\mu - 2.33\sigma$. $\square$

8.4 Функції від випадкових величин

Вправа 8.4.1. Нехай щільність випадкової величини X дорівнює $f_X(x) = e^{-x} \cdot \mathbb{1}\{x > 0\}$. Чому дорівнює щільність випадкової величини $Y = \sqrt{X}$? $\square$

Розв'язання. У нашому випадку носієм є $\text{supp}(X) = \{x : x > 0\}$, до того ж можна бачити, що $\text{supp}(Y) = g(\text{supp}(U)) = \{y : y > 0\}$.

На своєму носії f_X неперервна, а функція $g(x) = \sqrt{x}$ є взаємно однозначна, до того ж $g^{-1}(y) = y^2$ неперервно диференційовна на $\text{supp}(Y)$:

$$\left| \frac{d}{dy} g^{-1}(y) \right| = 2y.$$

Отже за (КЛ7.4.3) маємо

$$f_Y(y) = f_U(g^{-1}(y)) \cdot \left| \frac{d}{dy} g^{-1}(y) \right| \cdot \mathbb{1}\{y \in g(\text{supp}(X))\} = e^{-y^2} \cdot 2y \cdot \mathbb{1}\{y > 0\}.$$

$\square$

Вправа 8.4.2. Нехай випадкова величина $U \sim U((0; 1])$. Який розподіл має випадкова величина $Y = U^2$? $\square$

Розв'язання. На перший погляд може здатися, що функція $g(u) = u^2$ не є взаємно однозначною. Проте в умовах Теорему КЛ7.4.4 стверджується, що функція повинна бути взаємно однозначною на носії випадкової величини. У нашому випадку носієм є $\text{supp}(U) = (0; 1]$, і на цьому інтервалі функція $g(u) = u^2$ взаємно однозначна.

Тому можна сміливо застосувати Теорему КЛ7.4.4. У нашому випадку $g(u) = u^2$, $g^{-1}(y) = \sqrt{y}$, $\text{supp}(Y) = g(\text{supp}(U)) = (0; 1]$. До того ж на цьому носії $g^{-1}(y)$ неперервно диференційовна, і тому маємо

$$\left| \frac{d}{dy} g^{-1}(y) \right| = \frac{1}{2\sqrt{y}}.$$

Тоді за (КЛ7.4.3) маємо

$$f_Y(y) = f_U(g^{-1}(y)) \cdot \left| \frac{d}{dy} g^{-1}(y) \right| \cdot \mathbb{1}_{\{y \in g(\text{supp}(U))\}} = \frac{1}{2\sqrt{y}} \cdot \mathbb{1}_{\{y \in (0; 1]\}}.$$

Також цей результат можна дістати через відповідну функцію розподілу згідно з (КЛ7.4.1):

$$\begin{aligned} F_Y(y) &= \mathbb{P}_Y(Y \leq y) = \mathbb{P}_U(U^2 \leq y) \\ &= \begin{cases} 0, & y < 0 \\ \mathbb{P}_U(-\sqrt{y} \leq U \leq \sqrt{y}), & y \in [0; 1) \\ 1, & y \geq 1 \end{cases} \\ &= \begin{cases} 0, & y < 0 \\ \mathbb{P}_U(0 \leq U \leq \sqrt{y}), & y \in [0; 1) \\ 1, & y \geq 1 \end{cases} \\ &= \begin{cases} 0, & y < 0 \\ \sqrt{y}, & y \in [0; 1) \\ 1, & y \geq 1 \end{cases} \end{aligned}$$

звідки щільність можна дістати як похідну функції розподілу і побачити, що це буде точно такий вираз, як вище. $\square$

Вправа 8.4.3. Нехай частинку вистрілюють із початку координат за прямою траєкторією під випадковим кутом до осі абсцис φ . Нехай Y = «ордината точки перетину прямої траєкторії з прямою $x = 1$ » (Рис. 8.4.1). Нехай $\varphi \sim U((-\pi/2; \pi/2])$.

Яку щільність має випадкова величина Y ? $\square$

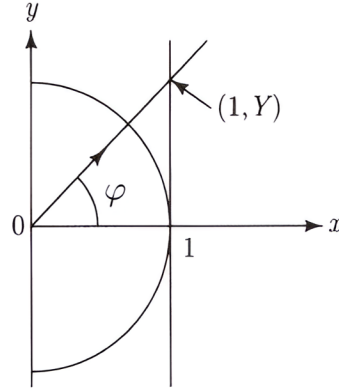


Рис. 8.4.1: Ілюстрація до Вправи 8.4.3

Розв'язання. Найпростіше відповісти на питання за допомогою (КЛ7.4.1), помітивши, що $Y = \operatorname{tg} \varphi$, і на носії $\operatorname{supp}(\varphi)$ тангенс взаємно однозначний:

$$F_Y(y) = \mathbb{P}_Y(Y \leq y) = \mathbb{P}_\varphi(\operatorname{tg} \varphi \leq y) = \mathbb{P}_\varphi(\varphi \leq \operatorname{arctg} y) = F_\varphi(\operatorname{arctg} y).$$

Функція розподілу φ , згідно з (КЛ7.5.5), дорівнює

$$F_\varphi(x) = \begin{cases} 0, & x < -\frac{\pi}{2} \\ \frac{x + \frac{\pi}{2}}{\frac{\pi}{2} + \frac{\pi}{2}}, & -\frac{\pi}{2} \leq x < \frac{\pi}{2} \\ 1, & x \geq \frac{\pi}{2} \end{cases}$$

$$= \begin{cases} 0, & x < -\frac{\pi}{2} \\ \frac{1}{2} + \frac{x}{\pi}, & -\frac{\pi}{2} \leq x < \frac{\pi}{2} \\ 1, & x \geq \frac{\pi}{2} \end{cases}.$$

Оскільки $-\frac{\pi}{2} \leq \operatorname{arctg} y \leq \frac{\pi}{2}$, маємо

$$F_Y(y) = \frac{1}{2} + \frac{\operatorname{arctg} y}{\pi}.$$

Це є функцією розподілу Коші з Вправи 7.2.3. Його щільність дорівнює

$$f_Y(y) = F'_Y(y) = \frac{1}{\pi(1 + y^2)}.$$

Цей же результат можна дістати за формулою заміни змінних. Так, щільність f_φ на цьому носії є неперервною, $g(\varphi) = \operatorname{tg} \varphi$, і на носії $\operatorname{supp}(\varphi)$ це взаємно однозначна функція з оберненою $g^{-1}(y) = \operatorname{arctg} y$, яка неперервно диференційовна на своєму носії $\operatorname{supp}(Y) = g(\operatorname{supp}(\varphi)) = \operatorname{tg}((-\pi/2; \pi/2]) = \mathbb{R}$

з

$$\left| \frac{d}{dy} g^{-1}(y) \right| = \frac{1}{1+y^2} .$$

Отже за (КЛ7.4.3) маємо

$$f_Y(y) = f_\varphi(g^{-1}(y)) \cdot \left| \frac{d}{dy} g^{-1}(y) \right| \cdot \mathbb{1}_{\{y \in g(\text{supp}(\varphi))\}} = \frac{1}{\pi} \cdot \frac{1}{1+y^2} ,$$

що є тотожним попередньому результату. $\square$

Вправа 8.4.4. Нехай $U \sim U((0; 1])$, а

$$Y_1 = \max\{U, 1-U\} , \quad Y_2 = \min\{U, 1-U\} .$$

Який розподіл мають випадкові величини Y_1 та Y_2 ? Чому дорівнює сподівана різниця між ними $\mathbb{E}[Y_1 - Y_2]$? $\square$

Розв'язання. Строго формально ми маємо справу з випадковим вектором $(U, 1-U)^\top$, і теоретично ми повинні були б спочатку дістати розподіл такого вектора. Проте ми вивчатимемо це в наступних лекціях, а наразі можемо застосувати (КЛ7.4.1):

$$\begin{aligned} F_{Y_1}(t) &= \mathbb{P}_{Y_1}(Y_1 \leq t) = \mathbb{P}_{\max\{U, 1-U\}}(\max\{U, 1-U\} \leq t) \\ &= \mathbb{P}_U((U \leq t) \cap (1-U \leq t)) \\ &= \mathbb{P}_U((U \leq t) \cap (1-t \leq U)) . \end{aligned}$$

Оскільки $\text{supp}(U) = (0; 1]$, можемо помітити, що функція розподілу дорівнюватиме 0 для всіх випадків таких, коли дві відповідні нерівності не можуть виконатися одночасно, тобто коли $1-t > t$, звідки випливає $t < \frac{1}{2}$. В інших випадках маємо

$$\begin{aligned} F_{Y_1}(t) &= \mathbb{P}_U(1-t \leq U \leq t) = \begin{cases} 0 , & t < \frac{1}{2} \\ F_U(t) - F_U(1-t) , & \frac{1}{2} \leq t < 1 \\ 1 , & t \geq 1 \end{cases} \\ &= \begin{cases} 0 , & t < \frac{1}{2} \\ 2t - 1 , & \frac{1}{2} \leq t < 1 \\ 1 , & t \geq 1 \end{cases} . \end{aligned}$$

Відповідно, щільність Y_1 дорівнює

$$f_{Y_1}(t) = 2 \cdot \mathbb{1} \left\{ \frac{1}{2} < t < 1 \right\} ,$$

де ми поклали $f_{Y_1}(t) = 0$ в точках недиференційовності F_{Y_1} .

Аналогічно можна показати, що

$$f_{Y_2}(t) = 2 \cdot \mathbb{1} \left\{ 0 < t < \frac{1}{2} \right\} .$$

Для цього потрібно помітити, що мінімум двох значень *більше* деякого t , якщо *обидва* значення більше t .

Отже Y_1 та Y_2 мають рівномірні розподіли на відповідних інтервалах.

Підрахунок сподівання $\mathbb{E}[Y_1 - Y_2]$ можна здійснити за визначенням та властивістю лінійності:

$$\mathbb{E}[Y_1 - Y_2] = \mathbb{E}[Y_1] - \mathbb{E}[Y_2] = \int_{\frac{1}{2}}^1 2t \, dt - \int_0^{\frac{1}{2}} 2t \, dt = 1 - \frac{1}{4} - \frac{1}{4} = \frac{1}{2} .$$

□

8.5 Додаткові вправи

Вправа 8.5.1. Нехай $X \sim U((-a; a])$, $a > 0$. Обчисліть значення параметра a , щоб

- $\mathbb{P}_X(-1 < X < 2) = 0.75$;
- $\mathbb{P}_X(|X| < 1) = \mathbb{P}_X(|X| > 2)$.

Для спрощення вважайте, що відповідні точки не є межами носія.

□

Розв'язання. Функція рівномірного розподілу дорівнює

$$F_X(x) = \begin{cases} 0, & x < -a \\ \frac{x+a}{2a}, & -a \leq x < a \\ 1, & x \geq a \end{cases} .$$

Звідси

$$\mathbb{P}_X(-1 < X < 2) = F_X(2) - F_X(-1) = \frac{2+a}{2a} - \frac{-1+a}{2a} = \frac{3}{2a} = 0.75 ,$$

звідки $a = 2$.

З аналогічних міркувань маємо

$$\mathbb{P}_X(|X| < 1) = \mathbb{P}_X(|X| > 2) \Leftrightarrow F_X(1) - F_X(-1) = 1 - (F_X(2) - F_X(-2)).$$

Оскільки

$$F_X(1) = \frac{1+a}{2a}, \quad F_X(-1) = \frac{a-1}{2a}, \quad F_X(2) = \frac{2+a}{2a}, \quad F_X(-2) = \frac{a-2}{2a},$$

маємо рівняння

$$\frac{1+a}{2a} - \frac{a-1}{2a} = 1 - \frac{2+a}{2a} + \frac{a-2}{2a},$$

звідки $a = 3$. □

Вправа 8.5.2. Нехай на виробництві виготовляють кулькові підшипники, які вважають придатними до використання, якщо їхній діаметр коливається в межах 0.5 ± 0.0006 см.

Перевірка денної партії підшипників виявила, що розподіл діаметрів продукції є нормальний із параметрами $\mu = 0.5007$ см та $\sigma = 0.0005$ см.

Яка є сподівана частка бракованих підшипників? □

Розв'язання. Імовірність того, що підшипник бракований, дорівнює

$$\begin{aligned} p &= \mathbb{P}_X((X > 0.5 + 0.0006) \cup (X < 0.5 - 0.0006)) \\ &= \mathbb{P}_X(X > 0.5006) + \mathbb{P}_X(X < 0.4994) \\ &= 1 - \mathbb{P}_X(X \leq 0.5006) + \mathbb{P}_X(X < 0.4994) \\ &= 1 - \mathbb{P}_Z\left(Z \leq \frac{0.5006 - 0.5007}{0.0005}\right) + \mathbb{P}_Z\left(Z < \frac{0.4994 - 0.5007}{0.0005}\right) \\ &= 1 - \mathbb{P}_Z(Z \leq -0.2) + \mathbb{P}_Z(Z < -2.6) \\ &= 1 - \Phi(-0.2) + \Phi(-2.6) \approx 0.584. \end{aligned}$$

Нехай Y = «загальне число бракованих підшипників». Вочевидь, $Y \sim \text{Віном}(n, p)$, де n — денний обсяг виробленої продукції. Сподіване число бракованих підшипників дорівнює $\mathbb{E}[Y] = np$, а частка, відповідно,

$$\frac{\mathbb{E}[Y]}{n} = p,$$

або приблизно 58.4%. □

Вправа 8.5.3. Нехай тривалість поїздки (у хв.) X від дому до офісу має нормальний розподіл із параметрами $\mu = 40$ та $\sigma = 7$. За скільки потрібно виїхати з дому, щоб імовірність запізнитися на зустріч, призначену на 13:00, була нижчою від 0.05? $\square$

Розв'язання. Імовірність запізнення дорівнює 0.05, якщо

$$0.05 = \mathbb{P}_X(X \geq x)$$

для деякого x , тобто якщо поїздка займе понад x хвилин. У термінах стандартизованої величини маємо

$$0.05 = 1 - \Phi\left(\frac{x - 40}{7}\right).$$

Можна обчислити, що $\Phi^{-1}(0.95) \approx 1.645$. Отже маємо рівняння

$$1.645 = \frac{x - 40}{7},$$

розв'язком якого є $x = 51.515$. Іншими словами, потрібно вийти з дому не пізніше від 12 години 8.485 хв. $\square$

Вправа 8.5.4. Нехай $Z \sim N(0, 1)$, а S — випадковий знак, тобто $S = \pm 1$ з однаковими ймовірностями. Який розподіл має величина SZ ? $\square$

Розв'язання. Розписуючи функцію розподілу випадкової величини SZ за формулою повної ймовірності та враховуючи $Z \perp\!\!\!\perp S$, маємо

$$\begin{aligned} \mathbb{P}_{SZ}(SZ \leq t) &= \mathbb{P}_{SZ|S}(SZ \leq t \mid S = 1) \cdot \mathbb{P}_S(S = 1) \\ &\quad + \mathbb{P}_{SZ|S}(SZ \leq t \mid S = -1) \cdot \mathbb{P}_S(S = -1) \\ &= \frac{1}{2} (\mathbb{P}_{Z|S}(Z \leq t \mid S = 1) + \mathbb{P}_{Z|S}(Z \geq -t \mid S = -1)) \\ &= \frac{1}{2} (\mathbb{P}_Z(Z \leq t) + \mathbb{P}_Z(Z \geq -t)) \\ &= \frac{1}{2} (\Phi(t) + 1 - \Phi(-t)) \\ &= \frac{1}{2} (\Phi(t) + \Phi(t)) = \Phi(t), \end{aligned}$$

де ми використали результат Твердження КЛ7.2.3.

Отже SZ має стандартний нормальний розподіл. $\square$

Вправа 8.5.5. Чому дорівнює розподіл випадкової величини $R = a \sin \theta$, де $\theta \sim U((-\pi/2; \pi/2])$, $a > 0$?

Ця величина виринає в балістичній теорії. Якщо снаряд випускають із початку координат під кутом α до землі зі швидкістю v , то точкою приземлення є $R = \frac{v^2}{g} \sin 2\alpha$, де g — прискорення вільного падіння. $\square$

Розв'язання. Можна помітити, що на носії $\text{supp}(\theta) = (-\pi/2; \pi/2]$ функція $g(\theta) = a \sin \theta$ є взаємно однозначною. Також маємо $g^{-1}(r) = \arcsin \frac{r}{a}$, яка на цьому носії є неперервно диференційовною з похідною

$$\left| \frac{d}{dr} g^{-1}(r) \right| = \frac{1}{a} \frac{1}{\sqrt{1 - \left(\frac{r}{a}\right)^2}}.$$

Також можна бачити, що $\text{supp}(R) = g(\text{supp}(X)) = (-a; a]$.

Отже за (КЛ7.4.3) маємо

$$\begin{aligned} f_R(r) &= f_X(g^{-1}(r)) \cdot \left| \frac{d}{dr} g^{-1}(r) \right| \cdot \mathbb{1}_{\{r \in g(\text{supp}(X))\}} \\ &= \frac{1}{\pi} \cdot \frac{1}{a} \frac{1}{\sqrt{1 - \left(\frac{r}{a}\right)^2}} \cdot \mathbb{1}_{\{-a < r < a\}} \\ &= \frac{1}{\pi \sqrt{a^2 - r^2}} \cdot \mathbb{1}_{\{-a < r < a\}}. \end{aligned}$$

$\square$

9 Окремі застосування неперервних випадкових величин

9.1 Основні теоретичні відомості

Детальний огляд понять, потрібний для розв’язання задач на цьому занятті, наведено в Курсі лекцій, Розд. КЛ8.

Один із важливих результатів, пов’язаних із неперервними випадковими величинами, є спосіб перетворення стандартного рівномірного розподілу у деякий довільний розподіл.

Теорема (Інтегральне перетворення ймовірностей (Probability integral transform), КЛ8.1.1). 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 $U \sim U((0; 1])$;
- (ii) якщо $U \sim U((0; 1])$, то випадкова величина $F_X^\dagger(U)$, де

$$F_X^\dagger(u) = \inf \{x \in \mathbb{R} : F_X(x) \geq u\}, \quad (\text{КЛ8.1.1})$$

має розподіл $\mathbb{P}_X$ ¹.

Теорему КЛ8.1.1 на практиці можна використовувати у двох випадках:

- для перетворення вибірки значень із невідомого розподілу у вибірку з рівномірним розподілом, потреба в чому виникає у деяких статистичних методах;
- для генерування випадкових чисел, розподілених за певним законом. Для цього достатньо згенерувати рівномірно розподілені випадкові числа на інтервалі $(0; 1]$ ² і застосувати до них функцію $F_X^\dagger$, відповідну бажаному розподілу.

¹Якщо 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 $F_X^\dagger(0) = -\infty$.

²У 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`.

Розгляньмо дуже цікавий розподіл та з'ясуймо, як його можна використовувати для оновлення інформації про розподіл деякого невідомого параметра.

Визначення (КЛ8.2.1). Випадкова величина X має *Бета-розподіл* (Beta distribution) із параметрами $a > 0$, $b > 0$, $X \sim \text{Beta}(a, b)$, якщо її щільність дорівнює

$$f(x) = \frac{1}{B(a, b)} x^{a-1} (1-x)^{b-1} \cdot \mathbb{1}\{0 < x < 1\}, \quad (\text{КЛ8.2.1})$$

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 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 $\text{Beta}(1, 1)$ тотожний стандартному рівномірному розподілу.

Завдяки своїй гнучкості, Бета-розподіл часто використовують для моделювання ймовірностей невідомих параметрів. Наприклад, нехай ми маємо деяку монетку, але не знаємо, із якою ймовірністю p вона випаде гербом. Ми воліємо з'ясувати цю ймовірність шляхом підкидання монетки n разів та аналізу випадів. Згідно з *частотним підходом* до інтерпретації ймовірностей, ми можемо порахувати середню частку гербів із-посеред n випадів, яка для великих n буде близька до справжнього значення.

У рамках галузи статистики, що має назву *Беесівського 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 = «число гербів із-посеред n підкидань монетки». Для деякого фіксованого p відомо, що $X | p \sim \text{Binom}(n, p)$. Тоді для пошуку апостеріорного розподілу можна застосувати теорему Беєса КЛ3.3.1, але замість функції ймовірності для p ми вико-

ристаємо щільність, оскільки p є неперервною випадковою величиною:

$$\begin{aligned} f_{p|X}(p | X = k) &= \frac{\mathbb{P}_X(X = k | p) f_p(p)}{\mathbb{P}_X(X = k)} \\ &= \frac{\binom{n}{k} p^k (1-p)^{n-k} \cdot \frac{1}{B(a, b)} p^{a-1} (1-p)^{b-1}}{\mathbb{P}_X(X = k)} \\ &= C \cdot p^k (1-p)^{n-k} \cdot p^{a-1} (1-p)^{b-1} \\ &= C \cdot p^{a+k-1} (1-p)^{b+n-k-1}, \end{aligned}$$

де ми об'єднали всі константи, які не залежать від p , в одну константу C . 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 апіорний 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 апіорним розподілом* (conjugate prior) для біномного розподілу.

Зауваження (КЛ8.2.3). 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. $\square$

Знайдімо сподівання та дисперсію Бета-розподілу.

Твердження (КЛ8.2.4). Якщо $X \sim B(a, b)$, то сподівання $\mathbb{E}[X] = \frac{a}{a+b}$, а дисперсія $\text{Var}(X) = \frac{ab}{(a+b)^2(a+b+1)}$.

Розгляньмо розподіли мінімуму, максимуму та інших подібних функцій.

Визначення (КЛ8.3.1). Нехай $X_1, \dots, X_n$ — деякі випадкові величини. Тоді *порядковими статистиками*³ (order statistics) будуть випадкові ве-

³Статистикою в загальному випадку називають будь-яку функцію від даних у деякій вибірці.

личини $X_{(1)}, \dots, X_{(n)}$ 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 (абсолютно) неперервних величинах, 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:

$$F_{X_{(n)}}(x) = \mathbb{P}_X(X_1 \leq x) \cdot \dots \cdot \mathbb{P}_X(X_n \leq x) = (F_X(x))^n . \quad (\text{КЛ8.3.1})$$

Щільність такого розподілу можна дістати як похідну⁴:

$$f_{X_{(n)}}(x) = n(F_X(x))^{n-1} f_X(x) . \quad (\text{КЛ8.3.2})$$

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:

$$F_{X_{(1)}}(x) = \mathbb{P}_{X_{(1)}}(X_{(1)} \leq x) = 1 - \mathbb{P}_{X_{(1)}}(X_{(1)} \geq x) = 1 - (1 - F_X(x))^n . \quad (\text{КЛ8.3.3})$$

Щільність такого розподілу можна дістати як похідну:

$$f_{X_{(1)}}(x) = n(1 - F_X(x))^{n-1} f_X(x) . \quad (\text{КЛ8.3.4})$$

У загальному випадку маємо таке Твердження.

Твердження (КЛ8.3.2). Нехай $X_1, \dots, X_n$ — незалежні неперервні величини з однаковим розподілом $\mathbb{P}_X$. Тоді функція і щільність розподілу по-

⁴Пам'ятаючи, що в точках недиференційовності функції розподілу можна покласти довільні значення щільності.

рядкової статистики $X_{(k)}$, $k = 1, \dots, n$, мають вид:

$$\begin{aligned} F_{X_{(k)}}(x) &= \sum_{j=k}^n \binom{n}{j} (F_X(x))^j (1 - F_X(x))^{n-j}, \\ f_{X_{(k)}}(x) &= n \binom{n-1}{k-1} f_X(x) (F_X(x))^{k-1} (1 - F_X(x))^{n-k}. \end{aligned} \quad (\text{КЛ8.3.5})$$

Експоненційний розподіл є аналогом геометричного розподілу в 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 λ на одиницю часу.

Визначення (КЛ8.4.1). Випадкова величина X має *експоненційний розподіл* (exponential distribution), що позначають через $X \sim \text{Exp}(\lambda)$, $\lambda > 0$, якщо її щільність розподілу має вид

$$f_X(x) = \lambda e^{-\lambda x} \cdot \mathbb{1}\{x > 0\}. \quad (\text{КЛ8.4.1})$$

□

Функція розподілу має вид

$$F_X(x) = \int_{-\infty}^x \lambda e^{-\lambda t} \cdot \mathbb{1}\{t > 0\} dt = \begin{cases} 0, & x < 0 \\ 1 - e^{-\lambda x}, & x \geq 0 \end{cases}.$$

Сподівання та дисперсія експоненційного розподілу пов'язані з параметром λ .

Твердження (КЛ8.4.3). Якщо $X \sim \text{Exp}(\lambda)$, то $\mathbb{E}[X] = \frac{1}{\lambda}$, а $\text{Var}(X) = \frac{1}{\lambda^2}$.

Експоненційний розподіл, як і геометричний, є розподілом без пам'яті, тобто тривалість t очікування першого успіху, за умови, що протягом часу s він іще не встиг настати, така сама, як тривалість очікування першого успіху «з самого початку»:

$$\mathbb{P}_X(X \geq s + t \mid X \geq s) = \mathbb{P}_X(X \geq t), \quad s, t \geq 0.$$

Більше того, експоненційний розподіл — *єдиний* додатний неперервний розподіл, який не має пам'яті.

Твердження (КЛ8.4.4). Нехай X — додатна неперервна випадкова величина без пам'яті. Тоді $X \sim \text{Exp}(\lambda)$.

Отже експоненційний розподіл є єдиним можливим варіантом для опису явищ, де час очікування появи деякої події не має пам'яті, наприклад:

- атоми радіоактивних ізотопів, зокрема вуглецю-14, урану-235, стронцію-90 у випадкові моменти часу раптово 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 зношуються з часом (принаймні в осяжній перспективі використання приладу), а виходять із ладу раптово і непередбачено.

Експоненціальний розподіл пов'язаний із розподілом Пуассона. Розгляньмо деяку послідовність подій, які стаються у випадкові моменти часу. Якщо послідовність має такі властивості:

- кількість подій N в інтервалі часу довжиною t має розподіл $N \sim \text{Pois}(\lambda t)$;
- поява події в одному часовому інтервалі не залежить від появи події в іншому часовому інтервалі без спільного перетину;

то її називають *процесом Пуассона* (Poisson process) з інтенсивністю настання подій λ . Ми 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 й «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 настанням другої події T_2 та часом настання першої події T_1 . Оскільки за умовами процесу Пуассона ймовірність настання події в одному інтервалі не залежить від імовірності настання події в іншому, розподіл $T_2 - T_1$ нічим не відрізняється від розподілу T_1 , тобто $T_2 - T_1 \sim \text{Exp}(\lambda)$. Відповідні міркування справедливі для різниць у часі між будь-якими сусідніми подіями.

9.2 Інтегральне перетворення ймовірностей

Вправа 9.2.1. Повернімося до розподілу Рейлі, розглянутого у Вправі 7.3.3:

$$F(x) = \left(1 - e^{-\frac{x^2}{2}}\right) \cdot \mathbb{1}\{x > 0\}.$$

Як можна згенерувати на комп'ютері вибірку випадкових чисел із розподілом Рейлі за допомогою інтегрального перетворення ймовірностей? $\square$

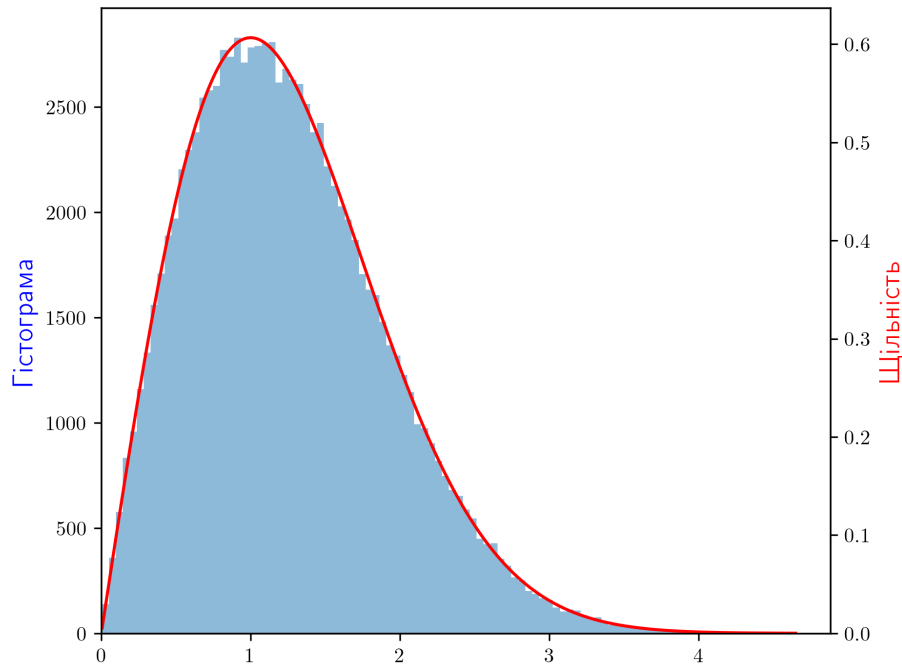


Рис. 9.2.1: Графік щільності розподілу Рейлі та гістограма 10^5 результатів застосування оберненої функції розподілу Рейлі до випадкових чисел для Вправи 9.2.1

Розв'язання. Оскільки функція розподілу Рейлі є строго монотонною на своєму носії $\text{supp}(X) = \{x > 0\}$, функція $F_X^\dagger(u) = \inf \{x \in \mathbb{R} : F_X(x) \geq u\}$ з Теорема КЛ8.1.1 є оберненою функцією:

$$F_X^\dagger(u) = F_X^{-1}(u) = \sqrt{-2 \ln(1 - u)}.$$

Щоб згенерувати вибірку з розподілу Рейлі, потрібно згенерувати на комп'ютері випадкові числа, розподілені рівномірно на $(0; 1)$, та застосувати до них функцію $F_X^{-1}(u) = \sqrt{-2 \ln(1 - u)}$.

Цей факт можна проілюструвати за допомогою відповідної симуляції за методом Монте-Карло. Ми згенеруємо 10^5 випадкових чисел на інтервалі $(0; 1)$, застосуємо до них функцію F_X^{-1} та побудуємо відповідну гістограму (Лістинг 9.2.1). Як можна бачити з Рис. 9.2.1, гістограма дуже близька до аналітичного графіка щільності розподілу.

Лістинг 9.2.1: Код для Вправи 9.2.1

```
import matplotlib.pyplot as plt
from matplotlib import rc
import numpy as np
from numpy.random import seed
5 from scipy.stats import uniform

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

```

seed(100)
u = uniform.rvs(size=10**5)
x = np.sort(np.sqrt(-2*np.log(1 - u)))
15 pdf = x * np.exp(- x**2 / 2) * (x > 0)

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.set_xlim(0)
25 ax1.yaxis.label.set_color("blue")

ax2 = ax1.twinx()

ax2.plot(x, pdf, color="red")
30 ax2.set_ylabel("Щільність", fontsize=15)
ax2.tick_params(axis='y', labelsize=10)
ax2.set_ylim(0)
ax2.set_xlim(0)
ax2.yaxis.label.set_color("red")
35 fig.tight_layout()
plt.savefig("../images/Chapter 9/rayleigh.png", dpi=300)

```

□

Вправа 9.2.2. Нехай $X \sim \text{Exp}(\lambda)$ і $a > 0$. Тоді величина $Y = ae^X$ має розподіл Парето (Pareto distribution). Цей розподіл часто застосовують у різних галузях статистичного моделювання.

Чому дорівнюють її функція та щільність розподілу? Як можна згенерувати на комп'ютері вибірку випадкових чисел із розподілом Парето за допомогою інтегрального перетворення ймовірностей? □

Розв'язання. Функцію розподілу Парето знайдемо за визначенням:

$$F_Y(y) = \mathbb{P}_Y(Y \leq y) = \mathbb{P}_X(ae^X \leq y) .$$

Варто помітити, що для $y \leq 0$ маємо $F_Y(y) = 0$, оскільки експонента завжди додатна. Для $y > 0$ маємо

$$F_Y(y) = \mathbb{P}_X\left(X \leq \ln \frac{y}{a}\right) .$$

Також варто помітити, що якщо $0 < y \leq a$, то логаритм буде недодатний. Проте експоненційна величина X завжди додатна, отже $F_Y(y) = 0$ і в цьому випадку також. Нарешті для $y > a$ маємо

$$F_Y(y) = 1 - e^{-\lambda \ln \frac{y}{a}} .$$

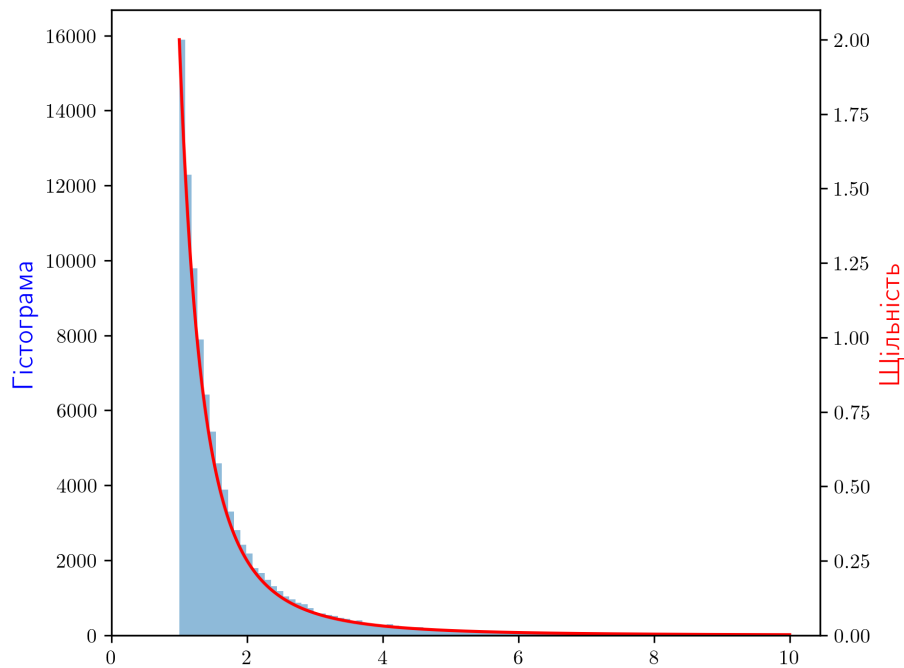


Рис. 9.2.2: Графік щільності розподілу Парето з параметрами $a = 1$, $\lambda = 2$ та гістограма 10^5 результатів застосування оберненої функції розподілу Парето до випадкових чисел для Вправи 9.2.2

Остаточно

$$F_Y(y) = \left(1 - \left(\frac{a}{y}\right)^\lambda\right) \cdot \mathbb{1}\{y > a\}.$$

Щільність розподілу Парето можна дістати шляхом диференціювання відповідної функції розподілу:

$$f_Y(y) = \lambda \left(\frac{a}{y}\right)^{\lambda-1} \cdot \frac{a}{y^2} \cdot \mathbb{1}\{y > a\} = \lambda a^\lambda y^{-(\lambda+1)} \cdot \mathbb{1}\{y > a\}.$$

Оскільки $u \in (0; 1)$, нас цікавить тільки та частина, що відповідає носію $\text{supp}(Y)$. Оскільки функція розподілу строго монотонна для $y > a$, для застосування інтегрального перетворення ймовірностей достатньо дістати обернену функцію розподілу:

$$F_Y^{-1}(u) = a(1 - u)^{-\frac{1}{\lambda}}.$$

Цей факт можна проілюструвати для випадку $a = 1$, $\lambda = 2$ за допомогою відповідної симуляції за методом Монте-Карло. Ми згенеруємо 10^5 випадкових чисел на інтервалі $(0; 1)$, застосуємо до них функцію F_Y^{-1} та побудуємо відповідну гістограму (Лістинг 9.2.2). Як можна бачити з Рис. 9.2.2, гістограма дуже близька до аналітичного графіка щільності розподілу.

Лістинг 9.2.2: Код для Вправи 9.2.2

```
import matplotlib.pyplot as plt
```

```

from matplotlib import rc
import numpy as np
from numpy.random import seed
5 from scipy.stats import uniform

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

a = 1
lmbd = 2

15 seed(100)
u = uniform.rvs(size=10**5)
x = np.sort(a * (1 - u)**(-1/lmbd))
x_max = (100*a)**(1/lmbd)
x = x[x < x_max]

20 pdf = lmbd * a**lmbd * x**(-(lmbd + 1)) * (x > a)

fig, ax1 = plt.subplots()

25 ax1.hist(x, bins=100, alpha=0.5)
ax1.set_ylabel("Гістограма", fontsize=15)
ax1.tick_params(axis='y', labelsize=10)
ax1.set_xlim(0)
ax1.set_ylim(0)
30 ax1.yaxis.label.set_color("blue")

ax2 = ax1.twinx()

ax2.plot(x, pdf, color="red")
35 ax2.set_ylabel("Щільність", fontsize=15)
ax2.tick_params(axis='y', labelsize=10)
ax2.set_xlim(0)
ax2.set_ylim(0)
ax2.yaxis.label.set_color("red")

40 fig.tight_layout()
plt.savefig("../images/Chapter 9/pareto.png", dpi=300)

```

□

Приклад 9.2.3. Нехай невід’ємна випадкова величина X має функцію розподілу F . Покажімо, що

$$\mathbb{E}[X] = \int_0^{\infty} (1 - F(t)) dt . \quad (9.2.1)$$

Спочатку потрібно помітити, що для будь-якого невід’ємного $x \geq 0$

$$x = \int_0^x dt = \int_0^{\infty} \mathbb{1}\{x > t\} dt ,$$

тому

$$X = \int_0^{\infty} \mathbb{1}\{X > t\} dt.$$

Візьмімо сподівання з обох боків та використаємо той факт, що сподівання — це також інтеграл:

$$\begin{aligned} \mathbb{E}[X] &= \mathbb{E}\left[\int_0^{\infty} \mathbb{1}\{X > t\} dt\right] = \int_0^{\infty} \mathbb{E}[\mathbb{1}\{X > t\}] dt \\ &= \int_0^{\infty} \mathbb{P}_X(X > t) dt = \int_0^{\infty} (1 - F(t)) dt. \end{aligned}$$

Проілюструймо цю тотожність для неперервної та строго монотонної функції стандартного логнормального розподілу. Розгляньмо Рис. 9.2.3. Зафарбовану область можна розглядати як інтеграл (9.2.1). З іншого боку цю ж область можна розглядати як інтеграл від оберненої функції розподілу F^{-1} . Для цього потрібно перевернути систему координат і розглянути ординату p як змінну інтегрування:

$$\begin{aligned} \int_0^{\infty} (1 - F(t)) dt &= \int_0^1 F^{-1}(p) dp = \int_{-\infty}^{\infty} F^{-1}(p) \cdot \mathbb{1}\{0 < p < 1\} dp \\ &= \mathbb{E}[F^{-1}(U)] = \mathbb{E}[X], \end{aligned}$$

де ми помітили, що відповідний інтеграл є інтегралом від функції F^{-1} , застосованої до випадкової величини $U \sim U((0; 1])$, а, згідно з Теоремою КЛ8.1.1, величина $F^{-1}(U) \stackrel{d}{=} X$.

□

9.3 Бета-розподіл та елементи Беєсівського виведення

Вправа 9.3.1. Маркетолога цікавить частка покупців p , які придбають новий товар у деякому магазині. Вона маркетологу невідома, і тому він моделює цей параметр як випадкову величину з апріорним стандартним рівномірним розподілом $p \sim U((0; 1])$. За такого підходу до моделювання він не має жодних *апріорних* переконань, переваг чи підозр, а отже всі можливі значення p для нього рівномірні.

По завершенні робочого дня маркетолог має дані, відповідно до яких тільки 6 із 25 покупців придбали новий товар. Припускаючи, що для будь-якого фіксованого p покупці ухвалюють рішення незалежно одне від одного, чому дорівнює апостеріорний розподіл параметра p ?

□

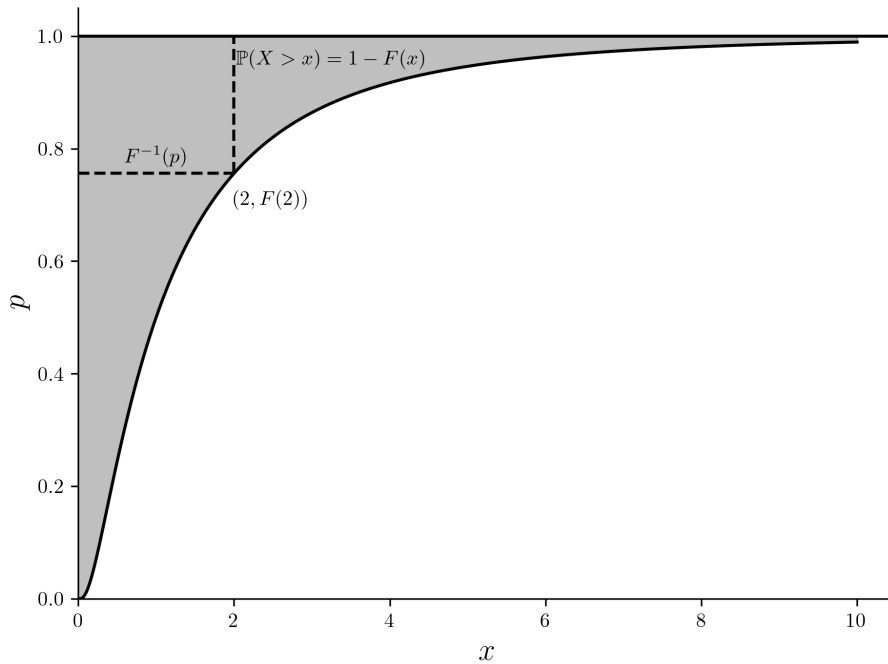


Рис. 9.2.3: Ілюстрація до Прикладу 9.2.3: площа над функцією стандартного логістичного розподілу F та під прямою $p = 1$ (зафарбована область) є одночасно інтегралом (9.2.1) та інтегралом від оберненої функції розподілу F^{-1}

Розв'язання. Оскільки апіорним розподілом p є стандартний рівномірний розподіл, який одночасно є розподілом Beta(1, 1), а кількість покупців X , які придбали товар, для деякого фіксованого p має біномний розподіл із параметрами $X \mid p \sim \text{Binom}(25, p)$, апостеріорний розподіл $p \mid X = 6$ також буде бета-розподілом, але вже з параметрами Beta(1 + 6, 1 + 19) = Beta(7, 20). $\square$

Приклад 9.3.2. У 1977 р. у США йшла судова справа *Castaneda v. Partida* щодо дискримінації американців мексиканського походження під час підбору присяжних. В американській судовій системі суд присяжних налічує 12 осіб. Протягом 2.5 років серед 220 осіб, які були присяжними, тільки 100 мали мексиканське походження, хоча місцеве населення на 79.1% складалося з нащадків мексиканців. Суть позову полягала в тому, що ці дані свідчать про дискримінацію за етнічною ознакою.

Суд порахував, що ймовірність вибрати тільки 100 представників меншини є дуже мала. Справді, якщо припустити, що відбір присяжних є абсолютно випадковий і відбір кожної окремої особи не залежить від інших, то число X присяжних мексиканського походження повинно мати біномний розподіл $X \sim \text{Binom}(220, 0.791)$. У цьому випадку

$$\mathbb{P}_X(X = 100) = \binom{220}{100} 0.791^{100} (1 - 0.791)^{120} \approx 6.287 \cdot 10^{-28}.$$

На підставі цих обчислень суд вирішив, що дискримінація мала місце.

Проте такі обчислення не зовсім коректні, адже вони *припускають*, що дискримінації *не було*, і з урахуванням цього обчислюють імовірність наявних даних. Що нас повинно більше цікавити — це ймовірність того, що дискримінації немає *за умови*, що 100 присяжних із 220 мають мексиканське походження. Нехай p — частка присяжних, які мають мексиканське походження. В ідеалі ми хотіли б, щоб $p = 0.791$. Ми можемо визначити, що має місце дискримінація, якщо p , спостережуване на практиці, менше, наприклад, $0.8 \cdot 0.791 = 0.6328$. Тобто нас цікавить імовірність того, що $p \leq 0.6328$ *за умови*, що $X = 100$.

Покладімо, що апріорний розподіл $p \sim \text{Beta}(\alpha, \beta)$. Тоді апостеріорний розподіл буде бета-розподілом із параметрами $\text{Beta}(\alpha + 100, \beta + 120)$. Як параметри α і β можна, звісно, взяти дві одиниці, що відповідатиме рівномірному розподілу, проте в нашому випадку розподіл очевидно не є рівномірний, адже ми знаємо, що 79.1% населення мають мексиканське походження. Іншими словами, згідно з Твердженням КЛ8.2.4, апіорно

$$\mathbb{E}[p] = \frac{\alpha}{\alpha + \beta} = 0.791 ,$$

тобто $\alpha = 3.785\beta$.

Використаймо цю інформацію для обчислення шуканої ймовірності:

$$\begin{aligned} \mathbb{P}_{p|X}(p \leq 0.6328 \mid X = 100) &= \frac{1}{B(3.785\beta + 100, \beta + 120)} \\ &\times \int_0^{0.6328} p^{3.785\beta + 99} (1 - p)^{\beta + 119} dp . \end{aligned}$$

Можна підрахувати значення цієї ймовірності чисельно для кожного β та дістати відповідний графік (Рис. 9.3.1). Як можна бачити з графіка, відповідна ймовірність менша від 0.5 (тобто дискримінація є радше мало-ймовірною, ніж високоюймовірною) для всіх $\beta \geq 51.5$, а отже для $\alpha \geq 194.9$.

Утім, усі бета-розподіли з відповідними параметрами є нереалістичними. Наприклад, нехай апіорний розподіл p насправді є $\text{Beta}(194.9, 51.5)$. Для такого розподілу апіорно $\mathbb{P}_p(p \leq 0.6328) \approx 3.28 \cdot 10^{-8}$, тобто апіорно ми оцінюємо ймовірність дискримінації як нульову (що дуже сумнівно, і навряд чи розумна людина так може вважати). Для всіх інших апіорних розподілів ця ймовірність іще менша. З іншого боку, якщо хтось апіорно вважає, що $\mathbb{E}[p] = 0.791$ і ймовірність дискримінації хоча б трішки більша від $3.28 \cdot 10^{-8}$, то після спостерігання $X = 100$ він обов'язково вважатиме, що дискримінація має ймовірність, вищу від 0.5.

Тому можна казати, що система підбору присяжних є несправедлива. $\square$

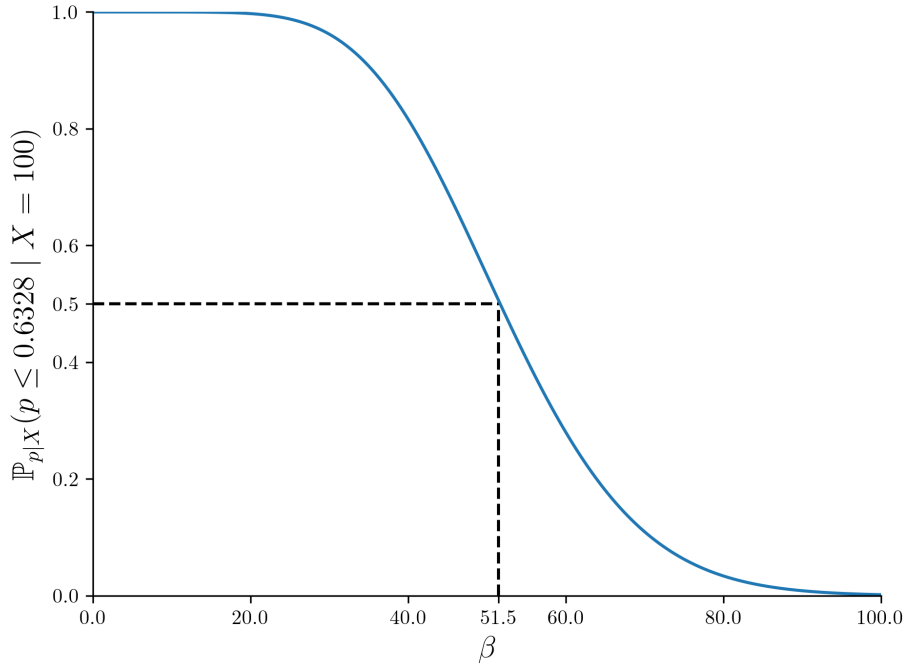


Рис. 9.3.1: Графік імовірності $\mathbb{P}_{p|X}(p \leq 0.6328 | X = 100)$ як функції від параметра β для Прикладу 9.3.2

Вправа 9.3.3. Нехай число дефектів на поверхні оптичного носія має розподіл Пуассона з параметром λ , який може бути або 1.0, або 1.5. Нехай апіорний розподіл параметра λ такий: $\mathbb{P}_\lambda(\lambda = 1.0) = 0.4$, $\mathbb{P}_\lambda(\lambda = 1.5) = 0.6$.

Якщо на оптичному носії знаходять 3 дефекти, який буде апостеріорний розподіл λ ? $\square$

Розв'язання. Згідно з Теоремою Беєса,

$$\mathbb{P}_{\lambda|X}(\lambda = 1.0 | X = 3) = \frac{\mathbb{P}_{X|\lambda}(X = 3 | \lambda = 1.0) \cdot \mathbb{P}_\lambda(\lambda = 1.0)}{\mathbb{P}_X(X = 3)}.$$

За формулою повної ймовірності,

$$\begin{aligned} \mathbb{P}_X(X = 3) &= \mathbb{P}_{X|\lambda}(X = 3 | \lambda = 1.0) \cdot \mathbb{P}_\lambda(\lambda = 1.0) \\ &\quad + \mathbb{P}_{X|\lambda}(X = 3 | \lambda = 1.5) \cdot \mathbb{P}_\lambda(\lambda = 1.5). \end{aligned}$$

Оскільки

$$\mathbb{P}_{X|\lambda}(X = 3 | \lambda) = \frac{e^{-\lambda} \lambda^3}{3!},$$

маємо

$$\mathbb{P}_{X|\lambda}(X = 3 | \lambda = 1.0) \cdot \mathbb{P}_\lambda(\lambda = 1.0) = \frac{e^{-1} 1^3}{3!} \cdot 0.4 \approx 0.0245,$$

$$\mathbb{P}_{X|\lambda}(X = 3 | \lambda = 1.5) \cdot \mathbb{P}_\lambda(\lambda = 1.5) = \frac{e^{-1.5} (1.5)^3}{3!} \cdot 0.6 \approx 0.0753.$$

Тому

$$\mathbb{P}_{\lambda|X}(\lambda = 1.0 \mid X = 3) = \frac{0.0245}{0.0245 + 0.0753} \approx 0.2455 ,$$

і, відповідно,

$$\mathbb{P}_{\lambda|X}(\lambda = 1.5 \mid X = 3) = 1 - \mathbb{P}_{\lambda|X}(\lambda = 1.0 \mid X = 3) = 0.7545 . \quad \square$$

9.4 Порядкові статистики

Вправа 9.4.1. Нехай X та Y — дві незалежні неперервні випадкові величини з функцією розподілу F та щільністю f . Чому дорівнює розподіл $M = \max\{X, Y\}$?

Нехай $X, Y \sim \text{Bern}(0.5)$, $X \perp Y$. Чому дорівнює функція ймовірності $M = \max\{X, Y\}$? Чи є різниця у відповідних формулах і чому? $\square$

Розв'язання. Максимум M буде менший від деякого x , якщо *обидві* випадкові величини менші від x , що через незалежність двох величин дорівнює

$$F_M(x) = \mathbb{P}_M(M \leq x) = \mathbb{P}_X(X \leq x) \mathbb{P}_Y(Y \leq x) = (F(x))^2 ,$$

а відтак щільність дорівнює $f_M(x) = 2F(x)f(x)$.

Розгляньмо тепер $X, Y \sim \text{Bern}(0.5)$. Імовірність того, що $M = 0$, еквівалентна ймовірності, що найбільша з двох величин дорівнює 0, тобто що вони обидві дорівнюють 0:

$$\mathbb{P}_M(M = 0) = \mathbb{P}_X(X = 0) \mathbb{P}_Y(Y = 0) = 0.5 \cdot 0.5 = 0.25 .$$

Тобто $M \sim \text{Bern}(0.75)$. Варто помітити, що в цьому випадку ми не можемо сказати, як у попередньому випадку, що $\mathbb{P}_M(M = x) = 2F(x)\mathbb{P}_X(X = x)$. Зокрема, для $x = 0$ маємо $\mathbb{P}_M(M = 0) = 0.25$, але $F(0) = 0.5$, $\mathbb{P}_X(X = 0) = 0.5$, а отже повинно було б бути $2 \cdot 0.5^2 = 0.5$. Різниця з неперервним випадком пов'язана з тим, що для дискретних випадкових величин допускається ненульова ймовірність однакових значень (наприклад, $X = 0$, $Y = 0$). $\square$

Вправа 9.4.2. Нехай складна система містить n компонент, тривалість роботи X_i , $i = 1, \dots, n$, кожної з яких має експоненційний розподіл $X_i \sim \text{Exp}(\lambda)$, а самі X_i незалежні між собою. Якщо компоненти з'єднано послідовно (тобто система функціонуватиме, тільки якщо всі компоненти працюють), чому дорівнює ймовірність, що система пропрацює понад t одиниць часу? Яка буде відповідь для паралельного з'єднання (тобто система функціонуватиме, якщо принаймні одна компонента працює)? Чому дорівнює сподіваний час роботи систем обох видів? $\square$

Розв'язання. У випадку послідовного з'єднання компонент система пропрацює понад t одиниць часу, якщо *найслабша* компонента пропрацює понад t одиниць часу. Відповідну ймовірність можна обчислити (для $t \geq 0$) з урахуванням незалежності відповідних випадкових величин:

$$\mathbb{P}_{X_{(1)}}(X_{(1)} \geq t) = \mathbb{P}_X(X_1 \geq t) \cdot \dots \cdot \mathbb{P}_X(X_n \geq t) = (1 - F_X(t))^n = e^{-\lambda t \cdot n}.$$

Іншими словами, $X_{(1)} \sim \text{Exp}(\lambda n)$, а отже сподіваний час роботи системи дорівнює $\mathbb{E}[X_{(1)}] = \frac{1}{\lambda n}$.

У випадку паралельного з'єднання система пропрацює понад t одиниць часу, якщо *найвитриваліша* компонента пропрацює понад t одиниць часу. Цю ймовірність можна обчислити (для $t \geq 0$) як

$$\begin{aligned} \mathbb{P}_{X_{(n)}}(X_{(n)} \geq t) &= 1 - \mathbb{P}_{X_{(n)}}(X_{(n)} \leq t) = 1 - \mathbb{P}_X(X_1 \leq t) \cdot \dots \cdot \mathbb{P}_X(X_n \leq t) \\ &= 1 - (F_X(t))^n = 1 - (1 - e^{-\lambda t})^n. \end{aligned}$$

Отже

$$F_{X_{(n)}}(t) = (1 - e^{-\lambda t})^n \cdot \mathbb{1}\{t \geq 0\}.$$

Для підрахунку сподівання тривалості роботи системи знайдімо щільність $X_{(n)}$ як похідну відповідної функції розподілу:

$$f_{X_{(n)}}(t) = F'_{X_{(n)}}(t) = n(1 - e^{-\lambda t})^{n-1} \lambda e^{-\lambda t} \cdot \mathbb{1}\{t > 0\}.$$

Сподівання цієї випадкової величини дорівнює

$$\mathbb{E}[X_{(n)}] = \int_0^\infty t \cdot n(1 - e^{-\lambda t})^{n-1} \lambda e^{-\lambda t} dt,$$

проте цей інтеграл не так просто обчислити.

Скористаймося натомість (9.2.1):

$$\begin{aligned} \mathbb{E}[X_{(n)}] &= \int_0^\infty \left(1 - (1 - e^{-\lambda t})^n\right) dt = \int_1^0 (1 - u^n) d\left(-\frac{1}{\lambda} \ln(1 - u)\right) \\ &= \frac{1}{\lambda} \int_0^1 \frac{1 - u^n}{1 - u} du \\ &= \frac{1}{\lambda} \int_0^1 (1 + u + u^2 + \dots + u^{n-1}) du \\ &= \frac{1}{\lambda} \left(u + \frac{u^2}{2} + \dots + \frac{u^n}{n}\right) \Big|_0^1 \\ &= \frac{1}{\lambda} \cdot \sum_{i=1}^n \frac{1}{i}. \end{aligned}$$

□

9.5 Експоненційний розподіл та процес Пуассона

Вправа 9.5.1. Тривалість роботи X акумулятора має експоненційний розподіл із параметром $\lambda = 1/3$. Чому дорівнює сподівана тривалість роботи акумулятора та відповідна дисперсія? Чому дорівнює ймовірність того, що акумулятор пропрацює від 2 до 4 одиниць часу? Якщо акумулятор *уже* пропрацював 3 одиниці часу, яка ймовірність, що він пропрацює *додатково* щонайменше 1 одиницю часу? $\square$

Розв'язання. Оскільки $X \sim \text{Exp}(1/3)$, за Твердженням КЛ8.4.3 маємо $\mathbb{E}[X] = 3$, $\text{Var}(X) = 9$. Ймовірність роботи акумулятора можна дістати за допомогою функції розподілу:

$$\mathbb{P}_X(2 \leq X \leq 4) = F_X(4) - F_X(2) = 1 - e^{-4\lambda} - (1 - e^{-2\lambda}) = e^{-\frac{2}{3}} - e^{-\frac{4}{3}} \approx 0.250.$$

Нарешті, оскільки експоненційний розподіл не має пам'яті, інформацію про попередню роботу можна відкинути. Тоді ймовірність, що він пропрацює додаткову одиницю часу, просто дорівнює

$$\mathbb{P}_X(X \geq 1) = e^{-\frac{1}{3}} \approx 0.717.$$

 $\square$

Вправа 9.5.2. Нехай час до падіння невеликого метеорита в пустелі має експоненційний розподіл зі сподіванням 10 днів. Наразі північ. Чому дорівнює ймовірність, що метеорит упаде з 6 ранку до 6 вечора? $\square$

Розв'язання. Згідно з умовами, X = «час до падіння метеорита» має розподіл $X \sim \text{Exp}(0.1)$, оскільки $\mathbb{E}[X] = \frac{1}{\lambda} = 10$. Помітивши, що 6 годин становить 0.25 доби, маємо

$$\mathbb{P}_X(0.25 \leq X \leq 0.75) = F_X(0.75) - F_X(0.25) = e^{-0.25 \cdot 0.1} - e^{-0.75 \cdot 0.1} \approx 0.048.$$

 $\square$

Вправа 9.5.3. Нехай на кол-центр дзвінки надходять відповідно до процесу Пуассона з інтенсивністю 3 дзвінки на хвилину.

Чому дорівнює ймовірність того, що в деякий інтервал тривалістю 2 хвилини не надійде жодного дзвінка?

Чому дорівнює ймовірність того, що перший дзвінок після $t = 0$ надійде пізніше, ніж за 2 хвилини?

Чому дорівнює ймовірність того, що в інтервалі від $t = 0$ до $t = 2$ не надійде жодного дзвінка, а з $t = 2$ по $t = 3$ надійде не більше 4 дзвінків?

Чому дорівнює ймовірність того, що четвертий дзвінок надійде не пізніше, ніж за 30 секунд після третього?

Чому дорівнює ймовірність того, що перший дзвінок після $t = 0$ надійде не пізніше ніж за 20 секунд, а час очікування між першим і другим дзвінком буде більший від 3 хвилин?

Чому дорівнює ймовірність того, що п'ятий дзвінок прийде пізніше, ніж за 2 хвилини? $\square$

Розв'язання. Для спрощення позначень розгляньмо $N(s, t)$ = «число дзвінків, що надійшли в інтервалі від s до t ». За побудовою процесу Пуассона, $N(s, t) \sim \text{Pois}(\lambda(t - s)) = \text{Pois}(3(t - s))$.

Тоді ймовірність, що в деякий інтервал тривалістю 2 хвилини не надійде жодного дзвінка, дорівнює

$$\mathbb{P}_N(N(0, 2) = 0) = \frac{e^{-3 \cdot 2} \cdot (3 \cdot 2)^0}{0!} \approx 0.0025.$$

Для підрахунку ймовірності, що перший дзвінок після $t = 0$ надійде пізніше, ніж за 2 хвилини, згадаймо, що відповідна випадкова величина $T_1 \sim \text{Exp}(3)$, а тому

$$\mathbb{P}_{T_1}(T_1 > 2) = e^{-3 \cdot 2} \approx 0.0025.$$

Ця відповідь ідентична попередній, оскільки це фактично однакові події.

Ймовірність, що в інтервалі від $t = 0$ до $t = 2$ не надійде жодного дзвінка, а з $t = 2$ по $t = 3$ надійде не більше 4 дзвінків, можна обчислити, враховувавши незалежність несумісних інтервалів між собою. Отже шукаємо ймовірність дорівнює

$$\mathbb{P}(A) = \mathbb{P}_N(N(0, 2) = 0) \cdot \mathbb{P}_N(N(2, 3) \leq 4) = e^{-3 \cdot 2} \cdot \sum_{j=0}^4 \frac{e^{-3 \cdot 1} (3 \cdot 1)^j}{j!} \approx 0.0020.$$

Ймовірність, що четвертий дзвінок надійде не пізніше, ніж за 30 секунд після третього, можна обчислити, враховувавши властивість експоненційного розподілу не мати пам'яті:

$$\mathbb{P}_{T_4 - T_3}(T_4 - T_3 \leq 0.5) = \mathbb{P}_{T_1}(T_1 \leq 0.5) = 1 - e^{-3 \cdot 0.5} \approx 0.7769.$$

Ймовірність, що перший дзвінок після $t = 0$ надійде не пізніше ніж за 20 секунд, а час очікування між першим і другим дзвінком буде більший від 3 хвилин, можна обчислити, враховувавши незалежність несумісних інтервалів між собою та властивість експоненційного розподілу не мати

пам'яти:

$$\mathbb{P}(B) = \mathbb{P}_{T_1} \left(T_1 \leq \frac{1}{3} \right) \cdot \mathbb{P}_{T_2-T_1} (T_2 - T_1 > 3) = \left(1 - e^{-3 \cdot \frac{1}{3}} \right) \cdot e^{-3 \cdot 3} \approx 0.00008 .$$

Нарешті ймовірність того, що п'ятий дзвінок прийде пізніше, ніж за 2 хвилини, можна обчислити, помітивши, що ця подія еквівалентна тому, що за 2 хвилини прийшло *щонайбільше 4 дзвінки*:

$$\mathbb{P}_{T_5} (T_5 > 2) = \mathbb{P}_N (N(0, 2) \leq 4) = \sum_{j=0}^4 \frac{e^{-3 \cdot 2} (3 \cdot 2)^j}{j!} \approx 0.2851 .$$

□

Приклад 9.5.4. На пошту незалежно одне від одного приходять клієнти двох типів. Ті, кому потрібно надіслати посилку, утворюють процес Пуассона з інтенсивністю λ_1 , а ті, кому потрібно отримати посилку, — з інтенсивністю λ_2 . Можемо показати, що *всі* клієнти пошти *обидвох* типів утворюють один великий процес Пуассона з інтенсивністю $\lambda_1 + \lambda_2$.

Перевірмо всі умови для виникнення процесу Пуассона, а саме:

- імовірність, що в інтервалі довжиною dt станеться точно одна подія, дорівнює приблизно $(\lambda_1 + \lambda_2) \cdot dt$;
- імовірністю, що в інтервалі довжиною dt станеться більше однієї події, можна знехтувати;
- імовірності появи подій у неперетинних інтервалах є незалежні одна від одної.

Вочевидь, імовірності появи подій у несумісних часових інтервалах незалежні для кожного з двох процесів, але оскільки самі процеси незалежні один від одного, то ця властивість зберігається і для об'єднаного потоку.

Розгляньмо деякий інтервал довжиною dt . Для обох процесів справедливо, що ймовірність появи однієї події дорівнює приблизно $\lambda_1 \cdot dt$ або $\lambda_2 \cdot dt$, залежно від того, із якого процесу сталася подія. Тоді для об'єднаного процесу ймовірність, що з'явилася *одна* подія, дорівнює

$$\begin{aligned} \mathbb{P}_N (N = 1) &= \mathbb{P}_{N|N_1, N_2} (N = 1 \mid N_1 = 1, N_2 = 0) \cdot \mathbb{P}_{N_1} (N_1 = 1) \cdot \mathbb{P}_{N_2} (N_2 = 0) \\ &\quad + \mathbb{P}_{N|N_1, N_2} (N = 1 \mid N_1 = 0, N_2 = 1) \cdot \mathbb{P}_{N_1} (N_1 = 0) \cdot \mathbb{P}_{N_2} (N_2 = 1) \\ &= 1 \cdot \lambda_1 \cdot dt (1 - \lambda_2 \cdot dt) + 1 \cdot (1 - \lambda_1 \cdot dt) \cdot \lambda_2 \cdot dt \\ &= (\lambda_1 + \lambda_2) \cdot dt - 2\lambda_1 \lambda_2 \cdot (dt)^2 \\ &= (\lambda_1 + \lambda_2) \cdot dt , \end{aligned}$$

де ми дозволили собі знехтувати нескінченно малими другого порядку.

Аналогічно впливає, що ймовірністю появи більше однієї події можна знехтувати. Справді, якщо в інтервал довжиною dt надійшло дві події з одного процесу, то цим можна знехтувати, адже кожний із двох процесів має відповідну властивість. Якщо ж в інтервал надійшло по одній події з кожного процесу, то ймовірність цього дорівнює $\lambda_1 \cdot dt \cdot \lambda_2 \cdot dt$, і ми можемо цим знехтувати як нескінченно малою другого порядку.

Нарешті, оскільки поява події в одному часовому інтервалі не залежить від появи події в іншому часовому інтервалі без спільного перетину, що впливає з властивостей двох початкових процесів та їхньої незалежності, можемо зробити висновок, що спільний процес буде процесом Пуассона з інтенсивністю $\lambda_1 + \lambda_2$.

Понад те, можемо обчислити ймовірність того, що деяка конкретна подія буде з першого процесу, наприклад, ймовірність того, що наступний клієнт прийде відправляти посилку. Оскільки ймовірність $\mathbb{P}(A)$ надходження однієї події в інтервал довжиною dt дорівнює $(\lambda_1 + \lambda_2)dt$, а ймовірність $\mathbb{P}(A_1)$ надходження однієї події з першого процесу, відповідно, дорівнює $\lambda_1 \cdot dt$, то шукана ймовірність дорівнює

$$\mathbb{P}(A_1 | A) = \frac{\mathbb{P}(A_1 \cap A)}{\mathbb{P}(A)} = \frac{\mathbb{P}(A_1)}{\mathbb{P}(A)} = \frac{\lambda_1}{\lambda_1 + \lambda_2}.$$

□

Вправа 9.5.5. На СТО надають послуги двох типів, A і B . Надходження клієнтів типу A (яким потрібно замінити колесо) становлять собою процес Пуассона з параметром $\lambda_A = 3$ (на хвилину), а клієнтів типу B (яким потрібно полагодити двигун) — із параметром $\lambda_B = 4$ (на хвилину).

Чому дорівнює функція ймовірності, сподівання та дисперсія загального числа клієнтів, що можуть прибути протягом інтервалу в 3 хвилини?

Якщо протягом 10-хвилинного інтервалу надійшло 10 клієнтів, яка ймовірність того, що точно три з них є типу A ?

Якщо на момент $t = 0$ СТО вільне, чому дорівнює функція ймовірності випадкової величини $Y =$ «число клієнтів типу B до моменту появи першого клієнта типу A »? Чому дорівнює сподіване число таких клієнтів? □

Розв'язання. Клієнти обох типів сукупно являють собою, відповідно до Прикладу 9.5.5, процес Пуассона з інтенсивністю $\lambda = \lambda_A + \lambda_B = 7$. Тоді в часовому інтервалі в 3 хвилини число подій має розподіл $N_3 \sim \text{Pois}(3 \cdot 7) = \text{Pois}(21)$. Отже сподівання та дисперсія загального числа клієнтів дорівнюють $\mathbb{E}[N_3] = \text{Var}(N_3) = 21$.

Що стосується 10 клієнтів за 10 хвилин, то незалежно від того, якої довжини часовий інтервал, ймовірність, що клієнт буде клієнтом типу A ,

дорівнює, відповідно до Прикладу 9.5.5,

$$p = \frac{\lambda_A}{\lambda_A + \lambda_B} = \frac{3}{7}.$$

Отже маємо схему Бернуллі з 10 випробувань, де успіх має ймовірність p , і тому ймовірність того, що точно три клієнти будуть типу A , дорівнює

$$\mathbb{P}(A_3) = \binom{10}{3} \left(\frac{3}{7}\right)^3 \left(\frac{4}{7}\right)^7 \approx 0.188.$$

Число появ клієнтів типу B до появи першого клієнту типу A має геометричний розподіл: $Y \sim \text{Geom}(p)$. Отже $\mathbb{E}[Y] = \frac{1-p}{p} = \frac{4}{3}$. $\square$

9.6 Додаткові вправи

Вправа 9.6.1. Як можна згенерувати на комп'ютері вибірку випадкових чисел з експоненційним розподілом із параметром λ за допомогою інтегрального перетворення ймовірностей? $\square$

Розв'язання. Оскільки функція експоненційного розподілу є строго монотонною на своєму носії $\text{supp}(X) = \{x > 0\}$, маємо, що функція $F_X^\dagger(u) = \inf\{x \in \mathbb{R} : F_X(x) \geq u\}$ з Теорема КЛ8.1.1 є просто оберненою функцією:

$$F_X^\dagger(u) = F_X^{-1}(u) = -\frac{1}{\lambda} \ln(1-u).$$

Понад те, можна помітити, що $1 - U \stackrel{d}{=} U$, а тому $F_X^\dagger(u) = -\frac{1}{\lambda} \ln u$.

Щоб згенерувати вибірку з експоненційного розподілу, потрібно згенерувати на комп'ютері випадкові числа, розподілені рівномірно на $(0; 1)$, та застосувати до них функцію $F_X^\dagger$. $\square$

Вправа 9.6.2. Нехай монетка може випасти гербом з імовірністю p , яка невідома і яку моделюють як випадкову величину з апіорним розподілом $p \sim \text{Beta}(a, b)$. На лекції ми розглянули спосіб дістати апостеріорний розподіл на основі спостережень за послідовністю підкидань монетки і аналізу випадкової величини $X = \text{«кількість випалих гербів»}$.

Нехай ми скористаємось альтернативним підходом і фіксуватимемо значення випадкової величини $G = \text{«кількість чисел до появи першого герба»}$. Очевидно, $G | p \sim \text{Geom}(p)$. Чому дорівнює апостеріорна ймовірність p за умови, що $G = k$, і як це відрізняється від першого підходу? $\square$

Розв'язання. Для пошуку апостеріорного розподілу можна застосувати Теорему Беєса КЛЗ.3.1. Нехай $p \sim \text{Beta}(a, b)$, тоді

$$\begin{aligned} f_{p|G}(p \mid G = k) &= \frac{\mathbb{P}_X(G = k \mid p) f_p(p)}{\mathbb{P}_X(G = k)} \\ &= C \cdot (1 - p)^k p \cdot p^{a-1} (1 - p)^{b-1} = C \cdot p^a (1 - p)^{b+k-1}, \end{aligned}$$

де C — деяка константа, яка потрібна, щоб інтеграл щільності був 1.

Із зовнішнього вигляду відповідної щільності можна зробити висновок, що $p \mid G = k \sim \text{Beta}(a + 1, b + k)$. Іншими словами, якщо апіорний розподіл є Бета-розподілом, а G має геометричний розподіл для деякого фіксованого p , то апостеріорний розподіл параметру p залишається Бета-розподілом. Отже, Бета-розподіл є спряженим апіорним розподілом (conjugate prior) не тільки для біномного, а й для геометричного розподілу.

Нескладно помітити, що насправді апостеріорні розподіли в обох випадках будуть однакові. Справді, нехай два дослідники мають однаковий апіорний розподіл, проте перший дослідник працює з біномним розподілом, а другий — із геометричним. Нехай послідовність випадів монетки така — 2 герби, 3 числа, 1 герб, 1 число, 1 герб, 2 числа, 1 герб.

Тоді перший дослідник може порахувати, що всього сталося 5 успіхів (герби) і 6 невдач (числа). Тому $p \mid X = 5 \sim \text{Beta}(a + 5, b + 6)$. Другий дослідник щоразу фіксує, скільки сталося випадів до появи чергового герба. Наведеній послідовності випадів монетки відповідають такі значення величини G : 0, 0, 3, 1, 2. Після кожного фіксування значення G дослідник оновлює апостеріорний розподіл згідно з вищевказаним правилом. Загалом параметр a збільшиться на 5 (5 ітерацій оновлення), а параметр b збільшиться на 6 (загальна сума значень G). Результат буде той самий — $\text{Beta}(a + 5, b + 6)$. $\square$

Вправа 9.6.3. Нехай дві команди, A та B , повинні взяти участь у змаганнях. Для деякого фіксованого параметра p результати ігор незалежні між собою: команда A може виграти з імовірністю p , а команда B — з імовірністю $q = 1 - p$. Також вважаємо, що протягом аналізованого періоду параметр p не змінювався. Проте сам параметр p невідомий, і його планується оцінити за допомогою Беєсівського виведення. Апіорним розподілом є стандартний рівномірний розподіл $p \sim U((0; 1])$.

Для уточнення розподілу перед початком серії ігор дослідник аналізує історичні дані про зустрічі двох команд. Хронологічно впорядкована послідовність переможців у попередніх 10 іграх така: $AAABVBAABV$. Чи має для формування апостеріорного розподілу параметра p значення, у якому порядку вигравала та чи та команда? Чому дорівнює апостеріорний розподіл параметра p на основі цих історичних даних? Чому дорівнює ймовірність перемоги команди A в новій грі (як сподіване значення відповідного

апостеріорного розподілу)? □

Розв'язання. Апостеріорний розподіл параметра p не залежить від порядку слідування випробувань, оскільки для деякого фіксованого параметра p результати ігор незалежні одна від одної, а тому ймовірність будь-якої перестановки однакового набору результатів мають однакову ймовірність.

Оскільки апіорним розподілом є стандартний рівномірний розподіл, що одночасно є розподілом $\text{Beta}(1, 1)$, а число перемог команди A у серії з 10 ігор для деякого фіксованого p має біномний розподіл із параметрами $\text{Binom}(10, p)$, апостеріорний розподіл також буде бета-розподілом, але вже з параметрами $\text{Beta}(1 + 6, 1 + 4) = \text{Beta}(7, 5)$, оскільки команда A в попередніх іграх виграла 6 разів, а команда B — 4 рази.

Сподіванням p , згідно з Твердженням КЛ8.2.4, є

$$\mathbb{E}[p \mid X = 6] = \frac{7}{7 + 5} = \frac{7}{12},$$

відповідно з такою ймовірністю команда A зможе виграти в новій грі. □

Вправа 9.6.4. Нехай маємо схему Бернуллі, випробування в якій стаються в моменти часу на відстані dt одиниць часу один від одного. Нехай ймовірність успіху в кожному такому випробуванні дорівнює $\lambda \cdot dt$, $\lambda > 0$. Нехай G = «число невдач перед першим успіхом», T = «час перед першим успіхом».

Яке просте рівняння поєднує випадкові величини G та T ? Чому дорівнює функція розподілу T ? Покажіть, що ця функція прямує до функції експоненційного розподілу, коли $dt \rightarrow 0$. □

Розв'язання. Можна бачити, що якщо сталося k невдач, то це зайняло $k \cdot dt$ часу. Отже $T = Gdt$.

Знайдімо функцію розподілу за визначенням:

$$F_T(t) = \mathbb{P}_T(T \leq t) = \mathbb{P}_G\left(G \leq \frac{t}{dt}\right).$$

Для $t \leq 0$ відразу бачимо, що $F_T(t) = 0$. Розгляньмо $t > 0$.

Для геометричного розподілу $\text{Geom}(p)$ справедливо, що

$$\mathbb{P}_G(G \geq k) = (1 - p)^k.$$

У нашому випадку маємо

$$F_T(t) = 1 - \mathbb{P}_G\left(G > \frac{t}{dt}\right) = 1 - \mathbb{P}_G\left(G \geq \left\lfloor \frac{t}{dt} \right\rfloor\right) = 1 - (1 - \lambda \cdot dt)^{\lfloor \frac{t}{dt} \rfloor},$$

де $\lfloor r \rfloor$ — ціла частина числа r .

Спрямувавши $dt \rightarrow 0$, маємо, за другою важливою границею,

$$\lim_{dt \rightarrow 0} F_T(t) = 1 - \lim_{dt \rightarrow 0} (1 - \lambda \cdot dt)^{\lfloor \frac{t}{dt} \rfloor} = 1 - e^{-\lambda t},$$

що відповідає функції експоненційного розподілу з параметром λ . $\square$

Приклад 9.6.5. На вузол у мережі надходить пакет даних. З імовірністю p це може бути локальний пакет, тобто який справді прямував на цей вузол, а з імовірністю $(1-p)$ — пакет, який випадково збився зі свого шляху. Пакети надходять згідно з процесом Пуассона з інтенсивністю λ , і кожний пакет може бути локальним незалежно від інших пакетів і від часу надходження.

Як виявляється, *тільки* локальні пакети утворюють свій власний процес Пуассона з інтенсивністю λp . Перевірмо спочатку, що виконуються всі умови для того, щоб можна було вважати, що число пакетів має розподіл Пуассона, а саме:

- імовірність, що в інтервалі довжиною dt станеться точно одна подія, дорівнює приблизно $\lambda \cdot dt$;
- імовірністю, що в інтервалі довжиною dt станеться більше однієї події, можна знехтувати;
- імовірності появи подій у неперетинних інтервалах є незалежні одна від одної.

Справді, імовірність того, що в інтервалі довжиною dt надійде один локальний пакет, дорівнює

$$\begin{aligned} \mathbb{P}((\text{пакет}) \cap (\text{локальний})) &= \mathbb{P}(\text{пакет} \mid \text{локальний}) \cdot \mathbb{P}(\text{локальний}) \\ &= \lambda \cdot dt \cdot p. \end{aligned}$$

Імовірність того, що в інтервалі довжиною dt буде більше одного *локального* пакета, можна відкинути, адже вона ще нижча, ніж імовірність появи більше одного пакета, *у тому числі* локального. Нарешті, очевидно, що ймовірності появи локальних пакетів у несумісних інтервалах незалежні, оскільки вони незалежні для всіх пакетів.

Додатково до того факту, що число появи локальних пакетів в інтервалі довжиною dt має розподіл $\text{Pois}(\lambda p \cdot dt)$, маємо, що поява події в одному часовому інтервалі не залежить від появи події в іншому часовому інтервалі без спільного перетину. Це впливає з того факту, що всі пакети являють собою процес Пуассона, для якого ця властивість виконується.

Отже маємо висновок, що локальні пакети утворюють власний процес Пуассона з інтенсивністю λp . $\square$

10 Випадкові вектори та незалежні випадкові величини. Частина 1

10.1 Основні теоретичні відомості

Детальний огляд понять, потрібний для розв'язання задач на цьому занятті, наведено в Курсі лекцій, Розд. КЛ9.

Розгляньмо декілька базових визначень для багатовимірних просторів.

Визначення (КЛ9.1.1). Нехай маємо два вимірні простори, $(X, \mathcal{X})$ і $(Y, \mathcal{Y})$. Добутком цих просторів (product space) називають простір

$$X \times Y = \{\omega = (\omega_1, \omega_2) : \omega_1 \in X, \omega_2 \in Y\}. \quad (\text{КЛ9.1.1})$$

□

Теорема (Теорема про добуток мір (Product measure theorem), КЛ9.1.6). Нехай $(X, \mathcal{X}, \mu)$ і $(Y, \mathcal{Y}, \nu)$ — два мірні простори з σ -скінченними мірами. На σ -алгебрі $\mathcal{X} \times \mathcal{Y}$ можна визначити міру π таку, що для будь-якої вимірної множини $E \in \mathcal{X} \times \mathcal{Y}$

$$\pi(E) = \int \nu(\{y : (x, y) \in E\}) d\mu = \int \mu(\{x : (x, y) \in E\}) d\nu. \quad (\text{КЛ9.1.2})$$

Тоді така міра є єдиною σ -скінченною мірою такою, що $\pi(A \times B) = \mu(A)\nu(B)$ для $A \in \mathcal{X}$, $B \in \mathcal{Y}$.

Визначення (КЛ9.1.7). Міру π з Теорема КЛ9.1.6 називають *добутком мір* (product measure) і позначають через $\pi = \mu \times \nu$. □

Теорема (Теорема Фубіні (Fubini Theorem), КЛ9.1.8). Нехай на добутку просторів $(X \times Y, \mathcal{X} \times \mathcal{Y}, \mu \times \nu \equiv \pi)$ з σ -скінченним добутком мір визначено випадкову величину $f : X \times Y \rightarrow \mathbb{R}$. Тоді:

- якщо $f \geq 0$ майже напевно¹, 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 справедливе твердження

$$\int_{X \times Y} f(x, y) d\pi = \int_X \left(\int_Y f(x, y) d\nu \right) d\mu = \int_Y \left(\int_X f(x, y) d\mu \right) d\nu. \quad (\text{КЛ9.1.3})$$

Інтеграл зліва називають *подвійним* (double), а два інші інтеграли — *повторними* (iterated);

- якщо f інтегровна відносно π (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 $\nu(Y \setminus B_0) = 0$ ², і (КЛ9.1.3) також виконується.

Відповідні міркування легко можна поширити на добутки просторів вищих порядків. Серед іншого, можна бачити, що міру Лебега в $\mathbb{R}^k$, λ_k , також можна розглядати як добуток одновимірних мір Лебега: $\lambda_k = \lambda \times \dots \times \lambda$.

Визначення (КЛ9.2.1). 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

$$\begin{aligned} \mathbb{P}_{\mathbf{X}}(A) &\equiv \mathbb{P}_{\mathbf{X}}((X_1, \dots, X_k)^\top \in A) \\ &= \mathbb{P}(\{\omega : (X_1(\omega), \dots, X_k(\omega))^\top \in A\}) , \quad A \in \mathcal{B}^k. \end{aligned} \quad (\text{КЛ9.2.1})$$

Такий розподіл називають *спільним розподілом* (joint distribution) випадкового вектора. $\square$

Як можна бачити, Визначення КЛ9.2.1 принципово нічим не відрізняється від Визначення КЛ4.1.6. Так само мало відрізняється і визначення функції розподілу.

Визначення (КЛ9.2.2). Нехай $\mathbf{X} = (X_1, \dots, X_k)^\top : (\Omega, \mathcal{A}, \mathbb{P}) \rightarrow (\mathbb{R}^k, \mathcal{B}^k)$ зі спільним розподілом $\mathbb{P}_{\mathbf{X}}$. Функція відповідного роз-

¹Цю частину, для невід'ємних функцій, називають також *теоремою Тонеллі* на честь італійського математика Леоніди Тонеллі (Leonida Tonelli, 1885–1946), який довів її у 1909 р.

²Тобто відповідні інтеграли можуть не існувати або бути нескінченними на деякій множині міри нуль.

поділу має вид

$$\begin{aligned} F_{\mathbf{X}}(x_1, \dots, x_k) &= \mathbb{P}_{\mathbf{X}}(X_1 \leq x_1, \dots, X_k \leq x_k) \\ &= \mathbb{P}_{\mathbf{X}}(\omega : (X_1(\omega) \leq x_1) \cap \dots \cap (X_k(\omega) \leq x_k)) \quad (\text{КЛ9.2.2}) \\ &= \mathbb{P}_{\mathbf{X}}(S_{\mathbf{x}}) , \end{aligned}$$

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 $\mathbf{x} = (x_1, \dots, x_k)^{\top} \in \mathbb{R}^k$.

Таку функцію розподілу називають *спільною функцією розподілу* (joint distribution function). $\square$

Як і функція розподілу деякої випадкової величини (КЛ6.1.1), спільна функція розподілу має три важливі властивості, відповідниками яких є властивості з Твердження КЛ6.1.2. Зокрема, спільна функція розподілу також є нормованою, їй властивий аналог неперервності справа, і для неї справедливе узагальнення неспадності. Формальний розгляд цих властивостей можна знайти в Розд. КЛ9.6.3.

За аналогією з одновимірним випадком можемо також розглянути поняття щільності випадкового вектора.

Визначення (КЛ9.2.3). Якщо розподіл деякого випадкового вектора $\mathbf{X} : (\Omega, \mathcal{A}, \mathbb{P}) \rightarrow (\mathbb{R}^k, \mathcal{B}^k)$ абсолютно неперервний відносно міри Лебега λ_k , то *спільною щільністю розподілу* (joint probability density function) є функція $f_{\mathbf{X}}$ така, що

$$\mathbb{P}_{\mathbf{X}}(A) = \int_A f_{\mathbf{X}} d\lambda_k, \quad \forall A \in \mathcal{B}^k, \quad (\text{КЛ9.2.3})$$

$\square$

За аналогією з (КЛ6.3.2) маємо властивість

$$F_{\mathbf{X}}(x_1, \dots, x_k) = \int_{-\infty}^{x_k} \dots \int_{-\infty}^{x_1} f_{\mathbf{X}}(t_1, \dots, t_k) dt_1 \dots dt_k. \quad (\text{КЛ9.2.4})$$

Відповідно, маємо узагальнення властивості (КЛ6.3.3)³:

$$f_{\mathbf{X}}(x_1, \dots, x_k) = \frac{\partial^k}{\partial x_1 \dots \partial x_k} F_{\mathbf{X}}(x_1, \dots, x_k). \quad (\text{КЛ9.2.5})$$

Визначення (КЛ9.2.4). 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

³Для абсолютно неперервних випадкових величин функція розподілу абсолютно неперервна, тому порядок диференціювання не має значення.

спільною функцією ймовірності (joint probability mass function) є

$$p_{\mathbf{X}}(x_1, \dots, x_k) = \mathbb{P}_{\mathbf{X}}(X_1 = x_1, \dots, X_k = x_k) \\ \times \mathbb{1} \{ (x_1, \dots, x_k)^{\top} \in \text{supp}(\mathbf{X}) \} . \quad (\text{КЛ9.2.6})$$

□

Варто зазначити, що випадковий вектор може містити як дискретні, так і неперервні випадкові величини як свої координати. Наприклад, нехай $\mathbf{X} = (X_1, X_2, X_3)^{\top}$, де X_1 є дискретною випадковою величиною з деяким не більш ніж зліченим носієм $\text{supp}(X_1)$, а X_2 і X_3 — неперервними величинами. Скажімо, нас може цікавити аналіз ціни X_3 квартири, яка має X_1 кімнат і загальну площу X_2 . У цьому випадку ми працюємо з добутком просторів $\text{supp}(X_1) \times \mathbb{R} \times \mathbb{R}$, на якому мірою є $\# \times \lambda \times \lambda$ — добуток лічної міри і двох мір Лебега. Якщо розподіл $\mathbb{P}_{\mathbf{X}}$ абсолютно неперервний відносно такого добутку мір, то також існуватиме щільність $f_{\mathbf{X}}$, із тим розумінням, що для всіх вимірних множин A

$$\mathbb{P}_{\mathbf{X}}(A) = \int_A f_{\mathbf{X}} \#(dx_1) \lambda(dx_2) \lambda(dx_3) \\ = \sum_{x_1 \in A \cap \text{supp}(X_1)} \iint_{x_2, x_3: (x_1, x_2, x_3)^{\top} \in A} f_{\mathbf{X}}(x_1, x_2, x_3) dx_2 dx_3 .$$

Як і в одновимірному випадку, спільна щільність⁴ має такі властивості.

Твердження (КЛ9.2.5). Будь-яка функція $f : \mathbb{R}^k \rightarrow [0; \infty)$ є спільною щільністю для деякого розподілу тоді й тільки тоді, коли

$$\int_{\mathbb{R}^k} f d\lambda_k = 1 .$$

Як і в одновимірному випадку, спільну щільність визначено з точністю до множини міри 0.

Як ми бачили в Твердженні КЛ7.3.6, 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 (КЛ9.2.7). 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).

Розподіл $\mathbb{P}_{j \in I}$ випадкового вектора $\mathbf{X}_{j \in I} \equiv (X_j)_{j \in I}$, $j \in I$, складеного з координат деякого випадкового вектора $\mathbf{X} = (X_1, \dots, X_k)^{\top}$, індекси яких

⁴ Аналогічні властивості справедливі і для функції ймовірності, оскільки функція ймовірності також є щільністю — відносно лічної міри.

належать множині I , також називають маржинальним розподілом, який одночасно є спільним розподілом для $\mathbf{X}_{j \in I}$. $\square$

Маржинальним розподілам відповідають *маржинальні функції розподілу* (marginal distribution function):

$$\begin{aligned} F_{X_j}(x) &= \mathbb{P}_{X_j}(X_j \leq x) = \mathbb{P}_{\mathbf{X}}(g_j(X_1, \dots, X_k) \leq x) \\ &= \mathbb{P}_{\mathbf{X}}(X_1 \in \mathbb{R}, \dots, X_{j-1} \in \mathbb{R}, X_j \leq x, X_{j+1} \in \mathbb{R}, \dots, X_k \in \mathbb{R}) \\ &= \mathbb{P}_{\mathbf{X}} \left(\lim_{\min\{x_1, \dots, x_{j-1}, x_{j+1}, \dots, x_k\} \rightarrow \infty} S_{(x_1, \dots, x_{j-1}, x_j, x_{j+1}, \dots, x_k)^\top} \right) \\ &= \lim_{\substack{x_1 \rightarrow \infty \\ \dots \\ x_{j-1} \rightarrow \infty \\ x_{j+1} \rightarrow \infty \\ \dots \\ x_k \rightarrow \infty}} F_{\mathbf{X}}(x_1, \dots, x_{j-1}, x_j, x_{j+1}, \dots, x_k) . \end{aligned} \quad (\text{КЛ9.2.8})$$

Також можна розглянути поняття маржинальної щільності.

Твердження (КЛ9.2.8). Якщо $\mathbb{P}_{\mathbf{X}}$ має щільність $f_{\mathbf{X}}$, то розподіл $\mathbb{P}_{X_j}$ має *маржинальну щільність розподілу* (marginal probability density function)

$$f_j(x) = \int_{\mathbb{R}^{k-1}} f_{\mathbf{X}}(x_1, \dots, x_{j-1}, x, x_{j+1}, \dots, x_k) d\lambda_{k-1} , \quad (\text{КЛ9.2.9})$$

де інтегрування відбувається за змінними x_i , $i \neq j$.

Як наслідок із цього твердження, можна розглянути спільну щільність $f_{j \in I}$ для випадкового вектора $\mathbf{X}_{j \in I} \equiv (X_j)_{j \in I}$, $j \in I$, 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 , $i \in \{1, \dots, k\} \setminus I$.

Аналогічно сподіванню випадкової величини, уведеному в Розд. КЛ6.4, можна розглянути поняття сподівання випадкового вектора.

Визначення (КЛ9.3.1). *Сподіванням* (expectation) випадкового вектора $\mathbf{X} = (X_1, \dots, X_k)^\top : (\Omega, \mathcal{A}, \mathbb{P}) \rightarrow \mathbb{R}^k$ є вектор зі сподівань його координат:

$$\mathbb{E}[\mathbf{X}] = \begin{pmatrix} \mathbb{E}[X_1] \\ \vdots \\ \mathbb{E}[X_k] \end{pmatrix} . \quad (\text{КЛ9.3.1})$$

$\square$

За властивістю лінійності (одновимірного) сподівання маємо, для деякої матриці $\mathbf{A} \in \mathbb{R}^n \times \mathbb{R}^k$ та деякого вектора $\mathbf{b} \in \mathbb{R}^n$:

$$\mathbb{E}[\mathbf{AX} + \mathbf{b}] = \mathbf{A}\mathbb{E}[\mathbf{X}] + \mathbf{b}. \quad (\text{КЛ9.3.2})$$

Справедливим є аналог Теорема КЛ6.4.2.

Теорема 10.1.1 (КЛ9.3.2). Нехай $\mathbf{X} = (X_1, \dots, X_k)^\top : (\Omega, \mathcal{A}, \mathbb{P}) \rightarrow \mathbb{R}^k$ — випадковий вектор, який породжує розподіл $\mathbb{P}_{\mathbf{X}}$, що абсолютно неперервний відносно міри μ 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

$$\begin{aligned} \int g(X_1(\omega), \dots, X_k(\omega)) d\mathbb{P} &= \int g(x_1, \dots, x_k) d\mathbb{P}_{\mathbf{X}} \\ &= \int g(x_1, \dots, x_k) \cdot f_{\mathbf{X}}(x_1, \dots, x_k) d\mu, \end{aligned}$$

у тому розумінні, що якщо існує один інтеграл, то існують усі інші, і вони дорівнюють один одному.

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 (КЛ9.3.4). Коваріацією (covariance) величин X і Y є

$$\text{Cov}(X, Y) = \mathbb{E}[(X - \mathbb{E}[X])(Y - \mathbb{E}[Y])] . \quad (\text{КЛ9.3.3})$$

□

За аналогією з дисперсією, коваріацію часто позначають як σ_{XY} .

За аналогією з одновимірним випадком можна довести деякі властивості коваріацій.

Твердження (КЛ9.3.5). Коваріація $\text{Cov}(X, Y)$ двох випадкових величин X та Y має такі властивості:

- $\text{Cov}(X, Y) = \text{Cov}(Y, X)$;
- $\text{Cov}(X, X) = \text{Var}(X)$;
- $\text{Cov}(X, Y) = \mathbb{E}[XY] - \mathbb{E}[X]\mathbb{E}[Y]$;
- $\text{Cov}(X, a) = 0, a \in \mathbb{R}$;

- $\text{Cov}(aX + bY, cV + dW) = ac\text{Cov}(X, V) + ad\text{Cov}(X, W) + bc\text{Cov}(Y, V) + bd\text{Cov}(Y, W)$;
- $\text{Var}(X + Y) = \text{Var}(X) + \text{Var}(Y) + 2\text{Cov}(X, Y)$.

З останньої властивості Твердження КЛ9.3.5 випливає

$$\text{Var}(X - Y) = \text{Var}(X) + \text{Var}(Y) - 2\text{Cov}(X, Y) .$$

Частою помилкою є використання знаку «мінус» перед $\text{Var}(Y)$, що позбавлене сенсу, адже $\text{Var}(-Y) = (-1)^2\text{Var}(Y) = \text{Var}(Y)$.

Відповідну властивість можна узагальнити на випадок k величин:

$$\text{Var}\left(\sum_{i=1}^k X_i\right) = \sum_{i=1}^k \text{Var}(X_i) + 2 \sum_{i < j} \text{Cov}(X_i, X_j) . \quad (\text{КЛ9.3.4})$$

За аналогією з середньоквадратичним відхиленням можна розглянути поняття коефіцієнта кореляції.

Визначення (КЛ9.3.6). *Коефіцієнтом кореляції* (correlation) двох випадкових величин X та Y є

$$\text{Corr}(X, Y) = \frac{\text{Cov}(X, Y)}{\sqrt{\text{Var}(X) \text{Var}(Y)}} . \quad (\text{КЛ9.3.5})$$

□

Коефіцієнт кореляції часто позначають як ρ_{XY} .

Твердження (). Коефіцієнт кореляції (КЛ9.3.5) між довільними випадковими величинами X та Y лежить на проміжку $[-1; 1]$.

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 $\text{Corr}(X, Y)$, ближчі за модулем до 1, свідчать про сильнішу лінійну залежність.

Визначення (КЛ9.3.9). Нехай $\mathbf{X} = (X_1, \dots, X_k)^\top : (\Omega, \mathcal{A}, \mathbb{P}) \rightarrow \mathbb{R}^k$ — випадковий вектор. Його *матрицею коваріацій* (covariance matrix) є матриця

$$\begin{aligned} \mathbf{Cov}(\mathbf{X}) &= \mathbb{E}[(\mathbf{X} - \mathbb{E}[\mathbf{X}])(\mathbf{X} - \mathbb{E}[\mathbf{X}])^\top] \\ &= \mathbb{E} \left[\begin{pmatrix} X_1 - \mathbb{E}[X_1] \\ \vdots \\ X_k - \mathbb{E}[X_k] \end{pmatrix} \cdot \begin{pmatrix} X_1 - \mathbb{E}[X_1], \dots, X_k - \mathbb{E}[X_k] \end{pmatrix} \right] \\ &= \begin{pmatrix} \text{Var}(X_1) & \text{Cov}(X_1, X_2) & \dots & \text{Cov}(X_1, X_k) \\ \text{Cov}(X_2, X_1) & \text{Var}(X_2) & \dots & \text{Cov}(X_2, X_k) \\ \vdots & \vdots & \ddots & \vdots \\ \text{Cov}(X_k, X_1) & \text{Cov}(X_k, X_2) & \dots & \text{Var}(X_k) \end{pmatrix}. \end{aligned} \quad (\text{КЛ9.3.6})$$

□

Для матриці коваріацій справедливі такі властивості.

Твердження (КЛ9.3.10). Матриця коваріацій $\mathbf{Cov}(\mathbf{X}) \equiv \Sigma$ деякого випадкового вектора $\mathbf{X}$ має такі властивості:

- (i) вона симетрична: $\Sigma^\top = \Sigma$;
- (ii) $\mathbf{Cov}(\mathbf{A}\mathbf{X} + \mathbf{b}) = \mathbf{A}\Sigma\mathbf{A}^\top$ ⁵;
- (iii) вона невід'ємно визначена⁶:

$$\mathbf{a}^\top \Sigma \mathbf{a} \geq 0, \quad \mathbf{a} \in \mathbb{R}^k.$$

Розгляньмо одне з найважливіших визначень теорії ймовірностей — визначення незалежних випадкових величин.

Визначення (КЛ9.4.1). Випадкові величини $X_1, \dots, X_k$ *незалежні* (independent), якщо

$$\begin{aligned} \mathbb{P}_{\mathbf{X}}(X_1 \in H_1, \dots, X_k \in H_k) &= \mathbb{P}_{X_1}(X_1 \in H_1) \cdot \dots \cdot \mathbb{P}_{X_k}(X_k \in H_k), \\ H_i &\in \mathcal{B}, \quad i = 1, \dots, k. \end{aligned} \quad (\text{КЛ9.4.1})$$

⁵Порівняйте з аналогічною властивістю для дисперсій: $\text{Var}(aX + b) = a^2 \text{Var}(X)$.

⁶Порівняйте з аналогічною властивістю для дисперсій: $\text{Var}(X) \geq 0$.

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

$$\mathbb{P}_{\mathbf{X}}(\mathbf{X}_1 \in H_1, \dots, \mathbf{X}_k \in H_k) = \mathbb{P}_{\mathbf{X}_1}(\mathbf{X}_1 \in H_1) \cdot \dots \cdot \mathbb{P}_{\mathbf{X}_k}(\mathbf{X}_k \in H_k) ,$$

$$H_i \in \mathcal{B}^{d_i}, i = 1, \dots, k .$$
(КЛ9.4.2)

Нескінченна множина випадкових величин чи векторів є незалежна, якщо незалежною є будь-яка скінченна підмножина цих величин чи векторів. $\square$

Визначення КЛ9.4.1 незалежності випадкових величин не має значної користи, оскільки перевірка (КЛ9.4.1) на практиці нетривіальна. Натомість із цієї характеристикації випливають декілька ключових результатів.

Теорема (КЛ9.4.3). Випадкові величини $X_1, \dots, X_k$ незалежні тоді й тільки тоді, коли

$$\mathbb{P}_{\mathbf{X}}(X_1 \leq x_1, \dots, X_k \leq x_k) = \prod_{i=1}^k \mathbb{P}_{X_i}(X_i \leq x_i) ,$$

$$x_i \in (-\infty; \infty) \cup \{\infty\} , i = 1, \dots, k .$$
(КЛ9.4.3)

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 (КЛ9.4.3) набуває форми

$$F_{\mathbf{X}}(x_1, \dots, x_k) = F_{X_1}(x_1) \cdot \dots \cdot F_{X_k}(x_k) , x_i \in (-\infty; \infty) \cup \{\infty\} , i = 1, \dots, k .$$
(КЛ9.4.4)

Якщо кожна випадкова змінна X_i має щільність f_i , то справедливий такий критерій.

Теорема (КЛ9.4.5). Випадкові величини $X_1, \dots, X_k$ зі щільностями розподілів $f_1, \dots, f_k$ відповідно незалежні тоді й тільки тоді, коли випадковий вектор $\mathbf{X} = (X_1, \dots, X_k)^\top$ має щільність $f_{\mathbf{X}}$ таку, що

$$f_{\mathbf{X}}(x_1, \dots, x_k) = f_1(x_1) \cdot \dots \cdot f_k(x_k) , \quad x_i \in \mathbb{R} , i = 1, \dots, k .$$
(КЛ9.4.6)

Також безпосереднім наслідком наведених вище міркувань є критерій незалежності дискретних випадкових величин:

$$\mathbb{P}_{\mathbf{X}}(X_1 = x_1, \dots, X_k = x_k) = \mathbb{P}_{X_1}(X_1 = x_1) \cdot \dots \cdot \mathbb{P}_{X_k}(X_k = x_k) .$$
(КЛ9.4.7)

Якщо випадкові величини незалежні, незалежні також є функції від них.

Теорема (КЛ9.4.9). Якщо величини $X_{i,j}$ незалежні для $i = 1, \dots, n, j = 1, \dots, m(i)$, а функції $f_i : \mathbb{R}^{m(i)} \rightarrow \mathbb{R}$ вимірні, то незалежні є випадкові величини $f_i(X_{i,1}, \dots, X_{i,m(i)})$.

Теорема (КЛ9.4.11). 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

$$\mathbb{E}[h(X, Y)] = \int_{\mathbb{R}} \int_{\mathbb{R}} h(x, y) d\mathbb{P}_X d\mathbb{P}_Y . \quad (\text{КЛ9.4.8})$$

Зокрема, якщо $h(x, y) = f(x)g(y)$, де $f, g : \mathbb{R} \rightarrow \mathbb{R}$ є вимірні функції, невід’ємні ($f, g \geq 0$) чи інтегровні ($\mathbb{E}[|f(X)|] < \infty$, $\mathbb{E}[|g(Y)|] < \infty$), то

$$\mathbb{E}[f(X)g(Y)] = \mathbb{E}[f(X)] \mathbb{E}[g(Y)] . \quad (\text{КЛ9.4.9})$$

Із щойно доведеної теореми безпосередньо випливає такий результат.

Теорема (КЛ9.4.12). 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

$$\mathbb{E} \left[\prod_{i=1}^k X_i \right] = \prod_{i=1}^k \mathbb{E}[X_i] , \quad (\text{КЛ9.4.10})$$

у тому розумінні, що сподівання зліва існує і дорівнює добутку справа.

Із того факту, що для незалежних випадкових величин сподівання добутку є добутком сподівань, негайно можна встановити таку властивість.

Твердження (КЛ9.4.13). Якщо випадкові величини X і Y незалежні, то $\text{Cov}(X, Y) = 0$.

Такі випадкові величини називають *некорельованими* (uncorrelated). При цьому дуже важливо пам’ятати, що *зворотне твердження в загальному випадку несправедливе*.

З урахуванням Твердження КЛ9.4.13 можна спростити підрахунок дисперсії суми випадкових величин за (КЛ9.3.4) у випадку незалежних величин:

$$\text{Var}(X_1 + \dots + X_k) = \sum_{i=1}^k \text{Var}(X_i) . \quad (\text{КЛ9.4.11})$$

10.2 Спільні функції та щільності розподілів

Приклад 10.2.1. Нехай у деякій громаді 15% родин не мають дітей, 20% родин мають одну дитину, 35% мають дві дитини, а 30% мають 3 дитини. Нехай імовірності народити хлопчика та дівчинку однакові, а стать однієї

Табл. 10.2.1: Таблиця спряженості для Прикладу 10.2.1

	$G = 0$	$G = 1$	$G = 2$	$G = 3$	Разом
$B = 0$	0.1500	0.1000	0.0875	0.0375	0.3750
$B = 1$	0.1000	0.1750	0.1125	0	0.3875
$B = 2$	0.0875	0.1125	0	0	0.2000
$B = 3$	0.0375	0	0	0	0.0375
Разом	0.3750	0.3875	0.2000	0.0375	1

народженої дитини не залежить від інших. Число хлопчиків B та число дівчаток G у навмання вибраній із цієї громади родині мають (спільну) функцію ймовірностей, наведену в таблиці спряженості Табл. 10.2.1.

Чому дорівнює ймовірність, що в родині немає дітей? Що в родині тільки одна дитина, і це дівчинка? Що в родині двоє дітей, і обидва хлопчики? Що в родині два хлопчики? Що в родині немає хлопчиків і щонайбільше дві дівчинки? $\square$

Розв'язання. Ймовірність того, що в родині немає дітей, дорівнює

$$\mathbb{P}_{B,G}(B = 0, G = 0) = 0.1500.$$

Ймовірність того, що в родині тільки одна дитина, і це дівчинка, дорівнює

$$\mathbb{P}_{B,G}(B = 0, G = 1) = 0.1000.$$

Ймовірність того, що в родині двоє дітей, і обидва хлопчики, дорівнює

$$\mathbb{P}_{B,G}(B = 2, G = 0) = 0.0875.$$

Ймовірність того, що в родині два хлопчики, передбачає пошук маржинальної функції ймовірностей. У Табл. 10.2.1 маржинальна функція ймовірностей для B займає останній стовпець, звідки $\mathbb{P}_B(B = 2) = 0.2000$.

Нарешті, ймовірність того, що в родині немає хлопчиків і щонайбільше два дівчинки, дорівнює

$$\begin{aligned}\mathbb{P}_{B,G}(B = 0, G \leq 2) &= \mathbb{P}_{B,G}(B = 0, G = 0) + \mathbb{P}_{B,G}(B = 0, G = 1) \\ &\quad + \mathbb{P}_{B,G}(B = 0, G = 2) \\ &= 0.1500 + 0.1000 + 0.0875 = 0.3375.\end{aligned}$$

Між іншим, інформація про те, що саме 15% родин не мають дітей, 20% родин мають одну дитину, 35% мають дві дитини, а 30% мають 3 дитини, зашифрована в побічних діагоналях таблиці спряженості. Наприклад, $0.2 = \mathbb{P}_{B,G}(B = 0, G = 1) + \mathbb{P}_{B,G}(B = 1, G = 0) = 0.1 + 0.1$. $\square$

Приклад 10.2.2. Нехай маємо дві різні правильні монетки, які підкидають n разів. Розгляньмо такі пари випадкових величин:

- X_1 — число гербів, випалих на першій монетці, Y_1 — число гербів, випалих на другій монетці;
- X_2 — число гербів, випалих на першій монетці, Y_2 — число гербів, випалих на *першій* монетці, тобто X_2 і Y_2 дорівнюють одна одній.

В усіх випадках, очевидно, величини мають однаковий (маржинальний) біномний розподіл $\text{Binom}(n, 0.5)$. Проте *спільні* розподіли суттєво різняться. Оскільки $X_1 \perp\!\!\!\perp Y_1$, їхня спільна функція ймовірності дорівнює добутку відповідних маржинальних функцій ймовірності:

$$\begin{aligned}\mathbb{P}_{X_1, Y_1}(X_1 = x, Y_1 = y) &= \mathbb{P}_{X_1}(X_1 = x) \cdot \mathbb{P}_{Y_1}(Y_1 = y) \\ &= \binom{n}{x} \cdot \frac{1}{2^x} \cdot \frac{1}{2^{n-x}} \cdot \mathbb{1}\{x \in \{0, 1, \dots, n\}\} \\ &\times \binom{n}{y} \cdot \frac{1}{2^y} \cdot \frac{1}{2^{n-y}} \cdot \mathbb{1}\{y \in \{0, 1, \dots, n\}\} \\ &= \binom{n}{x} \binom{n}{y} \frac{1}{2^{2n}} \cdot \mathbb{1}\{x \in \{0, 1, \dots, n\}\} \\ &\times \mathbb{1}\{y \in \{0, 1, \dots, n\}\} .\end{aligned}$$

У той же час, X_2 та Y_2 не можуть набувати різних значень, тому маємо таку спільну функцію ймовірності:

$$\begin{aligned}\mathbb{P}_{X_2, Y_2}(X_2 = x, Y_2 = y) &= \binom{n}{x} \frac{1}{2^x} \cdot \frac{1}{2^{n-x}} \cdot \mathbb{1}\{x \in \{0, 1, \dots, n\}\} \cdot \mathbb{1}\{x = y\} \\ &\neq \mathbb{P}_{X_2}(X_2 = x) \cdot \mathbb{P}_{Y_2}(Y_2 = y) .\end{aligned}$$

□

Вправа 10.2.3. Нехай спільна щільність розподілу деякого випадкового вектора $(X, Y)^\top$ дорівнює

$$f_{X, Y}(x, y) = 2e^{-x}e^{-2y} \cdot \mathbb{1}\{x > 0\} \cdot \mathbb{1}\{y > 0\} .$$

Чому дорівнює ймовірність $\mathbb{P}_{X, Y}(X > 1, Y < 1)$? Чому дорівнює ймовірність $\mathbb{P}_{X, Y}(X < Y)$? Чому дорівнює ймовірність $\mathbb{P}_X(X < a)$, $a > 0$? □

Розв'язання. Для обчислення будь-якої ймовірності потрібно проінтегрувати спільну щільність за відповідною множиною. З урахуванням невід'єм-

ности щільности можемо застосувати теорему Фубіні:

$$\begin{aligned}
 \mathbb{P}_{X,Y}(X > 1, Y < 1) &= \iint_{x>1, y<1} f_{X,Y}(x, y) dx dy \\
 &= \int_0^1 \left(\int_1^\infty e^{-x} dx \right) 2e^{-2y} dy \\
 &= \int_0^1 \left(-e^{-x} \Big|_1^\infty \right) 2e^{-2y} dy \\
 &= e^{-1} \int_0^1 2e^{-2y} dy = e^{-1} \left(-e^{-2y} \Big|_0^1 \right) = e^{-1} (1 - e^{-2}) .
 \end{aligned}$$

Імовірність, що $X < Y$, можна обчислити аналогічно:

$$\begin{aligned}
 \mathbb{P}_{X,Y}(X < Y) &= \iint_{x<y} f_{X,Y}(x, y) dx dy \\
 &= \int_0^\infty \left(\int_0^y e^{-x} dx \right) 2e^{-2y} dy \\
 &= \int_0^\infty \left(-e^{-x} \Big|_0^y \right) 2e^{-2y} dy \\
 &= \int_0^\infty (1 - e^{-y}) \cdot 2e^{-2y} dy \\
 &= \int_0^\infty 2e^{-2y} dy - \int_0^\infty 2e^{-3y} dy \\
 &= \left(-e^{-2y} \Big|_0^\infty \right) - \left(-\frac{2}{3}e^{-3y} \Big|_0^\infty \right) = 1 - \frac{2}{3} = \frac{1}{3} .
 \end{aligned}$$

Нарешті, ймовірність події $X < a$ еквівалентна події $(X < a) \cap (Y \in \mathbb{R})$:

$$\begin{aligned}
 \mathbb{P}_X(X < a) &= \mathbb{P}_{X,Y}(X < a, Y \in \mathbb{R}) = \iint_{x<a, y \in \mathbb{R}} f_{X,Y}(x, y) dx dy \\
 &= \int_0^\infty \left(\int_0^a e^{-x} dx \right) 2e^{-y} dy \\
 &= \int_0^\infty (1 - e^{-a}) 2e^{-2y} dy \\
 &= (1 - e^{-a}) \int_0^\infty 2e^{-2y} dy = 1 - e^{-a} .
 \end{aligned}$$

Ці обчислення, зокрема, показують, що X має маржинальний експоненційний розподіл $X \sim \text{Exp}(1)$.

Взагалі кажучи, можна було з самого початку помітити, що ми маємо справу з двома незалежними експоненційними величинами. Справді, спіль-

ну щільність можна подати як добуток двох маржинальних:

$$\begin{aligned} f_{X,Y}(x,y) &= 2e^{-x}e^{-2y} \cdot \mathbb{1}\{x > 0\} \cdot \mathbb{1}\{y > 0\} \\ &= e^{-x} \cdot \mathbb{1}\{x > 0\} \times 2e^{-2y} \cdot \mathbb{1}\{y > 0\} \equiv f_X(x)f_Y(y), \end{aligned}$$

де $X \sim \text{Exp}(1)$, $Y \sim \text{Exp}(2)$, до того ж $X \perp\!\!\!\perp Y$.

Звідси випливає, що першу шукану ймовірність $\mathbb{P}_{X,Y}(X > 1, Y < 1)$ можна було б дістати за визначенням незалежних величин як

$$\mathbb{P}_{X,Y}(X > 1, Y < 1) = \mathbb{P}_X(X > 1) \mathbb{P}_Y(Y < 1) = (1 - F_X(1))F_Y(1).$$

Проте ймовірність $\mathbb{P}_{X,Y}(X < Y)$ без інтегрування спільної щільності в загальному випадку знайти неможливо. В окремих випадках (наприклад, для незалежних нормальних величин) можна встановити розподіл величини $X - Y$ і розглянути ймовірність $\mathbb{P}_{X-Y}(X - Y < 0)$. Наприклад, якщо $X \sim N(0, 1)$ і $Y \sim N(1, 2)$, $X \perp\!\!\!\perp Y$, то $X - Y \sim N(-1, 3)$.

Проте для експоненційних величин (навіть незалежних) ми не можемо без додаткових обчислень сказати, який розподіл має $X - Y$. $\square$

Приклад 10.2.4. Розгляньмо клінічні випробування, у рамках яких пацієнти, хворі на депресію, проходять курс лікування, а потім лікарі дивляться, чи повернулася в них депресія, чи ні. Нехай X — індикаторна величина, яка відповідає тому, що депресія не повернулася. Також нехай P — частка учасників клінічних випробувань, у яких депресія не повернулася.

Цілком очевидно, що X — дискретна величина Бернуллі, а P — неперервна з носієм $\text{supp}(P) = [0; 1]$. Тобто в цьому випадку маємо справу з добутком просторів $\{0, 1\} \times \mathbb{R}$ та носієм відповідного вектора $\{0, 1\} \times [0; 1]$, на якому добутком мір буде $\# \times \lambda$, тобто добуток лічної міри й міри Лебега.

Нехай спільний розподіл X та P абсолютно неперервний відносно такого добутку мір і має щільність

$$f_{X,P}(x,p) = p^x(1-p)^{1-x} \cdot \mathbb{1}\{x \in \{0, 1\}\} \cdot \mathbb{1}\{0 < p < 1\}.$$

Можемо показати, що це справді щільність, тобто що її інтеграл дорівнює 1. Оскільки щільність невід'ємна, можемо застосувати теорему Фубіні, щоб перейти від подвійного інтегралу до повторного:

$$\begin{aligned} \iint f_{X,P}(x,p) \#(dx) \times \lambda(dp) &= \int_{-\infty}^{\infty} \sum_{x \in \{0,1\}} f_{X,P}(x,p) dp \\ &= \int_0^1 (f_{X,P}(0,p) + f_{X,P}(1,p)) dp \\ &= \int_0^1 (1-p+p) dp = 1. \end{aligned}$$

Можемо дістати вираз для маржинальної щільності для P , «інтегруючи» за змінною x :

$$\begin{aligned} f_P(p) &= \sum_{x \in \{0,1\}} p^x (1-p)^{1-x} \cdot \mathbb{1}\{x \in \{0,1\}\} \cdot \mathbb{1}\{0 < p < 1\} \\ &= (1-p+p) \cdot \mathbb{1}\{0 < p < 1\} = \mathbb{1}\{0 < p < 1\}, \end{aligned}$$

тобто $P \sim U((0; 1])$.

Аналогічно для X :

$$\begin{aligned} \mathbb{P}_X(X = x) &= \int_{-\infty}^{\infty} p^x (1-p)^{1-x} \cdot \mathbb{1}\{x \in \{0,1\}\} \cdot \mathbb{1}\{0 < p < 1\} dp \\ &= \int_0^1 p^x (1-p)^{1-x} \cdot \mathbb{1}\{x \in \{0,1\}\} dp. \end{aligned}$$

Зокрема,

$$\mathbb{P}_X(X = 0) = \int_0^1 (1-p) dp = \frac{1}{2}, \quad \mathbb{P}_X(X = 1) = \int_0^1 p dp = \frac{1}{2}.$$

Тобто $X \sim \text{Bern}(0.5)$.

Також можемо обчислити цікаві для нас імовірності, наприклад,

$$\begin{aligned} \mathbb{P}_{X,P}\left(X = 0, P \leq \frac{1}{2}\right) &= \iint_{x=0, 0 \leq p \leq \frac{1}{2}} p^x (1-p)^{1-x} \#(dx) \times \lambda(dp) \\ &= \int_0^{\frac{1}{2}} p^0 (1-p)^{1-0} dp = \frac{3}{8}, \end{aligned}$$

тобто ймовірність того, що в пацієнта повернулася депресія, а частка пацієнтів, у яких вона не повернулася, менше 50%. Ця ймовірність не має корисної інтерпретації, адже значно цікавіше дізнатися, яка ймовірність того, що депресія повернулася, *за умови*, що частка пацієнтів, у яких вона не повернулася, менше 50%. Для цього можна застосувати визначення умовної ймовірності:

$$\mathbb{P}_{X|P}\left(X = 0 \mid P \leq \frac{1}{2}\right) = \frac{\mathbb{P}_{X,P}\left(X = 0, P \leq \frac{1}{2}\right)}{\mathbb{P}_P\left(P \leq \frac{1}{2}\right)} = \frac{3/8}{1/2} = \frac{3}{4}.$$

□

10.3 Сподівання та коваріації, незалежність

Вправа 10.3.1. Нехай $\text{Corr}(X, Y) < 0$. Що можна сказати про $\text{Var}(X + Y)$ у порівнянні з $\text{Var}(X - Y)$? $\square$

Розв'язання. За формулою обчислення дисперсії суми та різниці випадкових величин маємо

$$\begin{aligned}\text{Var}(X + Y) &= \text{Var}(X) + \text{Var}(Y) + 2\text{Cov}(X, Y) , \\ \text{Var}(X - Y) &= \text{Var}(X) + \text{Var}(Y) - 2\text{Cov}(X, Y) .\end{aligned}$$

Звідси

$$\text{Var}(X + Y) - \text{Var}(X - Y) = 4\text{Cov}(X, Y) .$$

Оскільки коефіцієнт кореляції пропорційний до коваріації, $\text{Cov}(X, Y) < 0$, і звідси випливає, що $\text{Var}(X + Y) < \text{Var}(X - Y)$. $\square$

Вправа 10.3.2. Нехай $X_1, \dots, X_n$ — випадкові величини з дисперсією 1 та коваріацією між кожною парою величин $\frac{1}{4}$. Чому дорівнює дисперсія їх суми $\text{Var}(X_1 + \dots + X_n)$? $\square$

Розв'язання. За формулою дисперсії суми випадкових величин маємо

$$\begin{aligned}\text{Var}(X_1 + \dots + X_n) &= \sum_{i=1}^n \text{Var}(X_i) + 2 \sum_{i < j} \text{Cov}(X_i, X_j) \\ &= n + 2 \cdot \frac{n(n-1)}{2} \cdot \frac{1}{4} = n + \frac{n(n-1)}{4} ,\end{aligned}$$

оскільки в сумі коваріацій є $\binom{n}{2} = \frac{n(n-1)}{2}$ членів (число пар, які можна вибрати з n елементів без повторень і збереження порядку). $\square$

Вправа 10.3.3. Нехай $X \sim \text{Exp}(1)$, $Y \sim \text{Exp}(1)$, $X \perp\!\!\!\perp Y$. Чому дорівнює коефіцієнт кореляції між $M = \max\{X, Y\}$ та $L = \min\{X, Y\}$? $\square$

Розв'язання. У Прикладі КЛ8.5.3 ми зазначали, що

$$L \sim \text{Exp}(2) , \quad M - L \sim \text{Exp}(1) ,$$

а також що $M - L \perp L$. Тому з урахуванням цієї незалежності маємо

$$\begin{aligned}\operatorname{Cov}(M, L) &= \operatorname{Cov}(M - L + L, L) = \operatorname{Cov}(M - L, L) + \operatorname{Cov}(L, L) \\ &= 0 + \operatorname{Var}(L) = \frac{1}{4},\end{aligned}$$

а

$$\operatorname{Var}(M) = \operatorname{Var}(M - L + L) = \operatorname{Var}(M - L) + \operatorname{Var}(L) = 1 + \frac{1}{4} = \frac{5}{4}.$$

Звідси

$$\operatorname{Corr}(M, L) = \frac{\operatorname{Cov}(M, L)}{\sqrt{\operatorname{Var}(M) \operatorname{Var}(L)}} = \frac{1/4}{\sqrt{5/4 \cdot 1/4}} = \frac{1}{\sqrt{5}} > 0.$$

Варто зазначити, що в цій вправі ми використали цікаві властивості експоненційних розподілів. У загальному випадку ми можемо помітити, що $M \cdot L = X \cdot Y$ (тому що одна з величин є максимальною, а інша — мінімальною). Звідси випливає, що $\mathbb{E}[ML] = \mathbb{E}[XY]$. У нашому випадку, з урахуванням незалежності, маємо $\mathbb{E}[ML] = \mathbb{E}[X] \mathbb{E}[Y] = 1 \cdot 1 = 1$.

Також відразу бачимо, що $\mathbb{E}[L] = \frac{1}{2}$ і що $\mathbb{E}[M - L] = 1$, звідки $\mathbb{E}[M] = 1 + \frac{1}{2} = \frac{3}{2}$. Звідси

$$\operatorname{Cov}(M, L) = \mathbb{E}[ML] - \mathbb{E}[M] \mathbb{E}[L] = 1 - \frac{1}{2} \cdot \frac{3}{2} = \frac{1}{4},$$

що повністю відповідає раніше порахованому результату. $\square$

Вправа 10.3.4. Нехай $M = \max\{X, Y\}$ та $L = \min\{X, Y\}$ для довільних випадкових величин X та Y . Чи можна стверджувати, що

$$\operatorname{Cov}(M, L) = \operatorname{Cov}(X, Y) ?$$

Перевірте виконання цієї тотожності для двох незалежних стандартних рівномірних величин. $\square$

Розв'язання. На перший погляд може здатися, що така рівність має сенс, адже якщо маємо дві величини X та Y , то для кожної конкретної реалізації їхніх значень можна сказати, що одне число найбільше, а інше — найменше, а тому за комутативністю коваріації це буде те саме, що коваріація X та Y . Проте потрібно розуміти, що в цих міркуваннях ми вважаємо, що значення випадкових величин уже стали відомі. Але тоді коваріація взагалі повинна бути 0, адже це вже будуть не випадкові величини, а фіксовані дійсні числа.

Насправді в загальному випадку такої тотожності *не існує*. Найпростіший приклад пов'язаний із використанням двох незалежних величин $X \perp\!\!\!\perp Y$. Справді, якщо вони незалежні, то їхня коваріація $\text{Cov}(X, Y) = 0$. Проте цілком очевидно, що максимум і мінімум не можуть бути незалежними, оскільки вищі значення максимуму повинні означати вищі значення мінімуму, тобто між ними повинен існувати додатний, хоча, можливо, й слабкий, лінійний зв'язок.

Це можна проілюструвати, розглянувши стандартний рівномірний розподіл: $U_1, U_2 \sim U((0; 1])$, $U_1 \perp\!\!\!\perp U_2$. На лекції в Прикладі КЛ8.3.3 ми показали, що мінімум і максимум матимуть такі щільності:

$$f_L(u) = 2(1 - u) \cdot \mathbb{1}\{u \in (0; 1)\} \quad , \quad f_M(u) = 2u \cdot \mathbb{1}\{u \in (0; 1)\} \quad .$$

Як і в попередній вправі, можна порахувати $\mathbb{E}[ML] = \mathbb{E}[U_1] \mathbb{E}[U_2] = \frac{1}{4}$. Використовуючи ці дані, можна обчислити кореляцію безпосередньо. Проте заради цікавості виведемо спільну щільність максимуму та мінімуму.

Ми вже знаємо, що вони не можуть бути незалежні, тому підемо шляхом виведення спільної функції розподілу $F_{M,L}(x, y) = \mathbb{P}_{M,L}(M \leq x, L \leq y)$. Раніше ми неодноразово зазначали, що для проведення обчислень із мінімумами зручно переходити до ситуації, коли величина більше деякого значення, а не менше. Із властивостей спільних розподілів випливає:

$$\begin{aligned} \mathbb{P}_M(M \leq x) &= \mathbb{P}_{M,L}(M \leq x, L < \infty) \\ &= \mathbb{P}_{M,L}(M \leq x, L \leq y) + \mathbb{P}_{M,L}(M \leq x, L > y) \quad , \end{aligned}$$

звідки

$$F_{M,L}(x, y) = F_M(x) - \mathbb{P}_{M,L}(M \leq x, L > y) \quad .$$

Функція розподілу максимуму дорівнює

$$F_M(x) = \int_{-\infty}^x 2u \cdot \mathbb{1}\{u \in (0; 1)\} du = \begin{cases} 0, & x < 0 \\ x^2, & 0 \leq x < 1 \\ 1, & x \geq 1 \end{cases} \quad .$$

З урахуванням незалежності X та Y маємо

$$\begin{aligned} \mathbb{P}_{M,L}(M \leq x, L > y) &= \mathbb{P}_{X,Y}(X \leq x, Y \leq x, X > y, Y > y) \\ &= \mathbb{P}_X(y < X \leq x) \mathbb{P}_Y(y < Y \leq x) \\ &= \begin{cases} (x - y)^2, & 0 \leq y \leq x < 1 \\ x^2, & y < 0 \leq x < 1 \\ (1 - y)^2, & 0 \leq y < 1 \leq x \\ 1, & y < 0, 1 \leq x \\ 0, & \text{інакше} \end{cases} \quad . \end{aligned}$$

Отже

$$F_{M,L}(x, y) = \begin{cases} x^2 - (x - y)^2, & 0 \leq y \leq x < 1 \\ 1 - (1 - y)^2, & 0 \leq y < 1 \leq x \\ x^2, & 0 \leq x < 1, y > x \\ 1, & x \geq 1, y > x \\ 0, & \text{інакше} \end{cases}$$

Відповідна щільність дорівнює

$$f_{M,L}(x, y) = \frac{\partial^2}{\partial x \partial y} F_{M,L}(x, y) = 2 \cdot \mathbb{1} \{0 \leq y \leq x < 1\}.$$

Підрахуймо відповідну кореляцію:

$$\begin{aligned} \mathbb{E}[L] &= \int_0^1 u \cdot 2(1 - u) du = \int_0^1 (2u - 2u^2) du = 1 - \frac{2}{3} = \frac{1}{3}, \\ \mathbb{E}[L^2] &= \int_0^1 u^2 \cdot 2(1 - u) du = \int_0^1 (2u^2 - 2u^3) du = \frac{2}{3} - \frac{2}{4} = \frac{1}{6}, \\ \mathbb{E}[M] &= \int_0^1 u \cdot 2u du = \frac{2}{3}, \\ \mathbb{E}[M^2] &= \int_0^1 u^2 \cdot 2u du = \frac{1}{2}, \\ \mathbb{E}[ML] &= \int_0^1 \int_0^x xy \cdot 2 dy dx = \int_0^1 x \cdot x^2 dx = \frac{1}{4}, \\ \text{Corr}(M, L) &= \frac{\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} xy \cdot 2 \cdot \mathbb{1} \{0 \leq y \leq x < 1\} dx dy}{\sqrt{\text{Var}(M) \text{Var}(L)}} = \frac{\text{Cov}(M, L)}{\sqrt{\text{Var}(M) \text{Var}(L)}} \\ &= \frac{\mathbb{E}[ML] - \mathbb{E}[M] \mathbb{E}[L]}{\sqrt{(\mathbb{E}[M^2] - (\mathbb{E}[M])^2)(\mathbb{E}[L^2] - (\mathbb{E}[L])^2)}} = 0.5, \end{aligned}$$

тобто кореляція не просто додатна, а й не така вже й мала. $\square$

Вправа 10.3.5. Розгляньмо випадкове блукання на площині. Частинка починає свій рух із початку координат і на кожному кроці, незалежно від попередніх, переміщується на одиницю вправо, вліво, вгору та вниз з однаковою ймовірністю. Нехай поточну позицію частинки після n кроків описує випадковий вектор $(X_n, Y_n)^\top$, а відстань від початку координат дає випадкова величина R_n .

Чи є X_n незалежна від Y_n ? Чому дорівнює відповідна коваріація? Чому дорівнює $\mathbb{E}[R_n^2]$? $\square$

Розв'язання. Можемо розглянути відповідні випадкові величини як суми

$$X_n = \sum_{i=1}^n V_i, \quad Y_n = \sum_{i=1}^n W_i,$$

де V_i та W_i визначають рух за горизонталлю та вертикаллю відповідно. Вони мають однаковий розподіл:

$$p_V(x) = p_W(x) = \begin{cases} \frac{1}{4}, & x = -1 \\ \frac{1}{2}, & x = 0 \\ \frac{1}{4}, & x = 1 \end{cases}.$$

Чи є величини V_i та W_j незалежні? Якщо $i \neq j$, то так, оскільки частинка може рухатися в будь-якому напрямку незалежно від попередніх кроків. Проте якщо $i = j$, існує пряма залежність, адже якщо одна з цих величин не дорівнює 0, то інша *обов'язково* дорівнює 0, адже частинка не може рухатися одночасно і в горизонтальному, і у вертикальному напрямках.

Тому і X_n не може бути незалежна від Y_n . Зокрема, це добре видно в такому граничному випадку. Якщо частинка постійно рухається в одному напрямку, наприклад, усі n кроків були вгору, то обов'язково $X_n = 0$.

Проте коваріація цих величин, тим не менше, дорівнює 0:

$$\begin{aligned} \text{Cov}(X_n, Y_n) &= \text{Cov}\left(\sum_{i=1}^n V_i, \sum_{j=1}^n W_j\right) \\ &= \sum_{i=1}^n \sum_{j=1}^n \text{Cov}(V_i, W_j) \\ &= \sum_{i=1}^n \text{Cov}(V_i, W_i) = \sum_{i=1}^n (\mathbb{E}[V_i W_i] - \mathbb{E}[V_i] \mathbb{E}[W_i]) = 0, \end{aligned}$$

бо $V_i \perp W_j$ для $i \neq j$, $\mathbb{E}[V_i] = \mathbb{E}[W_i] = 0$, а $\mathbb{E}[V_i W_i] = 0$, бо $V_i W_i = 0$.

Що до $R_n^2 = X_n^2 + Y_n^2$, то через $X_n \stackrel{d}{=} Y_n$ та $V_i \perp V_j$ для $i \neq j$ маємо

$$\begin{aligned} \mathbb{E}[R_n^2] &= \mathbb{E}[X_n^2 + Y_n^2] = 2\mathbb{E}[X_n^2] = 2\mathbb{E}\left[\left(\sum_{i=1}^n V_i\right)^2\right] \\ &= 2\mathbb{E}\left[\sum_{i=1}^n V_i^2 + 2\sum_{i<j} V_i V_j\right] = 2\sum_{i=1}^n \mathbb{E}[V_i^2] + 2\sum_{i<j} \mathbb{E}[V_i] \mathbb{E}[V_j] \\ &= 2n\mathbb{E}[V_1^2] + 0 = 2n \cdot \frac{1}{2} = n. \end{aligned}$$

□

Приклад 10.3.6. Нехай інвестор планує інвестувати \$100 000, і задля цього він формує інвестиційний *портфель* (portfolio), тобто розподіляє всю суму між акціями різних компаній та іншими фінансовими інструментами. Нехай інвестор розглядає тільки акції двох компаній та депозит у деякому банку. Нехай s_1 та s_2 — сума, інвестована в акції двох компаній відповідно, а s_3 — сума, покладена на депозит.

Нехай акції першої та другої компаній мають випадкові дохідності (за рік) R_1 та R_2 відповідно. Нехай ставка депозиту дорівнює 3.6% за рік. Нехай R_1 має деякий розподіл зі сподіванням 6 та дисперсією 55, а R_2 має деякий розподіл зі сподіванням 4 та дисперсією 28. Тоді дохід (за рік) від такої інвестиції дорівнюватиме $s_1 R_1 + s_2 R_2 + 0.036 s_3$.

Нехай акції першої компанії коштують \$60 за штуку, а другої — \$48 за штуку. Тоді всі інвестиції повинні задовольняти обмеження $60s_1 + 48s_2 + s_3 = 100\,000$ та $s_1, s_2, s_3 \geq 0$.

Нехай $R_1 \perp R_2$. Тоді сподівання та дисперсія доходу дорівнюють

$$\begin{aligned}\mathbb{E}[s_1 R_1 + s_2 R_2 + 0.036 s_3] &= s_1 \mathbb{E}[R_1] + s_2 \mathbb{E}[R_2] + 0.036 s_3 \\ &= 6s_1 + 4s_2 + 0.036 s_3,\end{aligned}$$

$$\text{Var}(s_1 R_1 + s_2 R_2 + 0.036 s_3) = s_1^2 \text{Var}(R_1) + s_2^2 \text{Var}(R_2) + 0 = 55s_1^2 + 28s_2^2.$$

За яким критерієм можна підібрати портфель? У теорії портфелів розглядають поняття *ефективності* (efficiency). Ефективним називають такий портфель, який має найменшу дисперсію серед усіх портфелів із таким самим сподіванням. На Рис. 10.3.1 зображено область, у яку потрапляють усі можливі комбінації сподівань та дисперсій портфелів, які задовольняють обмеження, зазначені вище. Нижня межа цієї області показує множину ефективних портфелів.

Наприклад, якщо інвестор хоче, щоб його сподіваний дохід дорівнював \$7000, він повинен вибрати портфель із параметрами $s_1 = 524.7$, $s_2 = 609.7$, $s_3 = 39\,250$, оскільки в цьому випадку сподівання буде \$7000, а дисперсія буде найменша серед усіх портфелів із таким сподіванням, а саме 2.55×10^7 . Іншими словами, будь-який портфель зі сподіванням більше \$7000 повинен мати більшу дисперсію, а будь-який портфель із дисперсією менше 2.55×10^7 повинен мати сподівання менше \$7000.

Нехай тепер коефіцієнт кореляції між дохідністю двох акцій дорівнює -0.3 . Наприклад, це може означати, що акції двох компаній реагують на однакові події в протилежному напрямку. Оскільки

$$-0.3 = \frac{\text{Cov}(R_1, R_2)}{\sqrt{\text{Var}(R_1) \text{Var}(R_2)}} = \frac{\text{Cov}(R_1, R_2)}{\sqrt{55 \cdot 28}},$$

маємо

$$\text{Cov}(s_1 R_1, s_2 R_2) = s_1 s_2 \text{Cov}(R_1, R_2) = -0.3 \sqrt{55 \cdot 28} \cdot s_1 s_2.$$

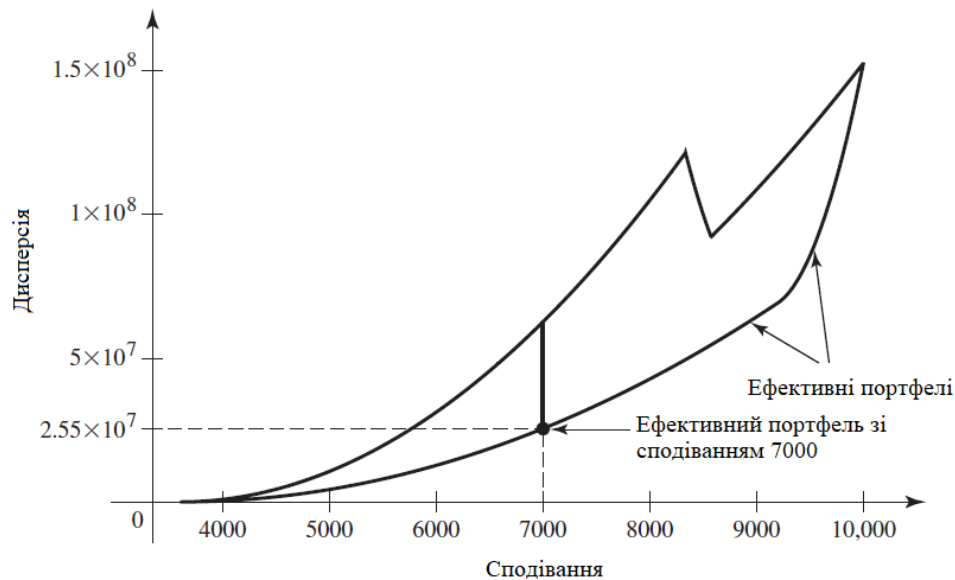


Рис. 10.3.1: Ілюстрація до Прикладу 10.3.6 (адаптовано з [7], Рис. 4.7)

Отже

$$\begin{aligned}\text{Var}(s_1 R_1 + s_2 R_2 + 0.036 s_3) &= s_1^2 \text{Var}(R_1) + s_2^2 \text{Var}(R_2) + 2s_1 s_2 \text{Cov}(R_1, R_2) \\ &= 55s_1^2 + 28s_2^2 - 2 \cdot 0.3 \cdot \sqrt{55 \cdot 28} \cdot s_1 s_2.\end{aligned}$$

Нескладно бачити, що дисперсія такого портфеля менша від дисперсії портфеля з аналогічними параметрами, але незалежними дохідностями.

Це інтуїтивно зрозуміло, адже якщо розкласти акції по компаніях із додатними кореляціями, то якщо ціни на акції впадуть, то вони впадуть (різною мірою) для всіх компаній із портфеля. Якщо ж кореляція від'ємна, то акції однієї компанії впадуть, а іншої — зростуть. Тому відхилення від сподівання буде не таке сильне, і такий портфель буде менш ризиковий. $\square$

10.4 Додаткові вправи

Приклад 10.4.1. Нехай у відділку деякого банку працюють два менеджери. Число клієнтів, які стоять у черзі до кожного з них, описує випадковий вектор $(X, Y)^\top$, де X — число клієнтів у черзі до першого менеджера, а Y — до другого. Нехай $\text{supp}(X) = \text{supp}(Y) = \{0, 1, 2, 3, 4\}$. Спільну функцію ймовірностей для цього випадкового вектора наведено в Табл. 10.4.1.

Чому дорівнює ймовірність $\mathbb{P}_{X,Y}(2 \leq X \leq 3, 0 \leq Y \leq 2)$? Чому дорівнює спільна функція розподілу в точці $(2, 1)^\top$? $\square$

Табл. 10.4.1: Таблиця спряженості для Прикладу 10.4.1

	$Y = 0$	$Y = 1$	$Y = 2$	$Y = 3$	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

Розв'язання. Для підрахунку відповідної ймовірності потрібно «проінтегрувати» щільність за відповідною множиною. У нашому випадку вектор дискретний, отже роль щільності виконує функція ймовірності, а роль інтегралу — сума:

$$\begin{aligned}
 \mathbb{P}_{X,Y}(2 \leq X \leq 3, 0 \leq Y \leq 2) &= \sum_{x=2}^3 \sum_{y=0}^2 \mathbb{P}_{X,Y}(X=x, Y=y) \\
 &= \mathbb{P}_{X,Y}(X=2, Y=0) + \mathbb{P}_{X,Y}(X=2, Y=1) \\
 &\quad + \mathbb{P}_{X,Y}(X=2, Y=2) \\
 &\quad + \mathbb{P}_{X,Y}(X=3, Y=0) + \mathbb{P}_{X,Y}(X=3, Y=1) \\
 &\quad + \mathbb{P}_{X,Y}(X=3, Y=2) \\
 &= 0 + 0.06 + 0.07 + 0 + 0 + 0.03 = 0.16 .
 \end{aligned}$$

Значення функції розподілу в деякій точці є сумою всіх значень функції ймовірності в усіх точках, що не перевищують шуканих значень:

$$\begin{aligned}
 F_{X,Y}(2, 1) &= \mathbb{P}_{X,Y}(X \leq 2, Y \leq 1) = \sum_{x=0}^2 \sum_{y=0}^1 \mathbb{P}_{X,Y}(X=x, Y=y) \\
 &= \mathbb{P}_{X,Y}(X=0, Y=0) + \mathbb{P}_{X,Y}(X=0, Y=1) \\
 &\quad + \mathbb{P}_{X,Y}(X=1, Y=0) \\
 &\quad + \mathbb{P}_{X,Y}(X=1, Y=1) + \mathbb{P}_{X,Y}(X=2, Y=0) \\
 &\quad + \mathbb{P}_{X,Y}(X=2, Y=1) \\
 &= 0.05 + 0.21 + 0.20 + 0.26 + 0 + 0.06 = 0.78 .
 \end{aligned}$$

□

Вправа 10.4.2. Нехай Y — інтенсивність надходження дзвінків на кол-центр (у дзвінках на хвилину). Нехай X — число дзвінків протягом 2 годин. Спільна щільність розподілу $(X, Y)^T$ дорівнює

$$f_{X,Y}(x, y) = \frac{(2y)^x}{x!} e^{-3y} \cdot \mathbb{1}\{y > 0\} \cdot \mathbb{1}\{x \in \mathbb{N} \cup \{0\}\} .$$

Чи справді це є щільність? Чому дорівнює $\mathbb{P}_X(X = 0)$? □

Розв'язання. Щільність, вочевидь, невід'ємна, тому потрібно перевірити, чи дорівнює 1 її інтеграл. Оскільки щільність невід'ємна, можемо застосувати теорему Фубіні, щоб перейти від подвійного інтегралу до повторного:

$$\begin{aligned} \iint f_{X,Y}(x,y) \#(dx) \times \lambda(dy) &= \int_0^\infty \left(\sum_{x=0}^\infty \frac{(2y)^x}{x!} e^{-3y} \right) dy \\ &= \int_0^\infty (e^{-3y} e^{2y}) dy \\ &= \int_0^\infty e^{-y} dy = 1, \end{aligned}$$

де ми використали розвинення e^{2y} в ряд Тейлора.

Тоді (маржинальну) ймовірність $X = 0$ можна дістати, інтегруючи спільну щільність за змінною y :

$$\mathbb{P}_X(X = 0) = \int_0^\infty \frac{(2y)^0}{0!} e^{-3y} dy = \frac{1}{3}. \quad \square$$

Вправа 10.4.3. Нехай $X, Y \sim U((0; 1])$, $X \perp\!\!\!\perp Y$. Чому дорівнює середньоквадратичне відхилення відстані між ними? □

Розв'язання. Нехай $W = |X - Y|$. Згідно з законом несвідомого статистика, ми можемо обчислити це сподівання як

$$\mathbb{E}[W] = \iint_{\mathbb{R}} |x - y| \cdot f_{X,Y}(x,y) dx dy.$$

Оскільки $X \perp\!\!\!\perp Y$, їхня спільна щільність є добутком щільностей маржинальних, а тому, використовуючи теорему Фубіні для невід'ємної підінтегральної функції, маємо

$$\begin{aligned} \mathbb{E}[W] &= \int_0^1 \int_0^1 |x - y| dx dy \\ &= \int_0^1 \left(\int_0^y (y - x) dx + \int_y^1 (x - y) dx \right) dy \\ &= \int_0^1 \left(y^2 - \frac{y^2}{2} + \frac{1}{2} - \frac{y^2}{2} - y + y^2 \right) dy = \frac{1}{3}. \end{aligned}$$

Для підрахунку дисперсії потрібно обчислити $\mathbb{E}[W^2]$. Це можна зробити за законом несвідомого статистика, але альтернативно можна помітити, що

$$\begin{aligned}\mathbb{E}[W^2] &= \mathbb{E}[|X - Y|^2] = \mathbb{E}[(X - Y)^2] = \mathbb{E}[X^2] - 2\mathbb{E}[XY] + \mathbb{E}[Y^2] \\ &= \mathbb{E}[X^2] - 2\mathbb{E}[X]\mathbb{E}[Y] + \mathbb{E}[Y^2],\end{aligned}$$

де ми використали незалежність X та Y . З урахуванням рівності розподілів X та Y маємо

$$\mathbb{E}[W^2] = 2\mathbb{E}[X^2] - 2(\mathbb{E}[X])^2 = 2\text{Var}(X) = \frac{1}{6}.$$

Остаточно

$$\text{sd}(W) = \sqrt{\text{Var}(W)} = \sqrt{\mathbb{E}[W^2] - (\mathbb{E}[W])^2} = \sqrt{\frac{1}{6} - \frac{1}{9}} = \frac{1}{3\sqrt{2}}.$$

□

Вправа 10.4.4. Чи може бути, що $\mathbb{E}[X] = 3$, $\mathbb{E}[Y] = 2$, $\mathbb{E}[X^2] = 10$, $\mathbb{E}[Y^2] = 29$, а $\mathbb{E}[XY] = 0$? □

Розв'язання. За визначенням коваріації,

$$\text{Cov}(X, Y) = \mathbb{E}[XY] - \mathbb{E}[X]\mathbb{E}[Y] = 0 - 3 \cdot 2 = -6.$$

Відповідні дисперсії дорівнюють

$$\begin{aligned}\text{Var}(X) &= \mathbb{E}[X^2] - (\mathbb{E}[X])^2 = 10 - 9 = 1, \\ \text{Var}(Y) &= \mathbb{E}[Y^2] - (\mathbb{E}[Y])^2 = 29 - 4 = 25.\end{aligned}$$

Тоді коефіцієнт кореляції повинен дорівнювати

$$\text{Corr}(X, Y) = \frac{\text{Cov}(X, Y)}{\sqrt{\text{Var}(X)\text{Var}(Y)}} = \frac{-6}{\sqrt{1 \cdot 25}} = -\frac{6}{5} < -1,$$

що неможливо. □

Вправа 10.4.5. Нехай X , Y та Z — деякі випадкові величини такі, що $\text{Var}(X) = 1$, $\text{Var}(Y) = 4$, $\text{Var}(Z) = 8$, $\text{Cov}(X, Y) = 1$, $\text{Cov}(X, Z) = -1$ і $\text{Cov}(Y, Z) = 2$.

Чому дорівнюють $\text{Var}(X + Y + Z)$ та $\text{Var}(3X - Y - 2Z + 1)$? □

Розв'язання. За визначенням дисперсії суми випадкових величин маємо

$$\begin{aligned}\operatorname{Var}(X + Y + Z) &= \operatorname{Var}(X) + \operatorname{Var}(Y) + \operatorname{Var}(Z) \\ &\quad + 2\operatorname{Cov}(X, Y) + 2\operatorname{Cov}(X, Z) + 2\operatorname{Cov}(Y, Z) \\ &= 1 + 4 + 8 + 2 - 2 + 4 = 17 .\end{aligned}$$

Так само можна підрахувати другу дисперсію, згадавши про властивість дисперсії $\operatorname{Var}(aX) = a^2\operatorname{Var}(X)$ та лінійність коваріації:

$$\begin{aligned}\operatorname{Var}(3X - Y - 2Z + 1) &= \operatorname{Var}(3X) + \operatorname{Var}(-Y) + \operatorname{Var}(-2Z) + \operatorname{Var}(1) \\ &\quad + 2\operatorname{Cov}(3X, -Y) + 2\operatorname{Cov}(3X, -2Z) + 2\operatorname{Cov}(3X, 1) \\ &\quad + 2\operatorname{Cov}(-Y, -2Z) + 2\operatorname{Cov}(-Y, 1) + 2\operatorname{Cov}(-2Z, 1) \\ &= 9\operatorname{Var}(X) + \operatorname{Var}(Y) + 4\operatorname{Var}(Z) + 0 \\ &\quad - 6\operatorname{Cov}(X, Y) - 12\operatorname{Cov}(X, Z) + 0 \\ &\quad + 4\operatorname{Cov}(Y, Z) + 0 + 0 \\ &= 9 \cdot 1 + 4 + 4 \cdot 8 - 6 \cdot 1 + 12 \cdot 1 + 4 \cdot 2 = 59 .\end{aligned}$$

□

11 Випадкові вектори та незалежні випадкові величини. Частина 2

11.1 Основні теоретичні відомості

Детальний огляд понять, потрібний для розв'язання задач на цьому занятті, наведено в Розд. 10, а також у Курсі лекцій, Розд. КЛ9.5. Зокрема, нас цікавлять такі результати.

Твердження (КЛ9.5.1). Нехай X та Y — дві незалежні випадкові величини зі щільностями розподілу f_X і f_Y відповідно. Тоді щільність суми $S = X + Y$ дорівнює

$$f_S(s) = \int_{-\infty}^{\infty} f_X(s-y) f_Y(y) dy. \quad (\text{КЛ9.5.1})$$

Інтеграл у (КЛ9.5.1) має назву *згортки* (convolution).

У застосуванні до дискретних величин формула (КЛ9.5.1) набуває виду

$$\mathbb{P}_S(S = s) = \sum_{j=-\infty}^{\infty} \mathbb{P}_X(X = s-j) \mathbb{P}_Y(Y = j). \quad (\text{КЛ9.5.2})$$

Цей результат можна застосувати до обчислення щільностей сум інших розподілів, які ми вивчали раніше в цьому курсі.

Твердження (КЛ9.5.4). 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 $S \sim N\left(\sum_{i=1}^k \mu_i, \sum_{i=1}^k \sigma_i^2\right)$.

11.2 Незалежність випадкових величин

Вправа 11.2.1. Нехай $X, Y, Z \sim U((0; 1])$, і всі вони незалежні. Чому дорівнює $\mathbb{P}_{X,Y,Z}(X \geq YZ)$? $\square$

Розв'язання. Оскільки відповідні величини незалежні, їхню спільну щільність можна розписати як добуток:

$$\begin{aligned} f_{X,Y,Z}(x, y, z) &= f_X(x)f_Y(y)f_Z(z) \\ &= \mathbb{1}\{0 < x < 1\} \cdot \mathbb{1}\{0 < y < 1\} \cdot \mathbb{1}\{0 < z < 1\}. \end{aligned}$$

Тоді відповідну ймовірність, як завжди, можна дістати як інтеграл за відповідною множиною. Оскільки підінтегральна функція невід'ємна, застосовна теорема Фубіні для переходу від кратного інтегралу до повторного:

$$\begin{aligned} \mathbb{P}_{X,Y,Z}(X \geq YZ) &= \iiint_{x \geq yz} f_{X,Y,Z}(x, y, z) dx dy dz \\ &= \int_0^1 \int_0^1 \int_{yz}^1 dx dy dz \\ &= \int_0^1 \int_0^1 (1 - yz) dy dz \\ &= \int_0^1 \left(1 - \frac{z}{2}\right) dz = \frac{3}{4}. \end{aligned}$$

$\square$

Вправа 11.2.2. Розгляньмо гру крепс (craps). Суть гри полягає у викиданні двох гральних кісточок згідно з певними правилами. Якщо сума чисел у першому викиді дорівнює 2, 3 або 12, гравець програє. Якщо сума дорівнює 7 або 11, гравець виграє. Якщо сума дорівнює деякому іншому значенню i , гравець продовжує викидати кісточки доти, доки сума не буде дорівнювати i або 7. Якщо сума дорівнює i , гравець виграє, а якщо 7, то гравець програє.

Нехай сума чисел у першому викиді дорівнює 4. Нехай N = «кількість викидів до появи суми в 4 або 7» і нехай X = «сума чисел на останньому викиді» (тобто $X \in \{4, 7\}$.) Чи можна казати, що $N \perp\!\!\!\perp X$? $\square$

Розв'язання. Суто теоретично дати відповідь на це питання можна, розглянувши спільну та маржинальні функції ймовірності N та X , тобто розглянувши всі можливі значення з відповідних носіїв та супутні ймовірності.

Це справа доволі невдячна, тому відповідь можна дістати за рахунок засадничої характеристики незалежних подій як таких, що інформація про настання однієї з них не впливає на ймовірність настання іншої.

До того ж, потрібно пам'ятати, що незалежність — це симетрична властивість, і інколи аналіз впливу однієї величини на іншу доцільно робити в один бік, а не в інший. Наприклад, на перший погляд важко сказати, чи проливає інформація про те, чому дорівнює сума останнього викиду (4 або 7), світло на значення випадкової величини N .

Розглянувши це питання в інший бік, бачимо, що відповідь очевидна. Справді, не викликає сумнівів, що інформація $N = n$ для деякого $n \in \mathbb{N}$ у жодний спосіб не впливає на суму чисел в останньому викиді, оскільки кожний викид не залежить від попереднього. Тому $N \perp\!\!\!\perp X$. $\square$

Вправа 11.2.3. Нехай маємо послідовність $X_1, X_2, \dots$ незалежних випадкових величин з однаковим розподілом. Якщо $X_n > X_i$, $i = 1, \dots, n-1$, таке значення називають *рекордним* (record value). Позначмо через A_n подію, що X_n має рекордне значення. Чи можна казати, що $A_{n+1} \perp\!\!\!\perp A_n$? $\square$

Розв'язання. Як і у Вправі 11.2.2, проаналізуємо цю ситуацію з міркувань впливу настання однієї події на ймовірність настання іншої. Нехай сталася подія A_n , тобто X_n має рекордне значення. Чи має інформація про цей факт вплив на ймовірність значень X_{n+1} , у тому числі на те, чи буде таке значення рекордним? Можливо, якщо X_n має дуже велике значення, то ймовірність, що X_{n+1} рекордна, стає меншою? На перший погляд відповідь не зовсім очевидна.

Проте розгляд цього питання з іншого боку суттєво спрощує міркування. Справді, нехай сталася подія A_{n+1} . Чи ця інформація в якийсь спосіб впливає на ймовірність настання події A_n ? Очевидно, що ні, адже нам відомо тільки те, що $X_{n+1} > X_i$, $i = 1, \dots, n$, але нам зовсім невідомо, яке саме з i -их значень було попереднім рекордним. Отже $A_n \perp\!\!\!\perp A_{n+1}$. $\square$

Вправа 11.2.4. Нехай спільна щільність випадкових величин X та Y є

$$f_{X,Y}(x, y) = 2xe^{-y} \cdot \mathbb{1}\{0 \leq x \leq 1\} \cdot \mathbb{1}\{y > 0\}.$$

Чи можна казати, що $X \perp\!\!\!\perp Y$? $\square$

Розв'язання. Згідно з Теоремою КЛ9.4.5, для незалежності двох величин потрібно, щоб їхня спільна щільність розпадалася в добуток двох маржинальних щільностей. У нашому прикладі спільна щільність очевидно є добутком двох функцій однієї змінної:

$$f_{X,Y}(x, y) = (2x \cdot \mathbb{1}\{0 \leq x \leq 1\}) \cdot (e^{-y} \cdot \mathbb{1}\{y > 0\}).$$

Це справді є щільності, оскільки їхні інтеграли дорівнюють 1:

$$\begin{aligned} \int_{-\infty}^{\infty} 2x \cdot \mathbb{1}\{0 \leq x \leq 1\} dx &= \int_0^1 2x dx = x^2 \Big|_0^1 = 1, \\ \int_{-\infty}^{\infty} e^{-y} \cdot \mathbb{1}\{y > 0\} dy &= \int_0^{\infty} e^{-y} dy = -e^{-y} \Big|_0^{\infty} = 1. \end{aligned} \quad \square$$

Як ми у Вправі 11.2.4 знали, що коефіцієнт 2 повинен належати саме функції від x , а не від y ? Заздалегідь цього передбачити, звісно, неможливо. Нехай, наприклад, X та Y мають деякі розподіли зі щільностями $ag(x)$ та $bh(y)$ відповідно, $X \perp Y$, $c = ab$. Тоді їхня спільна щільність дорівнює

$$f_{X,Y}(x, y) = cg(x)h(y).$$

У загальному випадку, не знаючи, яка функція відповідає якій величині, важко сказати, як «розділити» константу c між двома маржинальними щільностями, щоб інтеграл дорівнював 1. Проте завжди можна взяти інтеграли відповідних функцій окремо:

$$\int g(x) dx = \frac{1}{a}, \quad \int h(y) dy = \frac{1}{b},$$

і тоді стане зрозуміло, які коефіцієнти використати у відповідних маржинальних щільностях. Зокрема, у Вправі 11.2.4 можна було б порахувати

$$\int_{-\infty}^{\infty} x \cdot \mathbb{1}\{0 \leq x \leq 1\} dx = \frac{1}{2}$$

і зробити висновок, що двійка повинна належати саме щільності f_X .

Безпосередньо з цієї дискусії випливає результат, що дає змогу з'ясувати, чи є випадкові величини незалежними. Випадкові величини X та Y незалежні тоді й тільки тоді, коли для деяких функцій однієї змінної g та h

$$f_{X,Y}(x, y) \propto g(x)h(y).$$

Вправа 11.2.5. Нехай спільна щільність випадкових величин X та Y є

$$f_{X,Y}(x, y) = \frac{3}{2}y^2 \cdot \mathbb{1}\{0 \leq x \leq 2\} \cdot \mathbb{1}\{0 \leq y \leq 1\}.$$

Чи є $X \perp Y$? Чому дорівнюють маржинальні щільності X та Y ? $\square$

Розв'язання. Величини X та Y справді є незалежні, оскільки

$$f_{X,Y}(x, y) \propto g(x)h(y) = (\mathbb{1}\{0 \leq x \leq 2\}) \cdot (y^2 \cdot \mathbb{1}\{0 \leq y \leq 1\}).$$

Маржинальні щільності можна дістати або інтегруванням спільної щільності за кожною змінною окремо:

$$f_X(x) = \int_{-\infty}^{\infty} f_{X,Y}(x, y) dy = \int_0^1 \frac{3}{2} y^2 \cdot \mathbb{1}\{0 \leq x \leq 2\} dy = \frac{1}{2} \cdot \mathbb{1}\{0 \leq x \leq 2\} ,$$

$$f_Y(y) = \int_{-\infty}^{\infty} f_{X,Y}(x, y) dx = \int_0^2 \frac{3}{2} y^2 \cdot \mathbb{1}\{0 \leq y \leq 1\} dx = 3y^2 \cdot \mathbb{1}\{0 \leq y \leq 1\} ,$$

або інтегруючи відповідні функції g та h :

$$\int_{-\infty}^{\infty} g(x) dx = \int_0^2 dx = 2 ,$$

$$\int_{-\infty}^{\infty} h(y) dy = \int_0^1 y^2 dy = \frac{1}{3} ,$$

звідки

$$f_X(x) = \frac{1}{2} g(x) , \quad f_Y(y) = 3h(y) ,$$

що відповідає попереднім результатам. $\square$

Вправа 11.2.6. Нехай спільна щільність випадкових величин X та Y є

$$f_{X,Y}(x, y) = 24xy \cdot \mathbb{1}\{x \geq 0\} \cdot \mathbb{1}\{y \geq 0\} \cdot \mathbb{1}\{x + y \leq 1\} .$$

Чи можна казати, що $X \perp\!\!\!\perp Y$? $\square$

Розв'язання. На перший погляд відповідь ствердна, адже $24xy \propto xy$, проте це зовсім не відповідає дійсності, адже функція $f_{X,Y}$ містить у собі індикаторні функції, одну з яких, $\mathbb{1}\{x + y \leq 1\}$, неможливо розкласти в добуток функцій однієї змінної. Отже величини X та Y не є незалежні. $\square$

11.3 Обчислення сподівань через суму індикаторних величин

Розгляньмо поширений на практиці спосіб обчислення сподівань випадкових величин, розподіл яких або складно вивести, або сам по собі не є цікавий. Цей спосіб передбачає розкладення шуканої величини в суму деяких індикаторних, сподівання яких легко взяти за допомогою лінійності як сподівання випадкових величин Бернуллі.

Вправа 11.3.1. Повернімося до задачі де Монморта з Прикладу 3.3.1. Тоді ми розглядали колоду з n карт, пронумерованих від 1 до n . Гравець по черзі

перегортає карти і виграє у тому випадку, якщо i -та перегорнута карта мала нанесений на ній номер i . Ми з'ясували, що ймовірність виграшу приблизно дорівнює $1 - \frac{1}{e}$. Чому дорівнює сподіване число карт, для яких нанесений на карті номер відповідає позиції, на якій вона розташована? $\square$

Розв'язання. Розглянемо для кожної карти випадкову величину X_i = «карта на позиції i має нанесений на ній номер i », $i = 1, \dots, n$. Це буде величина Бернуллі з ймовірністю успіху $\mathbb{P}_{X_i}(X_i = 1) = \frac{1}{n}$ (відповідно до міркувань симетрії). Загальне число карт, для яких нанесений номер i номер позиції дорівнюють одне одному, є $X = X_1 + \dots + X_n$. Отже

$$\mathbb{E}[X] = \sum_{i=1}^n \mathbb{E}[X_i] = \sum_{i=1}^n \mathbb{P}_{X_i}(X_i = 1) = \sum_{i=1}^n \frac{1}{n} = 1. \quad \square$$

Вправа 11.3.2. Телефонні номери в кількості k штук зберігають у хеш-таблиці з n комірками. Чому дорівнює сподівана кількість комірок, у яких не міститься жодного телефонного номера? Один номер? Більше одного? Вважатимемо, що ймовірність потрапити в кожну комірку однакова. $\square$

Розв'язання. Нехай X_i = «комірка i порожня», $i = 1, \dots, n$. Тоді $X_i = 1$, якщо всі телефонні номери потрапили в деяку комірку, окрім i -ої. Потрапити в комірку i можна з ймовірністю $\frac{1}{n}$, а тому

$$\mathbb{P}_{X_i}(X_i = 1) = \left(1 - \frac{1}{n}\right)^k.$$

Тоді шукане сподівання дорівнює

$$\mathbb{E}[X] = \sum_{i=1}^n \mathbb{E}[X_i] = n \left(1 - \frac{1}{n}\right)^k.$$

Розглянемо тепер величини Y_i = «комірка i містить точно один телефонний номер», $i = 1, \dots, n$. Подія $Y_i = 1$ станеться, якщо всі номери, окрім одного, потрапили в інші комірки. Ймовірність, що $k-1$ номерів потрапили в інші комірки, за аналогією з попереднім випадком є $\left(1 - \frac{1}{n}\right)^{k-1}$, а ймовірність для деякого номера потрапити в конкретну комірку є $\frac{1}{n}$. Усього є k варіантів, за яких один номер потрапив у комірку i , а всі інші номери потрапили в інші комірки, тому остаточно за правилом множення маємо

$$\mathbb{P}_{Y_i}(Y_i = 1) = \frac{k}{n} \left(1 - \frac{1}{n}\right)^{k-1}.$$

Отже

$$\mathbb{E}[Y] = \sum_{i=1}^n \mathbb{E}[Y_i] = k \left(1 - \frac{1}{n}\right)^{k-1}.$$

Сподіване число комірок із понад одним номером $\mathbb{E}[W]$ можна обчислити як різницю сподіваного числа комірок із *будь-яким* числом номерів $\mathbb{E}[Z]$, яке, очевидно, дорівнює n , та щойно здобутими результатами:

$$\mathbb{E}[W] = \mathbb{E}[Z] - \mathbb{E}[X] - \mathbb{E}[Y] = n - n \left(1 - \frac{1}{n}\right)^k - k \left(1 - \frac{1}{n}\right)^{k-1}. \quad \square$$

11.4 Суми незалежних випадкових величин

Твердження КЛ9.5.4 доведено в Розд. КЛ9.6.5, проте для ліпшого розуміння згортки доцільно розглянути доведення окремих результатів тут.

Доведення. (i) Здійснімо доведення за індукцією. Нехай $X_1, X_2 \sim \text{Bern}(p)$. За формулою (КЛ9.5.2) для $S_2 = X_1 + X_2$ маємо:

$$\begin{aligned} \mathbb{P}_{S_2}(S_2 = s) &= \sum_{j=-\infty}^{\infty} \mathbb{P}_{X_1}(X_1 = s - j) \mathbb{P}_{X_2}(X_2 = j) \\ &= \mathbb{P}_{X_1}(X_1 = s - 0) \mathbb{P}_{X_2}(X_2 = 0) \\ &\quad + \mathbb{P}_{X_1}(X_1 = s - 1) \mathbb{P}_{X_2}(X_2 = 1) \\ &= (1 - p) \mathbb{P}_{X_1}(X_1 = s) + p \mathbb{P}_{X_1}(X_1 = s - 1). \end{aligned}$$

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

$$\mathbb{P}_{S_2}(S_2 = s) = \binom{2}{s} p^s (1 - p)^{2-s}, \quad s = 0, 1, 2,$$

а отже $S_2 \sim \text{Binom}(2, p)$.

Нехай $S_{k-1} = \sum_{i=1}^{k-1} X_i \sim \text{Binom}(k-1, p)$ і $X_k \sim \text{Bern}(p)$. Тоді для $S_k = S_{k-1} + X_k$ за формулою згортки маємо

$$\begin{aligned} \mathbb{P}_{S_k}(S_k = s) &= \sum_{j=-\infty}^{\infty} \mathbb{P}_{S_{k-1}}(S_{k-1} = s - j) \mathbb{P}_{X_k}(X_k = j) \\ &= \mathbb{P}_{S_{k-1}}(S_{k-1} = s - 0) \mathbb{P}_{X_k}(X_k = 0) \end{aligned}$$

$$\begin{aligned}
& + \mathbb{P}_{S_{k-1}}(S_{k-1} = s - 1) \mathbb{P}_{X_k}(X_k = 1) \\
& = (1 - p) \mathbb{P}_{S_{k-1}}(S_{k-1} = s) + p \mathbb{P}_{S_{k-1}}(S_{k-1} = s - 1) \\
& = (1 - p) \cdot \binom{k-1}{s} p^s (1 - p)^{k-1-s} \\
& + p \cdot \binom{k-1}{s-1} p^{s-1} (1 - p)^{k-1-(s-1)} \\
& = p^s (1 - p)^{k-s} \left(\binom{k-1}{s} + \binom{k-1}{s-1} \right) \\
& = \binom{k}{s} p^s (1 - p)^{k-s},
\end{aligned}$$

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 $S_k \sim \text{Binom}(k, p)$;

- (iv) нехай маємо дві незалежні величини $X_1 \sim N(0, \sigma^2)$, $X_2 \sim N(0, 1)$. Застосування (КЛ9.5.1) разом із прийомом із виділенням щільності деякого нормального розподілу, інтеграл від якої дорівнює 1, дає:

$$\begin{aligned}
f_{X_1+X_2}(s) &= \int_{-\infty}^{\infty} f_{X_1}(s-y) f_{X_2}(y) dy = \int_{-\infty}^{\infty} \frac{1}{2\pi\sigma} e^{-\frac{(s-y)^2}{2\sigma^2} - \frac{y^2}{2}} dy \\
&= \int_{-\infty}^{\infty} \frac{1}{2\pi\sigma} e^{-\frac{s^2 - 2sy + y^2 - y^2\sigma^2}{2\sigma^2}} dy \\
&= \int_{-\infty}^{\infty} \frac{1}{2\pi\sigma} e^{-\left(\frac{y^2(1+\sigma^2)}{2\sigma^2} - y\frac{s}{\sigma^2} + \frac{s^2}{2\sigma^2}\right)} dy \\
&= \int_{-\infty}^{\infty} \frac{1}{2\pi\sigma} e^{-\left(\left(\sqrt{\frac{1+\sigma^2}{2\sigma^2}}y - \frac{s}{2\sigma^2} \cdot \sqrt{\frac{2\sigma^2}{1+\sigma^2}}\right)^2 + \frac{s^2}{2\sigma^2} - \left(\frac{s}{2\sigma^2}\right)^2 \cdot \frac{2\sigma^2}{(1+\sigma^2)}\right)} dy
\end{aligned}$$

$$\begin{aligned}
&= \frac{1}{\sqrt{2\pi}\sigma} e^{-\left(\frac{s^2}{2\sigma^2} - \left(\frac{s}{2\sigma^2}\right)^2 \cdot \frac{2\sigma^2}{1+\sigma^2}\right)} \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{\left(y - \frac{s}{2\sigma^2} \cdot \frac{2\sigma^2}{1+\sigma^2}\right)^2}{2 \cdot \frac{\sigma^2}{1+\sigma^2}}} dy \\
&= \frac{\sigma}{\sqrt{1+\sigma^2}} \cdot \frac{1}{\sqrt{2\pi}\sigma} e^{-\left(\frac{s^2}{2\sigma^2} - \left(\frac{s}{\sigma^2}\right)^2 \cdot \frac{2\sigma^2}{4(1+\sigma^2)}\right)} \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi} \frac{\sigma}{\sqrt{1+\sigma^2}}} e^{-\frac{\left(y - \frac{s}{1+\sigma^2}\right)^2}{2 \cdot \frac{\sigma^2}{1+\sigma^2}}} dy \\
&= \frac{1}{\sqrt{2\pi(1+\sigma^2)}} e^{-\frac{s^2}{2(1+\sigma^2)}} ,
\end{aligned}$$

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 $X_1 \sim N(\mu_1, \sigma_1^2)$, $X_2 \sim N(\mu_2, \sigma_2^2)$, $X_1 \perp X_2$:

$$X_1 + X_2 = \sigma_2 \left(\frac{X_1 - \mu_1}{\sigma_2} + \frac{X_2 - \mu_2}{\sigma_2} \right) + \mu_1 + \mu_2 .$$

Відповідно до Зауваження КЛ7.5.9, $\frac{X_1 - \mu_1}{\sigma_2} \sim N\left(0, \frac{\sigma_1^2}{\sigma_2^2}\right)$, а $\frac{X_2 - \mu_2}{\sigma_2} \sim N(0, 1)$, тому використаємо попередній результат, помітивши, 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 Зауваження КЛ7.5.9,

$$X_1 + X_2 \sim N\left(\mu_1 + \mu_2, \sigma_2^2 \left(1 + \frac{\sigma_1^2}{\sigma_2^2}\right)\right) = N(\mu_1 + \mu_2, \sigma_1^2 + \sigma_2^2) .$$

За індукцією впливає, що

$$\sum_{i=1}^k X_i \sim N\left(\sum_{i=1}^k \mu_i, \sum_{i=1}^k \sigma_i^2\right) . \quad \square$$

Вправа 11.4.1. У великій корпорації заробітні платні працівників віком понад 30 років мають (приблизно) нормальний розподіл зі сподіванням \$60 000 та середньоквадратичним відхиленням \$10 000. Для працівників віком до 30 років зарплатні мають (приблизно) нормальний розподіл зі сподіванням \$40 000 та середньоквадратичним відхиленням \$10 000.

Нехай у випадковий спосіб вибрано двох працівників віком за 30 років. Яка ймовірність, що середнє значення їхніх зарплат перевищує \$65 000?

Нехай у випадковий спосіб вибрано по працівнику з обох вікових категорій. Яка ймовірність, що зарплата молодшого з них вища? Що менша з двох зарплат перевищує \$50 000? □

Розв'язання. Середнє між двома випадково вибраними зарплатами X_1 та X_2 дорівнює $\frac{1}{2}(X_1 + X_2)$. Оскільки $X_1 \perp\!\!\!\perp X_2$, згідно з Твердженням КЛ9.5.4,

$$\begin{aligned} \frac{1}{2}(X_1 + X_2) &\sim N\left(\frac{1}{2}(\mathbb{E}[X_1] + \mathbb{E}[X_2]), \frac{1}{2^2}(\text{Var}(X_1) + \text{Var}(X_2))\right) \\ &= N(60, 50) , \end{aligned}$$

де одиницею виміру зарплатні є тисяча доларів США. Зверніть увагу, що ми додаємо *дисперсії*, а не *середньоквадратичні відхилення*.

Тоді, переходячи до стандартного нормального розподілу, маємо

$$\mathbb{P}_{X_1, X_2}\left(\frac{1}{2}(X_1 + X_2) > 65\right) = \mathbb{P}_Z\left(Z > \frac{65 - 60}{\sqrt{50}}\right) = 1 - \Phi\left(\frac{1}{\sqrt{2}}\right) \approx 0.240 .$$

Розгляньмо тепер величини Y_1 = «зарплата у випадковий спосіб обраного працівника віком до 30» та Y_2 = «зарплата у випадковий спосіб обраного працівника віком за 30». Нас цікавить імовірність $Y_1 > Y_2$, або ж $Y_1 - Y_2 > 0$. Згідно з Твердженням КЛ9.5.4, враховуючи незалежність Y_1 та Y_2 , маємо

$$Y_1 - Y_2 \sim N(\mathbb{E}[Y_1] - \mathbb{E}[Y_2], \text{Var}(Y_1) + (-1)^2 \text{Var}(Y_2)) = N(-20, 200) ,$$

а отже

$$\mathbb{P}_{Y_1, Y_2}(Y_1 - Y_2 > 0) = \mathbb{P}_Z\left(Z > \frac{0 - (-20)}{\sqrt{200}}\right) = 1 - \Phi\left(\frac{2}{\sqrt{2}}\right) \approx 0.079 .$$

Менша з двох зарплатень, $L = \min\{Y_1, Y_2\}$, з урахуванням незалежності перевищуватиме \$50 000 з такою ймовірністю:

$$\begin{aligned} \mathbb{P}_L(L > 50) &= \mathbb{P}_{Y_1}(Y_1 > 50) \mathbb{P}_{Y_2}(Y_2 > 50) \\ &= \left(1 - \Phi\left(\frac{50 - 40}{10}\right)\right) \left(1 - \Phi\left(\frac{50 - 60}{10}\right)\right) \approx 0.133 . \quad \square \end{aligned}$$

Приклад 11.4.2. Нехай $S(n)$ — ціна деякого цінного паперу по завершенні $n \geq 1$ тижнів. На практиці використовують модель, згідно з якою ціна таких паперів має властивість, що $S(n)/S(n-1)$, $n \geq 1$ — незалежні випадкові величини з логнормальним розподілом (Приклад КЛ7.4.5).

Нехай $\mu = 0.02$, $\sigma = 0.07$. Підрахуймо ймовірність, що ціна паперів через два тижні буде вищою від сьогоднішньої:

$$\begin{aligned} \mathbb{P}_{S(0), S(2)}\left(\frac{S(2)}{S(0)} > 1\right) &= \mathbb{P}_{S(0), S(1), S(2)}\left(\frac{S(2)}{S(1)} \cdot \frac{S(1)}{S(0)} > 1\right) \\ &= \mathbb{P}_{S(0), S(1), S(2)}\left(\ln\left(\frac{S(2)}{S(1)}\right) + \ln\left(\frac{S(1)}{S(0)}\right) > 0\right) . \end{aligned}$$

Оскільки відношення двох цін має логнормальний розподіл із параметрами $\mu = 0.02$, $\sigma = 0.07$, його логаритм має нормальний розподіл із параметрами $\mu = 0.02$, $\sigma = 0.07$, а сума логаритмів, згідно з Твердженням КЛ9.5.4, має нормальний розподіл із параметрами $\mu' = 2 \cdot 0.02$, $(\sigma')^2 = 2 \cdot (0.07)^2$. Отже

$$\begin{aligned} \mathbb{P}_{S(0), S(2)} \left(\frac{S(2)}{S(0)} > 1 \right) &= \mathbb{P}_Z \left(Z > \frac{0 - 0.04}{\sqrt{2} \cdot 0.07} \right) \\ &= 1 - \Phi \left(-\frac{0.04}{\sqrt{2} \cdot 0.07} \right) \approx 0.657. \end{aligned} \quad \square$$

11.5 Додаткові вправи

Вправа 11.5.1. Розгляньмо групу з n осіб та підрахуймо сподіване число різних днів народження. Чому дорівнює сподіване число пар осіб з однаковим днем народження? Як завжди, вважатимемо, що в році 365 днів, і всі дні народження рівноймовірні. $\square$

Розв'язання. Розгляньмо індикаторні величини $X_i = \text{«серед } n \text{ осіб є принаймні одна, що народилася в день } i\text{»}$, $i = 1, \dots, 365$. Оскільки маємо слово «принаймні», імовірність $\mathbb{P}_{X_i}(X_i = 1)$ обчислімо через доповнення:

$$\mathbb{P}_{X_i}(X_i = 1) = 1 - \mathbb{P}_{X_i}(X_i = 0) = 1 - \left(\frac{364}{365} \right)^n,$$

оскільки дня народження *не буде* серед днів народження членів групи, якщо *всі* члени групи народяться в інші дні. Звідси

$$\mathbb{E}[X] = \sum_{i=1}^n \mathbb{E}[X_i] = 365 \left(1 - \left(\frac{364}{365} \right)^n \right).$$

Розгляньмо тепер пари осіб, що народилися в один день. Із групи в n осіб можна утворити $\binom{n}{2}$ пар. Для кожної пари уведемо індикаторну величину $Y_j = \text{«особи з пари } j \text{ мають однаковий день народження»}$, $j = 1, \dots, \binom{n}{2}$. Імовірність успіху для такої величини є ймовірністю того, що пара матиме за день народження один із 365 можливих, тому $\mathbb{P}_{Y_j}(Y_j = 1) = \frac{1}{365}$, а тому

$$\mathbb{E}[Y] = \sum_{i=1}^{\binom{n}{2}} \frac{1}{365} = \frac{\binom{n}{2}}{365}. \quad \square$$

Вправа 11.5.2. Кажемо, що перестановка $a_1, \dots, a_n$ цілих чисел від 1 до n має локальний максимум у позиції j , якщо $a_j > a_{j-1}$ та $a_j > a_{j+1}$, $j = 2, \dots, n-1$. Для $j = 1$ локальний максимум досягається, якщо $a_1 > a_2$, а для $j = n$ — якщо $a_{n-1} < a_n$. Наприклад, у перестановці 4, 2, 5, 3, 6, 1 є три локальні максимуми: у позиціях 1, 3 та 5.

Чому дорівнює сподіване число локальних максимумів у перестановках цілих чисел від 1 до n , $n \geq 2$, якщо всі перестановки рівноймовірні? $\square$

Розв'язання. Розгляньмо випадкові величини $X_i = \text{«у позиції } i \text{ досягається локальний максимум»}$, $i = 1, \dots, n$.

Для всіх $j = 2, \dots, n-1$ локальний максимум передбачає, що a_j повинно бути найбільшим із a_{j-1}, a_j, a_{j+1} , що може статися з імовірністю $\frac{1}{3}$, оскільки можливі перестановки рівноймовірні. У схожий спосіб локальний максимум у першій та останній позиціях може статися з імовірністю $\frac{1}{2}$. Тому

$$\mathbb{E}[X] = \sum_{i=1}^n \mathbb{E}[X_i] = \frac{1}{2} + (n-2) \cdot \frac{1}{3} + \frac{1}{2} = \frac{n+1}{3}. \quad \square$$

Вправа 11.5.3. Розгляньмо дві незалежні величини $X, Y \sim \text{Exp}(1)$. Покажіть, що $V = X - Y$ має розподіл Лапласа (Laplace distribution) зі щільністю

$$f_V(v) = \frac{1}{2} e^{-|v|}. \quad \square$$

Розв'язання. Оскільки величини незалежні, застосуємо формулу згортки (КЛ9.5.1). Для цього розгляньмо величину $-Y$ з функцією розподілу

$$\begin{aligned} \mathbb{P}_{-Y}(-Y \leq x) &= \mathbb{P}_Y(Y \geq -x) = 1 - \mathbb{P}_Y(-x) \\ &= 1 - \left(1 - e^{-(x)}\right) \cdot \mathbb{1}\{-x > 0\} \\ &= 1 - (1 - e^x) \cdot \mathbb{1}\{x < 0\} \end{aligned}$$

та щільністю

$$f_{-Y}(x) = e^x \cdot \mathbb{1}\{x < 0\}.$$

Тоді

$$\begin{aligned} f_V(v) &= \int_{-\infty}^{\infty} f_X(v-x) f_{-Y}(x) dx \\ &= \int_{-\infty}^{\infty} e^{-(v-x)} \cdot \mathbb{1}\{v-x > 0\} \cdot e^x \cdot \mathbb{1}\{x < 0\} dx \\ &= \int_{-\infty}^{\infty} e^{2x-v} \cdot \mathbb{1}\{x < \min\{0, v\}\} dx. \end{aligned}$$

Для $v < 0$ маємо

$$f_V(v) = \int_{-\infty}^v e^{2x-v} dx = e^{-v} \cdot \frac{1}{2} e^{2x} \Big|_{-\infty}^v = \frac{1}{2} e^v.$$

Для $v \geq 0$ маємо

$$f_V(v) = \int_{-\infty}^0 e^{2x-v} dx = e^{-v} \cdot \frac{1}{2} e^{2x} \Big|_{-\infty}^0 = \frac{1}{2} e^{-v}.$$

Об'єднуючи обидва результати, маємо

$$f_V(v) = \frac{1}{2} e^{-|v|}. \quad \square$$

Вправа 11.5.4. Нехай істинна вага невеличкої гирки дорівнює 10 г. Її зважують на двох незалежних вагах. Нехай похибка вимірювань на обох вагах має нормальний розподіл із параметрами $\mu = 0$ та $\sigma = 0.2$ г.

Чому дорівнює ймовірність, що результат на других вагах буде ближче до 10 г, аніж на перших? Чому дорівнює ймовірність, що результат на других вагах менше від результату на перших, але не більше від 0.2 г? $\square$

Розв'язання. Позначмо результати на обох вагах через X_1 та X_2 . Тоді нас цікавить ймовірність події $|X_1 - 10| < |X_2 - 10|$. Ми, звісно, можемо порахувати цю ймовірність, інтегруючи спільну щільність за відповідною множиною, проте через симетрію задачі відповідь буде дорівнювати 0.5.

Що стосується другого питання, то нас цікавить ймовірність події $(X_1 - X_2 > 0) \cap (X_1 - X_2 < 0.2) = 0 < X_1 - X_2 < 0.2$. Оскільки $X_1 \sim N(10, 0.2^2)$, $X_2 \sim N(10, 0.2^2)$, $-X_2 \sim N(-10, 0.2^2)$, за формулою згортки маємо $X_1 - X_2 \sim N(10 - 10, 0.2^2 + 0.2^2) = N(0, 0.08)$, а тому

$$\mathbb{P}_{X_1, X_2}(0 < X_1 - X_2 < 0.2) = \mathbb{P}_Z\left(0 < Z < \frac{0.2}{\sqrt{0.08}}\right) \approx 0.260. \quad \square$$

12 Функції від випадкових векторів та їхні розподіли

12.1 Основні теоретичні відомості

Детальний огляд понять, потрібний для розв'язання задач на цьому занятті, наведено в Курсі лекцій, Розд. КЛ10.

Раніше ми розглядали формули заміни змінних для дискретних і неперервних випадкових величин. Вони мають свої аналоги і для випадкових векторів. Зокрема, якщо маємо деякий дискретний 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 функція ймовірності відповідного перетворення $\mathbf{Y} = g(\mathbf{X})$ початкового вектора така:

$$\begin{aligned} \mathbb{P}_{\mathbf{Y}}(Y_1 = y_1, \dots, Y_n = y_n) &= \mathbb{P}_{\mathbf{X}}(g(X_1, \dots, X_k) = (y_1, \dots, y_n)^\top) \\ &= \sum_{\mathbf{x}: g(\mathbf{x})=\mathbf{y}} \mathbb{P}_{\mathbf{X}}(X_1 = x_1, \dots, X_k = x_k) \quad (\text{КЛ10.1.2}) \end{aligned}$$

Для неперервних векторів існує результат, аналогічний Теоремі КЛ7.4.4.

Теорема (Формула заміни змінних для випадкових векторів (Change of variables formula for random vectors), КЛ10.1.1). 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

$$f_{\mathbf{Y}}(\mathbf{y}) = f_{\mathbf{X}}(g_1(\mathbf{y}), \dots, g_k(\mathbf{y})) \cdot |J| \cdot \mathbb{1}_{\{\mathbf{y} \in T\}}. \quad (\text{КЛ10.1.3})$$

Якобіан у твердженні Теорема КЛ10.1.1 виконує роль «похідної оберненої функції».

Розгляньмо важливий розподіл, який є узагальненням експоненційного. У Розд. КЛ8.4 ми з'ясували, що величина з експоненційним розподілом $X \sim \text{Exp}(\lambda)$ відповідає часу очікування до настання певної події в процесі Пуассона з інтенсивністю λ . Розгляньмо розподіл, який характеризує час очікування до настання *декількох подій*.

Визначення (КЛ10.2.1). Додатна випадкова величина X має *Гамма-розподіл* (Gamma distribution) із параметрами $k > 0$, $\lambda > 0$, що позначають через $X \sim \text{Gamma}(k, \lambda)$, якщо її щільність дорівнює

$$f(x) = \frac{1}{\Gamma(k)} x^{k-1} e^{-\lambda x} \lambda^k \cdot \mathbb{1}\{x > 0\}, \quad (\text{КЛ10.2.1})$$

де $\Gamma(k)$ — гамма-функція:

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

□

Функція розподілу для Гамма-розподілу в загальному випадку не має простого аналітичного запису.

Нехай $X \sim \text{Gamma}(k, 1)$. Згідно з (КЛ7.5.2), для величини $Y = \frac{X}{\lambda}$, $\lambda > 0$, маємо:

$$\begin{aligned} f_Y(y) &= f_X(\lambda y) \cdot \lambda = \frac{1}{\Gamma(k)} (\lambda y)^{k-1} e^{-\lambda y} \cdot \mathbb{1}\{\lambda y > 0\} \cdot \lambda \\ &= \frac{1}{\Gamma(k)} y^{k-1} e^{-\lambda y} \lambda^k \cdot \mathbb{1}\{y > 0\}, \end{aligned}$$

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 (КЛ10.2.2). Якщо $X \sim \text{Gamma}(k, \lambda)$, то сподівання $\mathbb{E}[X] = \frac{k}{\lambda}$, а дисперсія $\text{Var}(X) = \frac{k}{\lambda^2}$.

Як можна помітити, розподіл $\text{Gamma}(1, \lambda)$ тотожний експоненційному розподілу з параметром λ . Додатковий параметр k Гамма-розподілу надає йому більше функційної гнучкості. Понад те, сума незалежних експоненційних розподілів має Гамма-розподіл. Це впливає з такого загального твердження.

Твердження (КЛ10.2.3). Нехай $X_i \sim \text{Gamma}(k_i, \lambda)$, $i = 1, \dots, n$. Якщо всі ці випадкові величини незалежні, то розподіл їх суми $S = \sum_{i=1}^n X_i$ такий: $S \sim \text{Gamma}(\sum_{i=1}^n k_i, \lambda)$.

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 λ .

Понад те, позначивши через N_t число подій, які сталися за певний період t , $N_t \sim \text{Pois}(\lambda t)$, маємо залежність:

$$\mathbb{P}_{T_r}(T_r > t) = \mathbb{P}_{N_t}(N_t < r), \quad r \in \mathbb{N}, t \geq 0.$$

Розгляньмо тепер декілька важливих на практиці розподілів випадкових векторів. Почнімо з узагальнення біномного розподілу.

Визначення (КЛ10.3.1). 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 випадковий вектор $\mathbf{X} = (X_1, \dots, X_k)^\top$ має *мультиномний розподіл* (multinomial distribution) із параметрами n і $\mathbf{p} = (p_1, \dots, p_k)^\top$, що позначають через $\mathbf{X} \sim \text{Mult}_k(n, \mathbf{p})$. $\square$

Знайдімо спільну функцію ймовірності такого випадкового вектора.

Твердження (КЛ10.3.2). Якщо випадковий вектор $\mathbf{X} \sim \text{Mult}_k(n, \mathbf{p})$, то його функція ймовірності дорівнює

$$\begin{aligned} \mathbb{P}_{\mathbf{X}}(X_1 = n_1, \dots, X_k = n_k) &= \frac{n!}{n_1! \dots n_k!} p_1^{n_1} \dots p_k^{n_k} \\ &\times \prod_{j=1}^k \mathbb{1}\{n_j \in \mathbb{Z}^+\} \cdot \mathbb{1}\{n_1 + \dots + n_k = n\}, \end{aligned} \quad (\text{КЛ10.3.1})$$

де $\mathbb{Z}^+ = \{0, 1, 2, \dots\}$.

Розгляньмо маржинальні розподіли на основі мультиномного.

Твердження (КЛ10.3.5). Нехай $\mathbf{X} \sim \text{Mult}_k(n, \mathbf{p})$. Розгляньмо випадкові величини $X_1, \dots, X_s$ і $Y = n - (X_1 + \dots + X_s)$ ¹. Нехай $q = 1 - (p_1 + \dots + p_s)$. Тоді $\mathbf{X}' = (X_1, \dots, X_s, Y)^\top \sim \text{Mult}_{s+1}(n, (p_1, \dots, p_s, q)^\top)$.

¹Нумерація величин, звісно, не грає жодної ролі через повну симетрію.

Зауваження (КЛ10.3.7). Із Твердження КЛ10.3.5 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 КЛ4.5.9 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*. Властивість суми біномних величин бути біномною величиною є особливістю мультиномного розподілу. $\square$

Відповідне спостереження можна використати для обчислення сподівання та коваріацій мультиномного розподілу.

Твердження (КЛ10.3.8). Якщо $\mathbf{X} \sim \text{Mult}_k(n, \mathbf{p})$, то

$$\mathbb{E}[\mathbf{X}] = \begin{pmatrix} np_1 \\ \vdots \\ np_k \end{pmatrix}, \quad \text{Cov}(\mathbf{X}) = \begin{pmatrix} np_1(1-p_1) & -np_1p_2 & \dots & -np_1p_k \\ -np_1p_2 & np_2(1-p_2) & \dots & -np_2p_k \\ \vdots & \vdots & \ddots & \vdots \\ -np_kp_1 & -np_kp_2 & \dots & np_k(1-p_k) \end{pmatrix}.$$

Варто звернути увагу, 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.

Інший надзвичайно важливий на практиці є багатовимірний нормальний розподіл, який узагальнює нормальний розподіл на k -вимірні простори.

Визначення (КЛ10.4.1). Випадковий вектор $\mathbf{X} = (X_1, \dots, X_k)^\top$ має *багатовимірний нормальний розподіл* (multivariate normal distribution), що позначають через $\mathbf{X} \sim N(\boldsymbol{\mu}, \Sigma)$, якщо

$$\mathbf{X} = \mathbf{A}\mathbf{Z} + \boldsymbol{\mu},$$

де $\Sigma = \mathbf{A}\mathbf{A}^\top$, $\mathbf{A} \in \mathbb{R}^k \times \mathbb{R}^n$, до того ж Σ невироджена², $\mathbf{Z} = (Z_1, \dots, Z_n)^\top$, $Z_i \sim N(0, 1)$, 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.

Визначення (КЛ10.4.2). Випадковий вектор $\mathbf{X} = (X_1, \dots, X_k)^\top \sim N(\boldsymbol{\mu}, \Sigma)$, якщо будь-яка лінійна комбінація

$$Y = \mathbf{c}^\top \mathbf{X} = \sum_{i=1}^k c_i X_i, \quad \mathbf{c} \in \mathbb{R}^k$$

²Якщо матриця вироджена, то розподіл не матиме щільності. Ми такі випадки не розглядаємо.

його координат є випадковою величиною з нормальним розподілом. $\square$

Ці визначення еквівалентні (Твердження КЛ10.4.3).

Із Визначення КЛ10.4.2 випливає, що маржинальними розподілами для багатовимірного нормального розподілу є також (одновимірні) нормальні розподіли. Зворотне твердження виконується далеко не завжди.

Знайдімо спільну щільність багатовимірного нормального розподілу.

Твердження (КЛ10.4.6). Нехай $\mathbf{X} = (X_1, \dots, X_k)^\top$ можна подати як

$$\mathbf{X} = \mathbf{A}\mathbf{Z} + \boldsymbol{\mu},$$

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

$$f_{\mathbf{X}}(x_1, \dots, x_k) = \frac{1}{\sqrt{(2\pi)^k |\Sigma|}} e^{-\frac{1}{2}(\mathbf{x} - \boldsymbol{\mu})^\top \Sigma^{-1}(\mathbf{x} - \boldsymbol{\mu})}, \quad (\text{КЛ10.4.2})$$

де $\mathbf{x} = (x_1, \dots, x_k)^\top$, $\Sigma = \mathbf{A}\mathbf{A}^\top$, до того ж Σ невироджена.

На практиці дуже часто застосовують двовимірний нормальний розподіл, щільність якого випливає з (КЛ10.4.2):

$$\begin{aligned} f_{X_1, X_2}(x_1, x_2) &= \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}} e^{-\frac{1}{2\sigma_1^2\sigma_2^2(1-\rho^2)} \left((x_1 - \mu_1, x_2 - \mu_2) \begin{pmatrix} \sigma_2^2 & -\rho\sigma_1\sigma_2 \\ -\rho\sigma_1\sigma_2 & \sigma_1^2 \end{pmatrix} \begin{pmatrix} x_1 - \mu_1 \\ x_2 - \mu_2 \end{pmatrix} \right)} \\ &= \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}} e^{-\frac{1}{2(1-\rho^2)} \left(\left(\frac{x_1 - \mu_1}{\sigma_1} \right)^2 - 2\rho \left(\frac{x_1 - \mu_1}{\sigma_1} \right) \left(\frac{x_2 - \mu_2}{\sigma_2} \right) + \left(\frac{x_2 - \mu_2}{\sigma_2} \right)^2 \right)}, \end{aligned}$$

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), (обсяг добив, урожайність), (рівень холестерину, тиск) тощо.

Як і в одновимірному випадку, де параметри розподілу μ і σ^2 відповідали сподіванню та дисперсії нормальної величини, так параметри $\boldsymbol{\mu}$ і Σ відіграють особливу роль.

Твердження (КЛ10.4.10). Якщо $\mathbf{X} \sim N(\boldsymbol{\mu}, \Sigma)$, то сподівання $\mathbb{E}[\mathbf{X}] = \boldsymbol{\mu}$, а матриця коваріацій $\text{Cov}(\mathbf{X}) = \Sigma$.

Особливістю багатовимірного нормального розподілу є не тільки той факт, що всі маржинальні розподіли є нормальними, а й що будь-яка підмножина його координат має багатовимірний нормальний розподіл.

Твердження (КЛ10.4.12). Якщо $\mathbf{X} = (X_1, \dots, X_k)^\top \sim N(\boldsymbol{\mu}, \Sigma)$, де

$$\boldsymbol{\mu} = \begin{pmatrix} \mu_1 \\ \mu_2 \\ \vdots \\ \mu_k \end{pmatrix}, \quad \Sigma = \begin{pmatrix} \text{Var}(X_1) & \text{Cov}(X_1, X_2) & \dots & \text{Cov}(X_1, X_k) \\ \text{Cov}(X_2, X_1) & \text{Var}(X_2) & \dots & \text{Cov}(X_2, X_k) \\ \vdots & \vdots & \ddots & \vdots \\ \text{Cov}(X_k, X_1) & \text{Cov}(X_k, X_2) & \dots & \text{Var}(X_k) \end{pmatrix}.$$

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

$$\boldsymbol{\mu}' = \begin{pmatrix} \mu_{i_1} \\ \mu_{i_2} \\ \vdots \\ \mu_{i_m} \end{pmatrix}, \quad \Sigma' = \begin{pmatrix} \text{Var}(X_{i_1}) & \text{Cov}(X_{i_1}, X_{i_2}) & \dots & \text{Cov}(X_{i_1}, X_{i_m}) \\ \text{Cov}(X_{i_2}, X_{i_1}) & \text{Var}(X_{i_2}) & \dots & \text{Cov}(X_{i_2}, X_{i_m}) \\ \vdots & \vdots & \ddots & \vdots \\ \text{Cov}(X_{i_m}, X_{i_1}) & \text{Cov}(X_{i_m}, X_{i_2}) & \dots & \text{Var}(X_{i_m}) \end{pmatrix}.$$

У Твердженні КЛ9.4.13 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. Виявляється, багатовимірний нормальний розподіл унікальний тим, що для нього зворотне твердження виконується.

Твердження (КЛ10.4.14). Нехай $\mathbf{X} \sim N(\boldsymbol{\mu}, \Sigma)$ можна подати як конкатенацію двох 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 $\mathbf{X}_1 \perp \mathbf{X}_2$.

Зауваження 12.1.1 (КЛ10.4.15). Застосування Твердження КЛ10.4.14 потребує умови, щоб вектори $\mathbf{X}_1$ і $\mathbf{X}_2$ мали спільний багатовимірний нормальний розподіл. Інакше з некорельованості незалежність впливатиме далеко необов'язково. $\square$

12.2 Розподіл функції від випадкового вектора

Вправа 12.2.1. Нехай $X \sim \text{Geom}(p)$, $Y \sim \text{Geom}(p)$, $X \perp Y$. Ми воліємо знайти спільний розподіл $U = \min\{X, Y\}$ та $V = X - Y$. Іншими словами, якщо в нас є дві різні схеми Бернуллі з однаковим параметром, то нас цікавить спільний розподіл найменшого «часу очікування» деякого успіху та різниці між «часом очікування» успіху в першій та другій схемах. Чи будуть незалежними U та V ?

Чому дорівнює спільний розподіл X та $X + Y$? $\square$

Розв'язання. Для виведення спільного розподілу U та V потрібно використати (КЛ10.1.2):

$$\mathbb{P}_{U,V}(U = u, V = v) = \mathbb{P}_{X,Y}(\min\{X, Y\} = u, X - Y = v) .$$

Згадаймо, що функція ймовірності геометричного розподілу дорівнює

$$\mathbb{P}_X(X = k) = p(1 - p)^k \cdot \mathbb{1}\{k \in \mathbb{Z}^+\} ,$$

де $\mathbb{Z}^+ = \{0, 1, 2, \dots\}$.

Розгляньмо дві ситуації: якщо $V < 0$, то $X < Y$, $\min\{X, Y\} = X$, отже

$$\begin{aligned} \mathbb{P}_{U,V}(U = u, V = v) &= \mathbb{P}_{X,Y}(X = u, Y = u - v) \\ &= p(1 - p)^u \cdot \mathbb{1}\{u \in \mathbb{Z}^+\} \cdot p(1 - p)^{u-v} \cdot \mathbb{1}\{u - v \in \mathbb{Z}^+\} \\ &= p^2(1 - p)^{2u-v} \cdot \mathbb{1}\{u \in \mathbb{Z}^+\} \cdot \mathbb{1}\{v \in \{-1, -2, \dots\}\} . \end{aligned}$$

Якщо ж $V \geq 0$, то $X \geq Y$, $\min\{X, Y\} = Y$, і з урахуванням $X \perp\!\!\!\perp Y$,

$$\begin{aligned} \mathbb{P}_{U,V}(U = u, V = v) &= \mathbb{P}_{X,Y}(X = u + v, Y = u) = \mathbb{P}_X(X = u + v) \mathbb{P}_Y(Y = u) \\ &= p(1 - p)^{u+v} \cdot \mathbb{1}\{u + v \in \mathbb{Z}^+\} \cdot p(1 - p)^u \cdot \mathbb{1}\{u \in \mathbb{Z}^+\} \\ &= p^2(1 - p)^{2u+v} \cdot \mathbb{1}\{u \in \mathbb{Z}^+\} \cdot \mathbb{1}\{v \in \mathbb{Z}^+\} . \end{aligned}$$

Ці два випадки можна об'єднати, наклавши на змінну v модуль:

$$\mathbb{P}_{U,V}(U = u, V = v) = p^2(1 - p)^{2u}(1 - p)^{|v|} \cdot \mathbb{1}\{u \in \mathbb{Z}^+\} \cdot \mathbb{1}\{v \in \mathbb{Z}\} .$$

Можна бачити, що $U \perp\!\!\!\perp V$, оскільки спільна функція ймовірності дорівнює добутку двох функцій однієї змінної. У Вправі 12.6.1 буде показано, як вивести маржинальну функцію ймовірності $\mathbb{P}_U(U = u)$ в інший спосіб, і там можна бачити, що коефіцієнт, який стоїть у маржинальній функції ймовірності, явно не фігурує у виразі для спільної функції ймовірності. Це підтверджує міркування з Розд. 11.2, що для встановлення факту незалежності двох величин необов'язково шукати їхні маржинальні функції ймовірності чи щільності, особливо якщо це зробити непросто.

Що ж стосується спільного розподілу X та $X + Y$, то застосування (КЛ10.1.2) дуже просте, адже

$$\begin{aligned} \mathbb{P}_{X,Y}(X = x, X + Y = t) &= \mathbb{P}_{X,Y}(X = x, Y = t - x) \\ &= p(1 - p)^x p(1 - p)^{t-x} \cdot \mathbb{1}\{x \in \mathbb{Z}^+\} \cdot \mathbb{1}\{t - x \in \mathbb{Z}^+\} \\ &= p^2(1 - p)^t \cdot \mathbb{1}\{x \in \{0, 1, \dots, t\}\} \cdot \mathbb{1}\{t \in \mathbb{Z}^+\} , \end{aligned}$$

де ми скористалися незалежністю X та Y .

Можна перевірити себе, обчисливши суму значень функції ймовірності:

$$\begin{aligned} \sum_{t=-\infty}^{\infty} \sum_{x=-\infty}^{\infty} p^2(1-p)^t \cdot \mathbb{1}\{x \in \{0, 1, \dots, t\}\} \cdot \mathbb{1}\{t \in \mathbb{Z}^+\} &= p^2 \sum_{t=0}^{\infty} (1-p)^t \sum_{x=0}^t 1 \\ &= p^2 \sum_{t=0}^{\infty} (t+1)(1-p)^t . \end{aligned}$$

Можна помітити, що

$$\begin{aligned} \sum_{t=0}^{\infty} (t+1)(1-p)^t &= \frac{d}{dp} \left(\sum_{t=0}^{\infty} -(1-p)^{t+1} \right) = \frac{d}{dp} \left(-(1-p) \frac{1}{1-(1-p)} \right) \\ &= \frac{d}{dp} \left(-\frac{1-p}{p} \right) = \frac{1}{p^2} . \end{aligned}$$

Отже

$$\sum_{t=-\infty}^{\infty} \sum_{x=-\infty}^{\infty} \mathbb{P}_{X,Y}(X=x, X+Y=t) = p^2 \cdot \frac{1}{p^2} = 1 .$$

□

Вправа 12.2.2. Нехай спільна щільність випадкових величин X та Y є

$$f_{X,Y}(x, y) = e^{-(x+y)} \cdot \mathbb{1}\{x > 0\} \cdot \mathbb{1}\{y > 0\} .$$

Чому дорівнює щільність випадкової величини $\frac{X}{Y}$?

□

Розв'язання. Перевірмо умови Теорема КЛ10.1.1. По-перше, щільність $f_{X,Y}$ неперервна на своєму носії $\text{supp}((X, Y)^\top) = \mathbb{R}^+ \times \mathbb{R}^+$. Далі нам потрібно утворити вектор $(U, V)^\top$ такий, що $U = h_1(X, Y) = \frac{X}{Y}$, а V підібрано з міркувань спрощення обчислень. Нехай $V = h_2(X, Y) = Y$. Тоді відповідне відображення є взаємно однозначне на нашому носії:

$$X = g_1(U, V) = UV , \quad Y = g_2(U, V) = V .$$

Образом носія під дією (h_1, h_2) є та сама множина $\mathbb{R}^+ \times \mathbb{R}^+$, адже перша координата є результатом ділення двох додатних чисел, а друга координата просто є додатним числом.

Відповідний якобіан можна обчислити як

$$J = \begin{vmatrix} \frac{\partial}{\partial u} g_1(u, v) & \frac{\partial}{\partial v} g_1(u, v) \\ \frac{\partial}{\partial u} g_2(u, v) & \frac{\partial}{\partial v} g_2(u, v) \end{vmatrix} = \begin{vmatrix} v & u \\ 0 & 1 \end{vmatrix} = v ,$$

що не дорівнює 0 на множині $\mathbb{R}^+ \times \mathbb{R}^+$, а всі частинні похідні неперервні.

Тоді застосування (КЛ10.1.3) дає

$$\begin{aligned} f_{U,V}(u, v) &= f_{X,Y}(g_1(u, v), g_2(u, v)) \cdot |J| \cdot \mathbb{1}\{(u, v)^\top \in \mathbb{R}^+ \times \mathbb{R}^+\} \\ &= f_{X,Y}(uv, v) \cdot |v| \cdot \mathbb{1}\{u > 0\} \cdot \mathbb{1}\{v > 0\} \\ &= e^{-(uv+v)} \cdot \mathbb{1}\{uv > 0\} \cdot \mathbb{1}\{v > 0\} \cdot |v| \cdot \mathbb{1}\{u > 0\} \cdot \mathbb{1}\{v > 0\} \\ &= v \cdot e^{-v(u+1)} \cdot \mathbb{1}\{u > 0\} \cdot \mathbb{1}\{v > 0\} . \end{aligned}$$

Маржинальну щільність $U = \frac{X}{Y}$ дістанемо інтегруванням частинами:

$$\begin{aligned} f_U(u) &= \int_{-\infty}^{\infty} v \cdot e^{-v(u+1)} \cdot \mathbb{1}\{u > 0\} \cdot \mathbb{1}\{v > 0\} dv \\ &= \mathbb{1}\{u > 0\} \cdot \int_0^{\infty} v \cdot e^{-v(u+1)} dv \\ &= \mathbb{1}\{u > 0\} \cdot \left(-\frac{1}{u+1}\right) \cdot \left(v \cdot e^{-v(u+1)} \Big|_0^{\infty} - \int_0^{\infty} e^{-v(u+1)} dv\right) \\ &= -\frac{1}{u+1} \cdot \mathbb{1}\{u > 0\} \cdot \left(-\frac{1}{u+1} e^{-v(u+1)} \Big|_0^{\infty}\right) \\ &= \frac{1}{(u+1)^2} \cdot \mathbb{1}\{u > 0\} . \end{aligned}$$

Можемо про всяк випадок пересвідчитися, що це справді щільність, адже інтеграл буде дорівнювати

$$\int_{-\infty}^{\infty} f_U(u) du = -\frac{1}{u+1} \Big|_0^{\infty} = 1 .$$

Замість використання формули заміни змінних ми могли б піти шляхом виведення функції розподілу:

$$F_U(u) = \mathbb{P}_{X,Y} \left(\frac{X}{Y} \leq u \right) = \iint_{\frac{x}{y} \leq u} f_{X,Y}(x, y) dx dy = \int_0^{\infty} \int_0^{uy} e^{-(x+y)} dx dy .$$

У цьому випадку для $u < 0$ ми б мали 0, а для $u \geq 0$ значення інтегралу дорівнювало б

$$\begin{aligned} F_U(u) &= \int_0^{\infty} e^{-y} \left(-e^{-x} \Big|_0^{uy} \right) dy = \int_0^{\infty} \left(e^{-y} - e^{-y(u+1)} \right) dy \\ &= \left(-e^{-y} + \frac{e^{-y(u+1)}}{u+1} \right) \Big|_0^{\infty} = 1 - \frac{1}{u+1} , \end{aligned}$$

звідки остаточно

$$F_U(u) = \left(1 - \frac{1}{u+1}\right) \cdot \mathbb{1}\{u \geq 0\} ,$$

а щільність дорівнює

$$f_U(u) = \frac{1}{(u+1)^2} \cdot \mathbb{1}\{u > 0\} ,$$

як і було виведено вище. □

Вправа 12.2.3. Нехай $U \sim U((0; 2\pi])$ і $T \sim \text{Exp}(1)$, $U \perp\!\!\!\perp T$. Чому дорівнює спільна щільність

$$X = \sqrt{2T} \cos U , \quad Y = \sqrt{2T} \sin U ?$$

Чи є вони незалежні? Чому дорівнюють маржинальні щільності X та Y ? □

Розв'язання. Через незалежність U та T спільна щільність дорівнює

$$f_{U,T}(u, t) = \frac{1}{2\pi} e^{-t} \cdot \mathbb{1}\{u \in (0; 2\pi]\} \cdot \mathbb{1}\{t > 0\} .$$

Перевірмо умови Теорема КЛ10.1.1.

По-перше, щільність $f_{U,T}$ неперервна на своєму носії $\text{supp}((U, T)^\top) = (0; 2\pi] \times \mathbb{R}^+$. Далі нам потрібно утворити вектор $(X, Y)^\top$ такий, що $X = h_1(U, T) = \sqrt{2T} \cos U$, а $Y = h_2(U, T) = \sqrt{2T} \sin U$. Образом носія під дією $(h_1, h_2) \in \mathbb{R}^2$, оскільки ми множимо деяке додатне число $(\sqrt{2t})$ на число, що лежить від -1 до 1 ($\cos u$ або $\sin u$).

Тоді відповідне відображення є взаємно однозначне на нашому носії. Це можна побачити, помітивши, що

$$X^2 + Y^2 = 2T (\cos^2 U + \sin^2 U) = 2T ,$$

звідки

$$T = \frac{1}{2} (X^2 + Y^2) , \quad U = \arccos \frac{X}{\sqrt{X^2 + Y^2}} = \arcsin \frac{Y}{\sqrt{X^2 + Y^2}} .$$

Це обернене відображення є неперервним на своєму носії $\mathbb{R}^2$, окрім, можливо, точки $(0, 0)^\top$, проте цією точкою можна в будь-якому випадку знехтувати як множиною міри нуль. Відповідні частинні похідні також будуть

неперервні, окрім, можливо, точки $(0, 0)^\top$, оскільки вони міститимуть вираз $x^2 + y^2$ у деякому ступені в знаменнику.

Проте знаходити ці похідні, а потім обчислювати якобіан є справою невдячною. Можемо згадати корисну властивість якобіанів, а саме що

$$J = \begin{vmatrix} \frac{\partial}{\partial x} g_1(x, y) & \frac{\partial}{\partial y} g_1(x, y) \\ \frac{\partial}{\partial x} g_2(x, y) & \frac{\partial}{\partial y} g_2(x, y) \end{vmatrix} = \left(\begin{vmatrix} \frac{\partial}{\partial u} h_1(u, t) & \frac{\partial}{\partial t} h_1(u, t) \\ \frac{\partial}{\partial u} h_2(u, t) & \frac{\partial}{\partial t} h_2(u, t) \end{vmatrix} \right)^{-1}$$

для бієктивних неперервно диференційовних відображень. Маємо:

$$J = \left(\begin{vmatrix} -\sqrt{2t} \sin u & \frac{\cos u}{\sqrt{2t}} \\ \sqrt{2t} \cos u & \frac{\sin u}{\sqrt{2t}} \end{vmatrix} \right)^{-1} = (-\sin^2 u - \cos^2 u)^{-1} = -1 \neq 0.$$

Тоді застосування (КЛ10.1.3) дає

$$\begin{aligned} f_{X,Y}(x, y) &= f_{U,T}(g_1(x, y), g_2(x, y)) \cdot |J| \cdot \mathbb{1} \{ (x, y)^\top \in \mathbb{R}^2 \} \\ &= f_{U,T} \left(\arccos \frac{x}{\sqrt{x^2 + y^2}}, \frac{1}{2} (x^2 + y^2) \right) \cdot 1 \\ &= \frac{1}{2\pi} \cdot \mathbb{1} \left\{ \arccos \frac{x}{\sqrt{x^2 + y^2}} \in (0; 2\pi) \right\} \\ &\quad \times e^{-\frac{1}{2}(x^2 + y^2)} \cdot \mathbb{1} \left\{ \frac{1}{2} (x^2 + y^2) > 0 \right\} = \frac{1}{2\pi} e^{-\frac{1}{2}(x^2 + y^2)}. \end{aligned}$$

Нескладно бачити, що ця спільна щільність розпадається в добуток двох маржинальних щільностей:

$$f_X(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}, \quad f_Y(y) = \frac{1}{\sqrt{2\pi}} e^{-\frac{y^2}{2}},$$

тобто $X \sim N(0, 1)$, $Y \sim N(0, 1)$, $X \perp\!\!\!\perp Y$, а самі X та Y мають спільний двовимірний нормальний розподіл.

Відповідне перетворення дає змогу генерувати випадкові величини зі стандартним нормальним розподілом. Це дуже корисний на практиці результат, адже застосування інтегрального перетворення ймовірностей із Теорема КЛ8.1.1 ускладнено через відсутність виразу функції стандартного нормального розподілу Φ . Цей результат можна сформулювати у формі

$$X = \sqrt{-2 \ln U_1} \cos(2\pi U_2), \quad Y = \sqrt{-2 \ln U_1} \sin(2\pi U_2), \quad (12.2.1)$$

де $U_1, U_2 \sim U((0; 1])$, $U_1 \perp\!\!\!\perp U_2$. Це перетворення має назву *перетворення*

Бокса-Мюллера (Box-Muller transform)³.

Те, що $U \stackrel{d}{=} 2\pi U_2$, очевидно, оскільки ці розподіли є рівномірними, але на різних носіях, тому до них застосовне перетворення відносно зсувмасштабування (КЛ7.5.1). Вираз під коренем можна утворити, застосувавши Теорему КЛ8.1.1 до величини T . Згідно з Вправою 9.6.1, $F_T^\dagger(u_1) = -\ln u_1$, звідки й випливає відповідний вираз для (12.2.1). $\square$

12.3 Гамма-розподіл

Вправа 12.3.1. Нехай загальна кількість води, зібрана в дощовому резервуарі за рік, має Гамма-розподіл зі сподіванням 1000 (л) та середньоквадратичним відхиленням 200 (л). Нехай кількості води, зібрані в різні роки, не залежать одна від одної.

Чому дорівнює ймовірність, що за 2 роки буде зібрано менше від 2500 л? Що протягом щонайменше 3 із найближчих 5 років буде зібрано більше середньої кількості води, яку зазвичай збирають за рік? $\square$

Розв'язання. Нам потрібно обчислити ймовірності за участю Гамма-розподілу, проте ми не знаємо його параметрів: ми тільки знаємо сподівання та корінь із дисперсії. Використовуючи Твердження КЛ10.2.2, можемо утворити таку систему рівнянь:

$$\begin{cases} \mathbb{E}[X] = \frac{k}{\lambda} = 1000, \\ \text{Var}(X) = \frac{k}{\lambda^2} = 200^2, \end{cases}$$

де X позначає випадкову величину з розподілом, описаним в умові задачі. Розв'язком цієї системи є

$$k = 25, \quad \lambda = \frac{1}{40}.$$

Нехай Y_i = «кількість води, зібрана за рік i ». Як ми встановили, $Y_i \sim \text{Gamma}(25, \frac{1}{40})$. Згідно з Твердженням КЛ10.2.3, сума незалежних величин $Y_1 + Y_2$ має розподіл $\text{Gamma}(25 + 25, \frac{1}{40}) = \text{Gamma}(50, \frac{1}{40})$. Отже перша шукана ймовірність дорівнює

$$\mathbb{P}_{Y_1+Y_2}(Y_1 + Y_2 \leq 2500) = F_{Y_1+Y_2}(2500) \approx 0.954,$$

що можна обчислити хіба що чисельно, оскільки функція Гамма-розподілу не має аналітичного запису.

³Названо так на честь британського статистика Джорджа Бокса (George Edward Pelham Box, 1919–2013) та американського математика Мервіна Мюллера (Mervin Edgar Muller, 1928–2018).

Імовірність зібрати більше середньої кількості води в щонайменше 3 із 5 років можна обчислити, помітивши, що перевищення середнього рівня можна вважати успіхом. Тоді через незалежність та однаковість розподілів $Y_i \stackrel{d}{=} X$ число таких років має біномний розподіл $\text{Binom}(5, \mathbb{P}_X(X \geq \mathbb{E}[X]))$. Відповідну ймовірність можна порахувати чисельно:

$$\mathbb{P}_X(X \geq \mathbb{E}[X]) = 1 - \mathbb{P}_X(X < 1000) \approx 0.473 .$$

Отже шукана ймовірність дорівнює

$$\mathbb{P}(A) = \sum_{i=3}^5 \binom{5}{i} (\mathbb{P}_X(X \geq \mathbb{E}[X]))^i (1 - \mathbb{P}_X(X \geq \mathbb{E}[X]))^{5-i} \approx 0.450 .$$

□

Вправа 12.3.2. Розгляньмо приклад застосування беєсівського виведення до процесу Пуассона. Нехай покупці відвідують магазин згідно з процесом Пуассона з інтенсивністю λ відвідувачів за годину. Справжнє значення λ невідоме, і ми хочемо його оцінити, розглядаючи як випадкову величину.

Нехай апіорний розподіл $\lambda \sim \text{Exp}(3)$. Нехай X = «число покупців, які відвідають магазин з 1 по 3 годину». Якщо ми спостерігаємо $X = 2$, яким стане апостеріорний розподіл параметру $\lambda \mid X = 2$? □

Розв'язання. Для розв'язання цієї задачі запишімо формулу Беєса:

$$f_{\lambda|X}(\lambda|2) = \frac{\mathbb{P}_{X|\lambda}(X = 2 \mid \lambda) \cdot f_{\lambda}(\lambda)}{\mathbb{P}_X(X = 2)} \propto \frac{e^{-2\lambda}(2\lambda)^2}{2!} \cdot 3e^{-3\lambda} \mathbb{1}\{\lambda > 0\} ,$$

де ми позбулися знаменника, оскільки він являє собою мультиплікативну константу, яка не впливає на форму апостеріорного розподілу. Для моделювання розподілу $X \mid \lambda$ ми використали розподіл Пуассона $\text{Pois}(2\lambda)$, оскільки ми розглядаємо часовий проміжок у 2 години. Зверніть увагу, що *безумовний* розподіл X не є розподілом Пуассона, оскільки він узагалі не повинен містити в собі λ . Проте оскільки це всього лише мультиплікативна константа, ми не повинні виводити цей розподіл.

Тоді ми бачимо, що

$$f_{\lambda|X}(\lambda|x) \propto \lambda^2 e^{-5\lambda} \cdot \mathbb{1}\{\lambda > 0\} .$$

Нескладно бачити, що маємо щільність Гамма-розподілу $\text{Gamma}(3, 5)$. Так, відомо, що повний запис апостеріорної щільності буде

$$f_{\lambda|X}(\lambda|x) = \frac{1}{\Gamma(3)} \lambda^2 e^{-5\lambda} 5^3 \cdot \mathbb{1}\{\lambda > 0\} = \frac{125}{2} \lambda^2 e^{-5\lambda} \cdot \mathbb{1}\{\lambda > 0\} .$$

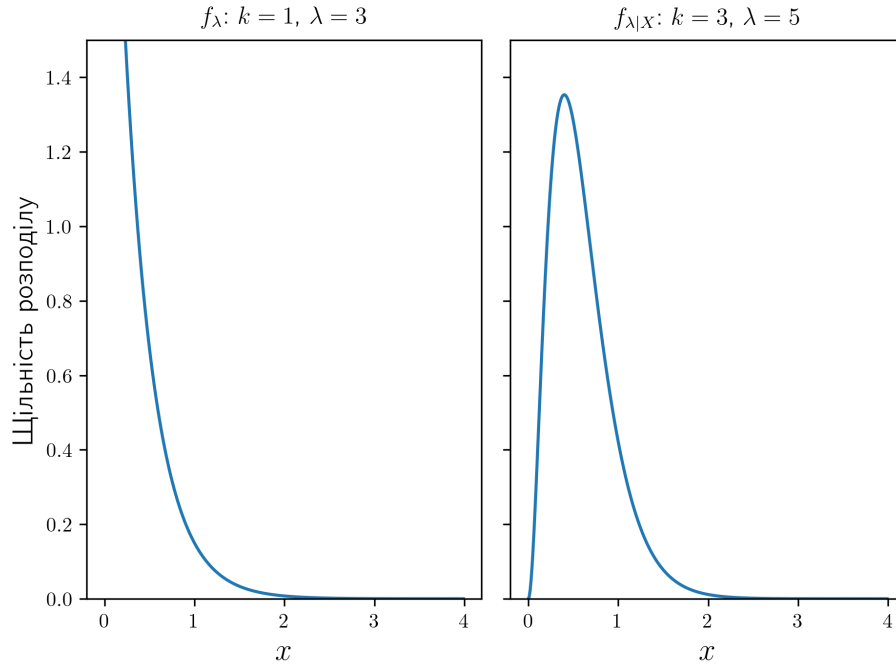


Рис. 12.3.1: Априорна щільність параметра λ , що відповідає розподілу Gamma (1, 3), та апостеріорна щільність параметра $\lambda \mid X = 2$, що відповідає розподілу Gamma (3, 5)

Априорну та апостеріорну щільності наведено на Рис. 12.3.1. Апостеріорний розподіл, під впливом спостереження $X = 2$, зсунувся вбік значення 2, але не дуже сильно. Із надходженням нових даних розподіл буде все точніше описувати значення, які може набувати невідомий параметр λ .

Оскільки априорний розподіл Exp (3) також є розподілом Gamma (1, 3), бачимо, що Гамма-розподіл є *спряженим априорним розподілом* (conjugate prior) до розподілу Пуассона. Справді, нехай $\lambda \sim \text{Gamma}(k, s)$, а $X \mid \lambda \sim \text{Pois}(t\lambda)$. Тоді якщо ми спостерігаємо значення $X = x$, ми маємо

$$f_{\lambda|X}(\lambda|x) \propto \frac{e^{-t\lambda}(t\lambda)^x}{x!} \cdot \frac{1}{\Gamma(k)} s^k \lambda^{k-1} e^{-s\lambda} \cdot \mathbb{1}\{\lambda > 0\} \propto e^{-(t+s)\lambda} \lambda^{x+k-1} \cdot \mathbb{1}\{\lambda > 0\},$$

тобто апостеріорний розподіл є $\lambda \mid X = x \sim \text{Gamma}(x + k, t + s)$, що є дуже простою формулою підрахунку апостеріорного розподілу, так само, як простою є формула підрахунку апостеріорного розподілу у випадку з біномним і Бета-розподілом, які ми розглядали в Розд. КЛ8.2. $\square$

12.4 Мультиномний розподіл

Вправа 12.4.1. У рамках генетичного експерименту схрещують особини двох видів. Нехай фенотип нащадків за деяким параметром може бути тільки одного з трьох типів, A , B й C . Відповідні ймовірності дорівнюють

$\frac{1}{12}$, $\frac{3}{12}$ і $\frac{8}{12}$. Фенотипи незалежні між собою. Вибірка складає 60 нащадків.

Чому дорівнює ймовірність, що серед нащадків 6 матимуть фенотип A , 18 — фенотип B , а 36 — фенотип C ? $\square$

Розв'язання. Оскільки маємо ситуацію незалежного розподілу об'єктів за категоріями, моделювати її потрібно за допомогою мультиномного розподілу. У нашому випадку вектор $\mathbf{X} = (X_A, X_B, X_C)^\top$, елементами якого є кількості нащадків із фенотипом кожного типу, має розподіл

$$\mathbf{X} \sim \text{Mult}_3 \left(60, \left(\frac{1}{12}, \frac{3}{12}, \frac{8}{12} \right)^\top \right).$$

Отже нам потрібно просто підставити відповідні числа у відповідну функцію ймовірності:

$$\mathbb{P}_{\mathbf{X}}(X_A = 6, X_B = 18, X_C = 36) = \frac{60!}{6!18!36!} \left(\frac{1}{12} \right)^6 \left(\frac{3}{12} \right)^{18} \left(\frac{8}{12} \right)^{36} \approx 0.011.$$

$\square$

Вправа 12.4.2. Нехай у деякому університеті розподіл студентів курсами такий: 20% навчаються на першому курсі, 23.75% — на другому, 31.25% — на третьому і 25% — на четвертому.

Нехай у випадковий спосіб відбирають 20 студентів. Чому дорівнює ймовірність, що у вибраній групі буде 4 першокурсники, 5 другокурсників, 6 третьокурсників і 5 четверокурсників? Що 5 студентів будуть четверокурсники? Що принаймні 11 студентів будуть третьо- чи четверокурсники?

Чому дорівнюють сподівання та коваріації між усіма цими величинами? $\square$

Розв'язання. Оскільки відбір студентів здійснюють у незалежний спосіб, маємо справу з мультиномним розподілом. Конкретніше, випадковий вектор $\mathbf{X} = (X_1, X_2, X_3, X_4)^\top$, елементами якого є кількості студентів відповідних курсів у вибірці, має розподіл

$$\mathbf{X} \sim \text{Mult}_4 \left(20, (p_1, p_2, p_3, p_4)^\top \right) = \text{Mult}_4 \left(20, (0.20, 0.2375, 0.3125, 0.25)^\top \right).$$

Тому перша ймовірність дорівнює просто

$$\mathbb{P}_{\mathbf{X}}(X_1 = 4, X_2 = 5, X_3 = 6, X_4 = 5) = \frac{20!}{4!5!6!5!} (0.20)^4 (0.2375)^5 (0.3125)^6 (0.25)^5 \approx 0.0108.$$

Згідно з Твердженням КЛ10.3.5, маржинально $X_4 \sim \text{Binom}(20, 0.25)$, оскільки можна розглянути схему Бернуллі, де поява четверокурсника у вибірці є успіхом, а студентів інших курсів — невдачею. Звідси

$$\mathbb{P}_{X_4}(X_4 = 5) = \binom{20}{5} (0.25)^5 (0.75)^{15} \approx 0.2024 .$$

Також згідно з Твердженням КЛ10.3.5, розподіл загального числа третьо- й четверокурсників $X_3 + X_4 \in \text{Binom}(20, p_3 + p_4) = \text{Binom}(20, 0.5625)$, оскільки можна розглянути схему Бернуллі, де поява третьо- чи четверокурсника у вибірці є успіхом, а студентів інших курсів — невдачею. Звідси

$$\mathbb{P}_{X_3, X_4}(X_3 + X_4 \geq 11) = \sum_{k=11}^{20} \binom{20}{k} (0.5625)^k (0.4375)^{20-k} \approx 0.6350 .$$

Згідно з Твердженням КЛ10.3.8, сподівання $\mathbb{E}[\mathbf{X}]$ можна обчислити як сподівання відповідних координат цього вектора:

$$\mathbb{E}[\mathbf{X}] = \mathbb{E}[(X_1, X_2, X_3, X_4)^\top] = (20p_1, 20p_2, 20p_3, 20p_4)^\top = (4, 4.75, 6.25, 5)^\top .$$

Відповідна матриця коваріацій дорівнює

$$\begin{aligned} \mathbf{Cov}(\mathbf{X}) &= \begin{pmatrix} \text{Var}(X_1) & \text{Cov}(X_1, X_2) & \text{Cov}(X_1, X_3) & \text{Cov}(X_1, X_4) \\ \text{Cov}(X_2, X_1) & \text{Var}(X_2) & \text{Cov}(X_2, X_3) & \text{Cov}(X_2, X_4) \\ \text{Cov}(X_3, X_1) & \text{Cov}(X_3, X_2) & \text{Var}(X_3) & \text{Cov}(X_3, X_4) \\ \text{Cov}(X_4, X_1) & \text{Cov}(X_4, X_2) & \text{Cov}(X_4, X_3) & \text{Var}(X_4) \end{pmatrix} \\ &= \begin{pmatrix} 20p_1(1-p_1) & -20p_1p_2 & -20p_1p_3 & -20p_1p_4 \\ -20p_1p_2 & 20p_2(1-p_2) & -20p_2p_3 & -20p_2p_4 \\ -20p_1p_3 & -20p_2p_3 & 20p_3(1-p_3) & -20p_3p_4 \\ -20p_1p_4 & -20p_2p_4 & -20p_3p_4 & 20p_4(1-p_4) \end{pmatrix} \\ &\approx \begin{pmatrix} 3.2 & -0.95 & -1.25 & -1 \\ -0.95 & 3.6219 & -1.4844 & -1.1875 \\ -1.25 & -1.4844 & 4.2969 & -1.5625 \\ -1 & -1.1875 & -1.5625 & 3.75 \end{pmatrix} . \end{aligned}$$

Окрім знаків коваріацій, ми не можемо нічого сказати про відповідні лінійні зв'язки між координатами нашого вектора, зокрема, наскільки сильними вони є. Для вимірювання сили лінійного зв'язку слугують коефіцієнти кореляції. Наприклад, коефіцієнт кореляції ρ_{X_1, X_2} дорівнює

$$\rho_{X_1, X_2} = \frac{\text{Cov}(X_1, X_2)}{\sqrt{\text{Var}(X_1) \text{Var}(X_2)}} = \frac{-0.95}{\sqrt{3.2 \cdot 3.6219}} \approx -0.0820 ,$$

тоді як коефіцієнт кореляції ρ_{X_1, X_4} дорівнює

$$\rho_{X_1, X_4} = \frac{\text{Cov}(X_1, X_4)}{\sqrt{\text{Var}(X_1) \text{Var}(X_4)}} = \frac{-1}{\sqrt{3.2 \cdot 3.75}} \approx -0.0833.$$

Оскільки ці значення дуже малі, лінійного зв'язку між відповідними величинами фактично немає. $\square$

12.5 Багатовимірний нормальний розподіл

Вправа 12.5.1. Нехай із деякої популяції одружених пар, у яких зріст дружини X та зріст чоловіка Y (у см) мають двовимірний нормальний розподіл, у випадковий спосіб вибирають одну пару. Нехай маржинально $X \sim N(170, 5^2)$, $Y \sim N(178, 5^2)$, а кореляція між ними дорівнює 0.68.

Чому дорівнює ймовірність, що дружина виявиться вищою? $\square$

Розв'язання. Нас цікавить ймовірність $\mathbb{P}_{X,Y}(X > Y)$, яку можна порахувати через відповідний інтеграл спільної щільності $f_{X,Y}$ за областю $X > Y$, але це буде нераціонально.

Випишімо повністю розподіл відповідного вектора. Для цього перейдімо від кореляції до коваріації:

$$\text{Cov}(X, Y) = \text{Corr}(X, Y) \cdot \sqrt{\text{Var}(X) \text{Var}(Y)} = 0.68 \cdot 5 \cdot 5 = 17.$$

Тоді

$$\begin{pmatrix} X \\ Y \end{pmatrix} \sim N \left(\begin{pmatrix} 170 \\ 178 \end{pmatrix}, \begin{pmatrix} 5^2 & 17 \\ 17 & 5^2 \end{pmatrix} \right).$$

Варто зазначити, що X та Y не є незалежними, тому формула згортки тут незастосовна. Тим не менше, ми знаємо, що для багатовимірного нормального розподілу за Визначенням КЛ10.4.2 будь-яка лінійна комбінація координат має нормальний розподіл. Зокрема, $X - Y$ має нормальний розподіл, і нас цікавить ймовірність $\mathbb{P}_{X-Y}(X - Y > 0)$.

Оскільки нормальний розподіл має два параметри — сподівання й дисперсію — достатньо їх порахувати для $X - Y$, щоб можна було визначити розподіл цієї величини. За властивостями сподівання та дисперсії маємо

$$\mathbb{E}[X - Y] = \mathbb{E}[X] - \mathbb{E}[Y] = 170 - 178 = -8,$$

$$\text{Var}(X - Y) = \text{Var}(X) + \text{Var}(Y) - 2\text{Cov}(X, Y) = 25 + 25 - 2 \cdot 17 = 16,$$

звідки $X - Y \sim N(-8, 16)$, а $\sigma_{X-Y} = \sqrt{16} = 4$.

Остаточно

$$\begin{aligned}\mathbb{P}_{X,Y}(X - Y > 0) &= \mathbb{P}_{X-Y}(X - Y > 0) = \mathbb{P}_Z\left(Z > \frac{0 - (-8)}{4}\right) \\ &= 1 - \Phi(2) = 0.0227.\end{aligned}$$

Фактично ми перейшли від обчислення двовимірного розподілу $\mathbb{P}_{X,Y}$ до обчислення розподілу (одновимірної) випадкової величини $X - Y$. $\square$

Вправа 12.5.2. Нехай $X, Y \sim N(0, 1)$, $X \perp Y$. Розгляньмо нову випадкову величину

$$W = \begin{cases} X, & XY \geq 0 \\ -X, & XY < 0 \end{cases}.$$

Який розподіл має W ? Чи має вектор $(W, Y)^\top$ двовимірний нормальний розподіл? $\square$

Розв'язання. За визначенням функції розподілу та формулою повної ймовірності маємо:

$$\begin{aligned}\mathbb{P}_W(W \leq w) &= \mathbb{P}_{W|X,Y}(W \leq w \mid XY \geq 0) \cdot \mathbb{P}_{X,Y}(XY \geq 0) \\ &\quad + \mathbb{P}_{W|X,Y}(W \leq w \mid XY < 0) \cdot \mathbb{P}_{X,Y}(XY < 0) \\ &= \mathbb{P}_{W|X,Y}(X \leq w \mid XY \geq 0) \cdot \mathbb{P}_{X,Y}(XY \geq 0) \\ &\quad + \mathbb{P}_{W|X,Y}(-X \leq w \mid XY < 0) \cdot \mathbb{P}_{X,Y}(XY < 0) \\ &= \mathbb{P}_{X,Y}(X \leq w, XY \geq 0) + \mathbb{P}_{X,Y}(X \geq -w, XY < 0).\end{aligned}$$

Нехай $w < 0$, тоді

$$\begin{aligned}\mathbb{P}_W(W \leq w) &= \mathbb{P}_{X,Y}(X \leq w, Y < 0) + \mathbb{P}_{X,Y}(X \geq -w, Y < 0) \\ &= \mathbb{P}_X(X \leq w) \mathbb{P}_Y(Y < 0) + \mathbb{P}_X(X \geq -w) \mathbb{P}_Y(Y < 0) \\ &= \Phi(w)\Phi(0) + (1 - \Phi(-w))\Phi(0) \\ &= \Phi(w)\Phi(0) + \Phi(w)\Phi(0) = 2\Phi(w)\Phi(0) = \Phi(w),\end{aligned}$$

де ми використали властивість (i) із Твердження КЛ7.2.3 та врахували, що $\Phi(0) = \frac{1}{2}$ за симетрією стандартного нормального розподілу відносно 0.

Аналогічно можна показати для $w \geq 0$, але для цього потрібно розглянути $\mathbb{P}_W(W \leq w) = 1 - \mathbb{P}_W(W > w)$. Ми також дістанемо $\mathbb{P}_W(W \leq w) = \Phi(w)$. Отже $W \sim N(0, 1)$.

Однак, незважаючи на те, що $Y \sim N(0, 1)$ також, разом вони не утворюють двовимірного нормального розподілу, адже W та Y завжди мають однаковий знак. Справді, якщо $W > 0$, то це означає, що або $X < 0$, $Y > 0$, або $X > 0$, $Y > 0$. Тобто носієм такого вектора не є весь простір $\mathbb{R}^2$, як мало би бути за Визначенням КЛ10.4.1. $\square$

12.6 Додаткові вправи

Вправа 12.6.1. Нехай $X \sim \text{Geom}(p)$, $Y \sim \text{Geom}(p)$, $X \perp\!\!\!\perp Y$. Виведіть функцію ймовірності випадкової величини $M = \min\{X, Y\}$. $\square$

Розв'язання. За визначенням функції ймовірності, потрібно обчислити значення ймовірностей $\mathbb{P}_M(M = m)$ для всіх можливих значень m . Якщо пригадати, що носієм геометричного розподілу є множина натуральних чисел $\text{supp}(X) = \text{supp}(Y) = \mathbb{N}$, то можна помітити, що $\text{supp}(M) = \mathbb{N}$ так само.

Спільною функцією ймовірності X та Y , з урахуванням $X \perp\!\!\!\perp Y$, є

$$\begin{aligned}\mathbb{P}_{X,Y}(X = x, Y = y) &= \mathbb{P}_X(X = x) \cdot \mathbb{P}_Y(Y = y) \\ &= p(1-p)^x \cdot p(1-p)^y \cdot \mathbb{1}\{x \in \mathbb{Z}^+\} \cdot \mathbb{1}\{y \in \mathbb{Z}^+\}.\end{aligned}$$

Тоді шукану ймовірність можна дістати згідно з (КЛ10.1.2) як

$$\begin{aligned}\mathbb{P}_M(M = m) &= \sum_{\min\{x,y\}=m} \mathbb{P}_{X,Y}(X = x, Y = y) \\ &= \sum_{\min\{x,y\}=m} p(1-p)^x \cdot p(1-p)^y \cdot \mathbb{1}\{x \in \mathbb{Z}^+\} \cdot \mathbb{1}\{y \in \mathbb{Z}^+\}.\end{aligned}$$

Можна помітити, що множина векторів $(x, y)^\top \in \mathbb{Z}^+ \times \mathbb{Z}^+$, для яких справедливо $\min\{x, y\} = m$, складається з трьох підмножин:

- вектор $(m, m)^\top$;
- множина векторів виду $(m, y)^\top$, де $y > m$;
- множина векторів виду $(x, m)^\top$, де $x > m$.

Другий і третій випадок симетричні, тому достатньо розглянути тільки один із них і помножити результат на 2.

Остаточно маємо для всіх $m \in \mathbb{Z}^+$:

$$\begin{aligned}\mathbb{P}_M(M = m) &= p(1-p)^m \cdot p(1-p)^m + 2 \sum_{y=m+1}^{\infty} p(1-p)^m \cdot p(1-p)^y \\ &= p^2(1-p)^{2m} + 2p^2(1-p)^m \sum_{y=m+1}^{\infty} (1-p)^y \\ &= p^2(1-p)^{2m} + 2p^2(1-p)^m \sum_{j=0}^{\infty} (1-p)^{j+m+1} \\ &= p^2(1-p)^{2m} + 2p^2(1-p)^{2m+1} \cdot \frac{1}{1-(1-p)}\end{aligned}$$

$$= (1 - p)^{2m}(p^2 + 2p(1 - p)) = p(2 - p)(1 - p)^{2m}.$$

Можна перевірити себе і просумувати всі значення функції ймовірності від 0 до нескінченності. Помітивши, що

$$\sum_{m=0}^{\infty} (1 - p)^{2m} = \frac{1}{1 - (1 - p)^2} = \frac{1}{p(2 - p)},$$

робимо висновок, що сума значень функції ймовірності M дорівнює 1. $\square$

Вправа 12.6.2. Нехай випадковий вектор $(X, Y)^\top$ має рівномірний розподіл на крузі з центром у початку координат та радіусом r . Чому дорівнює відповідна щільність розподілу? Чому дорівнюють маржинальні розподіли X та Y ? Чому дорівнює коефіцієнт кореляції між X та Y ?

Нехай D = «відстань від початку координат до випадково обраної точки». Чому дорівнюють функція та щільність розподілу D ? Сподівання D ? $\square$

Розв'язання. Випадковий вектор має рівномірний розподіл, якщо його щільність є константою на відповідному носії:

$$f_{X,Y}(x, y) = c \cdot \mathbb{1} \{x^2 + y^2 \leq R^2\}.$$

Оскільки інтеграл такої щільності дорівнює c , помножений на площу відповідного круга, πR^2 , маємо $c = \frac{1}{\pi R^2}$.

Щоб дістати маржинальні щільності, потрібно проінтегрувати спільну щільність за іншою координатою. Через симетрію X та Y достатньо розглянути тільки випадок з X :

$$\begin{aligned} f_X(x) &= \int_{\mathbb{R}} f_{X,Y}(x, y) dy = \frac{1}{\pi R^2} \int_{\mathbb{R}} \mathbb{1} \{x^2 + y^2 \leq R^2\} dy \\ &= \frac{1}{\pi R^2} \int_{-\sqrt{R^2 - x^2}}^{\sqrt{R^2 - x^2}} \mathbb{1} \{x^2 \leq R^2\} dy = \frac{2}{\pi R^2} \sqrt{R^2 - x^2} \cdot \mathbb{1} \{x^2 \leq R^2\}. \end{aligned}$$

Аналогічно для Y :

$$f_Y(y) = \frac{2}{\pi R^2} \sqrt{R^2 - y^2} \cdot \mathbb{1} \{y^2 \leq R^2\}.$$

Для обчислення коефіцієнта кореляції між X та Y потрібно обчислити відповідні коваріації й дисперсії. Оскільки $\text{Cov}(X, Y) = \mathbb{E}[XY] - \mathbb{E}[X] \mathbb{E}[Y]$,

обчислимо спочатку $\mathbb{E}[XY]$ за допомогою переходу в полярні координати:

$$\begin{aligned}\mathbb{E}[XY] &= \iint_{\mathbb{R}^2} xy \cdot f_{X,Y}(x,y) dx dy \\ &= \frac{1}{\pi R^2} \iint_{x^2+y^2 \leq R^2} xy dx dy \\ &= \frac{1}{\pi R^2} \iint_{\substack{0 \leq \rho \leq R \\ 0 \leq \varphi \leq 2\pi}} \rho \cos \varphi \cdot \rho \sin \varphi \cdot \rho d\rho d\varphi \\ &= \frac{1}{\pi R^2} \int_0^R \rho^3 \left(\int_0^{2\pi} \frac{1}{2} \sin 2\varphi d\varphi \right) d\rho = 0 ,\end{aligned}$$

оскільки інтеграл від $\sin 2\varphi$ за відповідним проміжком дорівнює 0. Ми дозволили собі перейти від подвійного інтегралу до повторного за теоремою Фубіні, незважаючи на те, що підінтегральна функція не є невід'ємною, адже вона є обмеженою, а тому вочевидь інтегровною.

Також маємо

$$\mathbb{E}[X] = \int_{-\infty}^{\infty} \frac{2}{\pi R^2} \sqrt{R^2 - x^2} \cdot \mathbb{1}\{x^2 \leq R^2\} dx = 0 ,$$

оскільки підінтегральна функція непарна, а проміжок інтегрування симетричний. За симетрією також маємо $\mathbb{E}[Y] = 0$. Отже $\text{Cov}(X, Y) = 0$, хоча відповідні величини, вочевидь, зовсім не є незалежними.

Розподіл D можна було б знайти за допомогою Теорема КЛ10.1.1, поклавши, наприклад, $h_1(x, y) = \sqrt{x^2 + y^2}$, $h_2(x, y) = y$, проте таке відображення не буде взаємно однозначним, та й наявність кореня не дає підстав очікувати, що застосування цього методу буде простим.

Дістаньмо функцію розподілу D за визначенням. Для $0 \leq a < R$

$$\begin{aligned}F_D(a) &= \mathbb{P}_D(D \leq a) = \mathbb{P}_{X,Y}(\sqrt{X^2 + Y^2} \leq a) \\ &= \mathbb{P}_{X,Y}(X^2 + Y^2 \leq a^2) \\ &= \iint_{x^2+y^2 \leq a^2} f_{X,Y}(x,y) dx dy \\ &= \frac{1}{\pi R^2} \iint_{x^2+y^2 \leq a^2} dx dy \\ &= \frac{\pi a^2}{\pi R^2} = \frac{a^2}{R^2} ,\end{aligned}$$

де ми дозволили собі не переходити від подвійного інтегралу до повторного, оскільки подвійний інтеграл від одиниці дає площу області інтегрування. Зрозуміло, що для $a < 0$ маємо 0, а для $a \geq R$ маємо 1.

Для сподівання можна помітити, що щільністю розподілу D є функція

$$f_D(a) = \frac{2a}{R^2} \cdot \mathbb{1}\{0 \leq a \leq R\} ,$$

а отже

$$\mathbb{E}[D] = \int_{-\infty}^{\infty} a \cdot f_D(a) da = \int_0^R \frac{2a^2}{R^2} da = \frac{2R}{3} . \quad \square$$

Вправа 12.6.3. Нехай маємо дві незалежні випадкові величини X та Y з однаковим розподілом $\text{Exp}(\lambda)$. Вони можуть відповідати, наприклад, часу очікування деякої події в процесі Пуассона з інтенсивністю λ та часу очікування наступної події після того, як перша з них уже сталася.

Чому дорівнює ймовірність, що час очікування між першою та другою подією буде вдвічі більший від часу очікування першої? $\square$

Розв'язання. Нас цікавить ймовірність $\mathbb{P}_{X,Y}(Y > 2X)$. Її можна обчислити як відповідний інтеграл спільної щільності, яку через незалежність можна утворити добутком двох маржинальних щільностей. Через невід'ємність підінтегральної функції можемо застосувати теорему Фубіні для переходу від подвійного інтегралу до повторного:

$$\begin{aligned} \mathbb{P}_{X,Y}(Y > 2X) &= \iint_{y > 2x} \lambda e^{-\lambda x} \cdot \mathbb{1}\{x > 0\} \cdot \lambda e^{-\lambda y} \cdot \mathbb{1}\{y > 0\} dx dy \\ &= \int_0^{\infty} \lambda e^{-\lambda y} \left(\int_0^{\frac{y}{2}} \lambda e^{-\lambda x} dx \right) dy \\ &= \int_0^{\infty} \lambda e^{-\lambda y} \left(1 - e^{-\frac{\lambda}{2}y} \right) dy \\ &= \int_0^{\infty} \lambda e^{-\lambda y} dy - \frac{2}{3} \int_0^{\infty} \frac{3\lambda}{2} e^{-\frac{3}{2}\lambda y} dy = 1 - \frac{2}{3} = \frac{1}{3} . \end{aligned}$$

Як альтернативу, або ж якщо ж нас цікавить розподіл величини $Y - 2X$ сам по собі, ми можемо використати формулу заміни змінних. Перевірмо умови Теорема КЛ10.1.1. По-перше, щільність $f_{X,Y}$ неперервна на своєму носії $\text{supp}((X, Y)^{\top}) = \mathbb{R}^+ \times \mathbb{R}^+$. Далі нам потрібно утворити вектор $(U, V)^{\top}$ такий, що $U = h_1(X, Y) = Y - 2X$, а V підібрано з міркувань спрощення обчислень. Нехай $V = h_2(X, Y) = Y$. Тоді відповідне відображення є взаємно однозначне на нашому носії:

$$X = g_1(U, V) = \frac{1}{2}(V - U) , \quad Y = g_2(U, V) = V .$$

Образом носія під дією $(h_1, h_2) \in \mathbb{R} \times \mathbb{R}^+$, адже перша координата є результатом віднімання двох додатних чисел, а тому може стати від'ємною.

Відповідний якобіан можна обчислити як

$$J = \begin{vmatrix} \frac{\partial}{\partial u} g_1(u, v) & \frac{\partial}{\partial v} g_1(u, v) \\ \frac{\partial}{\partial u} g_2(u, v) & \frac{\partial}{\partial v} g_2(u, v) \end{vmatrix} = \begin{vmatrix} -\frac{1}{2} & \frac{1}{2} \\ 0 & 1 \end{vmatrix} = -\frac{1}{2},$$

що не дорівнює 0 на множині $\mathbb{R} \times \mathbb{R}^+$, а всі частинні похідні неперервні.

Тоді застосування (КЛ10.1.3) дає

$$\begin{aligned} f_{U,V}(u, v) &= f_{X,Y}(g_1(u, v), g_2(u, v)) \cdot |J| \cdot \mathbb{1} \{(u, v)^\top \in \mathbb{R} \times \mathbb{R}^+\} \\ &= f_{X,Y}\left(\frac{1}{2}(v-u), v\right) \cdot \left|-\frac{1}{2}\right| \cdot \mathbb{1}\{v > 0\} \\ &= \lambda e^{-\frac{\lambda}{2}(v-u)} \cdot \mathbb{1}\left\{\frac{1}{2}(v-u) > 0\right\} \cdot \lambda e^{-\lambda v} \cdot \mathbb{1}\{v > 0\} \cdot \frac{1}{2} \cdot \mathbb{1}\{v > 0\} \\ &= \frac{\lambda^2}{2} e^{-\frac{\lambda}{2}u - \frac{3}{2}\lambda v} \cdot \mathbb{1}\{v > u\} \cdot \mathbb{1}\{v > 0\}. \end{aligned}$$

Маржинальну щільність $U = Y - 2X$ дістанемо інтегруванням

$$f_U(u) = \int_{-\infty}^{\infty} \frac{\lambda^2}{2} e^{-\frac{\lambda}{2}u - \frac{3}{2}\lambda v} \cdot \mathbb{1}\{v > u\} \cdot \mathbb{1}\{v > 0\} dv.$$

Зокрема, якщо $u < 0$, маємо

$$\begin{aligned} f_U(u) &= \int_0^{\infty} \frac{\lambda^2}{2} e^{-\frac{\lambda}{2}u - \frac{3}{2}\lambda v} dv = \frac{\lambda^2}{2} e^{-\frac{\lambda}{2}u} \int_0^{\infty} e^{-\frac{3}{2}\lambda v} dv \\ &= \frac{\lambda^2}{2} e^{-\frac{\lambda}{2}u} \left(-\frac{2}{3\lambda} e^{-\frac{3}{2}\lambda v} \Big|_0^{\infty} \right) = \frac{\lambda}{3} e^{\frac{\lambda}{2}u}. \end{aligned}$$

А якщо $u \geq 0$, маємо

$$\begin{aligned} f_U(u) &= \int_u^{\infty} \frac{\lambda^2}{2} e^{-\frac{\lambda}{2}u - \frac{3}{2}\lambda v} dv = \frac{\lambda^2}{2} e^{-\frac{\lambda}{2}u} \int_u^{\infty} e^{-\frac{3}{2}\lambda v} dv \\ &= \frac{\lambda^2}{2} e^{-\frac{\lambda}{2}u} \left(-\frac{2}{3\lambda} e^{-\frac{3}{2}\lambda v} \Big|_u^{\infty} \right) = \frac{\lambda}{3} e^{\frac{\lambda}{2}u} \cdot e^{-\frac{3}{2}\lambda u} = \frac{\lambda}{3} e^{-\lambda u}. \end{aligned}$$

Остаточно

$$f_U(u) = \begin{cases} \frac{\lambda}{3} e^{\frac{\lambda}{2}u}, & u < 0 \\ \frac{\lambda}{3} e^{-\lambda u}, & u \geq 0 \end{cases},$$

звідки

$$\mathbb{P}_{X,Y}(Y > 2X) = \mathbb{P}_U(U > 0) = \frac{\lambda}{3} \int_0^{\infty} e^{-\lambda u} du = \frac{1}{3}.$$

□

Вправа 12.6.4. Відомо, що новини о 8 дивляться 31.25% глядачів, документальну передачу — 25% глядачів, а інші програми — 43.75% глядачів. Глядачі вибирають, яку програму дивитися, незалежно одне від одного. У випадковий спосіб формують вибірку з 15 глядачів.

Чому дорівнює ймовірність, що 5 глядачів дивляться новини, 4 глядачі дивляться документальну передачу, а 6 глядачів дивляться решту програм? Що документальну передачу дивляться щонайменше 3 глядачі? Що не дивляться новин не більше 5 глядачів? $\square$

Розв'язання. Через незалежність категорій однієї від одної маємо справу з мультиномним розподілом. Розгляньмо вектор $\mathbf{X} = (X_1, X_2, X_3)^\top$, де координати відповідають числу глядачів, що дивляться новини, документальну передачу та решту програм відповідно. За умовами задачі

$$\mathbf{X} \sim \text{Mult}_3(15, (0.3125, 0.25, 0.4375)^\top) .$$

Тоді перша шукана ймовірність дорівнює

$$\mathbb{P}_{\mathbf{X}}(X_1 = 5, X_2 = 4, X_3 = 6) = \frac{15!}{5!4!6!} (0.3125)^5 (0.25)^4 (0.4375)^6 \approx 0.0515 .$$

Маржинальний розподіл X_2 є біномним $X_2 \sim \text{Binom}(15, 0.25)$, а отже

$$\mathbb{P}_{X_2}(X_2 \geq 3) = 1 - \mathbb{P}_{X_2}(X_2 < 3) = 1 - \sum_{k=0}^2 \binom{15}{k} (0.25)^k (0.75)^{15-k} \approx 0.7639 .$$

Що стосується глядачів, які не дивляться новин, то розподіл є $X_1 \sim \text{Binom}(15, 0.3125)$, і маємо справу з величиною $15 - X_1$, а тому

$$\begin{aligned} \mathbb{P}_{X_1}(15 - X_1 \leq 5) &= \mathbb{P}_{X_1}(X_1 \geq 10) = 1 - \mathbb{P}_{X_1}(X_1 < 10) \\ &= 1 - \sum_{k=0}^9 \binom{15}{k} (0.3125)^k (0.6875)^{15-k} \approx 0.0051 . \end{aligned}$$

Альтернативно можна міркувати, що глядачі, які не дивляться новин, дивляться все інше, тому через те, що $X_2 + X_3 \sim \text{Binom}(15, 0.6875)$, маємо

$$\mathbb{P}_{X_2+X_3}(X_2 + X_3 \leq 5) = \sum_{k=0}^5 \binom{15}{k} (0.6875)^k (0.3125)^{15-k} \approx 0.0051 . \quad \square$$

Вправа 12.6.5. Нехай

$$\mathbf{Y} = \begin{pmatrix} Y_1 \\ Y_2 \end{pmatrix} \sim N \left(\begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix}, \begin{pmatrix} \sigma_1^2 & \sigma_{12} \\ \sigma_{12} & \sigma_2^2 \end{pmatrix} \right) , \quad \mathbf{B} = \begin{pmatrix} 1 & -1 \\ -1 & 1 \end{pmatrix} .$$

Розгляньмо випадковий вектор $\mathbf{Z} = \mathbf{B}\mathbf{Y} = (Z_1, Z_2)^\top$. Покажіть, що Z_1 і Z_2 мають нормальний розподіл, але їхній спільний розподіл не нормальний. $\square$

Розв'язання. Оскільки будь-яка лінійна комбінація координат двовимірного нормального вектора є величиною з нормальним розподілом, маємо

$$\begin{pmatrix} Z_1 \\ Z_2 \end{pmatrix} = \begin{pmatrix} 1 & -1 \\ -1 & 1 \end{pmatrix} \cdot \begin{pmatrix} Y_1 \\ Y_2 \end{pmatrix} = \begin{pmatrix} Y_1 - Y_2 \\ Y_2 - Y_1 \end{pmatrix}.$$

Сподівання та дисперсія відповідних нормальних величин дорівнюють

$$\mathbb{E}[Z_1] = \mathbb{E}[Y_1] - \mathbb{E}[Y_2] = \mu_1 - \mu_2,$$

$$\mathbb{E}[Z_2] = \mathbb{E}[Y_2] - \mathbb{E}[Y_1] = \mu_2 - \mu_1,$$

$$\text{Var}(Z_1) = \text{Var}(Y_1) + \text{Var}(Y_2) - 2\text{Cov}(Y_1, Y_2) = \sigma_1^2 + \sigma_2^2 - 2\sigma_{12},$$

$$\text{Var}(Z_2) = \text{Var}(-Z_1) = \text{Var}(Z_1) = \sigma_1^2 + \sigma_2^2 - 2\sigma_{12}.$$

Отже $Z_1 \sim N(\mu_1, \sigma_1^2 + \sigma_2^2 - 2\sigma_{12})$, $Z_2 \sim N(\mu_2, \sigma_1^2 + \sigma_2^2 - 2\sigma_{12})$.

Проте спільний розподіл вектора $(Z_1, Z_2)^\top$ не є (двовимірним) нормальним. Справді, підрахуймо коваріацію цих величин:

$$\begin{aligned} \text{Cov}(Z_1, Z_2) &= \text{Cov}(Y_1 - Y_2, Y_2 - Y_1) \\ &= 2\text{Cov}(Y_1, Y_2) - \text{Var}(Y_1) - \text{Var}(Y_2) = 2\sigma_{12} - \sigma_1^2 - \sigma_2^2. \end{aligned}$$

Матриця коваріацій такого вектора дорівнює

$$\begin{aligned} \text{Cov} \left(\begin{pmatrix} Z_1 \\ Z_2 \end{pmatrix} \right) &= \begin{pmatrix} \sigma_1^2 + \sigma_2^2 - 2\sigma_{12} & 2\sigma_{12} - \sigma_1^2 - \sigma_2^2 \\ 2\sigma_{12} - \sigma_1^2 - \sigma_2^2 & \sigma_1^2 + \sigma_2^2 - 2\sigma_{12} \end{pmatrix} \\ &= (\sigma_1^2 + \sigma_2^2 - 2\sigma_{12}) \begin{pmatrix} 1 & -1 \\ -1 & 1 \end{pmatrix}. \end{aligned}$$

Детермінант цієї матриці дорівнює 0 (вона є вироджена), тому не виконується ключова умова з Визначення КЛ10.4.1. $\square$

13 Сподівання та інші характеристики випадкових величин

13.1 Основні теоретичні відомості

Детальний огляд понять, потрібний для розв'язання задач на цьому занятті, наведено в Курсі лекцій, Розд. КЛ11.

На попередніх заняттях ми детально розглядали поняття сподівання випадкової величини та вивчали методи його підрахунку. На цьому занятті розглянемо декілька властивостей сподівань та пов'язаних нерівностей.

Твердження (КЛ11.1.1). 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 $\mathbb{E}[X] = \mathbb{E}[Y]$.

Для формулювання надзвичайно важливої нерівності Єнсена потрібно згадати, які функції називають опуклими.

Визначення (КЛ11.1.4). Функція $\varphi : \mathbb{R} \rightarrow \mathbb{R}$ *опукла* (convex)¹ якщо

$$\lambda\varphi(x) + (1-\lambda)\varphi(y) \geq \varphi(\lambda x + (1-\lambda)y) , \quad \lambda \in (0; 1) , \quad x, y \in \mathbb{R} . \quad (\text{КЛ11.1.1})$$

□

¹Наведене формулювання опуклості, власне кажучи, і є його визначенням, оскільки працює для будь-яких функцій. Відомий із курсу «Математичний аналіз» спосіб перевірити опуклість функції взяттям її другої похідної працює тільки для функцій, які таку похідну мають.

Теорема (Нерівність Єнсена (Jensen's inequality)², КЛ11.1.5). 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

$$\mathbb{E}[\varphi(X)] \geq \varphi(\mathbb{E}[X]) . \quad (\text{КЛ11.1.2})$$

Зауваження (КЛ11.1.6). Декілька наслідків із нерівності Єнсена:

- 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 логаритм — 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.

□

Розгляньмо надзвичайно важливі результати, які нам знадобляться і в наступних заняттях.

Теорема (Нерівність Маркова (Markov's inequality)³, КЛ11.1.13). Нехай $g : \mathbb{R} \rightarrow [0; \infty)$ — неспадна вимірна функція. Тоді для будь-якої випадкової величини X має місце нерівність

$$\mathbb{P}_X(X \geq c) \leq \frac{\mathbb{E}[g(X)]}{g(c)} , \quad c > 0 . \quad (\text{КЛ11.1.5})$$

²Йоган Єнсен (Johan Ludvig William Valdemar Jensen, 1859–1925) — данський математик.

³Названа так на честь російського математика Андрея Маркова (1856–1922).

Зауваження (КЛ11.1.14). Класичне формулювання нерівності Маркова можна дістати, якщо покласти $X \geq 0$ майже напевно, $g(x) = x$:

$$\mathbb{P}_X (X \geq c) \leq \frac{\mathbb{E}[X]}{c}, \quad c > 0. \quad (\text{КЛ11.1.6})$$

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. $\square$

Чи не найважливішим наслідком є *нерівність Чебишова* (Chebyshev's inequality)⁴, яку можна дістати, поклавши $g(x) = x^2$ та розглянувши величину $|X - \mathbb{E}[X]|$:

$$\mathbb{P}_X (|X - \mathbb{E}[X]| \geq c) \leq \frac{\text{Var}(X)}{c^2}, \quad c > 0. \quad (\text{КЛ11.1.7})$$

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 відхилитися від свого сподівання на деяку відстань, що вимірюється в середньоквадратичних відхиленнях.

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.

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 цен-*

⁴Пафнутій Чебишов (1821–1894) — російський математик.

тральної тенденції (measures of central tendency)⁵. До таких характеристик, окрім сподівання, також належать медіана і мода.

Визначення (КЛ11.2.1). Число M називають *медіаною* (median) випадкової величини X із розподілом $\mathbb{P}_X$, якщо

$$\mathbb{P}_X(X \leq M) \geq \frac{1}{2}, \quad \mathbb{P}_X(X \geq M) \geq \frac{1}{2}. \quad (\text{КЛ11.2.1})$$

□

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.

Інтуїтивно медіану можна інтерпретувати як значення M таке, що справа і зліва від M зосереджено половину маси розподілу величини.

Варто зазначити, що медіана існує для будь-якого розподілу, чого не можна сказати про сподівання.

Визначення (КЛ11.2.2). 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.

Для медіан справедлива цікава властивість, якої не мають сподівання.

Твердження 13.1.1. 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 $r(X)$.

Для сподівань такої властивості не існує. Натомість ми маємо хіба що нерівність Єнсена.

Інколи для опису розподілу випадкової величини однієї числової характеристики, наприклад, сподівання, зовсім недостатньо, адже її значення може бути розкидано навколо сподівання. Це мотивує нас увести поняття дисперсії. Проте, як виявляється, навіть дисперсії може бути недостатньо, і тому для будь-якої випадкової величини можна увести поняття моменту.

⁵У цьому контексті слово «міра» використано в загальновживаному значенні, жодної мови про міру як функцію від множини тут немає.

Визначення (КЛ11.4.1). Для випадкової величини X *абсолютним моментом k -го порядку* (k-th absolute moment) є

$$\mathbb{E} [|X|^k] = \int |x|^k d\mathbb{P}, \quad k = 1, 2, \dots \quad (\text{КЛ11.4.1})$$

Якщо (КЛ11.4.1) скінченний для деякого k , кажемо, що випадкова величина X має *момент k -го порядку* (k-th moment), який дорівнює

$$\mathbb{E} [X^k] = \int x^k d\mathbb{P}, \quad k = 1, 2, \dots \quad (\text{КЛ11.4.2})$$

□

Очевидно, що відповідні формули для дискретних і (абсолютно) неперервних випадкових величин з функцією ймовірності $p_X(x)$ або щільністю $f_X(x)$ набувають виду

$$\begin{aligned} \mathbb{E} [X^k] &= \sum_{x \in \text{supp}(X)} x^k p_X(x), \\ \mathbb{E} [X^k] &= \int_{-\infty}^{\infty} x^k f_X(x) dx. \end{aligned} \quad (\text{КЛ11.4.3})$$

Твердження (КЛ11.4.3). Якщо абсолютний момент k -го порядку скінченний, то скінченний також і абсолютний момент j -го порядку, $j < k$.

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 розподілу Коші не існує сподівання, відтак у нього не існує жодних інших моментів вищих порядків.

Визначення (КЛ11.4.5). Для випадкової величини X *абсолютним центральним моментом k -го порядку* (k-th absolute central moment) є

$$\mathbb{E} [|X - \mathbb{E}[X]|^k] = \int |x - \mathbb{E}[X]|^k d\mathbb{P}, \quad k = 1, 2, \dots \quad (\text{КЛ11.4.4})$$

Якщо (КЛ11.4.4) скінченний для деякого k , кажемо, що випадкова величина X має *центральний момент k -го порядку* (k-th central moment):

$$\mathbb{E} [(X - \mathbb{E}[X])^k] = \int (x - \mathbb{E}[X])^k d\mathbb{P}, \quad k = 1, 2, \dots \quad (\text{КЛ11.4.5})$$

□

Так, зокрема, сподівання є моментом першого порядку, а дисперсія —

центральним моментом другого порядку. Центральний момент *третього* порядку має корисну інтерпретацію.

Визначення (КЛ11.4.6). Стандартизований центральний момент третього порядку стандартизованої випадкової величини X (якщо він існує) називають *коефіцієнтом асиметрії* (skewness):

$$\text{Skew}(X) = \mathbb{E} \left[\left(\frac{X - \mathbb{E}[X]}{\sqrt{\text{Var}(X)}} \right)^3 \right]. \quad (\text{КЛ11.4.6})$$

□

Стандартизація у Визначенні КЛ11.4.6 дає змогу не залежати ні від розташування, ні від масштабування випадкової величини, оскільки асиметричність розподілу від цих параметрів залежати не повинна.

Визначення (КЛ11.4.7). Розподіл деякої випадкової величини X є *симетричним* (symmetric) відносно деякого числа μ , якщо випадкові величини $X - \mu$ і $\mu - X$ мають однаковий розподіл. □

Для симетричних розподілів справедливо таке:

- $\mu = \mathbb{E}[X]$ (якщо сподівання існує);
- $M = \mathbb{E}[X]$.

Твердження (КЛ11.4.8). Якщо неперервна випадкова величина X має щільність f_X , її розподіл буде симетричний відносно μ тоді й тільки тоді, коли $f_X(x) = f_X(2\mu - x)$ для всіх $x \in \text{supp}(X)$.

Часто кажуть, що розподіл «просто» симетричний, маючи на увазі, що він симетричний відносно 0. У цьому випадку щільність (і функція ймовірності) буде парною функцією: $f_X(x) = f_X(-x)$.

Твердження (КЛ11.4.9). Нехай випадкова величина X має розподіл, симетричний відносно $\mathbb{E}[X]$. Тоді $\mathbb{E}[(X - \mathbb{E}[X])^m] = 0$, $m = 2k + 1$, $k = 0, 1, 2, \dots$ (якщо він існує).

На практиці перший центральний момент завжди дорівнює 0, тому третій стандартизований момент $\text{Skew}(X)$ розглядають як міру асиметричності розподілу. Якщо розподіл симетричний, то за Твердженням КЛ11.4.9 $\text{Skew}(X) = 0$. Відтак якщо $\text{Skew}(X) \neq 0$, то розподіл не є симетричний:

- 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.

Визначення (КЛ11.4.11). Стандартизований центральний момент четвертого порядку випадкової величини X (якщо він існує), зсунутий на 3, називають *коефіцієнтом ексцесу* (kurtosis):

$$\text{Kurt}(X) = \mathbb{E} \left[\left(\frac{X - \mathbb{E}[X]}{\sqrt{\text{Var}(X)}} \right)^4 \right] - 3. \quad (\text{КЛ11.4.7})$$

□

Коефіцієнт ексцесу дає змогу оцінити, наскільки швидко розподіл прямує до нуля. Потреба у відніманні трійки у (КЛ11.4.7) виникає з того, що коефіцієнт ексцесу стандартної нормальної випадкової величини, як було встановлено у Прикладі КЛ11.4.4, 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.

Окрім уже розглянутого набору числових характеристик випадкових величин розгляньмо додатково таке важливе поняття, як квантиль розподілу.

Визначення (КЛ11.5.1). Число x_p називають *p -им квантилем* (p th quantile) випадкової величини X із розподілом $\mathbb{P}_X$, якщо

$$\mathbb{P}_X(X \leq x_p) \geq p, \quad \mathbb{P}_X(X \geq x_p) \geq 1 - p. \quad (\text{КЛ11.5.1})$$

□

Очевидно, що медіана є 0.5-квантилем випадкової величини.

Якщо функція розподілу неперервна і строго зростаюча, то $x_p = F_X^{-1}(p)$. Проте в загальному випадку квантиль не обов’язково має бути єдиним.

Фактично, x_p ділить дійсну вісь на два проміжки:

- $(-\infty; x_p]$, 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 квантилем* (first quartile);

- $x_{0.5}$, як уже було зазначено, називають медіаною;
- $x_{0.75}$ називають *третім квантилем* (third quartile).

Квантиль x_p називають $100p\%$ -им персентилем ($100p\%$ th percentile).

13.2 Нерівності зі сподіваннями

Вправа 13.2.1. Для випадкової величини X справедливо, що $Z = \ln X$, і $\mathbb{E}[Z] = 0$. Що можна сказати про сподівання $\mathbb{E}[X]$ у порівнянні з 1? $\square$

Розв'язання. Оскільки $X = e^Z$, а $g(z) = e^z$ є опуклою функцією, за нерівністю Єнсена маємо, що

$$\mathbb{E}[X] = \mathbb{E}[e^Z] \geq e^{\mathbb{E}[Z]} = e^0 = 1.$$

Понад те, ця нерівність є строгою, оскільки похідна $g'(z) = e^z$ є строго зростаючою функцією. Справді, для того, щоб $\mathbb{E}[X] = 1$, потрібно, щоб $g(z)$ було лінійною на всій множині A , де $\mathbb{P}_Z(Z \in A) = 1$, що для експоненти вочевидь не виконується. $\square$

Приклад 13.2.2. У теорії інформації (information theory), заснованій американським математиком Клодом Шенноном (Claude Elwood Shannon, 1916–2001), розглядають поняття *власної інформації* (self-information, information content, surprisal value) випадкової величини. Її визначають так:

- певна подія (тобто та, що має ймовірність 1) несе нульову інформацію;
- що менш імовірна подія, то більшу інформацію вона несе;
- інформація від двох незалежних подій дорівнює сумі окремих власних інформацій.

Можна показати, що єдиною функцією від випадкової величини, яка має ці властивості, є

$$I(x) = -\log_b(\mathbb{P}_X(X = x)), \quad (13.2.1)$$

де основа $b > 1$ може бути різною залежно від сфери застосування. Часто кладуть $b = 2$, і тоді інформацію вимірюють у бітах.

Нехай X — дискретна величина з носієм $\text{supp}(X) = \{x_1, \dots, x_n\}$ та функцією ймовірності $p_X(x_i) = p_i$, $i = 1, \dots, n$. Тоді *ентропією* (entropy) X є сподівання її власної інформації:

$$H(X) = \mathbb{E}[I(X)] = -\sum_{i=1}^n \log_b(p_i) \cdot p_i. \quad (13.2.2)$$

Варто зазначити, що ентропія залежить не від самих значень, а тільки від їхніх імовірностей, тому, наприклад, $H(X^3) = H(X)$, або в загальному випадку $H(f(X)) = H(X)$ для будь-якої взаємно однозначної функції f .

Можемо показати, що найбільшу ентропію має величина з дискретним рівномірним розподілом. (Найменшу ентропію — нульову — має детермінована константа.) Для величини $X \sim U(\{x_1, \dots, x_n\})$ ентропія дорівнює

$$H(X) = - \sum_{i=1}^n \log_b \left(\frac{1}{n} \right) \cdot \frac{1}{n} = \log_b(n) .$$

Розгляньмо випадкову величину Y із тим же носієм, але функцією ймовірності $p_Y(1/p_i) = p_i$, $i = 1, \dots, n$. Якщо $p_i = p_j$ для $i \neq j$, то величина Y набуватиме значення $1/p_i = 1/p_j$ з імовірністю $p_i + p_j = 2p_i$. Тоді

$$H(Y) = \mathbb{E}[I(Y)] = - \sum_{i=1}^n \log_b(p_i) \cdot p_i = \sum_{i=1}^n \log_b \left(\frac{1}{p_i} \right) \cdot p_i = \mathbb{E}[\log_b Y] .$$

Оскільки

$$\mathbb{E}[Y] = \sum_{i=1}^n \frac{1}{p_i} \cdot p_i = n ,$$

за нерівністю Єнсена маємо

$$H(Y) = \mathbb{E}[\log_b Y] \leq \log_b(\mathbb{E}[Y]) = \log_b n = H(X) .$$

Як було зазначено вище, значення ентропії залежить тільки від імовірностей $p_1, \dots, p_n$, а не від значень $x_1, \dots, x_n$, яким вони відповідають. Тому у визначенні Y можемо замінити носій $\text{supp}(Y) = \{1/p_1, \dots, 1/p_n\}$ на будь-який довільний, і нерівність від цього не зміниться. Отже X має найбільшу ентропію. Інтуїтивно це можна зрозуміти, адже якщо всі значення рівноймовірні, то кожне з них несе однакову інформацію. $\square$

Приклад 13.2.3. Розгляньмо важливе поняття в статистиці, яке має назву *розходження Кульбака-Ляйблера* (Kullback-Leibler divergence)⁶. Інколи його також називають *відносною ентропією* (relative entropy).

Нехай $\mathbf{p} = (p_1, \dots, p_n)^\top$ і $\mathbf{r} = (r_1, \dots, r_n)^\top$ — імовірнісні вектори, тобто їхні елементи невід'ємні, а сума значень у векторі є 1. Ці вектори можна розглядати як функції ймовірності двох дискретних величин з однаковим

⁶Назване так на честь американських математиків Соломона Кульбака (Solomon Kullback, 1907–1994) та Ричарда Ляйблера (Richard A. Leibler, 1914–2003).

носієм. Тоді розходження Кульбака-Ляйблера між $\mathbf{p}$ і $\mathbf{r}$ дорівнює

$$D_{KL}(\mathbf{p} \parallel \mathbf{r}) = \sum_{i=1}^n p_i \log_b \left(\frac{p_i}{r_i} \right) = - \sum_{i=1}^n p_i \log_b (r_i) - \left(- \sum_{i=1}^n p_i \log_b (p_i) \right). \quad (13.2.3)$$

Можна увести аналог цього розходження для неперервних величин, замінивши функції ймовірності на щільності, а суму — на інтеграл.

Це розходження можна розглядати як різницю між сподіваною власною інформацією, коли справжні ймовірності дорівнюють $\mathbf{p}$, але ми вважаємо, що вони дорівнюють $\mathbf{r}$, та сподіваною власною інформацією, коли ми знаємо, що вони дорівнюють $\mathbf{p}$. Варто звернути уваги, що в загальному випадку $D_{KL}(\mathbf{p} \parallel \mathbf{r}) \neq D_{KL}(\mathbf{r} \parallel \mathbf{p})$.

Розходження Кульбака-Ляйблера має багато застосувань, але в контексті беєсівської статистики та машинного навчання можемо казати про *приріст інформації* (information gain), який досягається шляхом заміни (апріорного) $\mathbf{r}$ на (апостеріорне) $\mathbf{p}$. Іншими словами, $D_{KL}(\mathbf{p} \parallel \mathbf{r})$ показує, скільки інформації втрачається, якщо замість $\mathbf{p}$ використовувати $\mathbf{r}$.

Можемо показати, що розходження Кульбака-Ляйблера завжди невід'ємне⁷. Справді, за властивостями логаритмів, маємо, що

$$D_{KL}(\mathbf{p} \parallel \mathbf{r}) = - \sum_{i=1}^n p_i \log_b \left(\frac{r_i}{p_i} \right).$$

Розгляньмо випадкову величину Y із функцією ймовірності $p_Y(r_i/p_i) = p_i$, $i = 1, \dots, n$. Якщо $p_i = p_j$ для деяких $i \neq j$, то величина Y набуватиме значення $r_i/p_i = r_j/p_j$ з імовірністю $p_i + p_j = 2p_i$. Тоді $D_{KL}(\mathbf{p} \parallel \mathbf{r}) = -\mathbb{E}[\log_b Y]$.

Оскільки

$$\mathbb{E}[Y] = \sum_{i=1}^n \frac{r_i}{p_i} \cdot p_i = \sum_{i=1}^n r_i = 1,$$

за нерівністю Єнсена маємо

$$D_{KL}(\mathbf{p} \parallel \mathbf{r}) = -\mathbb{E}[\log_b Y] \geq -\log_b(\mathbb{E}[Y]) = -\log_b 1 = 0.$$

Рівність досягається, тільки якщо $\mathbf{p} = \mathbf{r}$. □

Вправа 13.2.4. Чому дорівнює ймовірність $\mathbb{P}_Z(|Z| > 3)$ для стандартного нормального розподілу? Як це порівнюється з оцінками, що випливають із нерівностей Маркова та Чебишова? □

⁷Цей результат також відомий під назвою нерівності Гіббса (Gibb's inequality), названої на честь американського вченого Вільярда Гіббса (Josiah Willard Gibbs, 1839–1903).

Розв'язання. Згідно з загальновідомим правилом трьох сигм, $\mathbb{P}_Z(|Z| > 3) = 2\Phi(-3) \approx 0.003$, що можна обчислити чисельно.

Нерівність Маркова дає оцінку

$$\mathbb{P}_Z(|Z| \geq 3) \leq \frac{\mathbb{E}[|Z|]}{3}.$$

Можна підрахувати, що сподіванням величини $|Z|$ буде

$$\begin{aligned} \mathbb{E}[|Z|] &= \int_{-\infty}^{\infty} |x| \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} dx \\ &= 2 \int_0^{\infty} x \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} dx \\ &= \frac{2}{\sqrt{2\pi}} \int_0^{\infty} e^{-u} du = \sqrt{\frac{2}{\pi}}. \end{aligned}$$

Отже

$$\mathbb{P}_Z(|Z| \geq 3) \leq \frac{1}{3} \cdot \sqrt{\frac{2}{\pi}} \approx 0.27.$$

Нерівність Чебишова дає таку оцінку:

$$\mathbb{P}_X(|Z| \geq 3) = \mathbb{P}_X(|Z - \mathbb{E}[Z]| \geq 3) \leq \frac{\text{Var}(Z)}{3^2} = \frac{1}{9} \approx 0.11.$$

Як можна бачити, оцінка з нерівності Чебишова вдвічі точніша, хоча все одно дуже далека від точної оцінки, яку можна дістати з функції розподілу. $\square$

Приклад 13.2.5. Виявляється, окрім нерівності Чебишова (КЛ11.1.7), можна розглянути *однобічну* (one-sided) нерівність Чебишова⁸. Для випадкової величини X зі сподіванням $\mathbb{E}[X] = 0$ та дисперсією $\text{Var}(X) = \sigma^2$

$$\mathbb{P}_X(X \geq c) \leq \frac{\sigma^2}{\sigma^2 + c^2} \quad (13.2.4)$$

для довільного $c > 0$. Вища точність оцінки у порівнянні з нерівністю Чебишова досягається тим, що ми фокусуємося на нерівності з одного боку. Вимога оцінити ймовірність в обох хвостах розподілу неодмінно веде до того, що оцінка буде слабшою.

Вивести цю нерівність можна так. Для $b > 0$ маємо

$$\mathbb{P}_X(X \geq c) = \mathbb{P}_X(X + b \geq c + b) \leq \mathbb{P}_X((X + b)^2 \geq (c + b)^2),$$

⁸Її також називають *нерівністю Кантеллі* (Cantelli's inequality).

що впливає з монотонності ймовірності, адже подія $(X + b)^2 \geq (c + b)^2$ включає в себе подію $X + b \geq c + b$ (інший варіант може бути $-(X + b) \leq -(c + b)$). Застосування нерівності Маркова дає

$$\mathbb{P}_X(X \geq c) \leq \frac{\mathbb{E}[(X + b)^2]}{(c + b)^2} = \frac{\mathbb{E}[X^2] + 2b\mathbb{E}[X] + b^2}{(c + b)^2} = \frac{\sigma^2 + b^2}{(c + b)^2}.$$

Вираз справа мінімізує значення b , яке можна дістати з умови першого порядку:

$$\frac{2b(c + b)^2 - (\sigma^2 + b^2) \cdot 2(c + b)}{(c + b)^4} = 0.$$

Після всіх спрощень маємо

$$b(c + b) - (\sigma^2 + b^2) = 0,$$

звідки $b = \frac{\sigma^2}{c}$. Що це мінімум, а не максимум, перевірте самостійно.

Підставмо це значення в нерівність для дістання якнайточнішої оцінки:

$$\mathbb{P}_X(X \geq c) \leq \frac{\sigma^2 + \left(\frac{\sigma^2}{c}\right)^2}{\left(c + \left(\frac{\sigma^2}{c}\right)\right)^2} = \frac{\sigma^2}{\sigma^2 + c^2}.$$

□

Вправа 13.2.6. Нехай число X виробів на заводі за тиждень має розподіл зі сподіванням $\mathbb{E}[X] = 100$ та дисперсією $\text{Var}(X) = 400$. Чому дорівнює оцінка зверху ймовірності того, що протягом заданого тижня буде вироблено щонайменше 120 штук виробів? □

Розв'язання. Для застосування одnobічної нерівності Чебишова (13.2.4) потрібно, щоб величина мала нульове сподівання. Тому

$$\mathbb{P}_X(X \geq 120) = \mathbb{P}_X(X - 100 \geq 20) \leq \frac{400}{400 + (20)^2} = \frac{1}{2}.$$

Аналогічна оцінка з нерівності Маркова (яку ми можемо дозволити собі застосувати, бо $X \geq 0$) дорівнює

$$\mathbb{P}_X(X \geq 120) \leq \frac{\mathbb{E}[X]}{120} = \frac{5}{6} \approx 0.83.$$

Така оцінка є значно слабшою. □

13.3 Міри центральної тенденції

Вправа 13.3.1. Розгляньмо дискретну випадкову величину X , розподіл якої подано в табличній формі:

x	1	2	3	4	5	6	7	8	9	10
$p_X(x)$	$\frac{2}{32}$	$\frac{1}{32}$	$\frac{5}{32}$	$\frac{3}{32}$	$\frac{4}{32}$	$\frac{1}{32}$	$\frac{2}{32}$	$\frac{6}{32}$	$\frac{2}{32}$	$\frac{6}{32}$

Чому дорівнюють її сподівання, мода та медіана?

□

Розв'язання. Сподівання дискретної величини можна порахувати як відповідне зважене середнє:

$$\mathbb{E}[X] = \sum_{i=1}^{10} x \cdot p_X(x) = \frac{1}{32} (1 \cdot 2 + 2 \cdot 1 + 3 \cdot 5 + \dots + 10 \cdot 6) \approx 6.16 .$$

Модою відповідної величини буде x таке, що має найбільшу ймовірність. У нашому випадку маємо дві моди — 8 і 10, які мають ймовірність $\frac{6}{32}$ кожна.

Нарешті, за визначенням медіани потрібно знайти значення M таке, що одночасно виконуються дві нерівності:

$$\begin{aligned} \mathbb{P}_X(X \leq M) &\geq \frac{1}{2} , \\ \mathbb{P}_X(X \geq M) &\geq \frac{1}{2} . \end{aligned}$$

Розгляньмо спочатку першу нерівність. Нас цікавить значення x таке, починаючи з якого відповідна ймовірність буде дорівнювати 0.5 або більше. Нескладно бачити, що це буде число $x = 6$, адже

$$\begin{aligned} \mathbb{P}_X(X \leq 5) &= \frac{2 + 1 + 5 + 3 + 4}{32} \approx 0.47 , \\ \mathbb{P}_X(X \leq 6) &= \frac{2 + 1 + 5 + 3 + 4 + 1}{32} = 0.5 . \end{aligned}$$

Вочевидь, для $x \geq 6$ матимемо $\mathbb{P}_X(X \leq x) \geq 0.5$.

Розгляньмо тепер другу нерівність:

$$\mathbb{P}_X(X \geq M) = 1 - \mathbb{P}_X(X < M) \geq \frac{1}{2} \quad \Rightarrow \quad \mathbb{P}_X(X < M) \leq \frac{1}{2} .$$

Нас цікавить значення x таке, завершуючи яким відповідна ймовірність

буде не перевищувати 0.5. Нескладно бачити, що це буде число $x = 7$:

$$\mathbb{P}_X(X < 6) = \frac{2 + 1 + 5 + 3 + 4}{32} \approx 0.47,$$

$$\mathbb{P}_X(X < 7) = \frac{2 + 1 + 5 + 3 + 4 + 1}{32} = 0.5.$$

Вочевидь, для $x \geq 8$ матимемо $\mathbb{P}_X(X < x) > 0.5$.

Отже першу нерівність задовольняють усі числа $M \in [6; \infty)$, а другу — $M \in (-\infty; 7]$. Тому медіаною є будь-яке число $M \in [6; 7]$. Будь-яке число з інтервалу $(6; 7)$ має інтерпретацію, що зліва і справа від нього зосереджено по половині маси ймовірностей X (це видно на Рис. 13.3.1)⁹. На практиці часто беруть середнє проміжку і кажуть, що медіаною є $M = 6.5$. $\square$

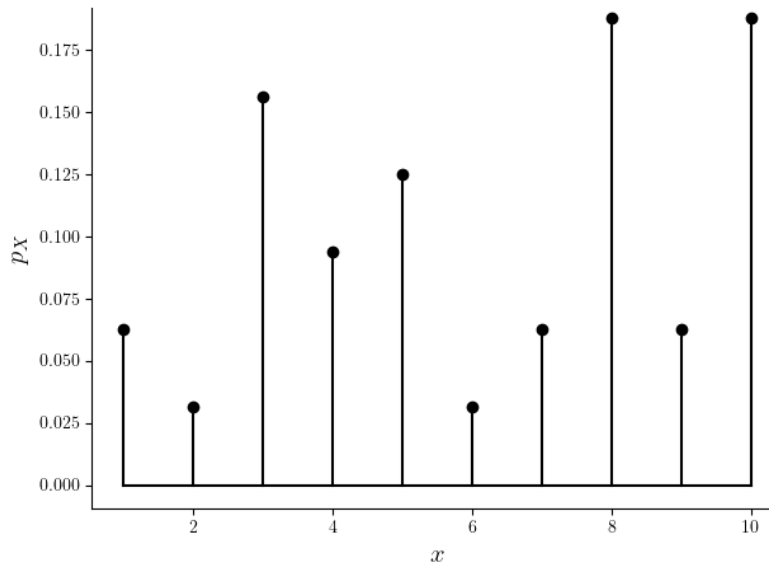


Рис. 13.3.1: Графік функції ймовірності p_X величини X для Вправи 13.3.1

Вправа 13.3.2. Чому дорівнює медіана випадкової величини з експоненційним розподілом $X \sim \text{Exp}(\lambda)$? $\square$

Розв'язання. Для неперервної випадкової величини умови

$$\mathbb{P}_X(X \leq M) \geq \frac{1}{2},$$

$$\mathbb{P}_X(X \geq M) \geq \frac{1}{2}$$

перетворюються в $\mathbb{P}_X(X \leq M) = \mathbb{P}_X(X \geq M) = \frac{1}{2}$, або ж $F_X(M) = \frac{1}{2}$.

⁹У самих точках 6 і 7 виконується загальна нерівність, тобто, наприклад, $\mathbb{P}_X(X \leq 6) = 0.5$, але $\mathbb{P}_X(X \geq 6) > 0.5$, і аналогічно для 7.

З урахуванням функції розподілу X нас цікавить значення M таке, що

$$F_X(M) = \frac{1}{2} \quad \Rightarrow \quad 1 - e^{-\lambda M} = \frac{1}{2}.$$

Тому медіаною є

$$M = \frac{\ln 2}{\lambda}.$$

Для порівняння сподіванням є $\mathbb{E}[X] = \frac{1}{\lambda} > M$.

Це проілюстровано на Рис. 13.3.2, де зафарбовані області зліва і справа від медіани мають однакову площу — 0.5. $\square$

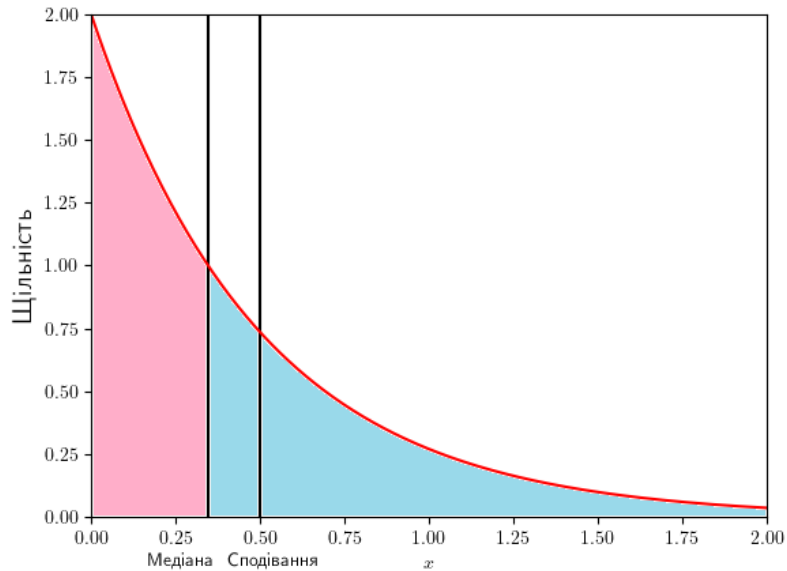


Рис. 13.3.2: Графік щільності випадкової величини $X \sim \text{Exp}(2)$ з медіаною $M \approx 0.346$ та сподіванням $\mathbb{E}[X] = 0.5$ для Вправи 13.3.2: зафарбовані області мають однакову площу 0.5

Вправа 13.3.3. Чому дорівнює медіана випадкової величини X , рівномірно розподіленої на проміжку $(a; b]$? $\square$

Розв'язання. Оскільки для рівномірного розподілу функція розподілу F_X на проміжку $(a; b]$ дорівнює

$$F_X(x) = \frac{x - a}{b - a},$$

маємо

$$\frac{1}{2} = F_X(M) = \frac{M - a}{b - a} \quad \Rightarrow \quad M = a + \frac{b - a}{2} = \frac{a + b}{2}.$$

Також можна помітити, що цей розподіл за Твердженням КЛ11.4.8 симетричний відносно свого сподівання $\mathbb{E}[X] = \frac{a+b}{2}$. Це добре видно на прикладі стандартного рівномірного розподілу на Рис. 13.3.3. Справді,

$$f_X(x) = \frac{1}{b-a} \cdot \mathbb{1}\{x \in (a; b)\}.$$

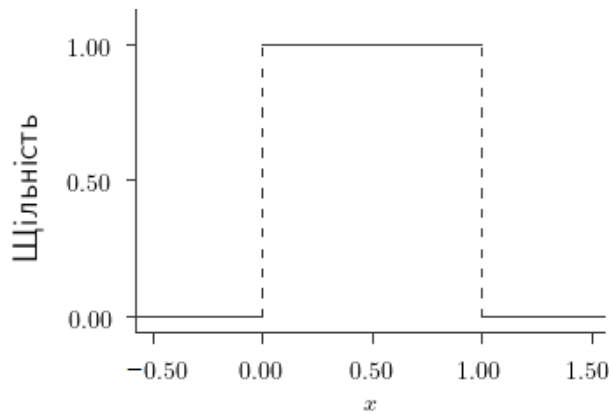


Рис. 13.3.3: Графік щільності випадкової величини X зі стандартним рівномірним розподілом, симетричним відносно $\mathbb{E}[X] = 0.5$

Розгляньмо тепер $f_X(2\mathbb{E}[X] - x)$:

$$\begin{aligned} f_X(2\mathbb{E}[X] - x) &= \frac{1}{b-a} \cdot \mathbb{1}\left\{2 \cdot \frac{a+b}{2} - x \in (a; b)\right\} \\ &= \frac{1}{b-a} \cdot \mathbb{1}\{a+b-x \in (a; b)\} \\ &= \frac{1}{b-a} \cdot \mathbb{1}\{x \in (a+b-b; a+b-a)\} = f_X(x). \end{aligned}$$

Отже для рівномірного розподілу медіана дорівнює сподіванню, тобто $M = \mathbb{E}[X] = \frac{a+b}{2}$. $\square$

Вправа 13.3.4. Чому дорівнює медіана випадкової величини з нормальним розподілом $N(\mu, \sigma^2)$? $\square$

Розв'язання. Нормальний розподіл є симетричний відносно свого сподівання, що впливає з Твердження КЛ11.4.8. Справді,

$$f_X(2\mu - x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(2\mu-x-\mu)^2}{2\sigma^2}} = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}.$$

Тому для нього медіана дорівнює сподіванню, тобто $M = \mu$ (Рис. 13.3.4). $\square$

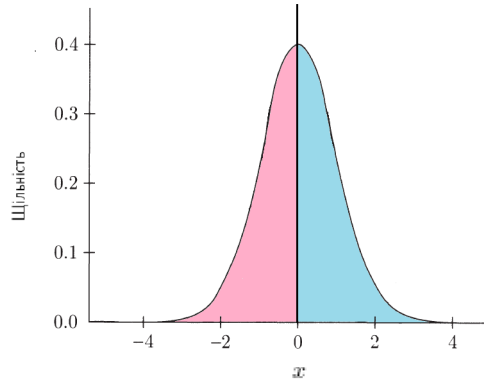


Рис. 13.3.4: Графік щільності випадкової величини X зі стандартним нормальним розподілом, симетричним відносно $\mathbb{E}[X] = 0$, для Вправи 13.3.4: зафарбовані області мають однакову площу 0.5

Вправа 13.3.5. Розгляньмо вибірку точок, розкиданих площиною згідно з розподілом Пуассона, тобто число X точок, які потрапили в деяку область, має розподіл Пуассона з інтенсивністю λ на одиницю площі. Нехай R = «відстань від початку координат до найближчої точки з вибірки».

Чому дорівнюють функція та щільність розподілу R ? Покажіть, що випадкова величина $\sqrt{2\lambda\pi}R$ має розподіл Рейлі. Чому дорівнюють сподівання, дисперсія, мода та медіана R ? $\square$

Розв'язання. За визначенням функції розподілу,

$$F_R(r) = \mathbb{P}_R(R \leq r) = 1 - \mathbb{P}_R(R > r) .$$

Подія $R > r$ означає, що в коло радіуса r не потрапило жодної точки з вибірки. Іншими словами,

$$\mathbb{P}_R(R > r) = \mathbb{P}_X(X = 0) = e^{-\lambda\pi r^2} ,$$

де ми використали те, що якщо на одиницю площі припадає в середньому λ точок, то на площу πr^2 припадає $\lambda\pi r^2$ точок, тобто $X \sim \text{Pois}(\lambda\pi r^2)$.

Звісно, для $r \leq 0$ матимемо $F_R(r) = 0$, оскільки відстань не може бути від'ємною. Отже

$$F_R(r) = \mathbb{P}_R(R \leq r) = \left(1 - e^{-\lambda\pi r^2}\right) \cdot \mathbb{1}\{r > 0\} ,$$

і тоді щільність розподілу дорівнює

$$f_R(r) = F'_R(r) = 2\lambda\pi r e^{-\lambda\pi r^2} \cdot \mathbb{1}\{r > 0\} .$$

Розгляньмо тепер випадкову величину $R_1 = \sqrt{2\lambda\pi}R$. Згідно з формулою

заміни змінних для перетворення зсуву-масштабування (КЛ7.5.2), маємо

$$\begin{aligned} f_{R_1}(r) &= f_R\left(\frac{r}{\sqrt{2\lambda\pi}}\right) \cdot \frac{1}{\sqrt{2\lambda\pi}} = 2\lambda\pi \cdot \frac{r}{\sqrt{2\lambda\pi}} \cdot e^{-\lambda\pi\left(\frac{r}{\sqrt{2\lambda\pi}}\right)^2} \cdot \frac{1}{\sqrt{2\lambda\pi}} \cdot \mathbb{1}\{r > 0\} \\ &= re^{-\frac{r^2}{2}} \cdot \mathbb{1}\{r > 0\} . \end{aligned}$$

Це справді щільність розподілу Рейлі, який ми розглядали у Прикладі 7.3.2.

Сподівання величини з розподілом Рейлі дорівнює

$$\mathbb{E}[R_1] = \int_{-\infty}^{\infty} r \cdot re^{-\frac{r^2}{2}} \cdot \mathbb{1}\{r > 0\} dr = \int_0^{\infty} r^2 e^{-\frac{r^2}{2}} dr .$$

Цей інтеграл можна не інтегрувати, адже це всього лише половина від $\sqrt{2\pi}\mathbb{E}[Z^2]$, де $Z \sim N(0, 1)$. Ми знаємо, що $\mathbb{E}[Z^2] = 1$, а тому

$$\mathbb{E}[R_1] = \frac{1}{2}\sqrt{2\pi} \cdot 1 = \sqrt{\frac{\pi}{2}} .$$

Для обчислення дисперсії величини з розподілом Рейлі спочатку знайдімо

$$\begin{aligned} \mathbb{E}[R_1^2] &= \int_0^{\infty} r^3 e^{-\frac{r^2}{2}} dr = \frac{1}{2} \int_0^{\infty} ue^{-\frac{u}{2}} du \\ &= -\frac{2}{2} \left(ue^{-\frac{u}{2}} \Big|_0^{\infty} - \int_0^{\infty} e^{-\frac{u}{2}} du \right) = \int_0^{\infty} e^{-\frac{u}{2}} du = 2 . \end{aligned}$$

Отже

$$\text{Var}(R_1) = \mathbb{E}[R_1^2] - (\mathbb{E}[R_1])^2 = 2 - \frac{\pi}{2} = \frac{4 - \pi}{2} .$$

Нарешті, сподівання й дисперсію величини R можна обчислити так:

$$\begin{aligned} \mathbb{E}[R] &= \frac{\mathbb{E}[R_1]}{\sqrt{2\lambda\pi}} = \frac{1}{2\sqrt{\lambda}} , \\ \text{Var}(R) &= \frac{\text{Var}(R_1)}{2\lambda\pi} = \frac{4 - \pi}{4\lambda\pi} . \end{aligned}$$

Модою R є число m , яке максимізує її щільність розподілу. Розгляньмо умову першого порядку:

$$f'_R(r) = 2\lambda\pi e^{-\lambda\pi r^2} - 4\lambda^2\pi^2 r e^{-\lambda\pi r^2} = 0 ,$$

звідки

$$m = r^* = \frac{1}{\sqrt{2\lambda\pi}} .$$

Те, що це саме точка максимуму, а не мінімуму, очевидно з того факту, що друга похідна в цій точці від'ємна. Можете перевірити це самостійно.

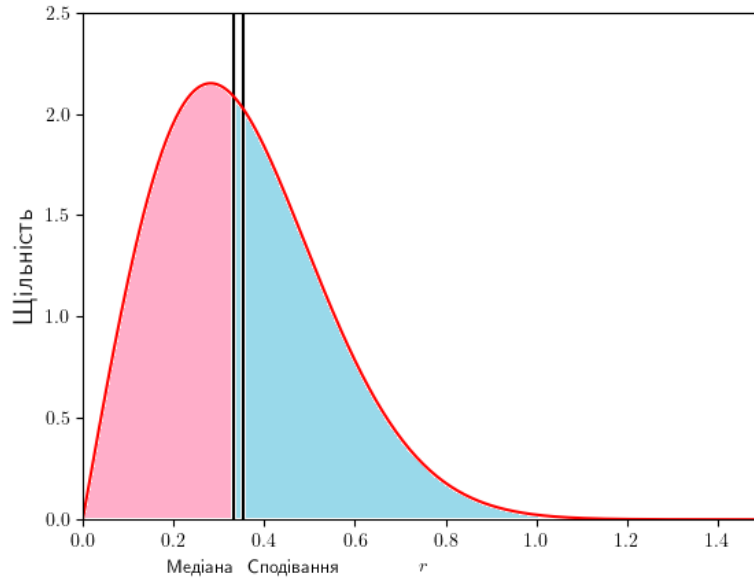


Рис. 13.3.5: Графік щільності випадкової величини R для випадку $\lambda = 2$ з медіаною $M \approx 0.332$ та сподіванням $\mathbb{E}[R] = 0.353$ для Вправи 13.3.5: зафарбовані області мають однакову площу 0.5

Медіану M можна обчислити з оберненої функції розподілу для R :

$$F_R(M) = 0.5 \quad \Rightarrow \quad 1 - e^{-\lambda\pi M^2} = 0.5 \quad \Rightarrow \quad M = \sqrt{\frac{\ln 2}{\lambda\pi}}.$$

Це проілюстровано на Рис. 13.3.5, де зафарбовані області зліва і справа від медіани мають однакову площу — 0.5.

□

13.4 Моменти вищих порядків та квантилі

Вправа 13.4.1. Нехай $X \sim U((a; b])$. Чи існують усі центральні моменти цієї випадкової величини і чому вони дорівнюють, якщо існують? □

Розв'язання. Для існування центрального моменту порядку k потрібно, щоб $\mathbb{E}[|X - \mathbb{E}[X]|^k] < \infty$. Оскільки $\mathbb{E}[X] = \frac{a+b}{2}$, а $\text{supp}(X) = (a; b]$, цілком очевидно, що всі такі інтеграли будуть скінченними. Тоді

$$\mathbb{E}[(X - \mathbb{E}[X])^k] = \mathbb{E}[Y^k],$$

де $Y = X - \frac{a+b}{2}$. Згідно з перетворенням відносно зсуву-масштабування,

$$Y \sim U\left(\left(a - \frac{a+b}{2}; b - \frac{a+b}{2}\right]\right) = U\left(\left(-\frac{b-a}{2}; \frac{b-a}{2}\right]\right).$$

Щоб спростити ситуацію, можна розглянути випадкову величину $Z = 2Y/(b-a)$ таку, що $Z \sim U((-1; 1])$. Тоді всі моменти дорівнюватимуть

$$\mathbb{E}[Z^k] = \int_{-1}^1 \frac{z^k}{2} dz = \begin{cases} 0, & k \text{ не парне} \\ \frac{1}{k+1}, & k \text{ парне} \end{cases}.$$

Остаточно

$$\mathbb{E}[Y^k] = \left(\frac{b-a}{2}\right)^k \mathbb{E}[Z^k] = \begin{cases} 0, & k \text{ не парне} \\ \left(\frac{b-a}{2}\right)^k \cdot \frac{1}{k+1}, & k \text{ парне} \end{cases}.$$

Отже це будуть моменти для випадкової величини Y , і в той же час центральні моменти — для випадкової величини X . $\square$

Вправа 13.4.2. Для величини зі Вправи 13.3.1 знайдіть перший кuartиль. $\square$

Розв'язання. За визначенням першого кuartиля потрібно знайти значення $x_{0.25}$ таке, що одночасно виконуються дві нерівності:

$$\begin{aligned} \mathbb{P}_X(X \leq x_{0.25}) &\geq \frac{1}{4}, \\ \mathbb{P}_X(X \geq x_{0.25}) &\geq \frac{3}{4}. \end{aligned}$$

Розгляньмо спочатку першу нерівність. Нас цікавить таке значення x , починаючи з якого відповідна ймовірність буде дорівнювати 0.25 або більше. Нескладно бачити, що це буде число $x = 3$, адже

$$\begin{aligned} \mathbb{P}_X(X \leq 2) &= \frac{2+1}{32} \approx 0.09, \\ \mathbb{P}_X(X \leq 3) &= \frac{2+1+5}{32} = 0.25. \end{aligned}$$

Вочевидь, для $x \geq 3$ матимемо $\mathbb{P}_X(X \leq x) \geq 0.25$.

Розгляньмо тепер другу нерівність:

$$\mathbb{P}_X(X \geq x_{0.25}) = 1 - \mathbb{P}_X(X < x_{0.25}) \geq \frac{3}{4} \Rightarrow \mathbb{P}_X(X < x_{0.25}) \leq \frac{1}{4}.$$

Нас цікавить значення x таке, завершуючи яким відповідна ймовірність

буде не перевищувати 0.25. Нескладно бачити, що це буде число $x = 4$:

$$\begin{aligned}\mathbb{P}_X(X < 3) &= \frac{2+1}{32} \approx 0.09, \\ \mathbb{P}_X(X < 4) &= \frac{2+1+5}{32} = 0.25.\end{aligned}$$

Вочевидь, для $x \geq 5$ матимемо $\mathbb{P}_X(X < x) > 0.25$.

Отже першу нерівність задовольняють усі числа $x_{0.25} \in [3; \infty)$, а другу — $x_{0.25} \in (-\infty; 4]$. Першим квантилем є будь-яке число $x_{0.25} \in [3; 4]$. Будь-яке число з інтервалу $(3; 4)$ має інтерпретацію, що зліва від нього зосереджено чверть маси ймовірностей X , а справа — три чверті (Рис. 13.3.1)¹⁰. $\square$

Вправа 13.4.3. Для випадкових величин зі Вправ 13.3.2–13.3.4 знайдіть вирази для всіх квантилів. $\square$

Розв'язання. Для неперервної випадкової величини умови

$$\begin{aligned}\mathbb{P}_X(X \leq x_p) &\geq p, \\ \mathbb{P}_X(X \geq x_p) &\geq 1 - p\end{aligned}$$

перетворюються в $\mathbb{P}_X(X \leq x_p) = 1 - \mathbb{P}_X(X \geq x_p) = p$, або ж $F_X(x_p) = p$.

Для експоненційного розподілу нас цікавить таке значення x_p , для якого

$$F_X(x_p) = p \quad \Rightarrow \quad 1 - e^{-\lambda x_p} = p.$$

Тому квантилем експоненційного розподілу є

$$x_p = -\frac{\ln(1-p)}{\lambda}.$$

Для рівномірного розподілу функція розподілу F_X на $(0; 1)$ дорівнює

$$F_X(x) = \frac{x-a}{b-a},$$

тому маємо

$$p = F_X(x_p) = \frac{x_p - a}{b - a} \quad \Rightarrow \quad x_p = a + p(b - a).$$

Нарешті, для нормального розподілу квантилем буде число, яке розв'язує рівняння $\Phi(x_p) = p$, і його можна обчислити тільки чисельно. $\square$

¹⁰У самих точках 3 і 4 виконується загальна нерівність, тобто, наприклад, $\mathbb{P}_X(X \leq 3) = 0.25$, але $\mathbb{P}_X(X \geq 3) > 0.25$, і аналогічно для 3.

Приклад 13.4.4. Розгляньмо *криву Лоренца* (Lorenz curve)¹¹. Це функція $L(p)$, що дорівнює частці сукупного доходу в державі, яка належить 100 p % найбіднішим працівникам. Розгляньмо квантиль x_p випадкової величини X_i = «дохід особи i » та індикаторну величину $I_p(X) = \mathbb{1}\{X < x_p\}$. Тоді частка сукупного доходу, який належить особам, дохід яких не перевищує x_p , дорівнює $\frac{\sum_{i=1}^n X_i I_p(X_i)}{\sum_{i=1}^n X_i}$, що за законом великих чисел, який ми розглянемо далі, прямує до

$$L(p) = \frac{\mathbb{E}[X I_p(X)]}{\mathbb{E}[X]}, \quad (13.4.1)$$

коли $n \rightarrow \infty$. Можна помітити, що

$$\mathbb{E}[X I_p(X)] = \int_0^\infty x I_p(x) \cdot f(x) dx = \int_0^{x_p} x f(x) dx,$$

а тому

$$L(p) = \frac{1}{\mathbb{E}[X]} \int_0^{x_p} x f(x) dx. \quad (13.4.2)$$

Можемо обчислити криву Лоренца у випадку, коли дохід працівників розподілено рівномірно на проміжку $(a; b)$, $0 \leq a < b$. Як ми показали у Вправі 13.4.3, квантилем буде $x_p = a + (b - a)p$, і з (13.4.2) маємо

$$\begin{aligned} L(p) &= \frac{2}{a+b} \int_0^{a+(b-a)p} x \cdot \frac{1}{b-a} \cdot \mathbb{1}\{x \in (a; b)\} dx \\ &= \frac{2}{(a+b)(a-b)} \int_a^{a+(b-a)p} x dx = \frac{2pa + (b-a)p^2}{a+b}. \end{aligned}$$

Зокрема, $\lim_{a \rightarrow b} L(p) = p$, тобто якщо всі отримують однаковий дохід, то частка доходу, яку отримують особи з доходом до x_p , є p , що очевидно.

Можна показати, що $L(p) \leq p$ ¹². Це дає змогу увести простий показник нерівності доходів — *індекс Джині* (Gini index)¹³. Оскільки $L(p) = p$ відповідає максимальній рівності доходів, площа між прямою $y = p$ та кривою $L(p)$ (Рис. 13.4.1) є мірою нерівності доходів. Індекс Джині G визначають як відношення цієї площі до площі трикутника під прямою, тобто

$$G = \frac{\frac{1}{2} - \int_0^1 L(p) dp}{\frac{1}{2}} = 1 - 2 \int_0^1 L(p) dp. \quad (13.4.3)$$

Наприклад, для випадку з рівномірним розподілом доходів, для якого ми

¹¹ Названа так на честь американського економіста Макса Лоренца (Max Otto Lorenz, 1876–1959), що її запропонував у 1905 р.

¹² Це випливає з $\mathbb{E}[f(X)g(X)] \geq \mathbb{E}[f(X)]\mathbb{E}[g(X)]$ для строго зростаючих f і g , застосованої до (13.4.1).

¹³ Названий так на честь італійського статистика Коррадо Джині (Corrado Gini, 1884–1965), що його запропонував у 1912 р.

порахували $L(p) = \frac{2pa + (b-a)p^2}{a+b}$, індекс Джині дорівнює

$$\begin{aligned} G &= 1 - 2 \int_0^1 \frac{2pa + (b-a)p^2}{a+b} dp = 1 - \frac{2}{a+b} \left(ap^2 \Big|_0^1 + \frac{b-a}{3} p^3 \Big|_0^1 \right) \\ &= 1 - \frac{2}{3} \cdot \frac{2a+b}{a+b} . \end{aligned}$$

□

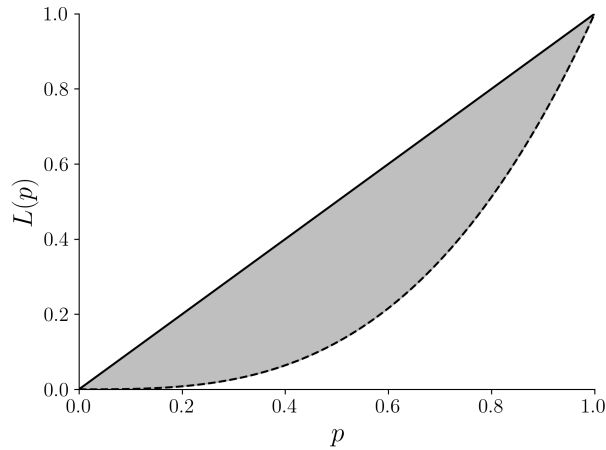


Рис. 13.4.1: Ілюстрація деякої кривої Лоренца $L(p)$

13.5 Додаткові вправи

Вправа 13.5.1. Розгляньмо тест із питаннями з декількома варіантами відповідей. Замість вибору одного варіанту студенту пропонують присвоїти кожному варіанту ймовірність, що він правильний. Тоді балом, здобутим за питання, є (натуральний) логаритм імовірності, що присвоєно відповіді, яка насправді правильна. Найліпший результат, який можна набрати за тест, дорівнює 0 — у випадку, коли всім правильним відповідям зіставлено одиничну ймовірність. Найгіршим результатом є $-\infty$, яку можна здобути, присвоївши нульову ймовірність принаймні одній правильній відповіді.

Нехай наші апіорні уявлення про правильність кожного варіанта відповіді дорівнюють $p_1, \dots, p_n$, де всі числа невід'ємні, а їх сума дорівнює 1. Покажіть, що сподіваний бал за питання максимізується, тільки якщо варіантам відповідей присвоєно правильні ймовірності. □

Розв'язання. Ця задача є повністю ізоморфною до Прикладу 13.2.3. Справді, сподіваний бал за питання дорівнює

$$\mathbb{E}[X] = \sum_{i=1}^n p_i \ln r_i ,$$

де r_i — імовірності, які присвоїв студент. Оскільки розходження Кульбака-Ляйблера завжди невід’ємне,

$$0 \leq D_{KL}(\mathbf{p} \parallel \mathbf{r}) = - \sum_{i=1}^n p_i \ln(r_i) - \left(- \sum_{i=1}^n p_i \ln(p_i) \right) ,$$

звідки

$$\sum_{i=1}^n p_i \ln(p_i) \geq \sum_{i=1}^n p_i \ln(r_i) ,$$

тобто сподіваний бал найвищий, якщо дати правильні відповіді. $\square$

Вправа 13.5.2. Нехай групу з 200 осіб (100 жінок і 100 чоловіків) у випадковий спосіб розбивають на 100 пар по 2 людини. Чому дорівнює оцінка зверху ймовірності того, що різностатевих пар буде щонайбільше 30? $\square$

Розв’язання. Розгляньмо індикаторну випадкову величину $X_i = \langle \text{чоловік } i \text{ має парою жінку} \rangle$. Тоді число різностатевих пар дорівнює $X = \sum_{i=1}^{100} X_i$.

Імовірність, що деякий чоловік матиме парою жінку, дорівнює $\mathbb{E}[X_i] = \mathbb{P}_X(X_i = 1) = \frac{100}{199}$, оскільки всього існує 199 кандидатів на формування пари, 100 з яких — жінки. У схожий спосіб можна помітити, що для $i \neq j$

$$\begin{aligned} \mathbb{E}[X_i X_j] &= \mathbb{P}_X(X_i = 1, X_j = 1) \\ &= \mathbb{P}_X(X_i = 1) \cdot \mathbb{P}_X(X_j = 1 \mid X_i = 1) = \frac{100}{199} \cdot \frac{99}{197} . \end{aligned}$$

Остаточно

$$\begin{aligned} \mathbb{E}[X] &= \sum_{i=1}^{100} \mathbb{E}[X_i] = \frac{100 \cdot 100}{199} \approx 50.25 , \\ \text{Var}(X) &= \sum_{i=1}^{100} \text{Var}(X_i) + 2 \sum_{i < j} \text{Cov}(X_i, X_j) \\ &= \sum_{i=1}^{100} \text{Var}(X_i) + 2 \binom{100}{2} (\mathbb{E}[X_i X_j] - \mathbb{E}[X_i] \mathbb{E}[X_j]) \\ &= 100 \cdot \frac{100}{199} \cdot \frac{99}{197} + 2 \binom{100}{2} \left(\frac{100}{199} \cdot \frac{99}{197} - \left(\frac{100}{199} \right)^2 \right) \approx 25.126 . \end{aligned}$$

Тоді за нерівністю Чебишова (КЛ11.1.7) маємо

$$\mathbb{P}_X(X \leq 30) \leq \mathbb{P}_X(|X - 50.25| \geq 20.25) \leq \frac{25.126}{(-20.25)^2} \approx 0.061 ,$$

де ми використали той факт, що подія $|X - 50.25| \geq 20.25$ включає в себе подію $X \leq 30$ (інший варіант може бути $X - 50.25 \geq 20.25$).

А за однобічною нерівністю Чебишова (13.2.4) маємо

$$\mathbb{P}_X(X \leq 30) \leq \mathbb{P}_X(X - 50.25 \leq -20.25) \leq \frac{25.126}{25.126 + (-20.25)^2} \approx 0.058 ,$$

що є точнішою оцінкою. $\square$

Вправа 13.5.3. Нехай еколог вважає, що водойму забруднюють дві речовини — миш'як і свинець. Справжні концентрації кожної з цих речовин є незалежними випадковими величинами X та Y (з однаковими одиницями виміру). Еколога цікавить частка свинцю серед забруднювачів, тобто він хоче дізнатися, чому дорівнює сподівання величини $R = Y/(X + Y)$.

Нехай нам відомі розподіли X та Y , проте з якихось міркувань виведення розподілу R на їхній основі є ускладненим, а підрахунок сподівання $\mathbb{E}[R]$ відповідно не є можливим. У цьому випадку значення сподівання можна дістати шляхом симуляції на комп'ютері. Справді, можна згенерувати велику вибірку $(X_1, Y_1), \dots, (X_n, Y_n)$ на комп'ютері з відповідними розподілами та підрахувати середнє вибіркове значення

$$\bar{R} = \frac{1}{n} \sum_{i=1}^n \frac{Y_i}{X_i + Y_i} .$$

Оскільки сподіванням такого значення буде $\mathbb{E}[\bar{R}] = \mathbb{E}[R]$, відповідна симуляція дасть змогу дістати непогану оцінку справжнього параметра, який напяму обчислити непросто.

Якого розміру потрібно взяти вибірку, щоб така апроксимація була достатньо точною? Оскільки $|R| \leq 1$,

$$\text{Var}(R) = \mathbb{E}[R^2] - \mathbb{E}[R]^2 \leq \mathbb{E}[R^2] = 1 ,$$

отже за нерівністю Чебишова маємо

$$\mathbb{P}_R(|\bar{R} - \mathbb{E}[R]| \geq \varepsilon) \leq \frac{1}{n\varepsilon^2} .$$

Наприклад, якщо ми хочемо, щоб середнє вибіркове відхилялося від $\mathbb{E}[R]$ не більше від 0.005 з імовірністю щонайменше 0.98, нам треба таке n , щоб

$$\mathbb{P}_R(|\bar{R} - \mathbb{E}[R]| \leq 0.005) \geq 1 - \frac{1}{n \cdot 0.005^2} \Rightarrow n > \frac{1}{0.02 \cdot 0.005^2} = 2\,000\,000 .$$

$\square$

Вправа 13.5.4. Згідно з моделлю газу Максвелла, молекули маси m рухаються зі швидкостями, проекції яких на координатні осі V_x , V_y та V_z незалежні одна від одної, а їхній спільний розподіл інваріантний відносно поворотів у тривимірному просторі. Максвелл показав, що $V_x, V_y, V_z \sim N(0, \sigma^2)$ для деякого $\sigma > 0$.

З урахуванням цього результату, чому дорівнює щільність кінетичної енергії молекули

$$K = \frac{1}{2}m (V_x^2 + V_y^2 + V_z^2) ?$$

Чому дорівнюють сподівання, медіана, мода розподілу кінетичної енергії? $\square$

Розв'язання. Потрібно помітити, що

$$\frac{2}{m\sigma^2}K = \frac{1}{\sigma} (V_x^2 + V_y^2 + V_z^2) = \left(\frac{V_x}{\sigma}\right)^2 + \left(\frac{V_y}{\sigma}\right)^2 + \left(\frac{V_z}{\sigma}\right)^2 \sim \chi_3^2$$

як сума квадратів трьох незалежних стандартних нормальних величин. Згідно з (КЛ10.2.3), щільність $\frac{2}{m\sigma^2}K$ дорівнює

$$f_{\frac{2}{m\sigma^2}K}(x) = \frac{1}{2^{\frac{3}{2}}\Gamma(\frac{3}{2})}x^{\frac{3}{2}-1}e^{-\frac{x}{2}} \cdot \mathbb{1}\{x > 0\} = \frac{1}{\sqrt{2\pi}}\sqrt{x}e^{-\frac{x}{2}} \cdot \mathbb{1}\{x > 0\} .$$

А тому з урахуванням перетворення масштабування маємо

$$\begin{aligned} f_K(x) &= f_{\frac{2}{m\sigma^2}K}\left(\frac{2}{m\sigma^2}x\right) \cdot \frac{2}{m\sigma^2} = \frac{1}{\sqrt{2\pi}}\sqrt{\frac{2}{m\sigma^2}x} \cdot e^{-\frac{1}{2} \cdot \frac{2}{m\sigma^2}x} \cdot \frac{2}{m\sigma^2} \cdot \mathbb{1}\{x > 0\} \\ &= \frac{2}{\sqrt{\pi}(m\sigma^2)^{3/2}}\sqrt{x}e^{-\frac{x}{m\sigma^2}} \cdot \mathbb{1}\{x > 0\} . \end{aligned}$$

Оскільки сподівання розподілу χ_3^2 дорівнює 3 як сподівання розподілу $\text{Gamma}(1.5, 0.5)$, сподівання кінетичної енергії дорівнює

$$\mathbb{E}[K] = 3 \cdot \frac{m\sigma^2}{2} = \frac{3}{2}m\sigma^2 .$$

Модой розподілу K буде число, яке максимізує відповідну щільність. Розгляньмо умову першого порядку для щільності:

$$f'_K(x) = \frac{2}{\sqrt{\pi}(m\sigma^2)^{3/2}} \left(\frac{1}{2\sqrt{x}}e^{-\frac{x}{m\sigma^2}} + \sqrt{x}e^{-\frac{x}{m\sigma^2}} \left(-\frac{1}{m\sigma^2} \right) \right) \cdot \mathbb{1}\{x > 0\} = 0 ,$$

звідки

$$\frac{1}{2\sqrt{x}} - \frac{\sqrt{x}}{m\sigma^2} = 0 ,$$

звідки $x = \frac{1}{2}t\sigma^2$.

Нарешті, медіана розподілу K буде такою ж, як і медіана розподілу χ_3^2 , що впливає з Твердження КЛ11.2.4 з урахуванням того, що K утворено з величини з розподілом χ_3^2 множенням на додатну мультиплікативну константу. Проте обчислити медіану можна хіба що чисельно, як таке значення M , для якого виконується

$$\int_0^M \frac{1}{\sqrt{2\pi}} \sqrt{x} e^{-\frac{x}{2}} dx = \frac{1}{2}.$$

Виявляється, це значення дорівнює $M \approx 2.366$. □

Вправа 13.5.5. Нехай моменти випадкової величини X дорівнюють $\mathbb{E}[X] = 1$, $\mathbb{E}[X^2] = 2$, $\mathbb{E}[X^3] = 5$. Чому дорівнює третій центральний момент X ? Який висновок можна зробити про симетричність розподілу X ? □

Розв'язання. За визначенням,

$$\begin{aligned} \mathbb{E}[(X - \mathbb{E}[X])^3] &= \mathbb{E}[(X - 1)^3] = \mathbb{E}[X^3 - 3X^2 + 3X - 1] \\ &= 5 - 3 \cdot 2 + 3 \cdot 1 - 1 = 1. \end{aligned}$$

Оскільки $1 > 0$, маємо, що розподіл скошений управо. □

Вправа 13.5.6. Покажіть, що

$$\mathbb{E}[(X - \mathbb{E}[X])^4] \geq (\text{Var}(X))^2$$

та зробіть висновок, що коефіцієнт ексцесу будь-якої випадкової величини не менший від -2 . □

Розв'язання. Можна застосувати нерівність Єнсена, помітивши, що $(X - \mathbb{E}[X])^4 = ((X - \mathbb{E}[X])^2)^2$, а тому

$$\mathbb{E}[(X - \mathbb{E}[X])^4] \geq (\mathbb{E}[(X - \mathbb{E}[X])^2])^2 = (\text{Var}(X))^2.$$

Звідси випливає, що коефіцієнт ексцесу

$$\text{Kurt}(X) = \mathbb{E} \left[\left(\frac{X - \mathbb{E}[X]}{\sqrt{\text{Var}(X)}} \right)^4 \right] - 3 \geq 1 - 3 = -2.$$

□

14 Умовні розподіли

14.1 Основні теоретичні відомості

Детальний огляд понять, потрібний для розв'язання задач на цьому занятті, наведено в Курсі лекцій, Розд. КЛ12.

Для будь-яких дискретних випадкових величин X і Z можна визначити функцію $p_{Z|X}(z | x) \equiv \mathbb{P}_{Z|X}(Z = z | X = x)$, яку називають *умовною функцією ймовірності* (conditional probability mass function) величини Z за умови $X = x$. Обчислити її можна за допомогою (КЛ5.5.1) 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.

Із цього визначення в повній аналогії з умовними ймовірностями випливає формула Беєса для функцій імовірності дискретних величин X та Z :

$$\mathbb{P}_{Z|X}(Z = k | X = n) = \frac{\mathbb{P}_{X|Z}(X = n | Z = k) \mathbb{P}_Z(Z = k)}{\mathbb{P}_X(X = n)}, \quad (\text{КЛ12.1.1})$$

якщо, звісно, знаменник не обертається в нуль.

Визначмо *умовну функцію розподілу* (conditional distribution function):

$$F_{Z|X}(z | x) = \mathbb{P}_{Z|X}(Z \leq z | X = x). \quad (\text{КЛ12.1.2})$$

Для дискретних величин її застосування доволі обмежене, як і функції розподілу безумовної.

Умовне сподівання дискретної величини, як безумовне, є зважене середнє

$$\mathbb{E}[Z | X = x] = \sum_{z \in \text{supp}(Z)} z \cdot p_{Z|X}(z | x). \quad (\text{КЛ12.1.3})$$

Варто звернути увагу, що умовне сподівання є функцією від x , а не сталою, тобто умовне сподівання є випадковою величиною.

Для (абсолютно) неперервних випадкових величин можна розглянути аналогічні поняття.

Твердження (КЛ12.4.1). Нехай випадкові величини X та Z мають абсолютно неперервний спільний розподіл зі щільністю $f_{Z,X}$. Нехай фун-

кція $g(Z)$ 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)

$$f_{Z|X}(z | x) = \begin{cases} \frac{f_{Z,X}(z, x)}{f_X(x)}, & f_X(x) \neq 0 \\ \xi(z), & f_X(x) = 0 \end{cases} \quad (\text{КЛ12.4.1})$$

у тому розумінні, що

$$\mathbb{E}[g(Z) | X] = h(X),$$

де

$$h(x) = \int_{-\infty}^{\infty} g(z) \cdot f_{Z|X}(z | x) dz. \quad (\text{КЛ12.4.2})$$

У визначенні умовної щільності ξ є деякою абсолютно довільною щільністю розподілу.

У випадку, коли $f_X(x) = 0$ у (КЛ12.4.1), ми можемо дозволити собі покласти $f_{Z|X}(z | x) = \xi(z)$, де ξ — 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 (КЛ12.4.1) має всі властивості будь-якої щільності, зокрема, вона очевидно є невід'ємною, а її інтеграл дорівнює 1:

$$\begin{aligned} \int_{-\infty}^{\infty} f_{Z|X}(z | x) dz &= \int_{-\infty}^{\infty} \frac{f_{Z,X}(z, x)}{f_X(x)} dz = \frac{1}{f_X(x)} \cdot \int_{-\infty}^{\infty} f_{Z,X}(z, x) dz \\ &= \frac{f_X(x)}{f_X(x)} = 1, \end{aligned}$$

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 ξ є деякою довільною, але все ж таки щільністю.

Зауваження (КЛ12.4.2). Умовна щільність (КЛ12.4.1) 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 (КЛ12.4.2) матимемо

$$\mathbb{P}(A | X = x) = \mathbb{E}[\mathbb{1}\{Z \in A\} | X = x] = \int_A f_{Z|X}(z | x) dz.$$

□

Для повноти картини ми також можемо розглянути поняття умовної функції розподілу.

Визначення (КЛ12.4.5). 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

$$F_{Z|X}(z | x) = \int_{-\infty}^z f_{Z|X}(u | x) du = \int_{-\infty}^z \frac{f_{Z,X}(u, x)}{f_X(x)} du . \quad (\text{КЛ12.4.3})$$

□

Зауваження (КЛ12.4.6). Потрібно розуміти, що обумовлення в записі умовної щільності $f_{Z|X}$ здійснюють цілою випадковою величиною X із відповідною σ -алгеброю, а не конкретною подією $X = x$, яка є нульовою, оскільки $\mathbb{P}_X(X = x) = 0$ для неперервних величин. Певну інтуїтивну інтерпретацію умовної щільності можна подати, розглянувши її не як похідну умовної функції розподілу, обчислену для $X = x$, тобто не як похідну $F_{Z|X}(z | x) = \mathbb{P}_{Z|X}(Z \leq z | X = x)$, 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 $f_X(x) > 0$. □

Для умовної щільності також, із цілком очевидних міркувань, справедлива теорема Беєса:

$$f_{Z|X}(z | x) = \frac{f_{X|Z}(x | z)f_Z(z)}{f_X(x)} , \quad f_X(x) > 0 . \quad (\text{КЛ12.4.4})$$

Також справедлива формула повної ймовірності для неперервного випадку:

$$\begin{aligned} \mathbb{P}_X(X \in A) &= \mathbb{E}[\mathbb{1}\{X \in A\}] = \mathbb{E}[\mathbb{E}[\mathbb{1}\{X \in A\} | Y]] = \mathbb{E}[\mathbb{P}_{X|Y}(X \in A | Y)] \\ &= \int_{-\infty}^{\infty} \mathbb{P}_{X|Y}(X \in A | Y = y) \cdot f_Y(y) dy . \end{aligned} \quad (\text{КЛ12.3.7})$$

Табл. 14.2.1: Спільна функція ймовірності для Вправи 14.2.1

	$Y = 0$	$Y = 1$	$Y = 2$	$Y = 3$	$Y = 4$	Разом
$X = 0$	0.129	0.298	0.161	0.280	0.108	0.976
$X = 1$	0.010	0.010	0.001	0.002	0.001	0.024
Разом	0.139	0.308	0.162	0.282	0.109	1

Розгляньмо задачу прогнозування значення деякої величини Y на основі спостережень за значенням іншої величини X . Позначмо предиктор величини Y через $g(X)$, підкресливши, 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 (КЛ12.3.16). Умовне сподівання випадкової величини Y зі скінченним 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:

$$\mathbb{E}[Y | X] = \arg \min_g \mathbb{E}[(Y - g(X))^2] . \quad (\text{КЛ12.3.8})$$

Можна довести, що *умовна медіана* (conditional median) $M_{Y|X}$ також є найліпшим предиктором, але в розумінні мінімізації *середньої абсолютної похибки* (mean absolute error):

$$M_{Y|X} = \arg \min_g \mathbb{E}[|Y - g(X)|] . \quad (\text{КЛ12.3.9})$$

За аналогією з умовним сподіванням також можна визначити умовні моменти довільного порядку. Зокрема, *умовною дисперсією* (conditional variance) величини Z за умови величини X є

$$\text{Var}(Z | X) = \mathbb{E}[(Z - \mathbb{E}[Z | X])^2 | X] = \mathbb{E}[Z^2 | X] - (\mathbb{E}[Z | X])^2 . \quad (\text{КЛ12.3.10})$$

14.2 Умовні функції ймовірності

Вправа 14.2.1. Страхова компанія веде статистику викрадень автомобілів різних марок. Спільну функцію ймовірності для випадкових величин X = «чи сталося викрадення» та Y = «марка автомобіля» наведено в Табл. 14.2.1.

Коли клієнт звертається по послуги страхової компанії для автомобіля певної марки, компанії хотілося б мати оцінку ймовірності, що саме цей

Табл. 14.2.2: Умовна функція ймовірностей $X | Y$ для Вправи 14.2.1

	$Y = 0$	$Y = 1$	$Y = 2$	$Y = 3$	$Y = 4$
$X = 0$	0.928	0.968	0.994	0.993	0.991
$X = 1$	0.072	0.032	0.006	0.007	0.009

Табл. 14.2.3: Результати опитування домогосподарств для Вправи 14.2.2

		X							
		1	2	3	4	5	6	7	8
Y	0	10	7	3	2	2	1	0	0
	1	12	21	25	30	25	15	5	1
	2	1	5	10	15	20	11	5	3
	3	0	2	3	5	5	3	2	1

автомобіль буде викрадено. Це може вплинути на вартість відповідного страхового полісу. Адже хоча загальна (маржинальна) ймовірність для деякого автомобіля бути викраденим дорівнює $\mathbb{P}_X(X = 1) = 0.024$, машини конкретних марок може бути викрадено з вищою ймовірністю.

Чому дорівнює умовна функція ймовірності $p_{X|Y}(x | y)$? □

Розв'язання. За формулою умовної ймовірності маємо

$$p_{X|Y}(x | y) = \frac{\mathbb{P}_{X,Y}(X = x, Y = y)}{\mathbb{P}_Y(Y = y)}.$$

Для нашої таблиці на практиці це означає, що для кожного $Y = y$ потрібно поділити значення в комірці з $X = x$, $Y = y$ на значення, що відповідає $Y = y$, в останньому рядку таблиці. Остаточно маємо результат, наведений у Табл. 14.2.2. Наприклад, $0.928 \approx 0.129/0.139$ тощо. □

Вправа 14.2.2. Результати опитування 250 домогосподарств, наведені в Табл. 14.2.3, показують, скільки автомобілів Y мають родини розміру X . Чому дорівнює умовне сподівання числа автомобілів в y випадковий спосіб вибраної родини з такої популяції, за умови, що ми спостерігаємо її розмір? □

Розв'язання. За визначенням умовного сподівання нас цікавить

$$\mathbb{E}[Y | X = x] = \sum_{y=0}^3 y \cdot \mathbb{P}_{Y|X}(Y = y | X = x) .$$

Відповідну умовну функцію ймовірності можна дістати як відношення спільної функції ймовірності до маржинальної:

$$\mathbb{P}_{Y|X}(Y = y | X = x) = \frac{\mathbb{P}_{X,Y}(X = x, Y = y)}{\mathbb{P}_X(X = x)} .$$

Оскільки всі сім'ї мають однакову ймовірність бути вибраними, спільна функція ймовірності є $t_{xy}/250$, де t_{xy} — це відповідний елемент таблиці, $x = 1, \dots, 8$, $y = 0, \dots, 3$. Маржинальну функцію ймовірності $\mathbb{P}_X(X = x)$ можна дістати, сумуючи всі Y для цього x та ділячи на 250. Наприклад,

$$\mathbb{P}_X(X = 1) = \frac{10 + 12 + 1 + 0}{250} \approx 0.092 .$$

Тоді, наприклад,

$$\mathbb{P}_{Y|X}(Y = 0 | X = 1) = \frac{10/250}{(10 + 12 + 1 + 0)/250} = \frac{10}{10 + 12 + 1 + 0} \approx 0.435 .$$

У схожий спосіб можна підрахувати й інші значення. Відповідну умовну функцію ймовірності наведено в Табл. 14.2.4. Тоді, наприклад

$$\begin{aligned} \mathbb{E}[Y | X = 1] &= 0 \cdot \mathbb{P}_{Y|X}(Y = 0 | X = 1) + 1 \cdot \mathbb{P}_{Y|X}(Y = 1 | X = 1) \\ &\quad + 2 \cdot \mathbb{P}_{Y|X}(Y = 2 | X = 1) + 3 \cdot \mathbb{P}_{Y|X}(Y = 3 | X = 1) \\ &\approx 0 \cdot 0.435 + 1 \cdot 0.522 + 2 \cdot 0.043 + 3 \cdot 0 \approx 0.609 . \end{aligned}$$

Рахуючи аналогічні вирази для інших значень X , дістаємо, що $Z \equiv \mathbb{E}[Y | X]$ є випадковою величиною такою, що

$$Z(x) = \begin{cases} 0.609 , & x = 1 \\ 1.057 , & x = 2 \\ 1.317 , & x = 3 \\ 1.442 , & x = 4 \\ 1.538 , & x = 5 \\ 1.533 , & x = 6 \\ 1.75 , & x = 7 \\ 2 , & x = 8 \end{cases} .$$

□

Табл. 14.2.4: Умовний розподіл $Y \mid X$ для Вправи 14.2.2

		X							
		1	2	3	4	5	6	7	8
Y	0	0.435	0.200	0.073	0.038	0.038	0.033	0	0
	1	0.522	0.600	0.610	0.577	0.481	0.500	0.417	0.2
	2	0.043	0.143	0.244	0.288	0.385	0.367	0.417	0.6
	3	0	0.057	0.073	0.096	0.096	0.100	0.167	0.2

Вправа 14.2.3. Нехай протягом дня до поштового відділку заходить $N \sim \text{Pois}(\lambda)$ клієнтів. Клієнт може виявитися чоловіком з імовірністю p . Нехай X = «число чоловіків за день», Y = «число жінок за день», $X + Y = N$.

Чому дорівнює спільна функція ймовірності для $(X, Y)^\top$? Чи є X та Y незалежні? Які їхні маржинальні розподіли? $\square$

Розв'язання. Зафіксувавши число клієнтів, $N = n$, кожного клієнта доречно розглядати як «випробування», яке може закінчитися «успіхом» (чоловік) з імовірністю p , тобто $X \mid N = n \sim \text{Binom}(n, p)$, $Y \mid N = n \sim \text{Binom}(n, 1 - p)$. Тоді за формулою повної ймовірності (КЛЗ.3.2):

$$\mathbb{P}_{(X,Y)^\top}(X = i, Y = j) = \sum_{n=0}^{\infty} \mathbb{P}_{(X,Y)^\top|N}(X = i, Y = j \mid N = n) \mathbb{P}_N(N = n) .$$

Можна помітити, що якщо $n \neq i + j$, то відповідна умовна ймовірність дорівнює 0, адже така подія неможлива. Тому вираз суттєво спрощується:

$$\mathbb{P}_{(X,Y)^\top}(X = i, Y = j) = \mathbb{P}_{(X,Y)^\top|N}(X = i, Y = j \mid N = i + j) \mathbb{P}_N(N = i + j) ,$$

до того ж події $\mathbb{P}_{X|N}(X = i \mid N = i + j)$ і $\mathbb{P}_{Y|N}(Y = j \mid N = i + j)$ — це одна й та сама подія, а тому

$$\begin{aligned} \mathbb{P}_{(X,Y)^\top|N}(X = i, Y = j \mid N = i + j) &= \mathbb{P}_{X|N}(X = i \mid N = i + j) \\ &= \binom{i+j}{i} p^i (1-p)^j . \end{aligned}$$

Отже,

$$\begin{aligned} \mathbb{P}_{(X,Y)^\top}(X = i, Y = j) &= \binom{i+j}{i} p^i (1-p)^j \cdot \frac{e^{-\lambda} \lambda^{i+j}}{(i+j)!} \\ &= \frac{e^{-\lambda p} (\lambda p)^i}{i!} \cdot \frac{e^{-\lambda(1-p)} (\lambda(1-p))^j}{j!} . \end{aligned}$$

Отже спільну функцію ймовірності можна подати як добуток двох функцій, які згідно з (КЛ9.4.7) мають інтерпретацію маржинальних функцій ймовірності. Тобто $X \perp\!\!\!\perp Y$, і до того ж $X \sim \text{Pois}(\lambda p)$, $Y \sim \text{Pois}(\lambda(1-p))$.

На перший погляд такий результат може здатися контрінтуїтивним, адже для деякого фіксованого $N = n$ кількість чоловіків i і жінок j повністю залежать одна від одної: $i = n - j$. Проте суть полягає в тому, що кількість клієнтів N не є фіксованою, і саме випадковість числа клієнтів уможливає незалежність змінних X і Y .

Також із цього прикладу можна зробити висновок, що сума двох незалежних величин із розподілами Пуассона з параметрами λ_1 і λ_2 є величиною з розподілом Пуассона з параметром $\lambda_1 + \lambda_2$, що ми вже вивчали раніше. $\square$

Вправа 14.2.4. Нехай величини X_1, X_2, X_3 мають спільну щільність

$$f_{X_1, X_2, X_3}(x_1, x_2, x_3) = cx_1^{1+x_2+x_3}(1-x_1)^{3-x_2-x_3} \\ \times \mathbb{1}\{0 < x_1 < 1\} \cdot \mathbb{1}\{x_2 \in \{0, 1\}\} \cdot \mathbb{1}\{x_3 \in \{0, 1\}\}.$$

Чому дорівнює константа c ? Чому дорівнює маржинальна функція ймовірності для $(X_2, X_3)^\top$? Умовна щільність X_1 за умови $X_2 = 1$, $X_3 = 1$? $\square$

Розв'язання. Для обчислення константи c ми, як завжди, прирівнюємо інтеграл спільної щільності до 1. Розподіл випадкового вектора абсолютно неперервний відносно добутку міри Лебега та двох лічливих мір, а отже шуканий інтеграл з допомогою теореми Фубіні перетвориться у такий вираз:

$$1 = \iiint_{[0,1] \times \{0,1\} \times \{0,1\}} cx_1^{1+x_2+x_3}(1-x_1)^{3-x_2-x_3} \\ \times \mathbb{1}\{0 < x_1 < 1\} \cdot \mathbb{1}\{x_2, x_3 \in \{0, 1\}\} \lambda(dx_1) \#(dx_2) \#(dx_3) \\ = \sum_{x_3 \in \{0,1\}} \sum_{x_2 \in \{0,1\}} \int_0^1 cx_1^{1+x_2+x_3}(1-x_1)^{3-x_2-x_3} dx_1.$$

На перший вигляд може здатися, що це складний інтеграл на кшталт Бета-функції, проте потрібно помітити, що x_2 та x_3 набувають тільки два значення кожна. Зокрема, якщо $x_2 = 0$, $x_3 = 0$, ми маємо

$$\int_0^1 cx_1(1-x_1)^3 dx_1 = \int_0^1 ct^3(1-t) dt = \frac{c}{4} - \frac{c}{5} = \frac{c}{20}.$$

Якщо $x_2 = 0$, $x_3 = 1$ або $x_2 = 1$, $x_3 = 0$, ми маємо

$$\int_0^1 cx_1^2(1-x_1)^2 dx_1 = \int_0^1 c(x_1^2 - 2x_1^3 + x_1^4) = \frac{c}{3} - \frac{2c}{4} + \frac{c}{5} = \frac{c}{30}.$$

Нарешті, для випадку $x_2 = 1, x_3 = 1$ ми маємо

$$\int_0^1 cx_1^3(1-x_1) dx_1 = \frac{c}{4} - \frac{c}{5} = \frac{c}{20}.$$

Сума всіх цих значень дорівнює

$$1 = \frac{c}{20} + \frac{c}{30} + \frac{c}{30} + \frac{c}{20},$$

звідки $c = 6$.

Щоб дістати маржинальну функцію ймовірності $(X_2, X_3)^\top$, ми повинні підставити c у вищевказані результати обчислення відповідних інтегралів:

$$p_{X_2, X_3}(x_2, x_3) = \begin{cases} \frac{6}{20}, & x_2 = 0, x_3 = 0 \\ \frac{6}{30}, & x_2 = 0, x_3 = 1 \\ \frac{6}{30}, & x_2 = 1, x_3 = 0 \\ \frac{6}{20}, & x_2 = 1, x_3 = 1 \end{cases}.$$

Умовну щільність можемо дістати, поділивши спільну щільність на маржинальну функцію ймовірності. Для $x_2 = 1, x_3 = 1$ ділити треба на $6/20$:

$$\begin{aligned} f_{X_1|X_2, X_3}(x_1 | 1, 1) &= \frac{6x_1^{1+x_2+x_3}(1-x_1)^{3-x_2-x_3} \cdot \mathbb{1}\{0 < x_1 < 1\}}{\frac{6}{20}} \\ &= 20x_1^3(1-x_1) \cdot \mathbb{1}\{0 < x_1 < 1\}. \end{aligned} \quad \square$$

14.3 Умовні щільності

Вправа 14.3.1. Виробничий процес складається з двох етапів. Перший етап триває Y одиниць часу, а загальний час на виконання всього процесу — X одиниць. Нехай X та Y мають спільний розподіл зі щільністю

$$f_{X,Y}(x, y) = e^{-x} \cdot \mathbb{1}\{0 \leq y \leq x\}.$$

Нехай ми знаємо, що перший етап зайняв $Y = 4$ одиниць часу. Чому тоді дорівнює ймовірність, що загальний час перевищить 9? $\square$

Розв'язання. Для обчислення потрібно дістати умовну щільність $X | Y$:

$$f_{X|Y}(x | y) = \frac{f_{X,Y}(x, y)}{f_Y(y)}.$$

Але для цього спочатку потрібно знати щільність маржинальну, яку можна дістати шляхом інтегрування спільної щільності за змінною x :

$$f_Y(y) = \int_{-\infty}^{\infty} e^{-x} \cdot \mathbb{1}\{0 \leq y \leq x\} dx = \int_y^{\infty} e^{-x} dx \cdot \mathbb{1}\{0 \leq y\} = e^{-y} \cdot \mathbb{1}\{0 \leq y\}.$$

Отже

$$f_{X|Y}(x | y) = \frac{e^{-x} \cdot \mathbb{1}\{0 \leq y \leq x\}}{e^{-y} \cdot \mathbb{1}\{0 \leq y\}} = \begin{cases} e^{y-x} \cdot \mathbb{1}\{y \leq x\}, & y \geq 0 \\ \xi(y), & y < 0 \end{cases}.$$

де, за Твердженням КЛ12.4.1, ми поклали, що там, де знаменник обертається в нуль ($y < 0$), щільність дорівнює деякій довільній (і нікому непотрібній) щільності ξ . Оскільки на практиці ніколи не може так бути, щоб перша частина процесу закінчилася за від'ємну кількість часу, це уточнення має суто формально теоретичний сенс.

Зверніть увагу, що ми не можемо написати, як зазвичай, відповідний вираз через індикатор, наприклад, як

$$f_{X|Y}(x | y) = e^{y-x} \cdot \mathbb{1}\{0 \leq y \leq x\},$$

адже тоді склалося б враження, що для, скажімо, $y = -1$, умовна щільність дорівнює 0, тобто це була б зовсім не щільність.

Отже шукана ймовірність дорівнює

$$\begin{aligned} \mathbb{P}_{X|Y}(X \geq 9 | Y = 4) &= \int_9^{\infty} f_{X|Y}(x | 4) dx = \int_9^{\infty} e^{4-x} \cdot \mathbb{1}\{4 \leq x\} dx \\ &= \int_9^{\infty} e^{4-x} dx = e^{-5} \approx 0.0067. \end{aligned}$$

Варто зазначити, що, як було сказано в Зауваженні КЛ12.2.10, ми не обумовлюємо подією $Y = 4$, яка вочевидь має ймовірність 0. Ми обумовлюємо цілою σ -алгеброю, яку породжує величина Y . Тому $\mathbb{P}_{X|Y}(X \geq 9 | Y = 4)$ — це просто зручне позначення, за яким, у неформальному сенсі, стоїть границя ймовірностей $\mathbb{P}_{X|Y}(X \geq 9 | Y \in (4 - h_n; 4 + h_n))$, $h_n \downarrow 0$. $\square$

Вправа 14.3.2. Нехай X — час, який потрібно прождати в черзі до деякого клерка. Звісно, що швидше працює клерк, то менший буде час очікування. Нехай Y — швидкість роботи клерка. Тоді $X | Y$ матиме умовний розподіл, який доречно моделювати як експоненційний: $X | Y = y \sim \text{Exp}(y)$. Чому дорівнює спільна щільність X та Y , якщо $Y \sim \text{Exp}(1)$? $\square$

Розв'язання. Умовна щільність — це відношення щільностей спільної й

маржинальної, тому спільна щільність є добуток умовної й маржинальної:

$$f_{X,Y}(x, y) = f_{X|Y}(x | y) \cdot f_Y(y) .$$

З урахуванням виразів для щільностей експоненційного розподілу маємо

$$f_{X,Y}(x, y) = ye^{-xy} \cdot \mathbb{1} \{x > 0\} \cdot e^{-y} \cdot \mathbb{1} \{y > 0\} = ye^{-y(x+1)} \cdot \mathbb{1} \{x > 0\} \cdot \mathbb{1} \{y > 0\} .$$

□

Вправа 14.3.3. Нехай у випадковий спосіб обирають точку з одиничного інтервалу: $X \sim U((0; 1])$. Нехай після цього вибирають, також у випадковий спосіб, точку Y із інтервалу $(x; 1]$, де x — реалізоване значення X . Чому дорівнює (маржинальна) щільність розподілу Y ? □

Розв'язання. За суттю задачі цілком очевидно, що *умовно* Y має рівномірний розподіл: $Y | X = x \sim U((x; 1])$, тобто для $x \geq 0$

$$f_{Y|X}(y | x) = \frac{1}{1-x} \cdot \mathbb{1} \{x < y < 1\} .$$

Маржинальну щільність Y можна дістати як інтеграл відповідної спільної щільності за змінною x , яку в свою чергу можна дістати добутком умовної щільності $Y | X$ та маржинальної щільності X :

$$f_{X,Y}(x, y) = f_{Y|X}(y | x) \cdot f_X(x) = \frac{1}{1-x} \cdot \mathbb{1} \{x < y < 1\} \cdot \mathbb{1} \{0 < x < 1\} .$$

Отже

$$\begin{aligned} f_Y(y) &= \int_{-\infty}^{\infty} \frac{1}{1-x} \cdot \mathbb{1} \{x < y < 1\} \cdot \mathbb{1} \{0 < x < 1\} dx \\ &= \mathbb{1} \{0 < y < 1\} \cdot \int_0^y \frac{1}{1-x} dx \\ &= \mathbb{1} \{0 < y < 1\} \cdot \left(-\ln |1-x| \Big|_0^y \right) = -\ln(1-y) \cdot \mathbb{1} \{0 < y < 1\} . \end{aligned}$$

□

Вправа 14.3.4. Розгляньмо чергу в деякій організації. Нехай у черзі до одного клерка стоять n клієнтів і нехай X_i = «час, потрібний на опрацювання клієнта i », $i = 1, \dots, n$. Нехай $X_i | Z = z \sim \text{Exp}(z)$ для деякої випадкової швидкості опрацювання клієнтів Z . Нехай $Z \sim \text{Exp}(2)$. Вважатимемо, що $X_1 \perp\!\!\!\perp X_2 | Z$, тобто час очікування для одного клієнта не залежить від

часу очікування іншого для деякого *фіксованого* z . Чому дорівнює умовний розподіл $Z \mid X_1, X_2$, тобто чому дорівнює розподіл швидкості роботи клерка, якщо ми побачили, як швидко він упорався з першими двома клієнтами? Чому дорівнює $\mathbb{P}_{Z \mid X_1, X_2}(Z \leq 1 \mid X_1 = 1, X_2 = 4)$? $\square$

Розв'язання. За формулою Беєса маємо

$$f_{Z \mid X_1, X_2}(z \mid x_1, x_2) = \frac{f_{X_1, X_2 \mid Z}(x_1, x_2 \mid z)f_Z(z)}{f_{X_1, X_2}(x_1, x_2)}.$$

Через умовну незалежність відразу маємо, для $z > 0$:

$$\begin{aligned} f_{X_1, X_2 \mid Z}(x_1, x_2 \mid z) &= f_{X_1 \mid Z}(x_1 \mid z)f_{X_2 \mid Z}(x_2 \mid z) \\ &= z^2 e^{-z(x_1 + x_2)} \cdot \mathbb{1}\{x_1 > 0\} \cdot \mathbb{1}\{x_2 > 0\}. \end{aligned}$$

(Для $z \leq 0$ маємо деяку абсолютно довільну щільність ξ , але ми на цьому не акцентуємо уваги, оскільки вона ніде не використовується, бо Z на практиці не може бути нульова або від'ємна.)

Маржинальну щільність у знаменнику можна дістати як інтеграл відповідної спільної щільності в чисельнику:

$$\begin{aligned} f_{X_1, X_2}(x_1, x_2) &= \int_{\mathbb{R}} f_{X_1, X_2 \mid Z}(x_1, x_2 \mid z)f_Z(z) dz \\ &= \mathbb{1}\{x_1 > 0\} \cdot \mathbb{1}\{x_2 > 0\} \cdot \int_0^\infty 2z^2 e^{-z(2+x_1+x_2)} dz. \end{aligned}$$

Підінтегральна функція відповідає щільності $\text{Gamma}(3, 2 + x_1 + x_2)$, проте повністю така щільність повинна бути

$$\frac{1}{\Gamma(3)} \cdot (2 + x_1 + x_2)^3 z^2 e^{-z(2+x_1+x_2)}.$$

Тому якщо ми помножимо підінтегральну функцію на відповідні коефіцієнти, її інтеграл буде одиниця. Остаточнo,

$$\begin{aligned} f_{X_1, X_2}(x_1, x_2) &= \frac{2\Gamma(3)}{(2 + x_1 + x_2)^3} \cdot \mathbb{1}\{x_1 > 0\} \cdot \mathbb{1}\{x_2 > 0\} \\ &= \frac{4}{(2 + x_1 + x_2)^3} \cdot \mathbb{1}\{x_1 > 0\} \cdot \mathbb{1}\{x_2 > 0\}. \end{aligned}$$

Отже для $x_1 > 0$ та $x_2 > 0$ маємо (для інших значень можна покласти

якусь щільність ξ і забути):

$$\begin{aligned} f_{Z|X_1, X_2}(z | x_1, x_2) &= \frac{z^2 e^{-z(x_1+x_2)} \cdot \mathbb{1}\{x_1 > 0\} \cdot \mathbb{1}\{x_2 > 0\} \cdot 2e^{-2z} \cdot \mathbb{1}\{z > 0\}}{\frac{4}{(2+x_1+x_2)^3} \cdot \mathbb{1}\{x_1 > 0\} \cdot \mathbb{1}\{x_2 > 0\}} \\ &= \frac{1}{2}(2+x_1+x_2)^3 z^2 e^{-z(2+x_1+x_2)} \cdot \mathbb{1}\{z > 0\} . \end{aligned}$$

Шукану ймовірність тоді можна обчислити, наприклад, інтегруванням:

$$\begin{aligned} \mathbb{P}_{Z|X_1, X_2}(Z \leq 1 | X_1 = 1, X_2 = 4) &= \int_0^1 \frac{1}{2}(2+1+4)^3 z^2 e^{-z(2+1+4)} dz \\ &= \frac{343}{2} \int_0^1 z^2 e^{-7z} dz \approx 0.9704 . \end{aligned}$$

□

Вправа 14.3.5. Нехай X та Y мають спільний рівномірний розподіл на крузі $x^2 + y^2 \leq 1$. Чому дорівнює умовне сподівання $\mathbb{E}[X | Y]$? □

Розв'язання. Для підрахунку умовного сподівання потрібно мати умовну щільність. Для рівномірного розподілу спільна щільність X та Y має запис

$$f_{X,Y}(x, y) = \frac{1}{\pi} \cdot \mathbb{1}\{x^2 + y^2 \leq 1\} .$$

Маржинальна щільність Y для $y^2 \leq 1$ дорівнює

$$\begin{aligned} f_Y(y) &= \int_{\mathbb{R}} f_{X,Y}(x, y) dx = \frac{1}{\pi} \int_{\mathbb{R}} \mathbb{1}\{x^2 + y^2 \leq 1\} dx \\ &= \frac{1}{\pi} \int_{-\sqrt{1-y^2}}^{\sqrt{1-y^2}} dy = \frac{2}{\pi} \sqrt{1-y^2} , \end{aligned}$$

а, відповідно, для всіх $y^2 > 1$ маємо 0.

Тоді умовна щільність $X | Y$ дорівнює, для $y^2 \leq 1$:

$$f_{X|Y}(x | y) = \frac{f_{X,Y}(x, y)}{f_Y(y)} = \frac{\frac{1}{\pi} \cdot \mathbb{1}\{x^2 + y^2 \leq 1\}}{\frac{2}{\pi} \sqrt{1-y^2} \cdot \mathbb{1}\{y^2 \leq 1\}} = \frac{\mathbb{1}\{x^2 + y^2 \leq 1\}}{2\sqrt{1-y^2}} .$$

Отже умовне сподівання дорівнює інтегралу від непарної функції:

$$\mathbb{E}[X | Y] = \int_{-\infty}^{\infty} x f_{X|Y}(x | y) dx = \int_{-\sqrt{1-y^2}}^{\sqrt{1-y^2}} \frac{x}{2\sqrt{1-y^2}} dx = 0 .$$

□

14.4 Умовні сподівання й медіана як найліпші предиктори

Вправа 14.4.1. Нехай (стандартизовані) оцінки студентів із математики та історії України позначено через X та Y відповідно. Спільна щільність розподілу оцінок для студентів із деякої популяції дорівнює

$$f_{X,Y}(x, y) = \frac{2}{5} (2x + 3y) \cdot \mathbb{1}\{0 \leq x \leq 1\} \cdot \mathbb{1}\{0 \leq y \leq 1\} .$$

Чому дорівнює найліпший (у розумінні мінімізації середньоквадратичної похибки) прогноз оцінки з історії України у випадковий спосіб вибраного студента? Чому дорівнює найліпший (у розумінні мінімізації середньої абсолютної похибки) прогноз його оцінки з математики?

Якщо ми спостерігаємо його оцінку з математики $X = 0.8$, чому дорівнює найліпший (у розумінні мінімізації середньоквадратичної похибки) прогноз його оцінки з історії України?

Якщо ми спостерігаємо його оцінку з історії України $Y = \frac{1}{3}$, чому дорівнює найліпший (у розумінні мінімізації середньої абсолютної похибки) прогноз його оцінки з математики? $\square$

Розв'язання. Для відповідей на питання потрібно знати маржинальні та умовні щільності. Спочатку обчислимо відповідні щільності маржинальні:

$$f_X(x) = \int_0^1 \frac{2}{5} (2x + 3y) \cdot \mathbb{1}\{0 \leq x \leq 1\} dy = \frac{1}{5} (4x + 3) \cdot \mathbb{1}\{0 \leq x \leq 1\} ,$$

$$f_Y(y) = \int_0^1 \frac{2}{5} (2x + 3y) \cdot \mathbb{1}\{0 \leq y \leq 1\} dx = \frac{2}{5} (1 + 3y) \cdot \mathbb{1}\{0 \leq y \leq 1\} .$$

Отже умовні щільності дорівнюють, для $0 \leq x \leq 1$ та $0 \leq y \leq 1$ відповідно:

$$\begin{aligned} f_{Y|X}(y | x) &= \frac{\frac{2}{5} (2x + 3y) \cdot \mathbb{1}\{0 \leq x \leq 1\} \cdot \mathbb{1}\{0 \leq y \leq 1\}}{\frac{1}{5} (4x + 3) \cdot \mathbb{1}\{0 \leq x \leq 1\}} \\ &= \frac{2(2x + 3y)}{4x + 3} \cdot \mathbb{1}\{0 \leq x \leq 1\} \cdot \mathbb{1}\{0 \leq y \leq 1\} , \\ f_{X|Y}(x | y) &= \frac{\frac{2}{5} (2x + 3y) \cdot \mathbb{1}\{0 \leq x \leq 1\} \cdot \mathbb{1}\{0 \leq y \leq 1\}}{\frac{2}{5} (1 + 3y) \cdot \mathbb{1}\{0 \leq y \leq 1\}} \\ &= \frac{2x + 3y}{1 + 3y} \cdot \mathbb{1}\{0 \leq x \leq 1\} \cdot \mathbb{1}\{0 \leq y \leq 1\} . \end{aligned}$$

Найліпший (безумовний) прогноз значення величини Y в розумінні міні-

мінімізації середньоквадратичної похибки є її (безумовне) сподівання:

$$\mathbb{E}[Y] = \int_{-\infty}^{\infty} y f_Y(y) dy = \int_0^1 y \cdot \frac{2}{5}(1+3y) dy = \frac{2}{5} \cdot \left(\frac{1}{2} + 1\right) = \frac{3}{5}.$$

Найліпший (безумовний) прогноз значення величини X в розумінні мінімізації середньої абсолютної похибки є її (безумовна) медіана M_X , тобто число, що розв'язує рівняння

$$\int_0^{M_X} \frac{1}{5}(4x+3) dx = \frac{1}{2},$$

тобто

$$\frac{1}{5}(2M_X^2 + 3M_X) = \frac{1}{2},$$

звідки

$$M_X = \frac{\sqrt{29} - 3}{4} \approx 0.596.$$

Найліпший прогноз значення величини Y за умови $X = 0.8$ в розумінні мінімізації середньоквадратичної похибки є її умовне сподівання:

$$\mathbb{E}[Y | X = 0.8] = \int_0^1 y \cdot \frac{2(2 \cdot 0.8 + 3y)}{4 \cdot 0.8 + 3} dy = \int_0^1 \frac{y(1.6 + 3y)}{3.1} dy = \frac{18}{31} \approx 0.581.$$

Найліпший прогноз значення величини X за умови $Y = \frac{1}{3}$ в розумінні мінімізації середньої абсолютної похибки є її умовна медіана $M_{X|Y=\frac{1}{3}}$, тобто число, що розв'язує рівняння

$$\int_0^{M_{X|Y=\frac{1}{3}}} \frac{2x + 3 \cdot \frac{1}{3}}{1 + 3 \cdot \frac{1}{3}} dx = \frac{1}{2},$$

тобто

$$\int_0^{M_{X|Y=\frac{1}{3}}} \frac{2x+1}{2} dx = \frac{1}{2},$$

звідки

$$M_{X|Y=\frac{1}{3}} = \frac{\sqrt{5} - 1}{2} \approx 0.618.$$

□

Вправа 14.4.2. Значення сигналу S потрібно передати з клієнта на сервер. Число R , що прийде на сервер, з урахуванням можливих похибок у мережі, має розподіл $R | S = s \sim N(s, 1)$. Якщо апріорний розподіл $S \sim N(\mu, \sigma^2)$, чому дорівнює найліпша (у розумінні мінімізації середньоквадратичної похибки) оцінка значення S , якщо на сервер надійшло число $R = r$? □

Розв'язання. Маємо справу з беєсівським виведенням. Для початку знайдімо умовну щільність $S \mid R$. Її можна знайти за формулою Беєса:

$$f_{S|R}(s \mid r) = \frac{f_{R|S}(r \mid s)f_S(s)}{f_R(r)} \propto e^{-\frac{(r-s)^2}{2}} \cdot e^{-\frac{(s-\mu)^2}{2\sigma^2}},$$

де ми вирішили не обчислювати знаменник, який є просто мультиплікативною константою, яка не залежить від s , та взагалі вирішили позбутися всіх констант, які не залежать від s .

Після розкриття всіх дужок та групування доданків маємо

$$\begin{aligned} -\frac{(r-s)^2}{2} - \frac{(s-\mu)^2}{2\sigma^2} &= s^2 \left(\frac{1}{2\sigma^2} + \frac{1}{2} \right) - s \left(\frac{\mu}{\sigma^2} + r \right) + \left(\frac{\mu^2}{\sigma^2} + \frac{r^2}{2} \right) \\ &= s^2 \left(\frac{1}{2\sigma^2} + \frac{1}{2} \right) - s \left(\frac{\mu}{\sigma^2} + r \right) + C_1, \end{aligned}$$

де ми знову позбулися частини виразу, яка не залежить від s , позначивши її через C_1 . Виділивши повний квадрат, маємо

$$\frac{1+\sigma^2}{2\sigma^2} \left(s - \frac{\mu+r\sigma^2}{1+\sigma^2} \right)^2 + C_2,$$

де C_2 — чергова константа, яка не залежить від s . Отже бачимо, що

$$f_{S|R}(s \mid r) \propto e^{-\frac{\left(s - \frac{\mu+r\sigma^2}{1+\sigma^2} \right)^2}{2\left(\frac{\sigma^2}{1+\sigma^2} \right)}},$$

тобто $S \mid R = r \sim N\left(\frac{\mu+r\sigma^2}{1+\sigma^2}, \frac{\sigma^2}{1+\sigma^2}\right)$. Отже ми бачимо, що якщо апіорний розподіл S є нормальний, і умовний розподіл $R \mid S$ також є нормальний, то апостеріорний розподіл $S \mid R$ також буде нормальний. Кажуть, що нормальний розподіл є *спряженим апіорним розподілом* (conjugate prior) для нормального ж розподілу¹.

Тоді найліпшим (у розумінні мінімізації середньоквадратичної похибки) предиктором значення $S \mid R = r$ буде умовне сподівання, яке дорівнює

$$\mathbb{E}[S \mid R = r] = \frac{\mu + r\sigma^2}{1 + \sigma^2} = \frac{1}{1 + \sigma^2}\mu + \frac{\sigma^2}{1 + \sigma^2}r = \frac{\frac{1}{\sigma^2}}{1 + \frac{1}{\sigma^2}}\mu + \frac{1}{1 + \frac{1}{\sigma^2}}r.$$

Як можна бачити, це умовне сподівання є зваженим середнім апіорного сподівання μ та отриманого значення r . Ваговими коефіцієнтами є вирази, пропорційні до значень, *обернених* до відповідних дисперсій (тобто фа-

¹Варто зазначити, що ми припустили, що дисперсія розподілу $R \mid S$ відома і дорівнює 1. Якби ця дисперсія також була невідома, то виведення було б трішки іншим, адже ми мали би оцінювати обидва невідомі параметри: сподівання $\mu_{R|S} \equiv s$ і дисперсію $\sigma_{R|S}^2$.

ктивно наскільки концентрованим є розподіл): 1 є (умовною) дисперсією отриманого сигналу, а σ^2 є апіорною дисперсією сигналу. $\square$

14.5 Додаткові вправи

Вправа 14.5.1. Нехай маємо схему Бернуллі з n випробувань та ймовірністю успіху p . Покажіть, що за умови, що сталося k успіхів, усі можливі *впорядковані* результати випробувань рівноймовірні. $\square$

Розв'язання. Розгляньмо деякий конкретний упорядкований результат $(x_1, x_2, \dots, x_n)$, де $x_i \in \{0, 1\}$, $i = 1, \dots, n$. Умовна ймовірність такого результату за умови $X = k$, де $X \sim \text{Binom}(n, p)$, за теоремою Беєса дорівнює

$$\begin{aligned} & \mathbb{P}_{X_1, \dots, X_n | X} (X_1 = x_1, \dots, X_n = x_n | X = k) \\ &= \frac{\mathbb{P}_{X | X_1, \dots, X_n} (X = k | X_1 = x_1, \dots, X_n = x_n) \mathbb{P}_{X_1, \dots, X_n} (X_1 = x_1, \dots, X_n = x_n)}{\mathbb{P}_X (X = k)} \\ &= \frac{1 \cdot p^k (1-p)^{n-k}}{\binom{n}{k} p^k (1-p)^{n-k}} = \frac{1}{\binom{n}{k}}. \end{aligned}$$

Ця ймовірність не залежить від $x_1, \dots, x_n$, тому всі такі *впорядковані* результати умовно рівноймовірні. $\square$

Вправа 14.5.2. Гравець у баскетбол планує здійснити $N = \text{Pois}(\lambda)$ кидків у завтрашній грі. Нехай X_j = «гравець влучив на j -ому кидку». Тоді $X = X_1 + \dots + X_N$ є загальним числом влучень у завтрашній грі (зокрема, $X = 0$, якщо $N = 0$).

Для моделювання невизначеності стосовно успішності кидка вважатимемо, що ймовірність влучення дорівнює $p \sim \text{Beta}(a, b)$. Вважатимемо, що успішності різних кидків (умовно) незалежні для деякого фіксованого p , і що N не залежить від p , X_1 , X_2 тощо.

Чому дорівнює (умовний) розподіл X для деяких фіксованих $p = p_0$ та $N = n$? Чому дорівнює (умовний) розподіл X для деякого фіксованого $p = p_0$? Чому дорівнює (умовний) розподіл X для деякого фіксованого $N = n$ для $a = b = 1$? Чому дорівнює умовний розподіл p для деяких фіксованих $X = x$ та $N = n$? $\square$

Розв'язання. Якщо p та N фіксовані, то X має біномний розподіл $X | N = n, p = p_0 \sim \text{Binom}(n, p_0)$.

Якщо N не є фіксоване, але p фіксоване, то ми маємо ситуацію, аналогічну Вправі 14.2.3, а отже $X | p = p_0 \sim \text{Pois}(\lambda p_0)$.

Якщо $N = n$, p не фіксоване, то за формулою повної ймовірності (КЛ12.3.7)

$$\begin{aligned}\mathbb{P}_{X|N}(X = k | N = n) &= \int_{-\infty}^{\infty} \mathbb{P}_{X|N,p}(X = k | N = n, p) \cdot f_p(p) dp \\ &= \int_0^1 \binom{n}{k} p^k (1 - p)^{n-k} dp = \frac{1}{n+1}.\end{aligned}$$

Нарешті, $p | X = x, N = n \sim \text{Beta}(a + x, b + n - x)$ через спряженість біномного та Бета-розподілів. $\square$

Вправа 14.5.3. Нехай X та Y мають спільну щільність

$$f_{X,Y}(x, y) = \frac{21}{4} x^2 y \cdot \mathbb{1}\{x^2 \leq y \leq 1\}.$$

Чому дорівнює умовна щільність $Y | X = x$? Чому дорівнює ймовірність, що $Y \geq 0.75$, якщо $X = 0.5$? $\square$

Розв'язання. Для підрахунку умовної щільності спочатку потрібно мати вираз для щільності маржинальної:

$$\begin{aligned}f_X(x) &= \int_{-\infty}^{\infty} f_{X,Y}(x, y) dy = \int_{-\infty}^{\infty} \frac{21}{4} x^2 y \cdot \mathbb{1}\{x^2 \leq y \leq 1\} dy \\ &= \frac{21}{4} x^2 \cdot \mathbb{1}\{x^2 \leq 1\} \cdot \int_{x^2}^1 y dy = \frac{21}{8} x^2 (1 - x^4) \cdot \mathbb{1}\{x^2 \leq 1\}.\end{aligned}$$

Остаточно для $x^2 \leq 1$ маємо

$$f_{Y|X}(y | x) = \frac{\frac{21}{4} x^2 y \cdot \mathbb{1}\{x^2 \leq y \leq 1\}}{\frac{21}{8} x^2 (1 - x^4) \cdot \mathbb{1}\{x^2 \leq 1\}} = \frac{2y}{1 - x^4} \cdot \mathbb{1}\{x^2 \leq y \leq 1\}.$$

(Для $x^2 > 1$ формально маємо деяку довільну щільність $\xi(y)$, яка ніде не використовується.)

Шукану умовну ймовірність можна дістати як інтеграл умовної щільності за відповідною множиною:

$$\mathbb{P}_{Y|X}\left(Y \geq \frac{3}{4} \mid X = \frac{1}{2}\right) = \int_{\frac{3}{4}}^1 \frac{2y}{1 - (\frac{1}{2})^4} dy = \frac{1 - 9/16}{1 - 1/16} = \frac{7}{15}. \quad \square$$

Вправа 14.5.4. Нехай спільна щільність розподілу X та Y дорівнює

$$f_{X,Y}(x, y) = (x + y) \cdot \mathbb{1}\{0 \leq x \leq 1\} \cdot \mathbb{1}\{0 \leq y \leq 1\}.$$

Чому дорівнюють $\mathbb{E}[Y | X]$ та $\text{Var}(Y | X)$? Якщо $X = \frac{1}{2}$, чому дорівнює прогнозне значення Y , яке мінімізує середньоквадратичну похибку? Яке буде значення цієї похибки? $\square$

Розв'язання. Для підрахунку умовного сподівання потрібно мати вираз умовної щільності. Вона дорівнює відношенню спільної та маржинальної щільностей. Маржинальна щільність дорівнює

$$f_X(x) = \int_0^1 (x + y) \cdot \mathbb{1}\{0 \leq x \leq 1\} dy = \left(x + \frac{1}{2}\right) \cdot \mathbb{1}\{0 \leq x \leq 1\}.$$

Отже умовна щільність дорівнює, для $\mathbb{1}\{0 \leq x \leq 1\}$,

$$\begin{aligned} f_{Y|X}(y | x) &= \frac{f_{X,Y}(x, y)}{f_X(x)} = \frac{x + y}{\frac{1}{2} + x} \cdot \mathbb{1}\{0 \leq x \leq 1\} \cdot \mathbb{1}\{0 \leq y \leq 1\} \\ &= \frac{2(x + y)}{2x + 1} \cdot \mathbb{1}\{0 \leq x \leq 1\} \cdot \mathbb{1}\{0 \leq y \leq 1\}. \end{aligned}$$

Отже

$$\begin{aligned} \mathbb{E}[Y | X] &= \int_0^1 y \cdot \frac{2(x + y)}{2x + 1} \mathbb{1}\{0 \leq x \leq 1\} dy = \frac{x + \frac{2}{3}}{2x + 1} \cdot \mathbb{1}\{0 \leq x \leq 1\}, \\ \mathbb{E}[Y^2 | X] &= \int_0^1 y^2 \cdot \frac{2(x + y)}{2x + 1} \mathbb{1}\{0 \leq x \leq 1\} dy = \frac{\frac{2}{3}x + \frac{1}{2}}{2x + 1} \cdot \mathbb{1}\{0 \leq x \leq 1\}, \\ \text{Var}(Y | X) &= \mathbb{E}[Y^2 | X] - (\mathbb{E}[Y | X])^2 \\ &= \left(\frac{4x + 3}{6(2x + 1)} - \left(\frac{3x + 2}{3(2x + 1)} \right)^2 \right) \cdot \mathbb{1}\{0 \leq x \leq 1\} \\ &= \frac{1}{36} \left(3 - \frac{1}{(2x + 1)^2} \right) \cdot \mathbb{1}\{0 \leq x \leq 1\}. \end{aligned}$$

Середньоквадратичну похибку для $X = \frac{1}{2}$ мінімізує умовне сподівання

$$\mathbb{E}\left[Y | X = \frac{1}{2}\right] = \frac{\frac{1}{2} + \frac{2}{3}}{2 \cdot \frac{1}{2} + 1} = \frac{7}{12},$$

і тоді середньоквадратична похибка дорівнює

$$\begin{aligned} \text{MSE}(Y | X) &= \mathbb{E}\left[\left(Y - \mathbb{E}\left[Y | X = \frac{1}{2}\right]\right)^2\right] = \text{Var}\left(Y | X = \frac{1}{2}\right) \\ &= \frac{1}{36} \left(3 - \frac{1}{(2 \cdot \frac{1}{2} + 1)^2} \right) = \frac{11}{144}. \end{aligned} \quad \square$$

15 Умовні сподівання

15.1 Основні теоретичні відомості

Детальний огляд понять, потрібний для розв’язання задач на цьому занятті, наведено в Курсі лекцій, Розд. КЛ12.

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 σ -алгебри $\mathcal{A}' \subset \mathcal{A}$, якщо $\{\omega : \mathbf{X}(\omega) \in H\} \in \mathcal{A}'$ для всіх $H \in \mathcal{B}^k$.

Визначення (КЛ12.2.8). Казатимемо, що σ -алгебру $\mathcal{A}(\mathbf{X})$ породжує вектор $\mathbf{X}$ (σ -algebra is generated by $\mathbf{X}$), якщо вона є найменшою σ -алгеброю, відносно якої він є вимірний.

Аналогічно, казатимемо, що σ -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 (КЛ12.2.9). Нехай маємо вимірний простір $(\Omega, \mathcal{A})$ і випадкову величину $X : (\Omega, \mathcal{A}) \rightarrow \mathbb{R}$, яка породжує σ -алгебру $\mathcal{A}(X)$. Тоді умовною ймовірністю (події A) за умови X (conditional probability given X) є випадкова величина $\mathbb{P}(A \mid \mathcal{A}(X)) \equiv \mathbb{P}(A \mid X)$. $\square$

Іншими словами, за умови X ми, хоча й не знаємо конкретного ω , 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 $\mathbf{X} \in \mathbb{R}^k$ і дістати $\mathbb{P}(A \mid \mathbf{X})$.

Ключовим результатом відповідної формальної побудови умовної ймовірності, обумовленої σ -алгеброю, є уможливлення обчислення умовних ймовірностей деяких подій за умови, що деяка випадкова величина набула деякого значення $X(\omega)$. Оскільки для неперервних випадкових величин подія $\mathbb{P}(X = x) = 0$, використання (КЛ3.1.1) було б просто неможливим.

Зауваження (КЛ12.2.10). 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 σ -алгеброю, яку породжує Y , ми можемо обчислити цю ймовірність як границю ймовірностей $\mathbb{P}_{X|Y}(X \leq 0 \mid Y \in (1 - h_n; 1 + h_n))$, де $h_n \downarrow 0$.

Тому до позначень на кшталт $\mathbb{P}_{X|Y}(X \leq 0 \mid Y = 1)$ потрібно ставитися акуратно, розуміючи, що насправді за ними ховається. $\square$

Визначення (КЛ12.3.1). Нехай $X : (\Omega, \mathcal{A}, \mathbb{P}) \rightarrow \mathbb{R}$ — інтегровна випадкова величина, а $\mathcal{A}' \subseteq \mathcal{A}$. Тоді існує випадкова величина $\mathbb{E}[X \mid \mathcal{A}']$, яка має назву *умовного сподівання X за умови $\mathcal{A}'$* (conditional expectation of X given $\mathcal{A}'$), якщо:

- (i) $\mathbb{E}[X \mid \mathcal{A}']$ вимірна відносно $\mathcal{A}'$ та інтегровна;
- (ii) її інтеграл дорівнює

$$\int_G \mathbb{E}[X \mid \mathcal{A}'] d\mathbb{P} = \int_G X d\mathbb{P}, \quad G \in \mathcal{A}'. \quad (\text{КЛ12.3.1})$$

$\square$

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 .

Можна зазначити таке:

- $\mathbb{E}[X \mid \{\emptyset, \Omega\}] = \mathbb{E}[X]$ майже напевно;
- $\mathbb{E}[X \mid \mathcal{A}] = X$ майже напевно;
- якщо X незалежна від σ -алгебри $\mathcal{A}'$, то $\mathbb{E}[X \mid \mathcal{A}'] = \mathbb{E}[X]$ майже напевно;
- умова (i) із Визначення КЛ12.3.1 стверджує, що ми можемо підраховувати умовне сподівання, навіть не знаючи ω , а тільки знаючи, для кожної події $G \in \mathcal{A}'$, належить їй ω чи не належить.

Нехай події $B_1, B_2, \dots$ утворюють не більш ніж зліченне розбиття Ω , яке породжує σ -алгебру $\mathcal{A}'$. Оскільки ми знаємо, якій події B_i , $i = 1, 2, \dots$, з розбиття належить ω , найліпшим прогнозним значенням X повинно бути

середнє значення X на відповідній B_i . Іншими словами, умовне сподівання $\mathbb{E}[X \mid \mathcal{A}]$ як функція від ω повинно бути сталим для $\omega \in B_i$ і дорівнювати

$$\mathbb{E}[X \mid \mathcal{A}] = \frac{1}{\mathbb{P}(B_i)} \int_{B_i} X d\mathbb{P}, \quad \omega \in B_i, \quad \mathbb{P}(B_i) > 0. \quad (\text{КЛ12.3.2})$$

Якщо $\mathbb{P}(B_i) = 0$, то (стале) значення $\mathbb{E}[X \mid \mathcal{A}]$ може бути довільне. 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.

Кажучи трішки неформально, ми показали, що умовне сподівання «за умови, що сталася подія B », дорівнює

$$\mathbb{E}[X \mid B] = \frac{1}{\mathbb{P}(B)} \int_B X d\mathbb{P}.$$

Із розгляду сподівання за умови подій із деякого розбиття безпосередньо впливає дуже важливе правило обчислення безумовних сподівань через сподівання умовні.

Твердження (Закон повного сподівання (Law of total expectation), КЛ12.3.4). Якщо події $B_1, B_2, \dots$ утворюють не більш ніж зліченне розбиття Ω , яке породжує σ -алгебру $\mathcal{A}'$, то безумовне сподівання випадкової величини X дорівнює

$$\mathbb{E}[X] = \sum_i \mathbb{E}[X \mid B_i] \mathbb{P}(B_i). \quad (\text{КЛ12.3.3})$$

Для індикаторної функції $\mathbb{1}_A(\omega)$ справедливо $\mathbb{E}[\mathbb{1}_A(\omega) \mid \mathcal{A}] = \mathbb{P}(A \mid \mathcal{A})$ майже напевно. Звідси впливає, що закон повної ймовірності із Теорема КЛ3.3.2 є частковим випадком Твердження КЛ12.3.4.

Визначення (КЛ12.3.7). Нехай маємо вимірний простір $(\Omega, \mathcal{A})$ і випадкову величину $X : (\Omega, \mathcal{A}) \rightarrow \mathbb{R}$, яка породжує σ -алгебру $\mathcal{A}(X)$. Тоді для випадкової величини Y *умовним сподіванням за умови X* (conditional expectation given X) є випадкова величина $\mathbb{E}[Y \mid \mathcal{A}(X)] \equiv \mathbb{E}[Y \mid X]$. $\square$

Використовуючи відповідні визначення, а також властивості інтегралу, можна довести такі очевидні властивості умовного сподівання.

Теорема (КЛ12.3.9). Нехай випадкові величини X, Y і $X_n, n = 1, 2, \dots$ інтегровні. Тоді

- (i) якщо $X = a$ майже напевно, то $\mathbb{E}[X \mid \mathcal{A}] = a$ майже напевно;
- (ii) для деяких $a, b \in \mathbb{R}$, $\mathbb{E}[aX + bY \mid \mathcal{A}] = a\mathbb{E}[X \mid \mathcal{A}] + b\mathbb{E}[Y \mid \mathcal{A}]$ 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 майже напевно $\lim_{n \rightarrow \infty} \mathbb{E}[X_n | \mathcal{A}'] = \mathbb{E}[X | \mathcal{A}']$;
- (v) якщо ϕ — опукла, то $\mathbb{E}[\phi(X) | \mathcal{A}'] \geq \phi(\mathbb{E}[X | \mathcal{A}'])$ майже напевно.

Зауваження (КЛ12.3.10). Надзвичайно важливо пам'ятати, що лінійність застосовна тільки для випадкових величин, сподівання яких шукаємо, а не для обумовлювальних величин. У загальному випадку $\mathbb{E}[Y | X_1 + X_2] \neq \mathbb{E}[Y | X_1] + \mathbb{E}[Y | X_2]$. $\square$

Твердження (КЛ12.3.11). Якщо X вимірна відносно $\mathcal{A}'$, а величини Y і XY інтегровні, то

$$\mathbb{E}[XY | \mathcal{A}'] = X \mathbb{E}[Y | \mathcal{A}'] \quad (\text{КЛ12.3.4})$$

майже напевно. Зверніть увагу, що X сама по собі не обов'язково повинна бути інтегровою.

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 (КЛ12.3.14, Закон ітерованих сподівань (Law of iterated expectations)). Якщо X інтегровна, а $\mathcal{A}_1$ і $\mathcal{A}_2$ — σ -алгебри такі, що $\mathcal{A}_1 \subset \mathcal{A}_2$, то

$$\mathbb{E}[\mathbb{E}[X | \mathcal{A}_2] | \mathcal{A}_1] = \mathbb{E}[X | \mathcal{A}_1] . \quad (\text{КЛ12.3.5})$$

Зокрема, якщо $\mathcal{A}_1 = \{\emptyset, \Omega\}$, а $\mathcal{A}_2 = \mathcal{A}$, то

$$\mathbb{E}[\mathbb{E}[X | \mathcal{A}_2] | \mathcal{A}_1] = \mathbb{E}[\mathbb{E}[X | \mathcal{A}]] = \mathbb{E}[X] . \quad (\text{КЛ12.3.6})$$

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:

$$\begin{aligned} \mathbb{P}_X(X \in A) &= \mathbb{E}[\mathbb{1}\{X \in A\}] = \mathbb{E}[\mathbb{E}[\mathbb{1}\{X \in A\} | Y]] \\ &= \mathbb{E}[\mathbb{P}_{X|Y}(X \in A | Y)] = \int_{-\infty}^{\infty} \mathbb{P}_{X|Y}(X \in A | Y = y) \cdot f_Y(y) dy . \end{aligned} \quad (\text{КЛ12.3.7})$$

Варто звернути увагу на умову інтегровності X для (КЛ12.3.6). Наприклад, для розподілу Коші $\mathbb{E}[X]$ не існує, хоча сподівання $\mathbb{E}[X | X > 0]$ та $\mathbb{E}[X | X < 0]$ існують і дорівнюють ∞ і $-\infty$ відповідно.

Для умовних дисперсій справедлива така властивість.

Теорема (Закон повної дисперсії (Law of total variance), КЛ12.3.18). Нехай X та Z — дві випадкові величини. Тоді дисперсію $\text{Var}(X)$, якщо вона існує, можна розписати так:

$$\text{Var}(X) = \mathbb{E}[\text{Var}(X | Z)] + \text{Var}(\mathbb{E}[X | Z]) . \quad (\text{КЛ12.3.11})$$

Приклад (КЛ12.4.3). Розгляньмо нормальний випадковий 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 нормальний розподіл, а й те, що умовні розподіли одних координат за умови інших також є нормальними. Так, зокрема, нехай

$$\mathbf{X} = \begin{pmatrix} \mathbf{X}_1 \\ \mathbf{X}_2 \end{pmatrix} \sim N \left(\begin{pmatrix} \boldsymbol{\mu}_1 \\ \boldsymbol{\mu}_2 \end{pmatrix}, \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix} \right) ,$$

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 Σ_{ij} , $i, j = 1, 2$ — це відповідні блоки матриці Σ .

Тоді простими, але тривалими алгебричними маніпуляціями спільної та маржинальної щільностей можна показати, що

$$\mathbf{X}_2 | \mathbf{X}_1 = \mathbf{x}_1 \sim N(\boldsymbol{\mu}_2 + \Sigma_{21}\Sigma_{11}^{-1}(\mathbf{x}_1 - \boldsymbol{\mu}_1), \Sigma_{22} - \Sigma_{21}\Sigma_{11}^{-1}\Sigma_{12}) . \quad \square$$

15.2 Обумовлення подіями з розбиття

Вправа 15.2.1. Нехай $T \sim \text{Exp}(\lambda)$. Чому дорівнює сподіване значення T , якщо $T > 1$? Наприклад, нас може цікавити сподіваний час (у годинах) очікування настання події за умови, що ми вже ждемо більше 1 години. $\square$

Розв'язання. Маємо обумовлення однією з двох подій, що утворюють розбиття $\Omega = \{\{\omega : T(\omega) > 1\}, \{\omega : T(\omega) \leq 1\}\}$, тому застосуємо (КЛ12.3.2):

$$\begin{aligned} \mathbb{E}[T | T > 1] &= \frac{1}{\mathbb{P}_T(T > 1)} \int_{T>1} t f_T(t) dt \\ &= \frac{1}{\int_1^\infty \lambda e^{-\lambda t} dt} \cdot \int_1^\infty t \lambda e^{-\lambda t} dt \\ &= e^\lambda \left(-te^{-\lambda t} \Big|_1^\infty + \int_1^\infty e^{-\lambda t} dt \right) \\ &= e^\lambda \left(e^{-\lambda} + \frac{1}{\lambda} e^{-\lambda} \right) = 1 + \frac{1}{\lambda} = 1 + \mathbb{E}[T] . \end{aligned}$$

Відповідний результат не повинен бути неочікуваним, адже ми знаємо, що експоненційний розподіл не має пам'яті, тому сподівання залишатиметься однаковим для будь-якої множини $T > a$, $a > 0$. $\square$

Приклад 15.2.2. Нехай нам на вибір пропонують два ідентичні конверти, у кожному з яких міститься деяка додатна сума грошей, але в одному сума вдвічі більша. Який конверт вибрати — перший або другий?

Позначмо суми грошей в обох конвертах через X та Y . Згідно з міркуваннями симетрії, $\mathbb{E}[X] = \mathbb{E}[Y]$, і жодної різниці між конвертами немає й бути не може. Проте можна розглянути такі міркування. Нехай $X = x$. Тоді в другому конверті може бути або $Y = 2x$, або $Y = 0.5x$. Сподіванням Y у цьому випадку, згідно з законом повного сподівання (КЛ12.3.3), буде

$$\mathbb{E}[Y] = \mathbb{E}[Y \mid Y = 2X] \cdot \frac{1}{2} + \mathbb{E}\left[Y \mid Y = \frac{X}{2}\right] \cdot \frac{1}{2}.$$

Звідси випливає доволі дивний результат:

$$\mathbb{E}[Y] = \frac{1}{2} \left(\mathbb{E}[2X] + \mathbb{E}\left[\frac{X}{2}\right] \right) = \frac{5}{4} \mathbb{E}[X],$$

який повністю суперечить міркуванням про симетрію!

Насправді перехід від виразу $\mathbb{E}[Y \mid Y = 2X]$ до виразу $\mathbb{E}[2X]$ є невмотивованим, адже Y не є незалежною від X . Насправді умова $Y = 2X$ означає, що X обов'язково дорівнює меншій із двох сум, і за цієї умови $\mathbb{E}[Y]$ повинно дорівнювати більшій із двох сум. Конкретніше, позначмо менше значення з двох конвертів через a . Тоді в іншому конверті повинно бути $2a$. Тоді $\mathbb{E}[Y \mid Y = 2X] = 2a$ і аналогічно $\mathbb{E}\left[Y \mid Y = \frac{X}{2}\right] = a$, звідки $\mathbb{E}[Y] = \frac{3}{2}a$.

Альтернативне пояснення таке: значення X та Y повинні реалізовуватися *одночасно*, тобто $(X, Y)^\top$ як *вектор* має спільну функцію ймовірності

$$\mathbb{P}_{X,Y}(X = x, Y = y) = \begin{cases} \frac{1}{2}, & x = a, y = 2a \\ \frac{1}{2}, & x = 2a, y = a \end{cases}.$$

Тоді, вочевидь, маємо

$$\mathbb{E}[X] = \frac{a}{2} + \frac{2a}{2} = \frac{3a}{2},$$

а для Y можна порахувати сподівання через закон повного сподівання так:

$$\begin{aligned} \mathbb{E}[Y] &= \mathbb{E}[Y \mid X = a] \mathbb{P}_X(X = a) + \mathbb{E}[Y \mid X = 2a] \mathbb{P}_X(X = 2a) \\ &= \frac{1}{2}(2a + a) = \frac{3a}{2}, \end{aligned}$$

тобто сподівання дорівнюють одне одному, і жодного парадоксу не виникає. $\square$

Вправа 15.2.3. Нехай нам пропонують зіграти в гру, у якій потенційно можна виграти грошовий приз, розмір якого (у тисячах гривень) невідомий і є випадковою величиною $V \sim U((0; 1])$. За правилами гри, спочатку потрібно зробити ставку b (у тисячах гривень). Якщо $b < \frac{2V}{3}$, то ставку відхиляють, і ми нічого не виграємо, але й не програємо. Натомість якщо $b \geq \frac{2V}{3}$, то ми одержуємо виграш $V - b$. Яким повинна бути оптимальна ставка з погляду максимізації сподіваного виграшу? $\square$

Розв'язання. Для обчислення сподіваного виграшу $\mathbb{E}[W]$ застосуємо закон повного сподівання:

$$\begin{aligned}\mathbb{E}[W] &= \mathbb{E}\left[W \mid b \geq \frac{2V}{3}\right] \mathbb{P}_V\left(b \geq \frac{2V}{3}\right) + \mathbb{E}\left[W \mid b < \frac{2V}{3}\right] \mathbb{P}_V\left(b < \frac{2V}{3}\right) \\ &= \mathbb{E}\left[V - b \mid b \geq \frac{2V}{3}\right] \mathbb{P}_V\left(b \geq \frac{2V}{3}\right) + 0 \\ &= \left(\mathbb{E}\left[V \mid V \leq \frac{3b}{2}\right] - b\right) \mathbb{P}_V\left(V \leq \frac{3b}{2}\right).\end{aligned}$$

Якщо $b \geq \frac{2}{3}$, то подія $V \leq \frac{3b}{2}$ має ймовірність 1. Тоді маємо

$$\mathbb{E}[W] = \mathbb{E}[V \mid V \leq 1] - b = \frac{1}{2} - b < 0.$$

Якщо ж $b < \frac{2}{3}$, то $V \mid V \leq \frac{3b}{2} \sim U\left(\left(0; \frac{3b}{2}\right]\right)$. Тому для цього випадку

$$\mathbb{E}[W] = \left(\frac{3b}{4} - b\right) \cdot \frac{3b}{2} = -\frac{3b^2}{8} < 0.$$

Отже $b = 0$ є оптимальною ставкою в будь-якому випадку (інакше сподіваний виграш від'ємний), тобто від такої гри ліпше утриматися.

Ця задача ілюструє так зване *прокляття переможця* (winner's curse), згідно з яким переможець в аукціоні з неповною інформацією дістає менше вигоди, ніж сподівалося. Це пов'язано з тим, що в багатьох випадках сподівана вартість виграного товару за умови виграшу є нижчою від безумовної сподіваної вартості. Як у нашому прикладі, $\mathbb{E}[V] = 0.5$, проте сподівана вартість за умови $V \leq \frac{3b}{2}$ виявилася $\frac{3b}{4} < 0.5$ для $b < \frac{2}{3}$. $\square$

Вправа 15.2.4. Мандрівник перебуває в печері, із якої є три виходи. Перший вихід веде в тунель, яким можна вибратися на поверхню за 3 години. Другий вихід веде в тунель, яким можна повернутися в цю ж печеру за 5 годин. Третій вихід веде в тунель, яким також можна повернутися в цю ж печеру, але вже за 7 годин.

Якщо мандрівник постійно вибиратиме один із трьох виходів навмання, за який сподіваний час він вибереться на поверхню? $\square$

Розв'язання. Позначмо через X час (у годинах) до виходу на поверхню. Якщо мандрівник вибере перший вихід, то $\mathbb{E}[X \mid \text{вихід } 1] = 3$. Якщо ж він вибере інший вихід, то сподіваний час буде дорівнювати сумі гарантованого часу блукання в тунелі та початкового сподівання. Конкретніше,

$$\begin{aligned}\mathbb{E}[X \mid \text{вихід } 2] &= 5 + \mathbb{E}[X] , \\ \mathbb{E}[X \mid \text{вихід } 3] &= 7 + \mathbb{E}[X] ,\end{aligned}$$

адже мандрівник повертається в печеру, і все починається спочатку.

Тоді за законом повного сподівання маємо

$$\begin{aligned}\mathbb{E}[X] &= \frac{1}{3}\mathbb{E}[X \mid \text{вихід } 1] + \frac{1}{3}\mathbb{E}[X \mid \text{вихід } 2] + \frac{1}{3}\mathbb{E}[X \mid \text{вихід } 3] \\ &= \frac{1}{3}(3 + (5 + \mathbb{E}[X]) + (7 + \mathbb{E}[X])) .\end{aligned}$$

Звідси $\mathbb{E}[X] = 15$. $\square$

Вправа 15.2.5. Нехай підкидають правильну монетку. Позначмо випад герба через 1, а числа — через 0. Чому дорівнює сподіване число підкидань до першої появи шаблону 10? До першої появи шаблону 11? $\square$

Розв'язання. Позначмо W_{10} = «число підкидань до першої появи шаблону 10». Можна помітити, що для появи шаблону 10 повинна мати місце послідовність спочатку з W_0 нулів, а потім — із W_1 одиниць. Іншого варіанту бути не може. Обидві ці величини мають розподіл FS(0.5), адже вони відповідають кількості невдач включно з першим успіхом кожна (тільки для W_0 невдачею є 0, а для W_1 невдачею є 1). Тоді маємо

$$\mathbb{E}[W_{10}] = \mathbb{E}[W_0] + \mathbb{E}[W_1] = 2 + 2 = 4 .$$

Аналіз величини W_{11} = «число підкидань до першої появи шаблону 11» трішки складніший. Очевидно, що перед першою 11 будуть стояти нулі. Проте серед цих нулів можуть мати місце одиниці. Адже якщо з'явиться одиниця, а відразу за нею — нуль, то відлік доводиться починати спочатку.

Це спостереження можна використати для обчислення шуканого $\mathbb{E}[W_{11}]$, здійснивши обумовлення тим фактом, що перший випад був одиниця:

$$\mathbb{E}[W_{11}] = \frac{1}{2}\mathbb{E}[W_{11} \mid \text{перша була } 1] + \frac{1}{2}\mathbb{E}[W_{11} \mid \text{перший був } 0] .$$

У цьому випадку

$$\mathbb{E}[W_{11} \mid \text{перший був } 0] = 1 + \mathbb{E}[W_{11}]$$

з урахуванням незалежності підкидань: ми починаємо відлік спочатку.

Що стосується $\mathbb{E}[W_{11} \mid \text{перша була } 1]$, то можна продовжувати обумовлення результатом другого підкидання:

$$\begin{aligned} \mathbb{E}[W_{11} \mid \text{перша була } 1] &= \frac{1}{2}\mathbb{E}[W_{11} \mid \text{перша й друга були } 1] \\ &\quad + \frac{1}{2}\mathbb{E}[W_{11} \mid \text{перша була } 1, \text{ а другий } - 0] . \end{aligned}$$

Як і вище, $\mathbb{E}[W_{11} \mid \text{перша була } 1, \text{ а другий } - 0] = 2 + \mathbb{E}[W_{11}]$, оскільки ми починаємо відлік спочатку. Але $\mathbb{E}[W_{11} \mid \text{перша й друга були } 1] = 2$, адже на цьому наш підрахунок завершується: шаблон 11 знайдено.

Остаточно

$$\mathbb{E}[W_{11}] = \left(2 \cdot \frac{1}{2} + (2 + \mathbb{E}[W_{11}]) \cdot \frac{1}{2}\right) + (1 + \mathbb{E}[W_{11}]) \cdot \frac{1}{2} .$$

Розв'язком цього рівняння є $\mathbb{E}[W_{11}] = 6$.

Різниця в значенні двох сподівань тільки на перший погляд може здатися дивною. Насправді жодної симетрії між шаблонами 10 і 11 немає, адже ждати шаблона 11 потрібно в певному сенсі довше, як було помітно з вищевикладеного аналізу: шаблон 10 «утворити» легше від шаблона 11. $\square$

15.3 Обумовлення випадковими величинами

Приклад 15.3.1. Нехай нам послідовно пропонують деякі об'єкти з набору в n штук, які становлять для нас різну цінність. По ознайомленню з черговим об'єктом ми повинні вирішити, забрати його чи продовжити перегляд. Якщо ми відмовляємося від об'єкта, ми його втрачаємо назавжди. Єдине, що ми можемо сказати про черговий об'єкт — чи ліпший він від усіх попередніх, чи ні. Наша задача полягає в максимізації ймовірності дістати найліпший об'єкт із-поміж n наявних. Вважатимемо, що всі можливі перестановки об'єктів є рівноймовірні, а серед об'єктів немає рівноцінних.

Ця задача належить до класу *навчання з підкріпленням* (reinforcement learning). Популярним підходом до розв'язання задач такого класу є поєднання етапів *розвідки* (exploration), коли ми перебираємо різні варіанти в пошуках найліпшого, та *експлуатації* (exploitation), коли ми максимально використовуємо розвідану інформацію замість пошуку нової.

Розгляньмо значення $0 \leq k < n$ таке, що ми перебираємо k об'єктів, після чого обираємо перший об'єкт, ліпший від цих k . Наприклад, якщо $k = 3$, то ми спочатку переглядаємо перші 3 об'єкти, а потім вибираємо той об'єкт, який першим виявиться ліпшим за перші 3. Це може бути, наприклад, об'єкт номер 6, якщо об'єкти 4 і 5 були гірші від перших 3. Звісно, може таке статися, що серед перших трьох об'єктів був найліпшим з усіх n . Тоді нам доведеться змиритися з тим об'єктом, який буде останній.

Розгляньмо ймовірність $\mathbb{P}_k(A)$ дістати найліпший об'єкт із-поміж n за такої стратегії. Виконаймо обумовлення подіями $X = i$, де X = «номер найліпшого об'єкта», $i = 1, \dots, n$. Тоді за формулою повної ймовірності

$$\mathbb{P}_k(A) = \sum_{i=1}^n \mathbb{P}_k(A \mid X = i) \mathbb{P}_X(X = i) = \frac{1}{n} \sum_{i=1}^n \mathbb{P}_k(A \mid X = i) .$$

Як зазначалося вище, якщо найліпший об'єкт був серед перших k , то ми його втратили назавжди: $\mathbb{P}_k(A \mid X = i) = 0$ для $i \leq k$.

Якщо ж $i > k$, то ймовірність дістати найліпший об'єкт на позиції i дорівнює ймовірності того, що до позиції $i - 1$ включно не буде жодного об'єкта, ліпшого від перших k . За умови, що $X = i$, усі перестановки об'єктів у решті позицій залишаються рівноймовірні, і найліпший об'єкт серед перших k може опинитися на будь-якій позиції. Отже

$$\mathbb{P}_k(A \mid X = i) = \frac{k}{i - 1} .$$

Тоді

$$\mathbb{P}_k(A) = \frac{k}{n} \sum_{i=k+1}^n \frac{1}{i - 1} .$$

Максимізувати вираз можна чисельно, але наближення можна дістати з

$$\frac{k}{n} \sum_{i=k+1}^n \frac{1}{i - 1} \approx \frac{k}{n} \int_{k+1}^n \frac{dx}{x - 1} = \frac{k}{n} \ln \frac{n - 1}{k} \approx \frac{k}{n} \ln \frac{n}{k} .$$

Для функції $g(x) = \frac{x}{n} \ln \frac{n}{x}$ умовою першого порядку є

$$\frac{1}{n} \ln \frac{n}{x} - \frac{x}{n} \cdot \frac{x}{n} \cdot \frac{n}{x^2} = 0 ,$$

і точка максимуму $x = \frac{n}{e}$. Що це точка максимуму, перевірте самостійно.

Отже найліпша стратегія полягає в перегляді перших $\frac{n}{e}$ об'єктів та виборі першого об'єкта, що буде ліпший від перших $\frac{n}{e}$. Понад те, імовірність дістати найліпший об'єкт у цьому випадку є приблизно $g\left(\frac{n}{e}\right) = \frac{1}{e} \approx 0.368$. $\square$

Вправа 15.3.2. Нехай X_1 і X_2 мають спільний двовимірний нормальний розподіл такий, що $\mathbb{E}[X_1] = 0$. Чому дорівнює $\mathbb{E}[X_1 X_2^2]$? $\square$

Розв'язання. За законом ітерованих сподівань,

$$\mathbb{E}[X_1 X_2^2] = \mathbb{E}[\mathbb{E}[X_1 X_2^2 \mid X_1]] = \mathbb{E}[X_1 \mathbb{E}[X_2^2 \mid X_1]] .$$

Із визначення умовної дисперсії випливає, що

$$\mathbb{E}[X_2^2 \mid X_1] = \text{Var}(X_2 \mid X_1) + (\mathbb{E}[X_2 \mid X_1])^2 .$$

За Прикладом КЛ12.4.3, якщо $(\mathbf{X}_1^\top, \mathbf{X}_2^\top)^\top$ багатовимірний нормальний,

$$\mathbf{X}_2 \mid \mathbf{X}_1 = \mathbf{x}_1 \sim N(\boldsymbol{\mu}_2 + \Sigma_{21}\Sigma_{11}^{-1}(\mathbf{x}_1 - \boldsymbol{\mu}_1), \Sigma_{22} - \Sigma_{21}\Sigma_{11}^{-1}\Sigma_{12}) .$$

У нашому випадку з урахуванням $\mu_1 = 0$ маємо

$$X_2 \mid X_1 = x_1 \sim N\left(\mu_2 + \frac{\sigma_{12}}{\sigma_1^2}x_1, \sigma_2^2 - \frac{\sigma_{12}^2}{\sigma_1^2}\right) .$$

Помітивши, що коефіцієнт кореляції є $\rho = \frac{\sigma_{12}}{\sigma_1\sigma_2}$, спростімо цей вираз:

$$X_2 \mid X_1 = x_1 \sim N\left(\mu_2 + \rho\frac{\sigma_2}{\sigma_1}x_1, (1 - \rho^2)\sigma_2^2\right) .$$

Отже

$$\mathbb{E}[X_2^2 \mid X_1] = (1 - \rho^2)\sigma_2^2 + \left(\mu_2 + \rho\frac{\sigma_2}{\sigma_1}X_1\right)^2 .$$

Остаточно

$$\begin{aligned} \mathbb{E}[X_1 \mathbb{E}[X_2^2 \mid X_1]] &= (1 - \rho^2)\sigma_2^2 \mathbb{E}[X_1] + \mu_2^2 \mathbb{E}[X_1] + 2\rho\mu_2\frac{\sigma_2}{\sigma_1} \mathbb{E}[X_1^2] \\ &\quad + \left(\rho\frac{\sigma_2}{\sigma_1}\right)^2 \mathbb{E}[X_1^3] . \end{aligned}$$

Оскільки $X_1 \sim N(0, \sigma_1^2)$, відразу маємо, що $\mathbb{E}[X_1^2] = \text{Var}(X_1) = \sigma_1^2$, а $\mathbb{E}[X_1^3] = 0$ за симетрією. Остаточно $\mathbb{E}[X_1 X_2^2] = 2\rho\mu_2\sigma_1\sigma_2$. $\square$

Вправа 15.3.3. Оцінки студентів за тест X та екзамен Y мають двовимірний нормальний розподіл з $\mu_X = 65$, $\mu_Y = 60$, $\sigma_X = 18$, $\sigma_Y = 20$, $\rho = 0.75$.

Чому дорівнює сподівана (середня) оцінка за екзамен для студентів, чия оцінка за МКР вища від середньої? Лежить між 60 і 80? $\square$

Розв'язання. В обох випадках нас цікавить $\mathbb{E}[Y \mid X \in (a; b)]$. Виведемо відповідну формулу, перейшовши до стандартних нормальних величин:

$$\begin{pmatrix} Z \\ W \end{pmatrix} = \begin{pmatrix} \frac{1}{\sigma_X} & 0 \\ 0 & \frac{1}{\sigma_Y} \end{pmatrix} \cdot \left(\begin{pmatrix} X \\ Y \end{pmatrix} - \begin{pmatrix} \mu_X \\ \mu_Y \end{pmatrix} \right).$$

Новоутворений вектор буде нормальний як лінійна комбінація нормальних величин. Можна обчислити його характеристики:

$$\begin{aligned} \mathbb{E}[Z] &= \mathbb{E}\left[\frac{X - \mu_X}{\sigma_X}\right] = 0, & \text{Var}(Z) &= \text{Var}\left(\frac{X - \mu_X}{\sigma_X}\right) = 1, \\ \mathbb{E}[W] &= \mathbb{E}\left[\frac{Y - \mu_Y}{\sigma_Y}\right] = 0, & \text{Var}(W) &= \text{Var}\left(\frac{Y - \mu_Y}{\sigma_Y}\right) = 1, \\ \text{Cov}(Z, W) &= \mathbb{E}\left[\frac{X - \mu_X}{\sigma_X} \cdot \frac{Y - \mu_Y}{\sigma_Y}\right] = \frac{1}{\sigma_X \sigma_Y} \text{Cov}(X, Y) = \rho. \end{aligned}$$

Тобто маємо вектор стандартних нормальних величин з коваріацією ρ .

Тоді з попередньої вправи ми знаємо, що $W \mid Z = z \sim N(\rho z, (1 - \rho^2))$, а тому $\mathbb{E}[W \mid Z] = \rho Z$. Нас цікавить $\mathbb{E}[W \mid Z \in (a; b)]$. Оскільки ми можемо утворити розбиття всього простору на ненульові події $Z \leq a$, $Z \in (a; b)$ та $Z \geq b$, застосовний вираз сподівання за умови деякої події:

$$\begin{aligned} \mathbb{E}[W \mid Z \in (a; b)] &= \frac{1}{\mathbb{P}_Z(Z \in (a; b))} \iint_{z \in (a; b)} w f_{Z,W}(z, w) dw dz \\ &= \frac{1}{\Phi(b) - \Phi(a)} \int_a^b \int_{-\infty}^{\infty} w f_{W|Z}(w, z) f_Z(z) dw dz \\ &= \frac{1}{\Phi(b) - \Phi(a)} \int_a^b \int_{-\infty}^{\infty} \frac{w}{\sqrt{2\pi(1 - \rho^2)}} e^{-\frac{(w - \rho z)^2}{2(1 - \rho^2)}} \cdot \frac{e^{-\frac{z^2}{2}}}{\sqrt{2\pi}} dw dz. \end{aligned}$$

Інтеграл

$$\int_{-\infty}^{\infty} \frac{w}{\sqrt{2\pi(1 - \rho^2)}} e^{-\frac{(w - \rho z)^2}{2(1 - \rho^2)}} dw = \rho z,$$

оскільки це просто сподівання випадкової величини $W \mid Z$, яке ми встановили, що дорівнює ρz . Тоді остаточно маємо

$$\mathbb{E}[W \mid Z \in (a; b)] = \frac{\rho}{\Phi(b) - \Phi(a)} \int_a^b z \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}} dz$$

$$= \frac{\rho}{\sqrt{2\pi}(\Phi(b) - \Phi(a))} \left(e^{-\frac{1}{2}a^2} - e^{-\frac{1}{2}b^2} \right) .$$

У нашій задачі $W \mid Z \sim N(0.75z, 1 - 0.75^2)$. Для питання про сподівання W серед усіх студентів, для яких оцінка за МКР вище середньої, тобто для яких $Z > 0$, маємо

$$\begin{aligned} \mathbb{E}[W \mid Z > 0] &= \mathbb{E}[W \mid Z \in (0; \infty)] = \frac{0.75}{\sqrt{2\pi}(\Phi(\infty) - \Phi(0))} (e^0 - e^{-\infty}) \\ &= \frac{1.5}{\sqrt{2\pi}} \approx 0.6 . \end{aligned}$$

Перейдімо в початкові одиниці виміру:

$$\mathbb{E}[Y \mid X > 65] = 60 + 20 \cdot 0.6 = 72 .$$

Аналогічно для другого питання маємо

$$\begin{aligned} \mathbb{E}[Y \mid X \in (60; 80)] &= \mathbb{E}\left[W \mid Z \in \left(\frac{60 - 65}{18}; \frac{80 - 65}{18}\right)\right] \\ &\approx \mathbb{E}[W \mid Z \in (-0.27; 0.83)] \\ &= \frac{0.75}{\sqrt{2\pi}(\Phi(0.83) - \Phi(-0.27))} \left(e^{-\frac{(-0.27)^2}{2}} - e^{-\frac{(0.83)^2}{2}} \right) \\ &\approx 0.19 . \end{aligned}$$

Перейдімо в початкові одиниці виміру:

$$\mathbb{E}[Y \mid X \in (60; 80)] = 60 + 20 \cdot 0.19 = 63.8 .$$

□

Вправа 15.3.4. Нехай станом на момент часу t число $N(t)$ пасажирів, що прибули на станцію метро для руху в конкретному напрямку, має розподіл Пуассона $\text{Pois}(\lambda t)$. Перший потяг прибуває на станцію, незалежно від пасажирів, у рівномірно розподілений на $(0; s]$ момент часу Y . Чому дорівнюють сподівання та дисперсія числа пасажирів, які зайдуть у потяг? □

Розв'язання. За умови $Y = t$ випадкова величина $N(Y) \mid Y = t \sim \text{Pois}(\lambda t)$. Нас цікавить $\mathbb{E}[N(Y)]$, яке обчислимо за законом ітерованих сподівань:

$$\mathbb{E}[N(Y)] = \mathbb{E}[\mathbb{E}[N(Y) \mid Y]] = \mathbb{E}[\lambda Y] = \lambda \mathbb{E}[Y] = \frac{\lambda s}{2} .$$

Для обчислення дисперсії застосуємо закон повної дисперсії:

$$\text{Var}(N(Y)) = \mathbb{E}[\text{Var}(N(Y) \mid Y)] + \text{Var}(\mathbb{E}[N(Y) \mid Y]) .$$

Умовна дисперсія для розподілу Пуассона є $\text{Var}(N(Y) | Y) = \lambda Y$, тому

$$\text{Var}(N(Y)) = \mathbb{E}[\lambda Y] + \text{Var}(\lambda Y) = \lambda \frac{s}{2} + \lambda^2 \frac{s^2}{12}. \quad \square$$

Вправа 15.3.5. Нехай $U_1, U_2, \dots$ — послідовність незалежних стандартних рівномірних величин. Чому дорівнює сподівання випадкової величини

$$N = \min \left\{ n : \sum_{i=1}^n U_i > 1 \right\} ? \quad \square$$

Розв'язання. Розгляньмо випадкову величину

$$N(x) = \min \left\{ n : \sum_{i=1}^n U_i > x \right\}, \quad x \in [0; 1],$$

яка відповідає числу незалежних стандартних рівномірних величин, які потрібно додати, щоб їх сума перевищила x . Виконаймо обумовлення випадковою величиною U_1 . Тоді за законом ітерованих сподівань маємо

$$\begin{aligned} \mathbb{E}[N(x)] &= \mathbb{E}[\mathbb{E}[N(x) | U_1]] = \int_{-\infty}^{\infty} \mathbb{E}[N(x) | U_1 = u] \cdot f_{U_1}(u) du \\ &= \int_0^1 \mathbb{E}[N(x) | U_1 = u] du. \end{aligned}$$

Якщо $u > x$, то вже U_1 достатньо, щоб перевищити x : $\mathbb{E}[N(x) | U_1 = u] = 1$. Якщо ж $u \leq x$, то можна звести задачу до попередньої: першої величини немає, а замість x новим пороговим значенням є $x - u$. Остаточно

$$\mathbb{E}[N(x) | U_1 = u] = \begin{cases} 1, & u > x \\ 1 + \mathbb{E}[N(x - u)], & u \leq x \end{cases}.$$

Тоді шукане сподівання дорівнює

$$\begin{aligned} \mathbb{E}[N(x)] &= \int_0^1 1 \cdot \mathbb{1}\{u > x\} du + \int_0^1 (1 + \mathbb{E}[N(x - u)]) \cdot \mathbb{1}\{u \leq x\} du \\ &= 1 + \int_0^x \mathbb{E}[N(x - u)] du = 1 + \int_0^x \mathbb{E}[N(y)] dy. \end{aligned}$$

Беручи похідну, дістаємо

$$\frac{d}{dx} \mathbb{E}[N(x)] = \mathbb{E}[N(x)].$$

Вочевидь, єдина функція, яка це задовольняє — $\mathbb{E}[N(x)] = e^x \cdot C$ для деякого $C \in \mathbb{R}$. Але оскільки $\mathbb{E}[N(0)] = 1$, бачимо, що $C = 1$, тому остаточно $\mathbb{E}[N(x)] = e^x$, і тоді відповідно $\mathbb{E}[N(1)] = e$. $\square$

15.4 Додаткові вправи

Вправа 15.4.1. Нехай X та Y — дві незалежні випадкові величини зі щільностями f_X та f_Y відповідно. Обчисліть $\mathbb{P}_{X,Y}(X < Y)$. $\square$

Розв'язання. Для обчислення такої ймовірності потрібно обчислити інтеграл від спільної щільності, яка в нашому випадку дорівнює $f_{X,Y}(x, y) = f_X(x)f_Y(y)$, за множиною $X < Y$. Проте цю ж задачу можна розглянути через призму формули повної ймовірності в неперервному випадку:

$$\begin{aligned}\mathbb{P}_{X,Y}(X < Y) &= \int_{-\infty}^{\infty} \mathbb{P}_{X|Y}(X < Y \mid Y = y) f_Y(y) dy \\ &= \int_{-\infty}^{\infty} \mathbb{P}_X(X < y) f_Y(y) dy \\ &= \int_{-\infty}^{\infty} F_X(y) f_Y(y) dy ,\end{aligned}$$

де ми використали незалежність X та Y . $\square$

Вправа 15.4.2. Нехай у вазі містяться n червоних та m синіх куль. Із вази у випадковий спосіб вибирають X куль, $1 \leq X \leq n$. Чому дорівнює ймовірність, що всі X куль — червоні? $\square$

Розв'язання. Виведемо результат за індукцією. Нехай $n = 1$, тобто у вазі містяться 1 червона й m синіх куль. Тоді, вочевидь, шукана ймовірність дорівнює просто $\mathbb{P}(A) = \frac{1}{m+1}$.

Нехай тепер відповідний результат виконується для $n - 1$ червоних і m синіх куль. Тоді за формулою повної ймовірності маємо

$$\begin{aligned}\mathbb{P}(A) &= \mathbb{P}(A \mid X = 1) \mathbb{P}_X(X = 1) + \mathbb{P}(A \mid X > 1) \mathbb{P}_X(X > 1) \\ &= \frac{n}{n+m} \cdot \frac{1}{n} + \mathbb{P}(A \mid X > 1) \cdot \frac{n-1}{n} ,\end{aligned}$$

де ймовірність $\mathbb{P}(A \mid X = 1)$ обчислено за правилами класичної ймовірності.

Для того, щоб усі кулі були червоні, коли $X > 1$, потрібно, щоб червона була перша куля, що станеться з імовірністю $\frac{n}{n+m}$, а також щоб червоні

були всі інші кулі з-посеред $X - 1$ решти. Але ймовірність цього дорівнює $\frac{1}{m+1}$ за індукційним припущенням. Отже

$$\mathbb{P}(A \mid X > 1) = \frac{n}{n+m} \cdot \frac{1}{m+1},$$

і остаточно

$$\mathbb{P}(A) = \frac{1}{n+m} + \frac{n}{n+m} \cdot \frac{1}{m+1} \cdot \frac{n-1}{n} = \frac{1}{m+1}.$$

□

Вправа 15.4.3. Нехай X та Y — незалежні неперервні величини зі щільностями f_X та f_Y відповідно. Чому дорівнює розподіл $X + Y$? □

Розв'язання. Відповідь на це питання, звісно, добре відома: розподіл суми двох незалежних випадкових величин можна обчислити за формулою згортки (КЛ9.5.1). Проте вивести її можна в альтернативний спосіб, за допомогою неперервного аналогу формули повної ймовірності:

$$\begin{aligned} \mathbb{P}_{X+Y}(X+Y < a) &= \int_{-\infty}^{\infty} \mathbb{P}_{X+Y|Y}(X+Y < a \mid Y=y) \cdot f_Y(y) dy \\ &= \int_{-\infty}^{\infty} \mathbb{P}_X(X < a-y) \cdot f_Y(y) dy \\ &= \int_{-\infty}^{\infty} F_X(a-y) f_Y(y) dy, \end{aligned}$$

де ми врахували, що

$$\begin{aligned} \mathbb{P}_{X+Y|Y}(X+Y < a \mid Y=y) &= \mathbb{P}_{X+Y|Y}(X+y < a \mid Y=y) \\ &= \mathbb{P}_X(X+y < a) \end{aligned}$$

через незалежність X та Y .

Щільність розподілу можна дістати як похідну:

$$\begin{aligned} f_{X+Y}(a) &= \frac{d}{da} \int_{-\infty}^{\infty} F_X(a-y) f_Y(y) dy = \int_{-\infty}^{\infty} \frac{d}{da} F_X(a-y) f_Y(y) dy \\ &= \int_{-\infty}^{\infty} f_X(a-y) f_Y(y) dy. \end{aligned}$$

Ця формула повністю дорівнює виразу (КЛ9.5.1).

□

Вправа 15.4.4. Розгляньмо n незалежних випробувань, кожний із яких може завершитися одним із k результатів із відповідною ймовірністю p_i , $i = 1, \dots, k$, $\sum_{i=1}^k p_i = 1$. Нехай N_i = «число випробувань, результатом яких був результат i ». Чому, для $i \neq j$, дорівнює $\mathbb{E}[N_j \mid N_i > 0]$? $\square$

Розв'язання. Згідно з законом повного сподівання,

$$\mathbb{E}[N_j] = \mathbb{E}[N_j \mid N_i = 0] \mathbb{P}_{N_i}(N_i = 0) + \mathbb{E}[N_j \mid N_i > 1] \mathbb{P}_{N_i}(N_i > 1) .$$

Бачимо, що результат описаного експерименту, вектор $(N_1, \dots, N_k)^\top$, має мультиномний розподіл $\text{Mult}_k(n, (p_1, \dots, p_k)^\top)$. Отже $N_j \sim \text{Binom}(n, p_j)$, і тому безумовно $\mathbb{E}[N_j] = np_j$.

За умови, що $N_i = r$, ймовірність, що випробування завершиться результатом j , за формулою умовної ймовірності дорівнює $p_j \mid N_i = r = \frac{p_j}{1-p_i}$. Отже умовний розподіл $N_j \mid N_i = r \sim \text{Binom}\left(n - r, \frac{p_j}{1-p_i}\right)$. З урахуванням того, що $\mathbb{P}_{N_i}(N_i = 0) = (1 - p_i)^n$, остаточно маємо

$$np_j = n \frac{p_j}{1-p_i} (1-p_i)^n + \mathbb{E}[N_j \mid N_i > 0] (1 - (1-p_i)^n) ,$$

звідки

$$\mathbb{E}[N_j \mid N_i > 0] = np_j \frac{1 - (1-p_i)^{n-1}}{1 - (1-p_i)^n} . \quad \square$$

Вправа 15.4.5. Нехай зріст сина деякого чоловіка зі зростом x має нормальний розподіл зі сподіванням $1 + x$ та дисперсією 4. Чому дорівнює найліпший (з погляду мінімізації середньоквадратичної похибки) прогноз зросту сина для чоловіка зі зростом 170? $\square$

Розв'язання. Як ми знаємо, найліпший прогноз у цьому випадку буде умовне сподівання $\mathbb{E}[Y \mid X]$, де X — зріст батька, а Y — зріст сина. За умовами задачі видно, що

$$Y \mid X = x \sim N(1 + x, 4) ,$$

а тому $\mathbb{E}[Y \mid X = 170] = 1 + 170 = 171$.

Цю модель також можна записати у формі простої регресії

$$Y = 1 + X + \varepsilon , \quad \varepsilon \sim N(0, 4) , \quad \varepsilon \perp\!\!\!\perp X .$$

У цьому випадку відповідь буде аналогічна:

$$\mathbb{E}[Y \mid X] = 1 + \mathbb{E}[X \mid X] + \mathbb{E}[\varepsilon \mid X] = 1 + X + 0 ,$$

де ми врахували незалежність ε та X . $\square$

Вправа 15.4.6. Обчисліть дисперсію величини з геометричним розподілом $N \sim \text{Geom}(p)$ обумовленням результатом першого випробування. $\square$

Розв'язання. Нехай $Y = 1$, якщо перше випробування було успіх, і 0 — у протилежному випадку. Обчислимо $\mathbb{E}[N^2]$ за допомогою закону ітерованих сподівань та закону повного сподівання:

$$\begin{aligned}\mathbb{E}[N^2] &= \mathbb{E}[\mathbb{E}[N^2 | Y]] \\ &= \mathbb{E}[\mathbb{E}[N^2 | Y = 1] \mathbb{P}_Y(Y = 1) + \mathbb{E}[N^2 | Y = 0] \mathbb{P}_Y(Y = 0)] .\end{aligned}$$

З одного боку, $\mathbb{E}[N^2 | Y = 1] = 0$, адже якщо перше випробування завершилося успіхом, то було 0 невдач. З іншого, $\mathbb{E}[N^2 | Y = 0] = \mathbb{E}[(1 + N)^2]$, адже через властивість геометричного розподілу не мати пам'яті ми можемо почати відлік спочатку.

Остаточно

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

Оскільки відомо, що $\mathbb{E}[N] = \frac{1-p}{p}$, маємо

$$\mathbb{E}[N^2] = (1 - p) \left(1 + 2\frac{1-p}{p} + \mathbb{E}[N^2] \right) ,$$

звідки

$$\mathbb{E}[N^2] = \frac{1-p}{p} + 2\frac{(1-p)^2}{p^2} .$$

Тоді дисперсія дорівнює

$$\text{Var}(N) = \mathbb{E}[N^2] - (\mathbb{E}[N])^2 = \frac{1-p}{p} + 2\frac{(1-p)^2}{p^2} - \frac{(1-p)^2}{p^2} = \frac{1-p}{p^2} . \quad \square$$

Вправа 15.4.7. У рамках дослідження поширення деякого захворювання в декількох містах регіону спочатку у випадковий спосіб обирають саме місто, а потім — вибірку з n жителів цього міста. Такого роду вибірки називають *кластерними вибірками* (cluster sampling).

Нехай Q — частка хворих в обраному місті, а X — число хворих у вибірці. Для різних міст ступінь поширення захворювання різний, тому Q є випадковою. Нехай $Q \sim U((0; 1])$. Вважатимемо, що за умови $Q = q$ кожна особа у вибірці може бути хворою з імовірністю q незалежно від інших.

Чому дорівнюють $\mathbb{E}[X]$ і $\text{Var}(X)$? $\square$

Розв'язання. Згідно з нашими припущеннями, $X \mid Q = q \sim \text{Binom}(n, q)$. Для біномного розподілу відомо, що $\mathbb{E}[X \mid Q = q] = nq$, $\text{Var}(X \mid Q = q) = nq(1 - q)$. Також нам відомо, що для стандартного рівномірного розподілу $\mathbb{E}[Q] = \frac{1}{2}$, $\text{Var}(Q) = \frac{1}{12}$. Тоді за законом ітерованих сподівань маємо

$$\mathbb{E}[X] = \mathbb{E}[\mathbb{E}[X \mid Q]] = \mathbb{E}[nQ] = \frac{n}{2}.$$

За законом повної дисперсії маємо

$$\begin{aligned} \text{Var}(X) &= \mathbb{E}[\text{Var}(X \mid Q)] + \text{Var}(\mathbb{E}[X \mid Q]) \\ &= \mathbb{E}[nQ(1 - Q)] + \text{Var}(nQ) \\ &= n\mathbb{E}[Q] - n\mathbb{E}[Q^2] + n^2\text{Var}(Q) \\ &= \frac{n}{6} + \frac{n^2}{12}. \end{aligned}$$

Понад те, можна помітити, що ця задача є ізоморфна ситуації з Зауваження КЛ8.2.2, отже ми знаємо не просто сподівання й дисперсію X , а навіть її розподіл: $X \sim U(\{0, 1, \dots, n\})$. Разом із тим, підхід із застосуванням закону ітерованих сподівань працює і в інших, складніших, випадках. Такими можуть бути ситуації, наприклад, коли розподіл Q нерівномірний або коли існують декілька рівнів єрархії (місто, область, країна), і розподіл X визначити не так просто. $\square$

16 Збіжність послідовностей випадкових величин.

Закон великих чисел

16.1 Основні теоретичні відомості

Детальний огляд понять, потрібний для розв'язання задач на цьому занятті, наведено в Курсі лекцій, Розд. КЛ13–КЛ14.

За аналогією з числовими послідовностями (Визначення КЛ7.3.11) можна визначити поняття границь деякої послідовності подій.

Визначення (КЛ13.1.1). 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

$$\liminf_{n \rightarrow \infty} A_n = \bigcup_{n=1}^{\infty} \bigcap_{k=n}^{\infty} A_k. \quad (\text{КЛ13.1.1})$$

Аналогічно, *верхньою границею* (limit superior) цієї послідовності є

$$\limsup_{n \rightarrow \infty} A_n = \bigcap_{n=1}^{\infty} \bigcup_{k=n}^{\infty} A_k. \quad (\text{КЛ13.1.2})$$

□

Словесна інтерпретація відповідних граничних подій 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 ω 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 ω належить усім множинам рано чи пізно, то зрозуміло, що він належить нескінченному числу подій. Зворотнє вкладення в загальному випадку не є справедливим: якщо елемент ω належить нескінченному числу подій, немає жодної гарантії, що він належатиме усім із них, окрім перших m штук.

Ми кажемо, що існує границя подій A , якщо нижня й верхня границі дорівнюють одна одній:

$$A = \lim_{n \rightarrow \infty} A_n = \liminf_{n \rightarrow \infty} A_n = \limsup_{n \rightarrow \infty} A_n . \quad (\text{КЛ13.1.3})$$

Для спрощення позначень писатимемо $A_n \rightarrow A$.

Можемо розглянути дві дуже важливі властивості послідовностей подій, які мають назви лем Бореля-Кантеллі.

Теорема (КЛ13.1.5, Перша лема Бореля-Кантеллі (First Borel-Cantelli lemma)¹). 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 маємо, що

$$\mathbb{P}(\{A_n \text{ н.ч.}\}) \equiv \mathbb{P}\left(\limsup_{n \rightarrow \infty} A_n\right) = 0.$$

Суть цієї леми полягає в тому, що якщо ймовірності подій наближаються до нуля, і відповідний ряд збіжний, то події A_n обов'язково припиняться.

Теорема (КЛ13.1.8, 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

$$\mathbb{P}(\{A_n \text{ н.ч.}\}) \equiv \mathbb{P}\left(\limsup_{n \rightarrow \infty} A_n\right) = 1.$$

Розгляньмо послідовності випадкових величин. Якщо границею числової послідовності є число, а границею послідовності подій — подія, то в контексті випадкових величин можна говорити про різні види збіжностей.

¹Франческо Кантеллі (Francesco Paolo Cantelli, 1875–1966) — італійський математик.

Визначення (КЛ13.2.3). Послідовність випадкових величин X_n *збігається* до випадкової величини X *майже напевно* (converges almost surely) або *з ймовірністю 1* (with probability 1), що позначають як $X_n \xrightarrow{\text{м.н.}} X$, якщо

$$\mathbb{P}\left(\left\{\omega \in \Omega : \lim_{n \rightarrow \infty} X_n(\omega) = X(\omega)\right\}\right) = 1, \quad (\text{КЛ13.2.2})$$

або, що те ж саме,

$$\mathbb{P}(\{\omega \in \Omega : \exists \varepsilon > 0 \quad \forall N \in \mathbb{N} \quad \exists n \geq N \quad |X_n(\omega) - X(\omega)| \geq \varepsilon\}) = 0. \quad (\text{КЛ13.2.3})$$

Для послідовності випадкових *векторів* $\mathbf{X}_1, \mathbf{X}_2, \dots$ маємо $\mathbf{X}_n \xrightarrow{\text{м.н.}} \mathbf{X}$, якщо виконується покоординатна збіжність майже напевно. $\square$

Зауваження (КЛ13.2.4). 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 розглядатимемо нижче. $\square$

Окрім збіжності майже напевно можна також розглянути варіант збіжності, за якого ми міняємо місцями взяття границі та обчислення ймовірності.

Визначення (КЛ13.2.8). 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

$$\forall \varepsilon > 0 \quad \lim_{n \rightarrow \infty} \mathbb{P}(\{\omega \in \Omega : |X_n(\omega) - X(\omega)| > \varepsilon\}) = 0. \quad (\text{КЛ13.2.4})$$

Для послідовності випадкових *векторів* $\mathbf{X}_1, \mathbf{X}_2, \dots$ маємо $\mathbf{X}_n \xrightarrow{p} \mathbf{X}$, якщо виконується покоординатна збіжність за ймовірністю. $\square$

Ключову відмінність збіжності за ймовірністю від збіжності майже напевно можна проілюструвати так. Нехай маємо деяке фіксоване $\varepsilon > 0$. Тоді збіжність майже напевно гарантує, що для кожної точки ω (окрім як із множини міри нуль) знайдеться таке n , починаючи з якого випадкова величина відхилятиметься від своєї границі менше від ε .

Натомість збіжність за ймовірністю нічого такого не гарантує. Вона гарантує, що ймовірності відповідних відхилень прямуватимуть до нуля. Тобто множина таких ω , для яких випадкова величина відхиляється від своєї границі більше від ε , зі збільшенням n стає все менш ймовірною, 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 (Теорема про неперервне відображення (Continuous mapping theorem), КЛ13.2.12). 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.

Однією з мотивацій розгляду збіжності майже напевно було спостереження, що інтеграли двох функцій, які відрізняються тільки на деякій множині міри нуль, однакові. Відповідно, можемо розглянути вид збіжності випадкових величин, визначений через їх інтеграли.

Визначення (КЛ13.2.14). 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

$$\lim_{n \rightarrow \infty} \mathbb{E}[|X_n - X|^p] = 0, \quad (\text{КЛ13.2.6})$$

якщо кожна випадкова величина належить простору L^p : $\mathbb{E}[|X|^p] < \infty$, $\mathbb{E}[|X_n|^p] < \infty$, $n \in \mathbb{N}$.

Для послідовності випадкових векторів $\mathbf{X}_1, \mathbf{X}_2, \dots$ маємо $\mathbf{X}_n \xrightarrow{L^p} \mathbf{X}$, якщо виконується покоординатна збіжність за u середньому порядку p . $\square$

Особливої ваги має випадок $p = 2$. Тоді кажуть, що послідовність випадкових величин збігається в *середньому квадратичному*, відтак можна стверджувати, що випадкові величини «близькі одна до одної», якщо квадрат їхньої різниці, у середньому, малий. Якщо ж $p = 1$, просто кажуть, що послідовність збігається в середньому.

Між різними видами збіжності існують такі зв'язки:

- 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 $k \rightarrow \infty$ (Твердження КЛ13.3.1);
- якщо $X_n \xrightarrow{L^p} X$, то $X_n \xrightarrow{p} X$ (Твердження КЛ13.3.3);
- між збіжністю майже напевно та збіжністю в середньому прямого зв'язку немає. Якщо виконуються умови теорем про монотонну збіжність КЛ11.1.2 та про мажоровану збіжність КЛ11.1.3, то збіжність майже напевно має наслідком збіжність у середньому для $p = 1$.

Середнє вибіркове $\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i$, розглянуте в Прикладі КЛ9.4.16, є чи не найважливішою числовою 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. Розгляньмо найпростіше його формулювання.

Теорема ((Слабкий) Закон великих чисел² ((Weak) Law of large numbers), КЛ14.1.1). Нехай маємо послідовність незалежних випадкових величин $X_1, X_2, \dots$ таких, що їхні сподівання існують, скінченні й дорівнюють $\mathbb{E}[X_i] = \mu_i, i \in \mathbb{N}$, а дисперсії також існують, скінченні й рівномірно обмежені, тобто $\text{Var}(X_i) \leq C < \infty, i \in \mathbb{N}$. Тоді

$$\frac{1}{n} \sum_{i=1}^n (X_i - \mathbb{E}[X_i]) \xrightarrow{p} 0. \quad (\text{КЛ14.1.1})$$

Зауваження (КЛ14.1.2). Випадкові величини насправді необов'язково повинні бути незалежні. Достатньо вимагати їхньої некорельованості.

Також нескладно бачити, що за умов Теорема КЛ14.1.1 можна довести не тільки збіжність за ймовірністю, а й збіжність у середньоквадратичному. Справді, оскільки

$$\mathbb{E} \left[\frac{1}{n^2} (S_n - \mathbb{E}[S_n])^2 \right] = \frac{1}{n^2} \text{Var}(S_n) \leq \frac{C}{n} \rightarrow 0,$$

маємо збіжність у середньоквадратичному.

Нарешті, можна завжди розглянути *центровані* величини $X_i^* = X_i - \mathbb{E}[X_i], i \in \mathbb{N}$, 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 величин з однаковим сподіванням.

□

Розгляньмо тепер *посилений* закон великих чисел (ПЗВЧ).

Теорема (Посилений закон великих чисел (Strong law of large numbers), КЛ14.2.1). Нехай маємо послідовність незалежних величин $X_1, X_2, \dots$ таких, що вони мають *однаковий розподіл* зі скінченним сподіванням $\mathbb{E}[X_i] = \mu, i \in \mathbb{N}$. Тоді

$$\frac{1}{n} \sum_{i=1}^n (X_i - \mu) \xrightarrow{\text{м.н.}} 0. \quad (\text{КЛ14.2.1})$$

Зауваження (КЛ14.2.2). Для доведення Теорема КЛ14.2.1 достатньо, щоб величини $X_i, i \in \mathbb{N}$, будуть усього лише *попарно* незалежними. □

Зауваження (КЛ14.2.3). Із Твердження КЛ13.3.1 відразу випливає, що за умов Теорема КЛ14.2.1 виконується не тільки ПЗВЧ, а й ЗВЧ. На практиці в більшості випадків збіжності за ймовірністю цілком достатньо, проте

²ЗВЧ в такому формулюванні також називають теоремою Чебишова.

приємно усвідомлювати, що збіжність майже напевно для однаково розподілених випадкових величин не вимагає додаткових посиленних вимог до відповідної послідовності. $\square$

16.2 Послідовності випадкових величин

Вправа 16.2.1. Нехай $X \sim \text{Exp}(1)$. Розгляньмо послідовність випадкових величин $Y_1, Y_2, \dots$, де $Y_n = \frac{X}{n}$, $n \in \mathbb{N}$.

До чого прямує така послідовність за ймовірністю? $\square$

Розв'язання. За визначенням, нам потрібно показати, що $\mathbb{P}_{Y_n}(|Y_n - Y| > \varepsilon)$ прямує до 0 для будь-якого $\varepsilon > 0$. Але для цього спочатку потрібно знати, чому дорівнює гранична випадкова величина Y . У нашому випадку можна зробити інтуїтивне припущення, що $Y = 0$. Перевірмо, чи ми праві:

$$\mathbb{P}_{Y_n}(|Y_n - Y| > \varepsilon) = \mathbb{P}_{Y_n}(|Y_n| > \varepsilon) = \mathbb{P}_X(X > n\varepsilon) = e^{-n\varepsilon}.$$

Цілком очевидно, що цей вираз прямує до 0, коли $n \rightarrow \infty$, для будь-якого $\varepsilon > 0$. Отже ми показали, що $Y_n \xrightarrow{p} 0$. $\square$

Вправа 16.2.2. Нехай маємо послідовність випадкових величин

$$X_n \sim U\left(\left(-\frac{1}{n}; \frac{1}{n}\right)\right), \quad n \in \mathbb{N}.$$

До чого прямує така послідовність за ймовірністю? $\square$

Розв'язання. Інтуїтивно можна очікувати, що границя послідовності $X = 0$. Отже потрібно розглянути для деякого $\varepsilon > 0$ ймовірність, що $|X_n - X| = |X_n| > \varepsilon$, і показати, що вона прямує до 0. Справді,

$$\mathbb{P}_{X_n}(|X_n| > \varepsilon) = 1 - \mathbb{P}_{X_n}(-\varepsilon \leq X_n \leq \varepsilon).$$

Якщо $\varepsilon \geq \frac{1}{n}$, вочевидь відразу матимемо 0. Якщо ж $\varepsilon < \frac{1}{n}$, то будемо мати

$$\mathbb{P}_{X_n}(|X_n| > \varepsilon) = 1 - \left(\frac{\varepsilon - (-\frac{1}{n})}{\frac{1}{n} - (-\frac{1}{n})} - \frac{-\varepsilon - (-\frac{1}{n})}{\frac{1}{n} - (-\frac{1}{n})} \right) = 1 - \varepsilon n.$$

Може здатися, що цей вираз не прямує до 0, проте потрібно помітити, що яке б не було $\varepsilon > 0$, завжди знайдеться таке велике n , що буде $\varepsilon \geq \frac{1}{n}$. Отже випадок $\varepsilon < \frac{1}{n}$ взагалі ні на що не впливає, і тому $X_n \xrightarrow{p} 0$. $\square$

Вправа 16.2.3. Нехай маємо послідовність незалежних випадкових величин $X_1, X_2, \dots \sim U((-1; 1])$.

Розгляньмо послідовність випадкових величин $Y_n = X_1 \cdot \dots \cdot X_n$, $n \in \mathbb{N}$.

До чого прямує така послідовність за ймовірністю?

Що можна сказати про послідовність $Z_n = \max\{X_1, \dots, X_n\}$, $n \in \mathbb{N}$? $\square$

Розв'язання. З урахуванням незалежності X_n , згідно з Теоремою КЛ9.4.12, сподівання та дисперсії величин Y_n дорівнюють

$$\begin{aligned}\mathbb{E}[Y_n] &= \mathbb{E}[X_1 \cdot \dots \cdot X_n] = \mathbb{E}[X_1] \dots \mathbb{E}[X_n] = 0, \\ \text{Var}(Y_n) &= \mathbb{E}[Y_n^2] = \mathbb{E}[X_1^2 \cdot \dots \cdot X_n^2] = \mathbb{E}[X_1^2] \dots \mathbb{E}[X_n^2] \\ &= (\text{Var}(X_1))^n = \left(\frac{4}{12}\right)^n.\end{aligned}$$

Можемо застосувати нерівність Чебишова:

$$\mathbb{P}_{Y_n}(|Y_n - \mathbb{E}[Y_n]| > \varepsilon) = \mathbb{P}_{Y_n}(|Y_n| > \varepsilon) \leq \frac{\text{Var}(Y_n)}{\varepsilon^2} = \left(\frac{4}{12}\right)^n \cdot \frac{1}{\varepsilon^2} \rightarrow 0.$$

Іншими словами, ми показали, що послідовність $Y_n \xrightarrow{p} 0$.

Такий спосіб доведення збіжності за ймовірністю (показати, що дисперсії прямують до 0) є зручним, якщо послідовність прямує до деякого значення μ , яке є спільним сподіванням для всіх Y_n , $n \in \mathbb{N}$.

Що стосується Z_n , то інтуїція підказує, що така послідовність повинна прямувати до 1. Тому потрібно показати, що до 0 прямує ймовірність $\mathbb{P}_{Z_n}(|Z_n - 1| > \varepsilon)$ (для будь-якого $\varepsilon > 0$). Ми не можемо застосувати нерівність Чебишова, адже $\mathbb{E}[Z_n] \neq 1$ для жодного n . Підемо іншим шляхом:

$$\begin{aligned}\mathbb{P}_{Z_n}(|Z_n - 1| > \varepsilon) &= \mathbb{P}_{X_1, \dots, X_n}(\max\{X_1, \dots, X_n\} \geq 1 + \varepsilon) \\ &\quad + \mathbb{P}_{X_1, \dots, X_n}(\max\{X_1, \dots, X_n\} \leq 1 - \varepsilon) \\ &= \mathbb{P}_{X_1, \dots, X_n}(X_1 \leq 1 - \varepsilon, \dots, X_n \leq 1 - \varepsilon) \\ &= (\mathbb{P}_{X_1}(X_1 \leq 1 - \varepsilon))^n = \left(1 - \frac{\varepsilon}{2}\right)^n.\end{aligned}$$

Ми використали незалежність X_n , а також очевидний факт, що величина з носієм $(-1; 1]$ не може перевищувати 1, а тим більше $1 + \varepsilon$.

Отже ми бачимо, що коли $n \rightarrow \infty$, шукана ймовірність прямує до 0, підтверджуючи той факт, що $Z_n \xrightarrow{p} 1$. $\square$

Вправа 16.2.4. Нехай маємо послідовність незалежних випадкових величин $Y_1, Y_2, \dots$, де

$$\mathbb{P}_{Y_n}(Y_n = y) = \begin{cases} 1 - \frac{1}{n}, & y = 0 \\ \frac{1}{n}, & y = n^2 \\ 0, & \text{інакше} \end{cases}, \quad n \in \mathbb{N}.$$

До чого прямує така послідовність за ймовірністю? Чи прямує вона до цієї ж граничної величини майже напевно? До чого прямує послідовність зі сподівань таких величин? $\square$

Розв'язання. Інтуїція підказує, що така послідовність може прямувати до 0. Справді, для деякого $\varepsilon > 0$ завжди можна підібрати достатньо велике n таке, що $\varepsilon < n^2$. Позначмо його n_ε . Отже

$$\mathbb{P}_{Y_n}(|Y_n| > \varepsilon) = \mathbb{P}_{Y_n}(Y_n = n^2) = \frac{1}{n} \rightarrow 0,$$

і $Y_n \xrightarrow{p} 0$.

Також можна проаналізувати, чи $Y_n \xrightarrow{\text{м.н.}} 0$. Справді, оскільки кожна величина Y_n дискретна, це означає, що прообразами значень із її носія є деякі множини $\{\omega : Y_n(\omega) = 0\}$ і $\{\omega : Y_n(\omega) = n^2\}$, що мають ймовірності $1 - \frac{1}{n}$ та $\frac{1}{n}$ відповідно. Іншими словами, $\mathbb{P}(\{\omega : Y_n(\omega) \notin \{0, n^2\}\}) = 0$.

Нехай $A_n = \{\omega : Y_n(\omega) = n^2\}$. Тоді верхня границя такої послідовності множин дорівнює

$$A = \limsup_n A_n = \bigcap_{n=1}^{\infty} \bigcup_{k=n}^{\infty} A_k$$

і містить усі такі точки $\omega \in \Omega$ з початкового абстрактного простору, для яких $Y_n(\omega)$ не збігається до 0. Це впливає з інтерпретації множини A як множини точок, для яких A_n стається нескінченно часто, тобто для яких нескінченне число елементів послідовності дорівнює не 0, а n^2 .

Нескладно бачити, що для будь-якого $\varepsilon > 0$

$$\sum_{n=1}^{\infty} \mathbb{P}(\{\omega : Y_n(\omega) \geq \varepsilon\}) = \sum_{n=1}^{\infty} \mathbb{P}_{Y_n}(Y_n \geq \varepsilon) = \sum_{n=n_\varepsilon}^{\infty} \frac{1}{n} = \infty.$$

Оскільки всі величини в послідовності незалежні, згідно з другою лемою Бореля-Кантеллі маємо, що

$$\mathbb{P}(\{\omega : Y_n(\omega) \geq \varepsilon \text{ н.ч.}\}) = 1,$$

тобто множина точок ω , у яких $Y_n(\omega)$ не прямує до 0 (тому що нескінченно

часто відповідна відстань $|Y_n(\omega) - 0|$ перевищує ε) має одиничну ймовірність. Отже *не можна* стверджувати, що $Y_n \xrightarrow{\text{м.н.}} 0$.

Що стосується сподівань, то

$$\mathbb{E}[Y_n] = \left(1 - \frac{1}{n}\right) \cdot 0 + \frac{n^2}{n} = n.$$

Тобто послідовність зі сподівань не те що не прямує до 0, а є розбіжною. Отже збіжності в середньому (до 0) в цьому випадку не спостерігається. $\square$

Вправа 16.2.5. Нехай маємо послідовність випадкових величин $X_1, X_2, \dots$. Покажіть, що ця послідовність збігається в середньоквадратичному до деякого $b \in \mathbb{R}$ тоді й тільки тоді, коли

$$\lim_{n \rightarrow \infty} \mathbb{E}[X_n] = b, \quad \lim_{n \rightarrow \infty} \text{Var}(X_n) = 0.$$

Нехай для $n \in \mathbb{N}$

$$\mathbb{P}_{X_n} \left(X_n = \frac{1}{n} \right) = 1 - \frac{1}{n^2}, \quad \mathbb{P}_{X_n}(X_n = n) = \frac{1}{n^2}.$$

До якої константи прямує ця послідовність за ймовірністю? Чи прямує вона до цієї ж константи в середньоквадратичному? $\square$

Розв'язання. Нам потрібно показати, що для того, щоб виконувалося

$$\lim_{n \rightarrow \infty} \mathbb{E}[(X_n - b)^2] = 0,$$

повинні виконуватися відповідні граничні переходи для сподівань та дисперсій. Можна помітити, що

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

Щоб ліва частина прямувала до 0, потрібно, щоб обидва доданки справа прямували до 0 (вони невід'ємні). Звідси випливає шуканий результат.

Розгляньмо тепер конкретну наведену послідовність. Інтуїція підказує, що така послідовність може прямувати за ймовірністю до $b = 0$. Справді, для будь-якого $\varepsilon > 0$ можна підібрати таке велике n , що $\varepsilon > \frac{1}{n}$, і тоді

$$\mathbb{P}_{X_n}(|X_n| > \varepsilon) = \mathbb{P}_{X_n}(X_n = n) = \frac{1}{n^2} \rightarrow 0,$$

отже $X_n \xrightarrow{p} 0$.

Також можна бачити, що

$$\mathbb{E}[X_n] = \frac{1}{n} \cdot \left(1 - \frac{1}{n^2}\right) + \frac{n}{n^2} = \frac{2}{n} - \frac{1}{n^3},$$

отже $\mathbb{E}[X_n] \rightarrow 0$. Тому якби послідовність прямувала в середньоквадратичному до деякої константи, це могла б бути тільки $b = 0$. Але тоді

$$\mathbb{E}[X_n^2] = \frac{1}{n^2} \cdot \left(1 - \frac{1}{n^2}\right) + \frac{n^2}{n^2} = 1 + \frac{1}{n^2} - \frac{1}{n^4},$$

і тоді

$$\text{Var}(X_n) = 1 + \frac{1}{n^2} - \frac{1}{n^4} - \left(\frac{2}{n} - \frac{1}{n^3}\right)^2 \rightarrow 1,$$

тобто збіжність у середньоквадратичному до $b = 0$ не виконується. $\square$

Вправа 16.2.6. Нехай $\lambda_n = \frac{1}{n}$, $n \in \mathbb{N}$, і нехай $X_n \sim \text{Pois}(\lambda_n)$. До чого прямує за ймовірністю послідовність X_n ? Послідовність nX_n ? $\square$

Розв'язання. Інтуїція підказує, що X_n повинна прямувати до 0, адже щоразу інтенсивність стає все меншою. Можна розглянути $\mathbb{P}_{X_n}(|X_n| > \varepsilon)$ для довільного $\varepsilon > 0$, проте в цьому випадку простіше піти за доведенням збіжності в середньоквадратичному. Справді,

$$\mathbb{E}[(X_n - 0)^2] = \mathbb{E}[X_n^2] = \text{Var}(X_n) + (\mathbb{E}[X_n])^2 = \lambda_n + \lambda_n^2 = \frac{1}{n} + \frac{1}{n^2} \rightarrow 0,$$

звідки випливає, що $X_n \xrightarrow{p} 0$.

На жаль, із величинами nX_n такий підхід не працює:

$$\mathbb{E}[(nX_n)^2] = n^2 \cdot \left(\frac{1}{n} + \frac{1}{n^2}\right) = n + 1.$$

Тому можна піти за визначенням. Для деякого $\varepsilon > 0$ завжди можна підібрати достатньо велике n таке, щоб $\frac{\varepsilon}{n} < 1$. Тоді

$$\begin{aligned} \mathbb{P}_{X_n}(nX_n \geq \varepsilon) &= \mathbb{P}_{X_n}\left(X_n \geq \frac{\varepsilon}{n}\right) = 1 - \mathbb{P}_{X_n}(X_n = 0) \\ &= 1 - e^{-\lambda_n} = 1 - e^{-\frac{1}{n}} \rightarrow 0, \end{aligned}$$

отже $nX_n \xrightarrow{p} 0$.

За такою схемою можна довести $n^k X_n \xrightarrow{p} 0$ для будь-якого $k > 0$. Іншими словами, X_n прямує до 0 швидше, ніж n^k прямує до нескінченності. $\square$

Вправа 16.2.7. Нехай $\mathbb{P}_X(X = 1) = \mathbb{P}_X(X = -1) = 0.5$. Нехай маємо послідовність $X_1, X_2, \dots$, де кожна X_n , $n \in \mathbb{N}$, може дорівнювати e^n з імовірністю $\frac{1}{n}$ і дорівнювати X з імовірністю $1 - \frac{1}{n}$.

Чи прямує X_n до X за ймовірністю? У середньоквадратичному? $\square$

Розв'язання. Як завжди, для будь-якого $\varepsilon > 0$ потрібно обчислити ймовірність $\mathbb{P}_{X_n, X}(|X_n - X| > \varepsilon)$. Може здатися, що цю ймовірність непросто обчислити, хоча насправді потрібно помітити, що $X_n - X$ може набувати всього двох значень, одне з яких $= 0$. За формулою повної ймовірності

$$\begin{aligned}\mathbb{P}_{X_n, X}(|X_n - X| > \varepsilon) &= \mathbb{P}_{X_n, X}(|X_n - X| > \varepsilon \mid X_n = X) \mathbb{P}_{X_n, X}(X_n = X) \\ &\quad + \mathbb{P}_{X_n, X}(|X_n - X| > \varepsilon \mid X_n = e^n) \mathbb{P}_{X_n}(X_n = e^n) \\ &= \frac{1}{n} \mathbb{P}_{X_n, X}(|X_n - X| > \varepsilon \mid X_n = e^n) \\ &= \frac{1}{n} \mathbb{P}_X(|e^n - X| > \varepsilon) .\end{aligned}$$

Ми змогли позбутися першого доданку, оскільки якщо $X_n = X$, то подія $|X_n - X| = 0 > \varepsilon$ є неімовірною. Також ми використали незалежність X_n та X , підставивши $X_n = e^n$ у вираз зліва від $|$ та позбувшись обумовлення.

Для будь-якого $\varepsilon > 0$ можна знайти достатньо велике n таке, що $|e^n - X| > \varepsilon$, оскільки $|X| \leq 1$. Отже для всіх $\varepsilon > 0$ шукана ймовірність буде рано чи пізно дорівнювати $\frac{1}{n}$, що у свою чергу прямує до 0. Отже $X_n \xrightarrow{p} X$.

Щоправда, збіжності в середньоквадратичному не буде. Суто формально можемо покласти, що X_n дорівнює e^n або X відповідно до значень деякої величини $W \sim \text{Bern}\left(\frac{1}{n}\right)$. Тоді можна використати закон ітерованих сподівань та закон повного сподівання:

$$\begin{aligned}\mathbb{E}[(X_n - X)^2] &= \mathbb{E}[\mathbb{E}[(X_n - X)^2 \mid W]] \\ &= \mathbb{E}[\mathbb{E}[(X_n - X)^2 \mid W = 0] \mathbb{P}_W(W = 0) \\ &\quad + \mathbb{E}[(X_n - X)^2 \mid W = 1] \mathbb{P}_W(W = 1)] \\ &= 0 \cdot \left(1 - \frac{1}{n}\right) + (\mathbb{E}[(e^n - X)^2 \mid W = 1]) \cdot \frac{1}{n} \\ &= 0 \cdot \left(1 - \frac{1}{n}\right) + (\mathbb{E}[(e^n - X)^2]) \cdot \frac{1}{n} \\ &= (e^{2n} - 2e^n \mathbb{E}[X] + \mathbb{E}[X^2]) \cdot \frac{1}{n} \\ &= (e^{2n} + 1) \cdot \frac{1}{n} \rightarrow \infty .\end{aligned}$$

Ми знову використали незалежність $X \mid W$, щоб позбутися обумовлення. Отже *не можемо* казати, що X_n прямує до X у середньоквадратичному. $\square$

16.3 Закон великих чисел

Вправа 16.3.1. Нехай $X_1, X_2, \dots$ — незалежні додатні випадкові величини з однаковим розподілом зі сподіванням 2, а $Y_1, Y_2, \dots$ — незалежні додатні випадкові величини з однаковим розподілом зі сподіванням 3. До чого (майже напевно) прямує послідовність

$$\frac{X_1 + X_2 + \dots + X_n}{Y_1 + Y_2 + \dots + Y_n} ?$$

□

Розв'язання. Оскільки обидві послідовності містять незалежні однаково розподілені величини зі скінченними сподіваннями, для кожної з них виконуються умови ПЗВЧ (Теорема КЛ14.2.1):

$$\frac{1}{n} \sum_{i=1}^n X_i \xrightarrow{\text{м.н.}} 2, \quad \frac{1}{n} \sum_{i=1}^n Y_i \xrightarrow{\text{м.н.}} 3.$$

Спочатку можемо привести відношення до потрібного вигляду, домноживши чисельник і знаменник на $\frac{1}{n}$:

$$\frac{X_1 + X_2 + \dots + X_n}{Y_1 + Y_2 + \dots + Y_n} = \frac{\frac{1}{n} \sum_{i=1}^n X_i}{\frac{1}{n} \sum_{i=1}^n Y_i}.$$

Отже за Теоремою про неперервне відображення КЛ13.2.12 (у застосуванні до функції ділення двох чисел) маємо

$$\frac{\frac{1}{n} \sum_{i=1}^n X_i}{\frac{1}{n} \sum_{i=1}^n Y_i} \xrightarrow{\text{м.н.}} \frac{2}{3}.$$

□

Вправа 16.3.2. Нехай $X_1, X_2, \dots$ — незалежні додатні випадкові величини з однаковим розподілом зі сподіванням μ . Нехай $W_n = \frac{X_1}{X_1 + \dots + X_n}$. До чого прямує послідовність nW_n , $n \in \mathbb{N}$, майже напевно? □

Розв'язання. Для послідовності X_n виконуються умови Теорема КЛ14.2.1, і тому $\frac{1}{n} \sum_{i=1}^n X_i \xrightarrow{\text{м.н.}} \mu$. Також цілком очевидно, що тривіально $X_1 \xrightarrow{\text{м.н.}} X_1$.

Випадкову величину W_n можна подати як

$$W_n = \frac{X_1/n}{\frac{1}{n} \sum_{i=1}^n X_i},$$

звідки за Теоремою про неперервне відображення КЛ13.2.12 маємо

$$nW_n = \frac{X_1}{\frac{1}{n} \sum_{i=1}^n X_i} \xrightarrow{\text{м.н.}} \frac{X_1}{\mu}.$$

Границею послідовності є випадкова величина, проте для *конкретної* реалізації границею є нормоване значення першого члена послідовності. $\square$

Вправа 16.3.3. Згідно з ПЗВЧ, середнє вибіркве $\bar{X}_n = \frac{1}{n} \sum_{i=1}^n X_i$ для вибірки $X_1, \dots, X_n$ незалежних однаково розподілених величин зі скінченним сподіванням μ прямує до μ майже напевно. Якщо додатково ці величини мають скінченну дисперсію σ^2 , то її оцінкою буде випадкова величина

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

Покажіть, що $\mathbb{E}[S_n^2] = \sigma^2$, що $S_n^2 \xrightarrow{\text{м.н.}} \sigma^2$, і що $\sqrt{S_n^2} \xrightarrow{\text{м.н.}} \sigma$. $\square$

Розв'язання. Цілком очевидно, що

$$X_i - \bar{X}_n = X_i - \mu - (\bar{X}_n - \mu) \equiv Y_i - \bar{Y}_n,$$

де $\mathbb{E}[Y_i] = 0$, $i = 1, \dots, n$, і тоді $\mathbb{E}[Y_i^2] = \mathbb{E}[(X_i - \mu)^2] = \text{Var}(X_i)$. Оскільки Y_i є функціями від X_i , вони також є незалежні між собою.

Підрахуймо відповідне сподівання:

$$\begin{aligned} \mathbb{E}[S_n^2] &= \frac{1}{n-1} \sum_{i=1}^n \left(\mathbb{E}[Y_i^2] - 2\mathbb{E}[Y_i \bar{Y}_n] + \mathbb{E}[(\bar{Y}_n)^2] \right) \\ &= \frac{1}{n-1} \sum_{i=1}^n \left(\mathbb{E}[Y_i^2] - \frac{2}{n} \sum_{j=1}^n \mathbb{E}[Y_i Y_j] + \frac{1}{n^2} \mathbb{E} \left[\left(\sum_{j=1}^n Y_j \right)^2 \right] \right) \\ &= \frac{1}{n-1} \sum_{i=1}^n \left(\mathbb{E}[Y_i^2] - \frac{2}{n} \mathbb{E}[Y_i^2] - \frac{2}{n} \sum_{j \neq i}^n \mathbb{E}[Y_i] \mathbb{E}[Y_j] \right. \\ &\quad \left. + \frac{1}{n^2} \sum_{j=1}^n \mathbb{E}[Y_j^2] + \frac{2}{n^2} \sum_{j > i}^n \mathbb{E}[Y_i] \mathbb{E}[Y_j] \right) \\ &= \frac{1}{n-1} \sum_{i=1}^n \left(\sigma^2 - \frac{2}{n} \sigma^2 + \frac{1}{n} \sigma^2 \right) = \frac{1}{n-1} \sum_{i=1}^n \left(\frac{n-1}{n} \sigma^2 \right) = \sigma^2. \end{aligned}$$

Що стосується збіжності за ймовірністю, то це випливає з ПЗВЧ та теореми про неперервне відображення. Справді,

$$S_n^2 = \frac{n}{n-1} \cdot \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X}_n)^2 = \frac{n}{n-1} \cdot \frac{1}{n} \sum_{i=1}^n \left(X_i^2 - 2X_i\bar{X}_n + (\bar{X}_n)^2 \right).$$

Оскільки X_i^2 незалежні, мають однаковий розподіл і скінченні сподівання,

$$\frac{1}{n} \sum_{i=1}^n X_i^2 \xrightarrow{\text{м.н.}} \mathbb{E}[X_i^2] = \text{Var}(X_i) + (\mathbb{E}[X_i])^2 = \sigma^2 + \mu^2.$$

Також маємо

$$-\frac{2}{n} \sum_{i=1}^n X_i \bar{X}_n = -2(\bar{X}_n)^2 \xrightarrow{\text{м.н.}} -2(\mathbb{E}[X_i])^2$$

за теоремою про неперервне відображення в застосуванні до функції піднесення до квадрату. Аналогічно й

$$\frac{1}{n} \sum_{i=1}^n (\bar{X}_n)^2 = (\bar{X}_n)^2 \xrightarrow{\text{м.н.}} (\mathbb{E}[X_i])^2.$$

Остаточно маємо

$$S_n^2 \xrightarrow{\text{м.н.}} 1 \cdot (\sigma^2 + \mu^2 - 2\mu^2 + \mu^2) = \sigma^2.$$

Нарешті, за тією ж таки теоремою про неперервне відображення маємо

$$\sqrt{S_n^2} \xrightarrow{\text{м.н.}} \sqrt{\sigma^2} = \sigma. \quad \square$$

Вправа 16.3.4. Нехай $X_1, X_2, \dots \sim \text{Bern}(p)$ незалежні. До чого прямують у середньоквадратичному (а отже й за ймовірністю) випадкові величини $\frac{1}{n} \sum_{i=1}^n X_i^2$, $n \in \mathbb{N}$? $\square$

Розв'язання. Відповідно до Зауваження КЛ14.1.2, за умовами з Теорему КЛ14.1.1, якщо послідовність містить незалежні величини зі скінченними сподіваннями та рівномірно обмеженими дисперсіями, то для них виконується не тільки слабкий ЗВЧ, а й збіжність у середньоквадратичному. Наші X_i^2 незалежні, мають скінченні сподівання $\mathbb{E}[X_i^2] = p$ та рівномірно обмежені дисперсії $\text{Var}(X_i^2) = \mathbb{E}[X_i^4] - (\mathbb{E}[X_i^2])^2 = p - p^2 \leq \frac{1}{4}$, тому

$$\frac{1}{n} \sum_{i=1}^n (X_i^2 - \mathbb{E}[X_i^2]) \xrightarrow{p} 0,$$

або ж

$$\frac{1}{n} \sum_{i=1}^n X_i^2 \xrightarrow{p} p ,$$

і аналогічно для збіжності в середньоквадратичному.

Понад те, виконуються всі умови Теорема КЛ14.2.1, тому маємо навіть збіжність майже напевно. $\square$

16.4 Додаткові вправи

Вправа 16.4.1. Нехай випадкова величина X має щільність

$$f(x) = \frac{\ln a}{x(\ln x)^2} \cdot \mathbb{1}\{x \geq a\} , \quad a > 1 .$$

Нехай маємо послідовність випадкових величин $X_n = \frac{X}{n}$, $n \in \mathbb{N}$. Покажіть, що ця послідовність прямує до 0 за ймовірністю, проте не прямує до 0 в середньому порядку p для будь-якого $p > 0$. $\square$

Розв'язання. Для деякого $\varepsilon > 0$ маємо

$$\begin{aligned} \mathbb{P}_{X_n}(|X_n| > \varepsilon) &= \mathbb{P}_X(X > n\varepsilon) \\ &= \int_{n\varepsilon}^{\infty} \frac{\ln a}{x(\ln x)^2} dx = \frac{\ln a}{\ln n\varepsilon} \rightarrow 0 , \end{aligned}$$

звідки безпосередньо випливає $X_n \xrightarrow{p} 0$.

З іншого боку, для будь-якого $p > 0$ маємо

$$\mathbb{E}[|X|^p] = \mathbb{E}\left[\left(\frac{X}{n}\right)^p\right] = \frac{\ln a}{n^p} \int_a^{\infty} \frac{dx}{x^{1-p}(\ln x)^2} = \infty ,$$

і X_n не прямує до 0 в середньоквадратичному жодного порядку $p > 0$. $\square$

Вправа 16.4.2. Нехай X має щільність f . Нехай існує послідовність щільностей $f_1, f_2, \dots$ така, що $f_n(x) \rightarrow f(x)$ поточково. Нехай маємо послідовність випадкових величин $X_n = \frac{f_n(X)}{f(X)}$, $n \in \mathbb{N}$. Покажіть, що $X_n \xrightarrow{\text{м.н.}} 1$. $\square$

Розв'язання. Нескладно бачити, що

$$\frac{f_n(x)}{f(x)} \rightarrow 1$$

поточково, окрім тих точок, де $f(x) = 0$. Тоді

$$\begin{aligned}\mathbb{P}_{X_n} \left(\left\{ \omega : \lim_{n \rightarrow \infty} X_n(\omega) = 1 \right\} \right) &\geq \mathbb{P}_X (f(X) > 0) = 1 - \mathbb{P}_X (f(X) = 0) \\ &= 1 - \int_{f(x)=0} f(x) dx = 1 - 0 = 1 .\end{aligned}$$

□

Вправа 16.4.3. Нехай маємо послідовність $X_1, X_2, \dots$, де $X_n \sim N\left(\frac{1}{n}, \frac{1}{n}\right)$, $n \in \mathbb{N}$. До чого прямує така послідовність за ймовірністю? □

Розв'язання. Інтуїція підказує, що така послідовність повинна прямувати до нуля, оскільки і сподівання, і дисперсія прямують до нуля. Справді, для будь-якого $\varepsilon > 0$ маємо

$$\begin{aligned}\mathbb{P}_{X_n} (|X_n| > \varepsilon) &= 1 - \mathbb{P}_{X_n} (|X_n| \leq \varepsilon) = 1 - \mathbb{P}_{X_n} \left(\left| \frac{X_n - \frac{1}{n}}{\frac{1}{\sqrt{n}}} \right| \leq \frac{\varepsilon - \frac{1}{n}}{\frac{1}{\sqrt{n}}} \right) \\ &= 1 - \left(\Phi \left(\sqrt{n} \left(\varepsilon - \frac{1}{n} \right) \right) \right) - \Phi \left(-\sqrt{n} \left(\varepsilon - \frac{1}{n} \right) \right) ,\end{aligned}$$

що прямує до $1 - (1 - 0) = 0$, коли $n \rightarrow \infty$, тобто $X_n \xrightarrow{p} 0$. □

17 Твірні функції моментів

17.1 Основні теоретичні відомості

Детальний огляд понять, потрібний для розв'язання задач на цьому занятті, наведено в Курсі лекцій, Розд. КЛ15.

Визначення (КЛ15.1.1). *Твірною функцією моментів* (moment generating function) деякої випадкової величини X є

$$M_X(t) = \mathbb{E} [e^{tX}] = \int e^{tX} d\mathbb{P}, \quad (\text{КЛ15.1.1})$$

яку визначено в усіх точках t , де $M_X(t) < \infty$. □

Аргумент t у цьому визначенні не має самостійного значення.

Якщо з контексту зрозуміло, про яку саме величину X мова, то відповідний індекс опускаємо. Можна помітити, що твірна функція моментів для різних випадкових величин з однаковим розподілом однакова, оскільки це сподівання, яке для величин $X \stackrel{d}{=} Y$ однакове. 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 перетворення випадкової величини.

Твердження (КЛ15.1.2). 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 $M_Y(t) = e^{tb} M_X(at)$, якщо M_X визначена в точці at .

Апарат твірних функцій моментів дає змогу суттєво спростити обчислення розподілів сум незалежних випадкових величин.

Твердження (КЛ15.1.3). Нехай X та Y — дві незалежні випадкові величини. Тоді в усіх точках t , де відповідні твірні функції моментів визначені, маємо

$$M_{X+Y}(t) = M_X(t) M_Y(t). \quad (\text{КЛ15.1.2})$$

Цілком очевидно, що за індукцією можна довести відповідний результат для суми n незалежних величин:

$$M_{X_1+\dots+X_n}(t) = M_{X_1}(t) \dots M_{X_n}(t).$$

Можна обчислити твірні функції моментів для окремих відомих розподілів.

Твердження 17.1.1 (КЛ15.1.4). 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 $t \in \left(-\infty; \ln \frac{1}{1-p}\right)$;
- (iv) якщо $X \sim \text{Pois}(\lambda)$, то $M_X(t) = e^{\lambda(e^t - 1)}$ для всіх $t \in \mathbb{R}$;
- (v) якщо $X \sim U((a; b])$, 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 $t < \lambda$.

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.

Теорема (КЛ15.2.1). 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

$$M_X^{(k)}(0) = \mathbb{E}[X^k]. \quad (\text{КЛ15.2.1})$$

Теорема КЛ15.2.1 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.

Проте потрібно розуміти, що Теорема КЛ15.2.1 дає тільки достатні, але не необхідні умови існування моментів.

У Розд. КЛ11.1.4 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 (Оцінка Черноффа¹ (Chernoff bound), КЛ15.3.1). Нехай випадкова величина X має твірну функцію моментів M_X . Тоді для будь-якого $t \in \mathbb{R}$ справедливо

$$\mathbb{P}_X(X \geq t) \leq \min_{s>0} e^{-st} M_X(s). \quad (\text{КЛ15.3.1})$$

Виявляється, якщо виконуються певні умови, твірна функція моментів однозначно відповідає деякому розподілу. Приймімо спочатку без доведення таку теорему.

Теорема (КЛ15.4.1). 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 Теорема КЛ15.4.1, має назву *розподілу, який визначають його моменти* (distribution determined by its moments).

Теорема (КЛ15.4.3). 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 $X \stackrel{d}{=} Y$.

17.2 Вправи

Вправа 17.2.1. Правильну гральну кісточку викидають двічі. Нехай результати обох викидів позначено через X та Y відповідно. Чому дорівнює твірна функція моментів випадкової величини $X + Y$? Використайте її для обчислення сподівання $X + Y$. $\square$

Розв'язання. Очевидно, що $X \perp Y$, тому за Твердженням КЛ15.1.3

$$M_{X+Y}(t) = M_X(t)M_Y(t) = \mathbb{E}[e^{tX}] \mathbb{E}[e^{tY}].$$

У свою чергу, за законом несвідомого статистика з урахування рівномірного розподілу результатів випадку гральної кісточки маємо

$$\mathbb{E}[e^{tX}] = \frac{1}{6} \sum_{k=1}^6 e^{kt},$$

¹Герман Чернофф (Herman Chernoff, нар. 1923) — американський математик.

звідки остаточно

$$M_{X+Y}(t) = \left(\frac{1}{6} \sum_{k=1}^6 e^{kt} \right)^2 .$$

Оскільки така твірна функція скінченна для всіх $t \in \mathbb{R}$, можемо обчислити моменти з Теорема КЛ15.2.1:

$$\begin{aligned} \mathbb{E}[X + Y] &= \left(\left(\frac{1}{6} \sum_{k=1}^6 e^{kt} \right)^2 \right)' \Big|_{t=0} \\ &= \frac{2}{6} \left(\sum_{k=1}^6 e^{kt} \right) \cdot \left(\frac{1}{6} \sum_{k=1}^6 k e^{kt} \right) \Big|_{t=0} = \frac{1}{18} \left(\sum_{k=1}^6 1 \right) \cdot \left(\sum_{k=1}^6 k \right) = 7 . \end{aligned}$$

Це відповідає нашим очікуванням від лінійності сподівання:

$$\mathbb{E}[X + Y] = \mathbb{E}[X] + \mathbb{E}[Y] = 3.5 + 3.5 = 7 .$$

□

Вправа 17.2.2. Нехай твірна функція моментів деякої випадкової величини X дорівнює

$$M_X(t) = e^{t^2+3t} .$$

Чому дорівнюють сподівання та дисперсія X ?

□

Розв'язання. Оскільки $M_X(t) < \infty$ для всіх $t \in \mathbb{R}$, тобто в тому числі в деякому відкритому околі 0, до неї застосовна Теорема КЛ15.2.1:

$$\begin{aligned} \mathbb{E}[X] &= \left(e^{t^2+3t} \right)' \Big|_{t=0} = e^{t^2+3t} (2t + 3) \Big|_{t=0} = 3 , \\ \mathbb{E}[X^2] &= \left(e^{t^2+3t} \right)'' \Big|_{t=0} = 2e^{t^2+3t} (2t + 3) + e^{t^2+3t} (2t + 3)^2 \Big|_{t=0} = 11 , \\ \text{Var}(X) &= \mathbb{E}[X^2] - (\mathbb{E}[X])^2 = 11 - 9 = 2 . \end{aligned}$$

Цей результат не є дивним, адже ця твірна функція моментів відповідає розподілу $N(3, 2)$.

□

Вправа 17.2.3. Нехай твірна функція моментів деяких незалежних випадкових величин X та Y дорівнює

$$M_X(t) = M_Y(t) = e^{t^2+3t} .$$

Чому дорівнює твірна функція моментів величини $Z = 2X - 3Y + 4$?

□

Розв'язання. З урахуванням Тверджень КЛ15.1.2–КЛ15.1.3 маємо

$$\begin{aligned} M_Z(t) &= e^{4t} M_{2X-3Y}(t) = e^{4t} M_{2X}(t) M_{-3Y}(t) \\ &= e^{4t} e^{t \cdot 0} M_X(2t) e^{t \cdot 0} M_Y(-3t) \\ &= e^{4t} e^{4t^2+6t} e^{9t^2-9t} = e^{13t^2+t}. \end{aligned}$$

Оскільки твірні функції моментів M_X та M_Y відповідають розподілу $N(3, 2)$, цей результат є цілком очікуваним. Справді, із виразу M_Z робимо висновок, що $Z \sim N(1, 26)$, але саме такі параметри повинна мати й величина $2X - 3Y + 4$ як лінійна комбінація незалежних нормальних величин:

$$2X - 3Y + 4 \sim N(2 \cdot 3 - 3 \cdot 3 + 4, 2^2 \cdot 2 + (-3)^2 \cdot 2) = N(1, 26). \quad \square$$

Вправа 17.2.4. Нехай $X, Y \sim \text{Pois}(\lambda)$, $X \perp Y$. Чи має величина $X + 2Y$ розподіл Пуассона? $\square$

Розв'язання. Через незалежність величин візьмімо Твердження КЛ15.1.3:

$$M_{X+2Y}(t) = M_X(t) M_{2Y}(t).$$

Із Твердження КЛ15.1.4 відомо, що

$$M_X(t) = e^{\lambda(e^t-1)}.$$

З урахуванням Твердження КЛ15.1.2 маємо

$$M_{2Y}(t) = e^{t \cdot 0} M_Y(2t) = e^{\lambda(e^{2t}-1)}.$$

Отже

$$M_{X+2Y}(t) = e^{\lambda(e^t-1)} \cdot e^{\lambda(e^{2t}-1)} = e^{\lambda(e^t+e^{2t}-2)}.$$

Твірна функція моментів для розподілу Пуассона завжди скінченна, тому за Теоремою КЛ15.4.3 її твірна функція моментів може відповідати тільки розподілу Пуассона. Твірна функція моментів, яку ми щойно дістали, не вкладається у функційну форму твірних функцій моментів для розподілу Пуассона, тому робимо висновок, що $X + 2Y$ не має розподілу Пуассона. $\square$

Вправа 17.2.5. Нехай твірна функція моментів деякої випадкової величини X дорівнює

$$M_X(t) = \frac{1}{5}e^t + \frac{2}{5}e^{4t} + \frac{2}{5}e^{8t}.$$

Який розподіл має X ? $\square$

Розв'язання. Якщо ми маємо три різні доданки у твірній функції моментів, усі з яких є експонентами, випадкова величина X повинна бути дискретною, а її носій повинен складатися з трьох елементів. Справді, нехай $\text{supp}(X) = \{x_1, x_2, \dots, x_n\}$, а функція ймовірності дорівнює $p_X(x_i) = p_i$, $i = 1, \dots, n$. Тоді

$$M_X(t) = \mathbb{E}[e^{tX}] = \sum_{k=1}^n e^{kt} p_X(x_k) = \sum_{k=1}^n p_k e^{kt}.$$

У нашому випадку присутні тільки експоненти для $k = 1, 4, 8$. Отже бачимо, що X має дискретний розподіл, а її функція ймовірності дорівнює

$$p_X(x) = \begin{cases} \frac{1}{5}, & x = 1 \\ \frac{2}{5}, & x = 4 \\ \frac{2}{5}, & x = 8 \end{cases}.$$

□

Вправа 17.2.6. Нехай твірна функція моментів деякої випадкової величини X дорівнює

$$M_X(t) = \frac{1}{6} (4 + e^t + e^{-t}).$$

Який розподіл має X ?

□

Розв'язання. За аналогією з попереднім випадком, X є дискретною випадковою величиною з функцією ймовірності

$$p_X(x) = \begin{cases} \frac{1}{6}, & x = -1 \\ \frac{4}{6}, & x = 0 \\ \frac{1}{6}, & x = 1 \end{cases}.$$

□

Вправа 17.2.7. Нехай випадкова величина X має щільність

$$f(x) = x^{-2} \cdot \mathbb{1}\{x > 1\}.$$

На якому проміжку скінченна її твірна функція моментів M_X ? Що можна сказати про існування моментів X ?

□

Розв'язання. Твірна функція моментів X дорівнює

$$M_X(t) = \mathbb{E}[e^{tX}] = \int_1^\infty \frac{e^{tx}}{x^2} dx.$$

Цей інтеграл розбіжний для всіх $t > 0$, але збіжний для всіх $t \leq 0$.

Теорема КЛ15.2.1 стверджує, що якби M_X була скінченна в околі 0, то існували б моменти всіх порядків. Наша твірна функція моментів не є скінченна в околі 0, проте для цього випадку теорема нічого не каже (згадаймо, що логнормальний розподіл має всі моменти, проте його твірна функція моментів також скінченна тільки для $t \leq 0$).

Проте в нашому випадку моменти справді не існують:

$$\mathbb{E}[X^n] = \int_1^\infty x^{n-2} dx = \infty$$

для всіх $n \in \mathbb{N}$. □

Вправа 17.2.8. Чому дорівнює твірна функція моментів для випадкової величини $X \sim \chi_k^2$? □

Розв'язання. Згідно з Прикладом КЛ10.2.5, $X = \sum_{i=1}^k Z_i^2$, де $Z_i \sim N(0, 1)$, $i = 1, \dots, k$, і всі величини незалежні.

Тоді за Твердженням КЛ15.1.3 маємо

$$M_X(t) = \prod_{i=1}^k M_{Z_i^2}(t).$$

Твірну функцію моментів для Z_i^2 знайдемо за визначенням:

$$\begin{aligned} M_{Z_i^2}(t) &= \mathbb{E}[e^{tZ_i^2}] = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{tx^2} e^{-\frac{x^2}{2}} dx \\ &= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-\frac{x^2}{2} \frac{1}{1-2t}} dx \\ &= \frac{1}{\sqrt{1-2t}} \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi} \frac{1}{\sqrt{1-2t}}} e^{-\frac{x^2}{2} \frac{1}{1-2t}} dx = \frac{1}{\sqrt{1-2t}}, \end{aligned}$$

оскільки під інтегралом стоїть щільність випадкової величини з розподілом $N(0, \frac{1}{1-2t})$, і тому інтеграл дорівнює 1.

Важливо розуміти, що відповідний інтеграл буде збіжний тільки тоді, коли $1 - 2t > 0$, тобто коли $t < \frac{1}{2}$. На інтуїтивному рівні це очікувано, бо дисперсія $\frac{1}{1-2t}$ повинна бути додатною. Строго формально, якщо $t \geq \frac{1}{2}$, то

показник експоненти під інтегралом буде додатний, а отже інтеграл буде розбіжний. Остаточно маємо

$$M_X(t) = \frac{1}{(1 - 2t)^{\frac{k}{2}}}.$$

Оскільки твірна функція моментів визначена для $t < \frac{1}{2}$, вона в тому числі визначена у відкритому околі 0, наприклад, $t \in (-\frac{1}{2}; \frac{1}{2})$. Отже існують моменти довільних порядків. $\square$

Вправа 17.2.9. Нехай $X_1, \dots, X_n \sim \text{Bern}(0.5)$, і всі величини незалежні. Чому дорівнює оцінка Черноффа ймовірності, що середнє вибіркове не відхилятиметься від свого сподівання більше від 0.1? Порівняйте з оцінкою з нерівності Чебишова. $\square$

Розв'язання. Оскільки $\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i$, а $\sum_{i=1}^n X_i \sim \text{Binom}(n, 0.5)$, сподівання середнього вибіркового дорівнює 0.5. Отже нас цікавить ймовірність

$$\begin{aligned} \mathbb{P}_{\bar{X}} \left(\left| \bar{X} - \frac{1}{2} \right| \geq \frac{1}{10} \right) &= \mathbb{P}_{\bar{X}} \left(\bar{X} - \frac{1}{2} \geq \frac{1}{10} \right) + \mathbb{P}_{\bar{X}} \left(\bar{X} - \frac{1}{2} \leq -\frac{1}{10} \right) \\ &= \mathbb{P}_{\bar{X}} \left(\bar{X} \geq \frac{1}{2} + \frac{1}{10} \right) + \mathbb{P}_{\bar{X}} \left(\bar{X} \leq \frac{1}{2} - \frac{1}{10} \right) \\ &= \mathbb{P}_{\bar{X}} \left(\sum_{i=1}^n X_i \geq \frac{3n}{5} \right) + \mathbb{P}_{\bar{X}} \left(-\sum_{i=1}^n X_i \geq -\frac{2n}{5} \right). \end{aligned}$$

Для використання оцінки Черноффа нам потрібно знати твірну функцію моментів випадкової величини, ймовірність якої оцінюємо. Для біномної величини твірна функція моментів дорівнює

$$M(t) = \left(1 + \frac{1}{2} (e^t - 1) \right)^n = \left(\frac{1}{2} (e^t + 1) \right)^n.$$

За нерівністю Черноффа першу ймовірність можна оцінити так:

$$\mathbb{P}_{\bar{X}} \left(\sum_{i=1}^n X_i \geq \frac{3n}{5} \right) \leq \min_{s>0} e^{-\frac{3ns}{5}} \left(\frac{1}{2} (e^s + 1) \right)^n.$$

Спочатку прологаритмуємо вираз справа, оскільки тоді буде простіше шукати умову першого порядку:

$$\ln \left(e^{-\frac{3ns}{5}} \left(\frac{1}{2} (e^s + 1) \right)^n \right) = n \left(\ln \frac{1}{2} + \ln(e^s + 1) - \frac{3s}{5} \right).$$

Умова першого порядку має вид

$$n \left(\frac{e^s}{e^s + 1} - \frac{3}{5} \right) = 0 ,$$

звідки $s = \ln \frac{3}{2}$. Перевірте, що друга похідна в цій точці буде додатна, отже це справді точка мінімуму. Підставивши в нерівність Черноффа, дістаємо

$$\mathbb{P}_{\bar{X}} \left(\sum_{i=1}^n X_i \geq \frac{3n}{5} \right) \leq e^{-\frac{3n \ln \frac{3}{2}}{5}} \left(\frac{1}{2} \left(e^{\ln \frac{3}{2}} + 1 \right) \right)^n \approx (0.98)^n .$$

Другу ймовірність можна оцінити, помітивши, що для $-\sum_{i=1}^n X_i$ твірну функцію моментів потрібно взяти з аргументом $-t$:

$$\mathbb{P}_{\bar{X}} \left(-\sum_{i=1}^n X_i \geq -\frac{2n}{5} \right) \leq \min_{s>0} e^{\frac{2ns}{5}} \left(\frac{1}{2} (e^{-s} + 1) \right)^n .$$

Логаритмуючи, дістаємо

$$\ln \left(e^{\frac{2ns}{5}} \left(\frac{1}{2} (e^{-s} + 1) \right)^n \right) = n \left(\ln \frac{1}{2} + \ln(e^{-s} + 1) + \frac{2s}{5} \right) .$$

Умова першого порядку має вид

$$n \left(\frac{-e^{-s}}{e^{-s} + 1} + \frac{2}{5} \right) = 0 ,$$

звідки $s = \ln \frac{3}{2}$. Перевірте, що друга похідна в цій точці буде додатна, отже це справді точка мінімуму. Підставивши в нерівність Черноффа, дістаємо

$$\mathbb{P}_{\bar{X}} \left(-\sum_{i=1}^n X_i \geq -\frac{2n}{5} \right) \leq e^{\frac{2n \ln \frac{3}{2}}{5}} \left(\frac{1}{2} \left(e^{-\ln \frac{3}{2}} + 1 \right) \right)^n \approx (0.98)^n .$$

Отже сумарна оцінка дорівнює

$$\mathbb{P}_{\bar{X}} \left(\left| \bar{X} - \frac{1}{2} \right| \geq \frac{1}{10} \right) \leq (0.98)^n + (0.98)^n = 2(0.98)^n .$$

Оцінка з нерівності Чебишова натомість дорівнює

$$\mathbb{P}_{\bar{X}} \left(\left| \bar{X} - \frac{1}{2} \right| \geq \frac{1}{10} \right) \leq \frac{\text{Var}(\bar{X})}{1/10^2} = \frac{100}{4n} = \frac{25}{n} ,$$

тобто оцінка Черноффа *значно* точніша.

□

Вправа 17.2.10. До цього моменту ми розглядали твірні функції моментів окремих величин. У цілком схожий спосіб можна розглянути твірні функції моментів випадкових векторів. Нехай $\mathbf{X} = (X_1, \dots, X_n)^\top$ має деякий спільний розподіл. Тоді його *спільною твірною функцією моментів* (joint moment generating function) є

$$M_{\mathbf{X}}(t_1, \dots, t_n) = \mathbb{E} [e^{t_1 X_1 + \dots + t_n X_n}] . \quad (17.2.1)$$

Тоді твірна функція окремої координати X_i , $i = 1, \dots, n$, дорівнює

$$M_{X_i}(t) = M(0, \dots, 0, t, 0, \dots, 0) = \mathbb{E} [e^{0 + \dots + 0 + t X_i + 0 + \dots + 0}] = \mathbb{E} [e^{t X_i}] .$$

Вважаємо, що $M_{\mathbf{X}}$ визначена, якщо відповідне сподівання скінченне.

Можна показати², що якщо твірна функція моментів визначена на деякому околі нуля, то вона однозначно задає відповідний (спільний) розподіл. Використайте цей результат для доведення такого твердження.

Випадкові величини $X_1, \dots, X_n$ незалежні тоді й тільки тоді, коли

$$M_{\mathbf{X}}(t_1, \dots, t_n) = M_{X_1}(t_1) \dots M_{X_n}(t_n) , \quad (17.2.2)$$

якщо твірні функції моментів визначені на деякому околі нуля кожна. $\square$

Розв'язання. Нехай величини $X_1, \dots, X_n$ незалежні. Тоді

$$\begin{aligned} M_{\mathbf{X}}(t_1, \dots, t_n) &= \mathbb{E} [e^{t_1 X_1 + \dots + t_n X_n}] \\ &= \mathbb{E} [e^{t_1 X_1} \dots e^{t_n X_n}] \\ &= \mathbb{E} [e^{t_1 X_1}] \dots \mathbb{E} [e^{t_n X_n}] \\ &= M_{X_1}(t_1) \dots M_{X_n}(t_n) . \end{aligned}$$

Нехай тепер виконується (17.2.2). Тоді твірна функція моментів $M_{\mathbf{X}}$ дорівнює добутку твірних функцій моментів окремих величин M_{Y_i} , $i = 1, \dots, n$, де $Y_i \stackrel{d}{=} X_i$. Оскільки спільна твірна функція моментів однозначно задає спільний розподіл, виходить, що Y_i , $i = 1, \dots, n$, мають спільний розподіл, який задає $M_{\mathbf{X}}$. А отже вони між собою незалежні, як і X_i . $\square$

Вправа 17.2.11. Нехай $X \perp\!\!\!\perp Y$ — дві величини з однаковим нормальним розподілом $N(\mu, \sigma^2)$. Покажіть, використовуючи апарат твірних функцій моментів, що $X + Y \perp\!\!\!\perp X - Y$. $\square$

²Проте це виходить за рамки нашого курсу.

Розв'язання. Нескладно бачити, що

$$\begin{aligned} M_{X+Y, X-Y}(s, t) &= \mathbb{E} \left[e^{t(X+Y)+s(X-Y)} \right] = \mathbb{E} \left[e^{(t+s)X+(t-s)Y} \right] \\ &= \mathbb{E} \left[e^{(t+s)X} \right] \mathbb{E} \left[e^{(t-s)Y} \right] = M_X(t+s)M_Y(t-s), \end{aligned}$$

де ми використали незалежність $X \perp Y$. Згідно з Твердженням КЛ15.1.4,

$$M_{X+Y, X-Y}(s, t) = e^{\mu(t+s)+\sigma^2 \frac{(t+s)^2}{2}} e^{\mu(t-s)+\sigma^2 \frac{(t-s)^2}{2}} = e^{2\mu t + \sigma^2 t^2} e^{\sigma^2 s^2}.$$

Цілком очевидно, що ми маємо добуток двох твірних функцій моментів, для випадкових величин $N(2\mu, 2\sigma^2)$ і $N(0, 2\sigma^2)$ відповідно. Але оскільки це повинна бути спільна твірна функція моментів деякого розподілу, і оскільки спільна твірна функція моментів однозначно задає розподіл (вона скінченна на всій площині), бачимо, що випадкові величини $X+Y \perp X-Y$, адже їхня спільна твірна функція моментів дорівнює добутку двох окремих твірних функцій моментів. Понад те, робимо висновок, що $X+Y \sim N(2\mu, 2\sigma^2)$, $X-Y \sim N(0, 2\sigma^2)$. $\square$

Вправа 17.2.12. Нехай число подій X має розподіл Пуассона з параметром λ . Нехай кожному з цих подій можуть, незалежно від інших, у деякий спосіб помітити з імовірністю p . Покажіть, що число помічених подій і число непомічених подій є незалежними пуассонівськими величинами зі сподіваннями λp і $\lambda(1-p)$. $\square$

Розв'язання. Позначмо число помічених подій через X_c . Нас цікавить спільна твірна функція моментів для величин X_c і $X - X_c$ (число непомічених подій). Якщо ми знаємо загальне число подій n , то вочевидь маємо ситуацію зі схемою Бернуллі, у якій успіхом є помічена подія. Тоді $X_c \mid X = n \sim \text{Binom}(n, p)$ і можна обчислити спільну твірну функцію моментів за допомогою закону ітерованих сподівань:

$$\begin{aligned} M_{X_c, X-X_c}(s, t) &= \mathbb{E} \left[e^{sX_c+t(X-X_c)} \right] \\ &= \mathbb{E} \left[\mathbb{E} \left[e^{sX_c+t(X-X_c)} \mid X = n \right] \right] \\ &= \mathbb{E} \left[e^{tX} \mathbb{E} \left[e^{(s-t)X_c} \mid X = n \right] \right]. \end{aligned}$$

З урахуванням виразу для твірної функції моментів для біномного розподілу $\text{Binom}(n, p)$ з Твердження КЛ15.1.4, маємо

$$M_{X_c, X-X_c}(s, t) = \mathbb{E} \left[e^{tX} (1 + p(e^{s-t} - 1))^X \right] = \mathbb{E} \left[(pe^s + (1-p)e^t)^X \right].$$

Оскільки $X \sim \text{Pois}(\lambda)$, відповідно до Твердження КЛ15.1.4 маємо

$$\mathbb{E} \left[(pe^s + (1-p)e^t)^X \right] \equiv \mathbb{E} \left[e^{aX} \right] = M_X(a) = e^{\lambda(e^a - 1)},$$

отже

$$M_{X_c, X-X_c}(s, t) = e^{\lambda((pe^s + (1-p)e^t) - 1)} = e^{\lambda p(e^s - 1)} e^{\lambda(1-p)(e^t - 1)}.$$

Як і у Вправі 17.2.11, робимо висновок, що $X_c \sim \text{Pois}(p\lambda)$, $X - X_c \sim \text{Pois}((1-p)\lambda)$, і що $X_c \perp\!\!\!\perp X - X_c$. $\square$

Література

- [1] J. K. Blitzstein and J. Hwang, *Introduction to Probability*, 2nd ed. Boca Raton: Taylor & Francis Group, LLC, 2019, 636 pp.
- [2] M. D. DiSorbo. “Probability!” (2022), [Online]. Available: <https://bookdown.org/probability> (visited on 09/04/2022).
- [3] S. S. Venkatesh, *The Theory of Probability: Explorations and Applications*. Cambridge: Cambridge University Press, 2013, 832 pp.
- [4] S. Ross, *A First Course in Probability*, 10th ed. Harlow: Pearson, 2020, 530 pp.
- [5] A. Johnson, *Geometric Probability*. Lexington: COMAP, Inc., 1995, 60 pp.
- [6] G. Casella and R. Berger, *Statistical Inference*, English, 2nd ed. Pacific Grove: Duxbury Resource Center, 2002, 686 pp.
- [7] M. H. DeGroot and M. J. Schervish, *Probability and statistics*, 4th ed. London: Pearson, 2014, 911 pp.
- [8] J. Pitman, *Probability*. New York: Springer-Verlag, 1993, 564 pp.

Предметний показник

А

абсолютна неперервність відносно міри, 121
 абсолютні моменти, 253
 абсолютні центральні моменти, 253
 абсолютно неперервна випадкова величина, 121
 σ -адитивна міра, 46
 алгебра, 26
 σ -алгебра, 27
 σ -алгебра, яку породжує випадковий вектор, 295
 апостеріорна ймовірність, 65
 апріорна ймовірність, 65
 асоціативність, 8

Б

багатовимірний нормальний розподіл, 227
 Беєсівське виведення, 162
 Бета-розподіл, 162
 біномна теорема, 17
 біномний коефіцієнт, 16
 біномний розподіл, 90
 Борелева σ -алгебра, 28
 Борелева дійсна вісь, 29
 Борелева множина, 28
 Борелів простір, 29

В

вартість під ризиком, 152
 верхня границя послідовности подій, 314
 вибір, 12
 вибір без збереження порядку, 12
 вибір без повторень, 12

вибір зі збереженням порядку, 12
 вибір із повтореннями, 12
 вимірна множина, 27
 вимірна функція, 85
 вимірний простір, 27
 випадкова величина, 85
 випадковий експеримент, 8
 випробування Бернуллі, 89
 від'ємний біномний розподіл, 106
 власна інформація, 256

Г

Гамма-розподіл, 225
 геометрична ймовірність, 10
 геометричний розподіл, 104
 гіпергеометричний розподіл, 106

Д

де Моргана закони, 9
 дискретна випадкова величина, 87
 дисперсія, 88
 дистрибутивність, 9
 добуток мір, 185
 добуток просторів, 185
 доповнення множини, 9
 друга лема Бореля-Кантеллі, 315

Е

експоненційний розподіл, 165
 елементарна подія, 8
 ентропія, 256

З

задача Бертрана про вибори, 32
 задача зіставлень де Монморта, 52
 задача Монті Голла, 72
 задача Ньютона-Піпса, 18

задача про голку Бюффона, 37
 задача про дні народження, 13
 задача про зустріч, 39
 задача про колекціонера купонів, 114
 задача про розорення гравця, 75
 закон ітерованих сподіван, 298
 Закон несвідомого статистика, 88
 закон несвідомого статистика, 124
 закон повного сподівання, 297
 закон повної дисперсії, 299
 закон повної ймовірності, 66
 збіжність з ймовірністю 1, 316
 збіжність за ймовірністю, 316
 збіжність майже напевно, 316
 збіжність у середньому порядку p , 317
 згортка, 211

I

імовірнісна міра, 47
 індекс Джині, 270
 інтегральне перетворення ймовірностей, 161

K

квантиль, 255
 квартиль, 255
 класична ймовірність, 9
 класичне визначення ймовірності, 9
 коваріація, 190
 коефіцієнт асиметрії, 254
 коефіцієнт ексцесу, 255
 коефіцієнт кореляції, 191
 комбінаторна тотожність Ферма, 42
 комутативність, 8
 крива Лоренца, 270

L

логістичний розподіл, 129

M

маржинальна функція розподілу, 189

маржинальна щільність розподілу, 189
 маржинальний розподіл, 188
 матриця коваріацій, 192
 медіана, 252
 метод відбиття, 32
 міра, 46
 міри центральної тенденції, 251
 мода, 252
 моменти, 253
 монотонно неспадна міра, 47
 мультиномний розподіл, 226

N

Названий так на честь французького математика П'єра-Симона де Лапласа (Pierre-Simon, marquis de Laplace, 1749–1827)., 222
 незалежні події, 67
 незалежність випадкових величин, 192
 некорельовані випадкові величини, 194
 неможлива подія, 9
 неперервна міра, 48
 неперервність зверху, 48
 неперервність знизу, 48
 неперервність у порожній множині, 48
 нерівність Була, 48
 нерівність Єнсена, 250
 нерівність Маркова, 250
 нерівність Чебишова, 251
 несумісні події, 9
 нижня границя послідовності подій, 314
 нормальний розподіл, 146
 носій випадкової величини, 87, 123

O

об'єднання множин, 9
 образ міри, 86
 одностороння нерівність Чебишова, 259

опукла функція, 249
оцінка Чернова, 332

П

певна подія, 9
перестановка, 13
перетворення Бокса-Мюллера, 235
перетворення зсуву та
масштабування, 143
перетин множин, 9
персентиль, 256
перша лема Бореля-Кантеллі, 315
повторний інтеграл, 186
подвійний інтеграл, 186
подія, 8
породжена алгебра, 28
породжена σ -алгебра, 28
порожня множина, 9
порядкові статистики, 163
посилений закон великих чисел,
318
похідна Радона-Никодима, 121
правило множення, 11
приріст інформації, 258
прокляття переможця, 301
прообраз, 85
простір елементарних подій, 8
процес Пуассона, 166
Пуассонівська парадигма, 104

Р

рівність за розподілом, 86
рівномірний розподіл, 89, 144
різниця множин, 9
розбиття, 8
розподіл без пам'яті, 105
розподіл Бернуллі, 89
розподіл випадкової величини, 86
розподіл Лапласа, 222
розподіл Парето, 168
розподіл Пуассона, 103
розподіл Рейлі, 130
розподіл, який визначають його
моменти, 332

розходження Кульбака-Ляйблера,
257

С

середня абсолютна похибка, 279
середньоквадратичне відхилення,
88
симетричний розподіл, 254
сім'я розподілів відносно
зсуву-масштабування, 143
скінченна міра, 46
 σ -скінченна міра, 47
скінченно-адитивна міра, 47
слабкий закон великих чисел, 318
спільна твірна функція моментів,
339
спільна функція ймовірності, 188
спільна функція розподілу, 187
спільна щільність розподілу, 187
спільний розподіл, 186
сподівання, 87, 123, 189
спряжений апріорний розподіл,
163, 237, 291
стандартний нормальний
розподіл, 145
стандартний рівномірний
розподіл, 144
стандартний розподіл Коші, 127
 σ -субадитивна міра, 47
схема Бернуллі, 90

Т

твірна функція моментів, 330
теорема Беєса, 66
теорема Каратеодорі, 48
теорема про добуток мір, 185
теорема про неперервне
відображення, 317
теорема Радона-Никодима, 121
теорема Фубіні, 185
теорія інформації, 256

У

умовна дисперсія, 279
умовна ймовірність, 65, 295

умовна медіана, 279
 умовна функція ймовірности, 108, 276
 умовна функція розподілу, 276, 278
 умовна щільність розподілу, 277
 умовне сподівання, 296, 297
 умовно незалежні події, 67
 універсальна множина, 9

Ф

формула заміни змінних, 142
 формула заміни змінних для випадкових векторів, 224
 функція від множини, 46
 функція ймовірности, 87
 функція розподілу, 119

Ц

центральні моменти, 253

Щ

щільність розподілу, 122

А

absolute central moments, 253
 absolute moments, 253
 absolutely continuity with respect to measure, 121
 absolutely continuous random variable, 121
 σ -additive measure, 46
 algebra, 26
 σ -algebra, 27
 σ -algebra generated by a random vector, 295
 associativity, 8

В

Bayes' Theorem, 66
 Bayesian inference, 162
 Bernoulli distribution, 89
 Bernoulli scheme, 90
 Bernoulli trial, 89
 Bertrand's ballot problem, 32
 Beta distribution, 162

binomial coefficient, 16
 Binomial distribution, 90
 binomial theorem, 17
 birthday problem, 13
 Boole's inequality, 48
 Borel σ -algebra, 28
 Borel real line, 29
 Borel set, 28
 Borel space, 29
 Box-Muller transform, 235
 Buffon's needle problem, 37

С

Carathéodory theorem, 48
 central moments, 253
 certain event, 9
 change of variables formula, 142
 change of variables formula for random vectors, 224
 Chebyshev's inequality, 251
 Chernoff bound, 332
 classical probability, 9
 commutativity, 8
 conditional distribution function, 276, 278
 conditional expectation, 296, 297
 conditional median, 279
 conditional probability, 65, 295
 conditional probability density, 277
 conditional probability mass function, 108, 276
 conditional variance, 279
 conditionally independent events, 67
 conjugate prior, 163, 237, 291
 continuous mapping theorem, 317
 continuity at the empty set, 48
 continuity from above, 48
 continuity from below, 48
 continuous measure, 48
 convergence almost surely, 316
 convergence in probability, 316
 convergence in p th mean, 317
 convergence with probability 1, 316
 convex function, 249

convolution, 211
 correlation, 191
 coupon collector's problem, 114
 covariance, 190
 covariance matrix, 192

D

de Montmort's matching problem, 52
 de Morgan's laws, 9
 discrete random variable, 87
 distributivity, 9
 distribution determined by its moments, 332
 distribution function, 119
 distribution of a random variable, 86
 double integral, 186

E

empty set, 9
 entropy, 256
 equality in distribution, 86
 expectation, 87, 123, 189
 exponential distribution, 165

F

Fermat's combinatorial identity, 42
 finite measure, 46
 σ -finite measure, 47
 finitely additive measure, 47
 first Borel-Cantelli lemma, 315
 Fubini theorem, 185

G

gambler's ruin problem, 75
 Gamma distribution, 225
 generated algebra, 28
 generated σ -algebra, 28
 geometric distribution, 104
 geometric probability, 10
 Gini index, 270

H

hypergeometric distribution, 106

I

impossible event, 9
 independent events, 67
 independent random variables, 192
 induced measure, 86
 information content, 256
 information gain, 258
 information theory, 256
 inverse image, 85
 iterated integral, 186

J

Jensen's inequality, 250
 joint distribution, 186
 joint distribution function, 187
 joint moment generating function, 339
 joint probability density function, 187
 joint probability mass function, 188

K

Kullback-Leibler divergence, 257
 kurtosis, 255

L

Laplace distribution, 222
 law of iterated expectations, 298
 law of total expectation, 297
 Law of total probability, 66
 law of total variance, 299
 Law of unconscious statistician, 88
 law of unconscious statistician, 124
 limit inferior of a sequence of events, 314
 limit superior of a sequence of events, 314
 location-scale distribution family, 143
 location-scale transformation, 143
 logistic distribution, 129
 Lorenz curve, 270

M

marginal distribution, 188

marginal distribution function, 189
 marginal probability density
 function, 189
 Markov's inequality, 250
 mean absolute error, 279
 measurable function, 85
 measurable set, 27
 measurable space, 27
 measure, 46
 measures of central tendency, 251
 median, 252
 meeting problem, 39
 memoryless distribution, 105
 mode, 252
 moment generating function, 330
 moments, 253
 monotonically nondecreasing, 47
 Monty Hall problem, 72
 multinomial distribution, 226
 multiplication rule, 11
 multivariate normal distribution,
 227
 mutually exclusive events, 9

N

naive probability, 9
 negative Binomial distribution, 106
 Newton-Pepys problem, 18
 normal distribution, 146

O

one-sided Chebyshev inequality, 259
 order statistics, 163

P

Pareto distribution, 168
 partition, 8
 percentile, 256
 permutation, 13
 Poisson distribution, 103
 Poisson paradigm, 104
 Poisson process, 166
 posterior probability, 65
 prior probability, 65
 probability density function, 122

probability integral transform, 161
 probability mass function, 87
 probability measure, 47
 product measure, 185
 product measure theorem, 185
 product space, 185

Q

quantile, 255
 quartile, 255

R

Radon-Nikodym derivative, 121
 Radon-Nikodym theorem, 121
 random experiment, 8
 random variable, 85
 Rayleigh distribution, 130
 reflection method, 32

S

sample, 8
 sample space, 8
 sampling, 12
 sampling with replacement, 12
 sampling without replacement, 12
 second Borel-Cantelli lemma, 315
 self-information, 256
 set complement, 9
 set difference, 9
 set function, 46
 set intersection, 9
 set union, 9
 skewness, 254
 standard Cauchy distribution, 127
 standard deviation, 88
 standard normal distribution, 145
 standard uniform distribution, 144
 strong law of large numbers, 318
 σ -subadditive measure, 47
 support of a random variable, 87,
 123
 surprisal value, 256
 symmetric distribution, 254

U

uncorrelated random variables, 194

uniform distribution, 89, 144

universal set, 9

V

value at risk, 152

variance, 88

W

weak law of large numbers, 318

winnder's curse, 301

Z

Z-score, 147

Z-оцінка, 147