Лекції з теорії ймовірностей I

Adam Osękowski (Адам Осенковський)

19 лютого 2023 р.

Вступ

Мета цього вступу полягає в тому, щоб надати короткий історичний огляд і описати інтуїційні підхіди, пов'язані з формальним визначенням ймовірності. Читач, який хоче відразу перейти до суті справи, може пропустити цей розділ і перейти до глави 1.

Обчислення ймовірностей - це сфера, яка супроводжує людство з давніх часів (поняття випадковості або ризику природним чином виникає в контексті азартних ігор). Історія розвитку цієї галузі заслуговує на окрему монографію, тому ми обмежимося лише елементарними відомостями щодо цього питання. Перші спроби формального, математичного підходу до цієї галузі датуються 15 століттям: у 1494 році було надруковано перший підручник під авторством Л. Пачолі *Summa de arithmetica, geometria, proportioni e proportionalita*, певні частини якого стосуються теорії ймовірностей. У 16 столітті Дж. Кардано, натхненний цим твором, написав книгу *Liber de Ludo Aleae*, присвячену азартним іграм. Подальший розвиток галузі відбувався у зв'язку з жвавим листуванням П. Ферма і Б. Паскаля щодо проблеми справедливого поділу ставки (перша половина 17 ст). Серед багатьох видатних математиків, які займалися створенням теорії ймовірностей та її застосувань, слід згадати Я. Бернуллі (*Ars Conjectandi*, 1713 р.), А. де Муавра (*The Doctrine of Changes*, 1718 р.), П.-С. Лапласа (*Théorie analytique des probabilités*, 1812 р.), слід згадати також роботи Л. Больцмана і Дж. Ґіббса зі статистичної механіки (кінець 19 ст). Цікаво, що до середини 20 століття поняття ймовірності не було повністю визначено математично. Це призвело до багатьох незручних і несподіваних парадоксів (серед інших Парадокс Дж. Бертрана з кінця 19 століття, див. Розділ 1 нижче), що підривало сенс трактування ймовірності як розділу математики. Ці проблеми зникли, коли в 1933 році А. Колмогоров подав строгу аксіоматику; цей рік можна вважати початком сучасней теорії ймовірностей. Тут варто зазначити, що аксіоми Колмогорова визначають ймовірність як міру, нормалізовану на вимірюваному просторі; тому в 1940-х і 1950-х роках панувала думка, що теорія ймовірностей насправді є окремим випадком теорії міри та статистики. Лише роботи Дж. Л. Дуба цього періоду, які запровадили певний інноваційний апарат понять і аргументів, дозволили сформувати обчислення ймовірностей як окрему та незалежну галузь математики.

На цьому ми закінчуємо короткий історичний нарис і переходимо до ілюстрації основних проблем, які виникають при аналізі навіть найпростіших питань. Наведені тут міркування будуть неточними, формальні визначення з'являться в наступному розділі. Почнемо з наступного простого завдання. Припустімо, ми підкидаємо монету один раз і нас цікавлять шанси отримати голову (реверс). Ми починаємо аналіз із переліку всіх можливих результатів, які ми можемо отримати. У нашому випадку є два результати: голова („O”) або решка („R”); позначимо множину цих результатів літерою Ω. Цим результатам відповідають шанси їх отримання, p_O та p_R, які є числами в діапазоні $[0, 1]$, що задовольняють умову $p_O + p_R = 1$. Як визначити ці числа? Очевидно, що нам потрібно знати щось більше про монету, яку ми підкидаємо. Легко уявити, що монета настільки збалансована, що завжди випадає головою вгору; тоді отримання голови - це певна подія, а отримання решки - це неможлива подія. У цій особливій ситуації ми кладемо $p_O = 1$, $p_R = 0$. Подібним чином, якщо результатом підкидання монети завжди є решка, ми повинні прийняти $p_O = 0$, $p_R = 1$. Отже без додаткової інформації щодо „фізичних” властивостей монети ми не можемо нічого сказати про шанси p_O та p_R. У багатьох типових прикладах монета вважається „правильною” (іноді також використовується термін „симетрична”), тобто і голова і решка мають однаковий шанс появи. Тоді $p_O = p_R$, і умова $p_O + p_R = 1$ означає, що кожен із цих результатів має ймовірність $1/2$.

Щойно отримана відповідь має дуже зручну інтерпретацію в „частотному” контексті. А саме, щоб визначити ймовірність отримання голів, ми могли б діяти інакше. Підкинемо фіксовану монету багато разів, скажімо, 10 000 разів, і порахуємо, скільки голів ми отримаємо. Природно припустити, що ймовірність отримання голів має бути *близькою* до дрібу

$$\frac{\text{кількість підкидань, в яких були отримані голови}}{10000}.$$

На практиці виявляється, що кількість отриманих голів (отримана в серії підкидань по 10 000 разів) коливається близько 5 000, тому наведене вище співвідношення є близьким до $1/2$. Як зазначалося раніше, це пов'язано з частотною інтерпретацією ймовірності (а також з так званими законами великих чисел), ми розглянемо це більш детально в подальших розділах.

Тепер розглянемо трохи складніший приклад. Припустимо, ми кидаємо гральну кість один раз і нас цікавить ймовірність отримання трійки на верхній грані. Як і в попередньому прикладі, ми починаємо зі списку всіх можливих результатів експерименту. Ми можемо отримати один, два, три, чотири, п'ять або шість, тому ми беремо $\Omega = \{1, 2, 3, 4, 5, 6\}$. Кожному результатові $j \in \Omega$ відповідає свій шанс p_j; числа p_1, p_2, ..., p_6 є невід'ємними і в сумі дають 1. Очевидно, що ми не можемо сказати нічого більше про ці числа, якщо нічого не знаємо про фізичні властивості кубика (наприклад, його можна збалансувати так, щоб завжди було шість, тоді $p_1 = p_2 = \ldots = p_5 = 0$, $p_6 = 1$). У багатьох прикладах кубик вважається *правильним*, тобто $p_1 = p_2 = \ldots = p_6$. Тоді з умови $p_1 + p_2 + \ldots + p_6 = 1$ випливає рівність $p_1 = p_2 = \ldots = p_6 = 1/6$. Варто також зазначити, що розглянутий тут експеримент може призвести до інших питань. Наприклад, ми можемо запитати, які шанси отримати число, що ділиться на три; отримати число не менше ніж чотири; і т.д... У першому з цих запитань нас цікавить ймовірність того, що результат буде належати множині $\{3, 6\}$; в другім, множині $\{4, 5, 6\}$; загалом, кожне питання відноситься до деякої підмножини множини Ω (ми будемо називати такі підмножини *подіями*). Щоб надати відповіді, ми складаємо ймовірності „одиничних” результатів у цих множинах. Тож у першому випадку ми отримуємо шанс $1/6 + 1/6 = 1/3$, а в другому $1/6 + 1/6 + 1/6 = 1/2$.

У наступному прикладі ми трохи ускладнюємо експеримент і розглядаємо *подвійне* кидання гральної кості. Припустімо, ми хочемо знайти ймовірність того, що ми отримаємо однакове число в обох підкиданнях. Цього разу одиничним результатом експерименту є пара чисел (i, j), i, $j \in \{1, 2, 3, 4, 5, 6\}$ (перший елемент пари - це кількість пунктів, отриманих при першому підкиданні кубика, а другий елемент пари - кількість пунктів при другому підкиданні). Множина Ω усіх таких пар є 36-елементовою, і якщо кубик є симетричним, кожен результат має однакову ймовірність $1/36$. В тим прикладі нас цікавить ймовірність того, що отримана пара чисел буде належати множині $\{(1,1),(2,2),(3,3),(4,4),(5,5),(6,6)\}$. Ця подія складається з шести елементів, кожен з яких має шанс $1/36$; отже, відповідь $6 \cdot 1/36 = 1/6$.

Переходимо до наступного прикладу. Розглянемо урну, яка містить одну білу, дві чорні та три зелені кулі. Яка ймовірність того, що, витягнувши одну кулю з урни, ми витягнемо зелену кулю? Ми наведемо два міркування, що ведуть до відповіді. Як і раніше, почнемо з „технічного" припущення щодо кульок: ми припускаємо, що вони однакового розміру та ретельно перемішані (ідея полягає в тому, що кожна кулька має однаковий шанс бути витягнутою). Є три можливі результати: взяти білу („B"), чорну („C") або зелену („Z") кулю, тому беремо $\Omega = \{B, C, Z\}$. Зауважимо, що, на відміну від попередніх прикладів, ймовірності p_B, p_C, p_Z, що відповідають цим результатам, тепер *нерівні*. Оскільки чорних куль удвічі більше, ніж білих, ми очікуємо, що ймовірність витягнути чорну кулю буде вдвічі більшою ніж ймовірність витягнути білу кулю: $p_C = 2p_B$. Аналогічно отримуємо тотожність $p_Z = 3p_B$, що в поєднанні з рівністю $p_B + p_C + p_Z = 1$ дає $p_B = 1/6$, $p_C = 1/3$ та $p_Z = 1/2$. Отже, відповідь $1/2$.

Другий шлях розв'язання вищевказаної проблеми - це введення „колірних відтінків"; іншими словами, ми погоджуємося, що в межах фіксованого кольору кульки трохи відрізняються. Тепер якщо ми витягнемо одну кулю з урни, у нас буде *шість* можливих результатів: $\Omega = \{B, C_1, C_2, Z_1, Z_2, Z_3\}$, кожен з яких відповідає рівно одній кулі, тому індивідуальні результати однаково ймовірні. Отже $p_B = p_{C_1} = p_{C_2} = p_{Z_1} = p_{Z_2} = p_{Z_3} = 1/6$; нас цікавить шанс, що результат буде належати множині $\{Z_1, Z_2, Z_3\}$, так що отримуємо відповідь $3 \cdot 1/6 = 1/2$.

У наступному прикладі множина потенційних результатів експерименту буде нескінченною (але зліченною). А саме, припустімо, що ми підкидаємо правильну монету, поки не випаде голова („O"). Який шанс, що ми будемо підкидати монету щонайбільше п'ять разів? Як і вище, ми починаємо з переліку всіх можливих результатів експерименту. Наш випадковий експеримент передбачає підкидання монети поки не випаде голова, тому поодинчим результатом є послідовність решок, що закінчується головою; наприклад, послідовність (R, R, R, R, R, O) означає, що ми викинули п'ять разів решки і тільки у шостому відкиданні отримали голову. Отже

$$\Omega = \{(O), (R, O), (R, R, O), (R, R, R, O), \ldots\}$$

і ця множина нескінченна. Спробуємо визначити ймовірність $p_{(R,R,\ldots,R,O)}$ результату $(R, R, \ldots, R, O)$, де маємо рівно k решек (k — фіксоване ціле невід'ємне число). Досить очевидно, що на відміну від перших трьох прикладів, різні результати мають *різні* ймовірності: наприклад, $p_{(O)}$ більше ніж $p_{(R,R,R,O)}$. Щоб обчислити $p_{(R,R,\ldots,R,O)}$, розглянемо дещо інший експеримент: розглянемо $k+1$-кратне підкидання монети і шанс отримати першу голову під час останнього підкидання. Цілком зрозуміло, що ця ймовірність дорівнює числу $p_{(R,R,\ldots,R,O)}$, яке ми шукаємо. Перевага цього модифікованого експерименту полягає в тому, що він має скінченну кількість результатів: ми кидаємо $k+1$ разів, отже

$$\overline{\Omega} = \big\{(a_1, a_2, \ldots, a_{k+1}) : a_i \in \{O, R\},\ i = 1,\ 2,\ \ldots,\ k+1\big\},$$

і кожна послідовність має такий самий шанс бути отриманою. Оскільки $\overline{\Omega}$ містить 2^{k+1} елементів, ми робимо висновок, що ймовірність, яку ми шукаємо в початковому експерименті, дорівнює

$$p_{(R,R,\ldots,R,O)} = \frac{1}{2^k}.$$

В задачі нас цікавить ймовірність того, що ми кинемо щонайбільше п'ять разів, тобто ймовірність події

$$\{(O), (R, O), (R, R, O), (R, R, R, O), (R, R, R, R, O)\}.$$

Отже, ця ймовірність дорівнює

$$p_{(O)} + p_{(R,O)} + p_{(R,R,O)} + p_{(R,R,R,O)} + p_{(R,R,R,R,O)} = \frac{1}{2} + \frac{1}{4} + \frac{1}{8} + \frac{1}{16} + \frac{1}{32} = \frac{31}{32}.$$

Останній приклад буде трохи складнішим. Припустімо, ми вибираємо випадкове число в діапазоні $[-1, 2]$ і хочемо знайти ймовірність того, що вибране число більше 1. Як і раніше, ми починаємо аналіз із подання множини всіх можливих результатів експерименту: $\Omega = [-1, 2]$. Зауважте, що множина Ω є нескінченною і навіть незліченною (це викликає певні технічні ускладнення, див. Розділ 1 нижче). Цікава для нас подія – витягування числа, більшого за 1 – відповідає діапазону $[1, 2]$. Як визначити ймовірність цієї події?

Як і в попередніх прикладах, неможливо дати відповідь без додаткових знань про то *як виконується розіграш* (відповідь може бути неоднозначною). Наприклад, розіграш може бути таким: ми спочатку підкидаємо правильну монету, потім вибираємо 0 якщо отримали голову, і 1 якщо отримали решку. Інший приклад: ми кидаємо правильний кубик, а потім ділимо число на 3. Як бачите, в обох цих прикладах ми отримуємо число з $[-1, 2]$ (навіть більше: у першому випадку ми вибираємо число з множини $\{0, 1\}$, а в другий - із множини $\{1/3, 2/3, \ldots, 2\}$). Читач легко поставить інші, більш складні експерименти, результати яких належать до множини $[-1, 2]$.

Тому нам необхідно уточнити питання. Коли ми пишемо „ми вибираємо випадкове число з відрізка $[-1, 2]$" (без будь-якої додаткової інформації), ми зазвичай маємо на увазі, що вибране число „має однаковий шанс з'явитися в кожній частині відрізка". Больш точно, це означає, що ймовірність того, що результат належить даній підмножині $A \subset \Omega$, пропорційна мірі Лебега цієї підмножини: ймовірність A дорівнює $c|A|$ для певної константи c, що залежить лише від Ω. Щоб знайти цю константу, ми підставляємо $A = \Omega$. Це певна подія (кожна вибрана точка, звичайно, належить Ω), тому вона має імовірність 1. Отже, ми отримуємо $1 =$ ймовірність $\Omega = c|\Omega|$, тобто $c = 1/|\Omega|$ і

$$\text{ймовірність } A = \frac{|A|}{|\Omega|}.$$

Отже, повертаючись до нашого конкретного прикладу, ми отримуємо $|[1, 2]|/|[-1, 2]| = 1/3$.

1 Аксіоми теорії ймовірностей

Метою цього розділу є формальне введення основних понять теорії ймовірностей і дослідження їх основних властивостей. Припустимо, ми проводимо якийсь випадковий експеримент. Відразу виникає питання: як це описати математично?

Як ми зуважили у вступі, ми можемо говорити про потенційні „найдрібніші" результати, які ми називатимемо *елементарними подіями* і позначатимемо ω. Множину всіх елементарних подій ми позначаємо буквою Ω.

Приклади:

1. Підкидання однієї монети. Можливі два результати: $\Omega = \{O, R\}$.
2. Підкидання гральної кості. Можливі шість результатів: $\Omega = \{1, 2, 3, 4, 5, 6\}$.

Як ми бачили в попередньому розділі, ми часто зацікавлені не стільки в конкретному результаті ω, скільки в тому чи належить він до визначеної підмножини Ω. Такі підмножини називаються *подіями* і позначаються літерами A, B, C, ...

Приклади:

3. Кидаємо кубик двічі, A – загальна сума дорівнює 4. Маємо

$$\Omega = \big\{(i,j) : i,\, j \in \{1,2,3,4,5,6\}\big\} \quad \text{та} \quad A = \{(1,3),(2,2),(3,1)\}.$$

4. Підкидаємо монету доки не отримуємо голову, A – потребуємо для цього максимум три підкидання. Один досвід зводиться до серії підкидань монети, що закінчується головою. Як поодинчий результат можемо взяти послідовність результатів, отриманих у послідовних відкидуваннях:

$$\Omega = \big\{(O),(R,O),(R,R,O),(R,R,R,O),\dots\big\}.$$

Подія, що нас цікавить, відповідає підмножині

$$A = \big\{(O),(R,O),(R,R,O)\big\}.$$

5. Обертання колеса рулетки, A – стрілка зупиняється в другій чверті. Тоді $\Omega = [0,2\pi)$ і $A = [\pi/2,\pi]$.

Szczególne zdarzenia, interpretacje działań i relacji na zdarzeniach:

· Ω – певна подія,

· $\emptyset$ – неможлива подія,

· $A \cap B$ – відбулися обидві події A, B,

· $A \cap B = \emptyset$ – події несумісні,

· $A \cup B$ – відбулося A або B,

· A' – заперечення A (A' – доповненням множини A),

· $A \setminus B = A \cap B'$ – A відбулося, а B – ні,

· $A \subseteq B$ – A спричиняє B.

Перейдемо до іншого важливого питання. Припустимо, що Ω є фіксованою множиною, і спробуємо подумати про те, які підмножини Ω будуть/можуть зацікавити нас у подальших міркуваннях, позначимо клас цих „допустимих” підмножин через $\mathcal{F}$. На перший погляд, здається, що ця проблема не має сенсу: чому ми не можемо просто перевірити всі можливі підмножини, тобто чому б не поставити $\mathcal{F} = 2^\Omega$? Що ж, виявляється, що цей вибір добре працює, коли Ω не більше ніж лічильна. З іншого боку, для $|\Omega| > \aleph_0$ клас 2^Ω загалом занадто великий – у багатьох випадках виникають проблеми з визначенням ймовірностей для нього. Це, як наслідок, змушує вибрати конкретну підродину 2^Ω. Як вибрати таку підродину? Розумний клас $\mathcal{F}$ повинен бути замкнений відносно взяття лічильних сум, добутків та протилежних подій, щоб мати можливість виконувати основні операції над подіями (див. список операцій вище). Це, у свою чергу, призводить до постулату, що $\mathcal{F}$ є деякою виділеною σ-алгеброю підмножин Ω. Нагадаємо відповідне визначення.

Визначення 1.1. Родина $\mathcal{F}$ підмножин Ω називається σ-алгеброю, якщо

$$\begin{aligned} &(\mathrm{i}) \quad \emptyset \in \mathcal{F}, \\ &(ii) \quad A \in \mathcal{F} \Rightarrow A' \in \mathcal{F}, \\ &(iii) \quad A_1,\, A_2,\, \ldots \in \mathcal{F} \Rightarrow \bigcup_{n=1}^{\infty} A_n \in \mathcal{F}. \end{aligned}$$

Пару $(\Omega, \mathcal{F})$ ми називаємо *вимірюваним простіром.*

Переходимо тепер до визначення ймовірності: цей об'єкт буде заданий рядом властивостей і постулатів. Щоб отримати деяку інтуїцію щодо цієї концепції, а також зрозуміти, звідки беруться відповідні припущення, зручно спочатку розглянути т.зв. частоту подій (див. попередній розділ). Припустимо, що в експерименті нас цікавить ймовірність певної події A. Повторимо цей експеримент n разів і визначимо

$$\rho_n(A) = \frac{\text{кількість експериментів, в яких відбулася } A}{n}.$$

Це відносна частота A в серії n експериментів. Ми очікуємо, що для великих n число $\rho_n(A)$ має бути близьким до шансу отримати A в одному експерименті. Як легко перевірити, ρ_n приймає значення в діапазоні $[0, 1]$ і має такі властивості:

$$\begin{aligned} &(\mathrm{i}) \quad \rho_n(\Omega) = 1, \\ &(\mathrm{ii}) \text{ якщо } A_1,\, A_2, \ldots \text{ попарно несумісні, то } \rho_n\left(\bigcup_{k=1}^{\infty} A_k\right) = \sum_{k=1}^{\infty} \rho_n(A_k). \end{aligned}$$

Це призводить до наступного визначення.

Визначення 1.2 (Аксіоматика Колмогорова)**.** Нехай $(\Omega, \mathcal{F})$ – фіксований вимірний простір. $\mathbb{P} : \mathcal{F} \to [0, 1]$ називається *імовірністю*, якщо

(i) $\mathbb{P}(\Omega) = 1$,

(ii) для будь-яких подій $A_1,\, A_2,\, \ldots \in \mathcal{F}$ несумісних парами, маємо

$$\mathbb{P}\left(\bigcup_{k=1}^{\infty} A_k\right) = \sum_{k=1}^{\infty} \mathbb{P}(A_k).$$

Трійка $(\Omega, \mathcal{F}, \mathbb{P})$ називається *імовірнісним простором.*

Зауваження:

1. Таким чином, ймовірність є нормалізованою невід'ємною мірою на $(\Omega, \mathcal{F})$. Іноді ми говоримо, що $\mathbb{P}$ є *імовірнісною мірою.*

2. Слід пам'ятати, що при моделюванні конкретного випадкового експерименту вибір імовірнісного простору залежить тільки від нас. У багатьох ситуаціях з умов експерименту випливають певні постулати, які більш-менш однозначно дають трійку $(\Omega, \mathcal{F}, \mathbb{P})$; однак іноді це не так (пор. парадокс Бертрана нижче).

Теорема 1.1 (Властивості ймовірності)**.** *Нехай* $(\Omega, \mathcal{F}, \mathbb{P})$ *— імовірнісний простір,* $A,\, B,\, A_1,\, A_2,\, \ldots \in \mathcal{F}$.

Тоді

(i) $\mathbb{P}(\emptyset) = 0$.

(ii) Якщо $A_1, A_2, \ldots, A_n$ *попарно несумісні, то* $\mathbb{P}\left(\bigcup_{i=1}^{n} A_i\right) = \sum_{i=1}^{n} \mathbb{P}(A_i)$.

(iii) $\mathbb{P}(A') = 1 - \mathbb{P}(A)$.

(iv) Якщо $A \subseteq B$, *то* $\mathbb{P}(B \setminus A) = \mathbb{P}(B) - \mathbb{P}(A)$ *та* $\mathbb{P}(A) \leqslant \mathbb{P}(B)$.

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

(vi) $\mathbb{P}\left(\bigcup_{i=1}^{\infty} A_i\right) \leqslant \sum_{i=1}^{\infty} \mathbb{P}(A_i)$.

Властивість (v) із наведеної вище теореми можна узагальнити на випадок скінченної кількості множин. Справедливим є наступний факт.

Теорема 1.2 (Формула включення-виключення)**.** *Якщо* $A_1, A_2, \ldots, A_n \in \mathcal{F}$, *то*

$$\mathbb{P}(A_1 \cup A_2 \cup \ldots \cup A_n) = \sum_{i=1}^{n} \mathbb{P}(A_i) - \sum_{i<j} \mathbb{P}(A_i \cap A_j) + \sum_{i<j<k} \mathbb{P}(A_i \cap A_j \cap A_k) - \ldots$$
$$+ (-1)^{n+1} \mathbb{P}(A_1 \cap A_2 \cap \ldots \cap A_n).$$

Докази двох наведених вище теорем дуже прості й базуються на аксіомах Колмогорова. Деталі залишаємо Читачеві.

Теорема 1.3 (Теорема неперервності)**.** *Нехай* $(\Omega, \mathcal{F}, \mathbb{P})$ *– імовірнісний простір, а* $(A_n)_{n=1}^{\infty}$ *– послідовність подій.*

(i) Якщо події A_n *не спадають (тобто* $A_1 \subseteq A_2 \subseteq \ldots$*), то*

$$\mathbb{P}\left(\bigcup_{n=1}^{\infty} A_n\right) = \lim_{n\to\infty} \mathbb{P}(A_n).$$

(ii) Якщо події A_n *не зростають (тобто* $A_1 \supseteq A_2 \supseteq \ldots$*), то*

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

Доведення: (i) Розглянемо послідовність $(B_n)_{n\geqslant 1}$ подій, задану як

$$B_1 = A_1,\ \ B_2 = A_2 \setminus A_1,\ \ B_3 = A_3 \setminus A_2,\ \ \ldots.$$

Легко перевірити, що події $B_1, B_2, \ldots$ несумісні, $\bigcup_{n=1}^{k} B_n = A_k$ для будь-кого $k \geqslant 1$ та $\bigcup_{n=1}^{\infty} B_n = \bigcup_{n=1}^{\infty} A_n$. Тоді

$$\mathbb{P}\left(\bigcup_{n=1}^{\infty} A_n\right) = \mathbb{P}\left(\bigcup_{n=1}^{\infty} B_n\right)$$
$$= \sum_{n=1}^{\infty} \mathbb{P}(B_n) = \lim_{k\to\infty} \sum_{n=1}^{k} \mathbb{P}(B_n) = \lim_{k\to\infty} \mathbb{P}\left(\bigcup_{n=1}^{k} B_n\right) = \lim_{k\to\infty} \mathbb{P}(A_k),$$

де в другому переході ми використали лічильну адитивність міри $\mathbb{P}$, а в четвертому переході ми використали Теорему 1(ii).

(ii) Додаткова послідовність $(A'_n)_{n\geqslant 1}$ є зростаючою, тому, використовуючи (i) і закони де Моргана, ми маємо

$$\begin{aligned}\mathbb{P}\left(\bigcap_{n=1}^{\infty} A_n\right) &= 1-\mathbb{P}\left(\left(\bigcap_{n=1}^{\infty} A_n\right)'\right)\\ &= 1-\mathbb{P}\left(\bigcup_{n=1}^{\infty} A'_n\right) = 1-\lim_{n\to\infty}\mathbb{P}(A'_n) = \lim_{n\to\infty}\mathbb{P}(A_n).\square\end{aligned}$$

Представимо деякі базові приклади, які зустрічаються в багатьох природних і поширених проблемах.

Приклади:

1. (Класична схема, класична ймовірність). Припустимо, Ω — скінченна множина, $\mathcal{F} = 2^{\Omega}$ і всі одноелементні події однаково ймовірні. Тоді, як легко перевірити, для будь-яких $A \in \mathcal{F}$,

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

2. Припустимо, що $\Omega = \{\omega_1, \omega_2, \dots\}$ не більше ніж зліченна множина, а p_1, p_2, ... - невід'ємні числа, що сумі дають 1. Тоді вибір $\mathcal{F} = 2^{\Omega}$ і $\mathbb{P}(\{\omega_i\}) = p_i$, $i = 1, 2, \dots$ однозначно встановлює ймовірнісний простір $(\Omega, \mathcal{F}, \mathbb{P})$: для кожного $A \in \mathcal{F}$ ми маємо

$$\mathbb{P}(A) = \sum_i 1_A(\omega_i)p_i,$$

де 1_A – *індикатор* множини A:

$$1_A(\omega) = \begin{cases} 1 & \text{якщо } \omega \in A, \\ 0 & \text{якщо } \omega \notin A. \end{cases}$$

3. (Геометрична ймовірність). Припустимо, що $\Omega \in \mathcal{B}(\mathbb{R}^d)$, тобто Ω є борелівською підмножиною $\mathbb{R}^d$ і $0 < |\Omega| < \infty$ (тут $|\cdot|$ – міра Лебега в $\mathbb{R}^d$). Нехай $\mathcal{F} = \mathcal{B}(\Omega)$ – σ-алгебра борелівських підмножин Ω, а $\mathbb{P}$ визначається як

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

Тоді трійка $(\Omega, \mathcal{F}, \mathbb{P})$ є ймовірнісним простором. Ми використовуємо цей простір для моделювання експерименту, що відповідає вибору навмання точки з множини Ω.

4. (Парадокс Бертрана) В колі радіуса 1 проведено навмання хорду AB. Яка ймовірність того, що вона буде довшою за сторону рівностороннього трикутника, вписаного в це коло?

Ми представимо три рішення.

I) Через інваріантність кола до поворотів ми можемо прирівняти вибір навмання хорди AB до вибору навмання центрального кута $\alpha = \angle AOB \in [0, 2\pi)$. Отже, $\Omega = [0, 2\pi)$, $\mathcal{F} = \mathcal{B}(\Omega)$, і $\mathbb{P}$ є геометричною ймовірностю. Хорда задовольняє умовам задачі тоді і тільки тоді, коли $\alpha \in (2\pi/3, 4\pi/3)$, тому ймовірність, яку ми шукаємо, дорівнює

$$\mathbb{P}((2\pi/3, 4\pi/3)) = \frac{|(2\pi/3, 4\pi/3)|}{|[0, 2\pi)|} = \frac{1}{3}.$$

II) Вибір хорди можна ототожнити з вибором її центра. Отже, ми маємо $\Omega = B(0,1)$, $\mathcal{F} = \mathcal{B}(\Omega)$, а $\mathbb{P}$ – це відповідня геометрична ймовірність. Хорда задовольнятиме бажану умову тоді і тільки тоді, коли її центр лежить всередині кола радіуса $1/2$, концентричного до даного кола, тому ймовірність, яку ми шукаємо, дорівнює

$$\mathbb{P}([0,1/2)) = \frac{|B(0,1/2)|}{B(0,1)} = \frac{1}{4}.$$

III) Як і в попередньому розв'язанні, ми розглядаємо вибір центру хорди, але цього разу дивимося на його відстань від центру кола. Отже, $\Omega = [0,1]$, $\mathcal{F} = \mathcal{B}(\Omega)$ і $\mathbb{P}$ – це геометрична ймовірність. Хорда задовольнятиме умови задачі, якщо її центр знаходиться на відстані менше $1/2$ від центру кола. Отже, ймовірність, яку ми шукаємо, дорівнює

$$\mathbb{P}([0,1/2)) = \frac{|[0,1/2)|}{|[0,1]|} = \frac{1}{2}.$$

Отже, ми отримали три різні результати, звідси слово „парадокс" у назві вище. Однак тут немає жодного протиріччя – ми використовували три різні ймовірнісні простори для опису одного випадкового експерименту. Загалом, теорія ймовірності не диктує, яку модель вибрати; вона дозволяє обчислювати ймовірності подій лише тоді, коли конкретна трійка $(\Omega, \mathcal{F}, \mathbb{P})$ вже вказана.

2 Завдання

1. Скількома способами можна розташувати шість одиниць, п'ять двійок і чотири трійки?

2. Знайти кількість розв'язків рівняння $x_1 + x_2 + x_3 + x_4 = 50$

a) у невід'ємних цілих числах $x_1,\ x_2,\ x_3,\ x_4$,

b) в натуральних числах $x_1,\ x_2,\ x_3,\ x_4$.

3. Скільки можно вибрати шісток (різних) чисел з $1, \ldots, 49$, в яких немає послідовних чисел?

4. З колоди в 52 карти витягується 13 карт. Яка ймовірність того, що а) рівно сім; б) рівно шість карток мають той самий колір?

5. У класі 15 учнів. На кожному уроці вчитель вибирає навмання одного учня, щоб опросити. Обчислити ймовірність того, що кожен учень буде опитаний протягом 16 уроків.

6. У шафі n пар взуття. Виймаємо навмання k черевиків ($k \leqslant n$). Обчисліть ймовірність того, що

a) серед вибраних черевиків є принаймні одна пара,

b) серед вибраних черевиків є рівно одна пара.

7. $(\Omega, \mathcal{F}, P)$ є ймовірнісним простором, $A,\ B,\ C \in \mathcal{F}$.

a) Припустимо, що $P(A \cup B) = 1/2,\ P(A \cap B) = 1/4,\ P(A \backslash B) = P(B \backslash A)$. Обчислити $P(A)$ і $P(B \backslash A)$.

b) Припустимо, що $A \cup B \cup C = \Omega,\ P(B) = 2P(A),\ P(C) = 3P(A),\ P(A \cap B) = P(A \cap C) = P(B \cap C)$. Доведіть, що $1/6 \leqslant P(A) \leqslant 1/4$.

c) Припустимо, що $P(A) \geqslant 2/3,\ P(B) \geqslant 2/3,\ P(C) \geqslant 2/3,\ P(A \cap B \cap C) = 0$. Обчислити $P(A)$.

8. 52 карти роздаються чотирьом гравцям, по 13 карт кожному. Яка ймовірність того, що кожен гравець має принаймні одну піку?

9. Є N листів і N конвертів з різними адресами. Кожен лист відповідає точно одній адресі і навпаки. Листи складали в конверти навмання, по одному в кожен конверт. Обчисліть ймовірність того, що жоден лист не попаде в правильний конверт.

10. Доведіть, що кожна нескінченна σ-алгебра незліченна.

11. Паличка випадково ламається в двох точках. Яка ймовірність того, що три відрізки можуть утворити трикутник?

12. Монету діаметром $\frac{2}{3}$ кидають на нескінченну шахову дошку зі стороною клітинок 1. Яка ймовірність того, що (а) монета повністю опиниться всередині одної з клітин; б) перетинатиметься з двома сторонами клітини?

13. Голка довжиною ℓ ($\ell < d$) навмання кинута на площину, поділену на нескінченні смуги шириною d. Знайти ймовірність того, що голка перетне край якоїсь смужки.

3 Умовна ймовірність і незалежність подій

3.1 Умовна ймовірність

У попередньому розділі ми бачили, як визначити ймовірності подій, якщо дано ймовірнісний простір. Тепер ми переходимо до наступного моменту: іноді, досліджуючи ймовірності певної події, ми маємо додаткову інформацію, яка суттєво змінює умови. Найкраще це можна проілюструвати на прикладі.

Приклад 3.1. Урна містить п'ять білих куль з номерами 1, 2, 3, 4, 5 і три чорні кулі з номерами 1, 2, 3. Витягується навмання одна куля.

а) Яка ймовірність того, що число на ній парне?

б) Відомо, що витягнута куля біла. Яка ймовірність того, що число на ній парне?

У (а) відповідь 3/8: у нас є вісім куль, три з яких парні. В б) мами додаткову інформацію: витягнута куля біла. Звичайно, ця інформація змінює відповідь: є п'ять білих куль, дві з яких парні, з чого ми отримуємо 2/5. Формально ми маємо справу з класичною ймовірністю на

$$\Omega = \{(1, b), (2, b), \ldots, (5, b), (1, c), (2, c), (3, c)\}.$$

Визначимо події: А – витягнута куля парна, Б – витягнута куля біла; тому

$$A = \{(2, b), (4, b), (2, c)\}, \qquad B = \{(1, b), (2, b), \ldots, (5,)\}$$

і ми маємо

$$\frac{2}{5} = \frac{|A \cap B|}{|B|} = \frac{|A \cap B|/|\Omega|}{|B|/|\Omega|} = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)}.$$

З наведених міркувань випливає наступне визначення.

Визначення 3.1. Нехай $(\Omega, \mathcal{F}, \mathbb{P})$ є ймовірнісним простором, A, $B \in \mathcal{F}$, $\mathbb{P}(B) > 0$. Умовною ймовірністю A за умови B називається число

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

Зауваження: Легко перевірити, що для довольної події B такої що $\mathbb{P}(B) > 0$, умовна ймовірність $\mathbb{P}(\cdot|B)$ є новою імовірнісною мірою на $(\Omega, \mathcal{F})$.

Теорема 3.1 (Імовірність перерізу подій)**.** *Нехай $(\Omega, \mathcal{F}, \mathbb{P})$ — ймовірнісний простір, а A_1, A_2, ..., A_n — події, що задовольняють умову $\mathbb{P}(A_1 \cap A_2 \cap \ldots \cap A_n) > 0$. Тоді*

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

Доведення. Застосуйте визначення умовної ймовірності. □

Приклад 3.2. Урна містить $n - 1$ білих кульок і одну чорну кульку. Витягуємо навмання одну кульку за раз, поки не витягнемо чорну кульку. Яка ймовірність того, що ми витягнемо k кульок, якщо а) витягуємо без повернення, б) вітягуємо з поверненням?

Позначимо білі кульки b_1, b_2, …, b_{n-1}, а чорну кульку – c. Маємо

$$\Omega = \{(c), (b_1, c), (b_2, c), \ldots, (b_{n-1}, c), (b_1, b_1, c), \ldots\},$$

$\mathcal{F} = 2^{\Omega}$, а ймовірність задана шляхом визначення мас окремих одноелементних подій (пор. приклад 2 із попередньої лекції).

Розглянемо подію A_i – i-та куля біла, $i = 1, 2, \ldots$. З наведеної вище теореми випливає

$$\begin{aligned}&\mathbb{P}(A_k' \cap A_{k-1} \cap A_{k-2} \cap \ldots \cap A_1)\\ &= \mathbb{P}(A_k'|A_{k-1} \cap \ldots \cap A_1)\mathbb{P}(A_{k-1}|A_{k-2} \cap \ldots \cap A_1)\ldots\mathbb{P}(A_2|A_1)\mathbb{P}(A_1).\end{aligned}$$

a) З умов задачі випливає, що

$$\mathbb{P}(A_i|A_{i-1} \cap \ldots \cap A_1) = \frac{n-i}{n-i+1}, \quad \mathbb{P}(A_k'|A_{k-1} \cap \ldots \cap A_1) = \frac{1}{n-k+1},$$

тому ймовірність, яку ми шукаємо, дорівнює

$$\frac{1}{n-k+1} \cdot \frac{n-k+1}{n-k+2} \cdot \frac{n-k+2}{n-k+3} \cdot \ldots \cdot \frac{n-2}{n-1} \cdot \frac{n-1}{n} = \frac{1}{n}.$$

b) Цього разу маємо

$$\mathbb{P}(A_i|A_{i-1} \cap A_{i-2} \cap \ldots \cap A_1) = \frac{n-1}{n},$$

тому шукана ймовірність дорівнює

$$\left(1 - \frac{n-1}{n}\right) \cdot \frac{n-1}{n} \cdot \frac{n-1}{n} \cdot \ldots \cdot \frac{n-1}{n} = \frac{1}{n}\left(\frac{n-1}{n}\right)^{k-1}.$$

Тепер ми звернемося до аналізу „багатоетапних” експериментів, у яких ми маємо справу з рандомізацією в кілька кроків, а ймовірнісний простір задається шляхом визначення умовних ймовірностей, пов'язаних з окремими кроками (див. приклад нижче). Почнемо з визначення.

Визначення 3.2. Ми кажемо, що родина подій $(B_k)_{k=1}^n$ є повною групою подій або розбиттям (скінченної) множини Ω, якщо $B_1 \cup B_2 \cup \ldots \cup B_n = \Omega$ і події B_1, B_2, …, B_n несумісні парами. Подібним чином ми визначаємо зліченне розбиття Ω.

Теорема 3.2 (Формула повної ймовірності)**.** *Нехай $(\Omega, \mathcal{F}, \mathbb{P})$ – імовірнісний простір, а $(B_k)_k$ – розбиття Ω (скінченне або зліченне), таке що $\mathbb{P}(B_k) > 0$ для всіх k. Для кожної події $A \in \mathcal{F}$ виконується рівність*

$$\mathbb{P}(A) = \sum_k \mathbb{P}(A|B_k)\mathbb{P}(B_k).$$

Доведення. Події $A \cap B_1$, $A \cap B_2$, … несумісні парами і в сумі дають A, тому

$$\mathbb{P}(A) = \sum_k \mathbb{P}(A \cap B_k) = \sum_k \mathbb{P}(A|B_k)\mathbb{P}(B_k). \qquad \square$$

Теорема 3.3 (Формула Байєса)**.** *З припущеннями, як вище, якщо $\mathbb{P}(A) > 0$, то для кожного k,*

$$\mathbb{P}(B_k|A) = \frac{\mathbb{P}(A|B_k)\mathbb{P}(B_k)}{\sum_n \mathbb{P}(A|B_n)\mathbb{P}(B_n)} \quad \left(= \frac{\mathbb{P}(A|B_k)\mathbb{P}(B_k)}{\mathbb{P}(A)}\right).$$

Доведення. Формула безпосередньо випливає з визначення умовної ймовірності та формули для повної ймовірності. □

Приклад 3.3. Дано урни I та II. В урні I є b_1 білих куль і c_1 чорних куль, а в урні II – b_2 білих куль і c_2 чорних куль. Ми вибираємо урну, а потім кулю з цієї урни.

а) Яка ймовірність того, що куля біла?

б) Припустимо, що витягнута куля біла. Яка ймовірність того, що її було витягнуто з першої урни?

Мами два етапи експерименту: вибір урни і вибір кулі з обраної урни. Введемо події A – витягнуто білу кулю, B_1 – обрано урну I, B_2 – обрано урну II. Маємо $B_1 \cap B_2 = \emptyset$, $B_1 \cup B_2 = \Omega$, тому сімейство $(B_i)_{i=1}^2$ є розбиттям Ω. З умов задачі випливає, що

$$\mathbb{P}(B_1) = \mathbb{P}(B_2) = \frac{1}{2} > 0, \quad \mathbb{P}(A|B_1) = \frac{b_1}{b_1 + c_1}, \quad \mathbb{P}(A|B_2) = \frac{b_2}{b_2 + c_2}.$$

а) Використовуючи формулу повної ймовірності, маємо

$$\mathbb{P}(A) = \mathbb{P}(A|B_1)\mathbb{P}(B_1) + \mathbb{P}(A|B_2)\mathbb{P}(B_2) = \frac{1}{2}\left(\frac{b_1}{b_1 + c_1} + \frac{b_2}{b_2 + c_2}\right).$$

b) За формулою Байєса

$$\mathbb{P}(B_1|A) = \frac{\mathbb{P}(A|B_1)\mathbb{P}(B_1)}{\mathbb{P}(A)} = \frac{b_1/(b_1 + c_1)}{b_1/(b_1 + c_1) + b_2/(b_2 + c_2)}.$$

3.2 Незалежність подій

Почнемо з інтуїтивних міркувань. Нехай A, $B \in \mathcal{F}$, $\mathbb{P}(B) > 0$. Тоді події A, B незалежні, якщо інформація про те, що подія B сталася, не впливає на ймовірність події A; тобто незалежність еквівалентна рівності $\mathbb{P}(A|B) = \mathbb{P}(A)$, або $\mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B)$. Ми приймаємо це як визначення.

Визначення 3.3. Нехай $(\Omega, \mathcal{F}, \mathbb{P})$ – імовірнісний простір. Події A, B незалежні, якщо

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

Зауваження: Якщо $\mathbb{P}(A) = 0$, то для будь-якої $B \in \mathcal{F}$ події A і B незалежні. Така сама теза справедлива, коли $\mathbb{P}(A) = 1$.

Визначимо тепер незалежність більшого числа подій. Почнемо зі скінченного випадку. Інтуїтивно зрозуміло, що події A_1, A_2, …, A_n незалежні, якщо кожна підмножина цих подій незалежна, і події A_n, $A_1 \cap A_2 \cap \ldots A_{n-1}$ незалежні. Легко побачити, що наведені вище умови забезпечують рівність

$$\mathbb{P}(A_{i_1} \cap A_{i_2} \cap \ldots \cap A_{i_k}) = \mathbb{P}(A_{i_1})\mathbb{P}(A_{i_2}) \ldots \mathbb{P}(A_{i_k})$$

для будь-яких $k = 2, 3, \ldots, n$ і будь-якої зростаючої послідовності $1 \leqslant i_1 < i_2 < \ldots < i_k \leqslant n$. Ми беремо це як визначення.

Визначення 3.4. Кажемо, що події A_1, A_2, …, A_n незалежні у сукупності, якщо для всіх $2 \leqslant k \leqslant n$ і будь-якої послідовності $1 \leqslant i_1 < i_2 < \ldots < i_k \leqslant n$ виконується рівність

$$\mathbb{P}(A_{i_1} \cap A_{i_2} \cap \ldots \cap A_{i_k}) = \mathbb{P}(A_{i_1})\mathbb{P}(A_{i_2}) \ldots \mathbb{P}(A_{i_k}).$$

Визначення 3.5. Кажемо, що події A_1, A_2, …, A_n є попарно незалежними, якщо для будь-яких різних i, $j \in \{1, 2, \ldots, n\}$, події A_i і A_j незалежні.

Звичайно, незалежність у сукупності подій A_1, A_2, …, A_n тягне за собою їх незалежність у парах. Зворотна імплікація невірна, як показано в наступному прикладі.

Приклад 3.4. Кидаємо кубик двічі. Нехай A – перше число парне, B – друге число парне, C – сума парна. Безпосередньо розраховуємо

$$\mathbb{P}(A) = \mathbb{P}(B) = \mathbb{P}(C) = \frac{1}{2}, \quad \mathbb{P}(A \cap B) = \mathbb{P}(B \cap C) = \mathbb{P}(C \cap A) = \frac{1}{4},$$

тому події A, B, C є попарно незалежними. Але вони залежні у сукупності: маємо $A \cap B \subset C$, тому $\mathbb{P}(A \cap B \cap C) = \mathbb{P}(A \cap B) = 1/4 \neq \mathbb{P}(A)\mathbb{P}(B)\mathbb{P}(C)$.

Для будь-якої (можливо, нескінченної) кількості подій ми визначаємо незалежність наступним чином.

Визначення 3.6. Нехай $\{A_i\}_{i \in I}$ є деяким набором подій. Кажемо, що ці події незалежні, якщо для кожного n і різних i_1, i_2, …, $i_n \in I$ події A_{i_1}, A_{i_2}, …, A_{i_n} є незалежними.

Тепер ми визначимо незалежність σ-алгебр.

Визначення 3.7. Нехай $(\Omega, \mathcal{F}, \mathbb{P})$ – ймовірнісний простір та $\mathcal{F}_1$, $\mathcal{F}_2$, …, $\mathcal{F}_n$ – σ-алгебри, що містяться в $\mathcal{F}$. Кажемо, що σ-алгебри незалежні, якщо для будь-яких $A_1 \in \mathcal{F}_1$, $A_2 \in \mathcal{F}_2$, …, $A_n \in \mathcal{F}_n$ виконується умова

$$\mathbb{P}(A_1 \cap A_2 \cap \ldots \cap A_n) = \mathbb{P}(A_1)\mathbb{P}(A_2)\ldots\mathbb{P}(A_n).$$

Теорема 3.4. *σ-алгебри $\mathcal{F}_1$, $\mathcal{F}_2$, …, $\mathcal{F}_n$ є незалежними тоді і тільки тоді, коли будь-які $A_1 \in \mathcal{F}_1$, $A_2 \in \mathcal{F}_2$, …, $A_n \in \mathcal{F}_n$ є незалежними.*

Доведення. $\Leftarrow$ очевидно.

$\Rightarrow$ Ми повинні довести, що для будь-яких $2 \leqslant k \leqslant n$ і будь-якої послідовності $1 \leqslant i_1 < i_2 < \ldots < i_k \leqslant n$ виконується рівність

$$(*) \qquad \mathbb{P}(A_{i_1} \cap A_{i_2} \cap \ldots \cap A_{i_k}) = \mathbb{P}(A_{i_1})\mathbb{P}(A_{i_2})\ldots\mathbb{P}(A_{i_k}).$$

Розглянемо події B_1, B_2, …, B_n, задані наступним чином

$$B_i = \begin{cases} A_i & \text{якщо } i = i_\ell \text{ для певного } \ell, \\ \Omega & \text{інакше.} \end{cases}$$

Тоді $B_1 \in \mathcal{F}_1$, $B_2 \in \mathcal{F}_2$, …, $B_n \in \mathcal{F}_n$, значить

$$\mathbb{P}(B_1 \cap B_2 \cap \ldots \cap B_n) = \mathbb{P}(B_1)\mathbb{P}(B_2)\ldots\mathbb{P}(B_n),$$

що еквівалентно (*). □

Приклад 3.5. Кидаємо кубик двічі. Введемо стандартний імовірнісний простір, що описує цей експеримент (див. попередню лекцію). Розглянемо σ-алгебри

$$\mathcal{F}_1 = \{A \times \{1,2,3,4,5,6\} : A \subseteq \{1,2,3,4,5,6\}\},$$

$$\mathcal{F}_2 = \{\{1,2,3,4,5,6\} \times B : B \subseteq \{1,2,3,4,5,6\}\}.$$

Тоді $\mathcal{F}_1$, $\mathcal{F}_2 \subset \mathcal{F}$ і $\mathcal{F}_1$, $\mathcal{F}_2$ є незалежними: справді, для будь-яких подій $A \times \{1,2,3,4,5,6\} \in \mathcal{F}_1$, $B \times \{1,2,3,4,5,6\}$ $\mathcal{F}_2$ маємо

$$\mathbb{P}(A \times \{1,2,3,4,5,6\}) = \frac{|A| \cdot 6}{36} = \frac{|A|}{6}, \quad \mathbb{P}(\{1,2,3,4,5,6\} \times B) = \frac{6 \cdot |B|}{36} = \frac{|B|}{6}$$

та

$$\mathbb{P}((A \times \{1,2,3,4,5,6\}) \cap (\{1,2,3,4,5,6\} \times B)) = \mathbb{P}(A \times B) = \frac{|A| \cdot |B|}{36}.$$

Приклад 3.6. Нехай $\sigma(A_1)$, $\sigma(A_2)$, ..., $\sigma(A_n)$ будуть σ-алгебрами, породженими подіями A_1, A_2, ..., A_n відповідно (нагадаємо, $\sigma(A) = \{A, A', \emptyset, \Omega\}$). Тоді, якщо A_1, A_2, ..., A_n незалежні, то $\sigma(A_1)$, $\sigma(A_2)$, ..., $\sigma(A_n)$ також незалежні. Щоб довести це, нам потрібно перевірити, що для будь-якого $B_i \in \mathcal{F}_i$, $i = 1, 2, \ldots, n$, виконується рівність

$$\mathbb{P}(B_1 \cap B_2 \cap \ldots \cap B_n) = \mathbb{P}(B_1)\mathbb{P}(B_2)\ldots\mathbb{P}(B_n).$$

Якщо принаймні одна з подій B_i є порожньою множиною, то наведена вище рівність виконується: обидві сторони дорівнюють 0. Якщо для деякого j ми маємо $B_j = \Omega$, то ми можемо пропустити цю подію з обох сторін. Таким чином, достатньо довести, що для будь-якої послідовності $1 \leqslant i_1 < i_2 < \ldots < i_k \leqslant n$ маємо

$$\mathbb{P}(B_{i_1} \cap B_{i_2} \cap \ldots \cap B_{i_k}) = \mathbb{P}(B_{i_1})\mathbb{P}(B_{i_2})) \ldots \mathbb{P}(B_{i_k}),$$

де для кожного j подія B_{i_j} дорівнює A_{i_j} або A'_{i_j}. Для цього достатньо показати, що

$$\mathbb{P}(A'_{i_1} \cap A_{i_2} \cap A_{i_3} \cap \ldots \cap A_{i_k}) = \mathbb{P}(A'_{i_1})\mathbb{P}(A_{i_2})\mathbb{P}(A_{i_3})\ldots\mathbb{P}(A_{i_k}),$$

і застосувати просту індукцію. Остання рівність випливає з незалежності подій A_1, A_2, ..., A_n: справді,

$$\begin{aligned}
&\mathbb{P}(A'_{i_1} \cap A_{i_2} \cap A_{i_3} \cap \ldots \cap A_{i_k}) \\
&= \mathbb{P}(A_{i_2} \cap A_{i_3} \cap \ldots \cap A_{i_k}) - \mathbb{P}(A_{i_1} \cap A_{i_2} \cap A_{i_3} \cap \ldots \cap A_{i_k}) \\
&= \mathbb{P}(A_{i_2})\mathbb{P}(A_{i_3})\ldots\mathbb{P}(A_{i_k}) - \mathbb{P}(A_{i_1})\mathbb{P}(A_{i_2})\mathbb{P}(A_{i_3})\ldots\mathbb{P}(A_{i_k}) \\
&= \mathbb{P}(A'_{i_1})\mathbb{P}(A_{i_2})\mathbb{P}(A_{i_3})\ldots\mathbb{P}(A_{i_k}).
\end{aligned}$$

Тепер розглянемо наступну проблему. Припустимо, що ми маємо N досвідів, де i-й досвід описується імовірнісним простором $(\Omega_i, \mathcal{F}_i, \mathbb{P}_i)$. Як ми можемо створити простір $(\Omega, \mathcal{F}, \mathbb{P})$ для досвіду *незалежного* проведення цих N експериментів?

Звичайно, ми беремо $\Omega_1 \times \Omega_2 \times \ldots \Omega_N$ як Ω. Щоб визначити $\mathcal{F}$, зауважимо, що σ-алгебра $\mathcal{F}_i$ представлена, у контексті множини Ω вище, класом

$$\mathcal{F}'_i = \{\Omega_1 \times \Omega_2 \times \ldots \times \Omega_{i-1} \times A_i \times \Omega_{i+1} \times \ldots \times \Omega_N : A_i \ \mathcal{F}_i\}.$$

Звідси природна ідея взяти $\mathcal{F} = \sigma(\mathcal{F}'_1, \mathcal{F}'_2, \ldots, \mathcal{F}'_N)$, σ-алгебру, згенеровану на подставі $\mathcal{F}'_1$, $\mathcal{F}'_2$, , ..., $\mathcal{F}'_N$. Іншими словами, для $\mathcal{F}$ ми беремо σ-алгебру $\mathcal{F}_1 \otimes \mathcal{F}_2 \otimes \ldots \otimes \mathcal{F}_N$. Переходимо до визначення ймовірнісної міри $\mathbb{P}$. З наведених вище постулатів, σ-алгебри $\mathcal{F}_1'$, $\mathcal{F}_2'$, ..., $\mathcal{F}_N'$ мають бути незалежними, тому ми шукаємо таку ймовірність $\mathbb{P}$, що для будь-яких $A_i \in \mathcal{F}_i$, $i = 1, 2, \ldots, N$,

$$\mathbb{P}(A_1 \times A_2 \times \ldots \times A_N) =$$
$$\mathbb{P}\big((A_1 \times \Omega_2 \times \ldots \times \Omega_N) \cap (\Omega_1 \times A_2 \times \ldots \times \Omega_N) \cap \ldots \cap (\Omega_1 \times \ldots \times \Omega_{N-1} \times A_N)\big)$$
$$= \prod_{i=1}^{N} \mathbb{P}(\Omega_1 \times \ldots \times A_i \times \ldots \times \Omega_N).$$

Крім того, ми хочемо аби $\mathbb{P}(\Omega_1 \times \ldots \times A_i \times \ldots \times \Omega_N) = \mathbb{P}_i(A_i)$ для кожного i. Отже шукаємо $\mathbb{P}$ таку, що для будь-яких подій A_1, A_2, ..., A_N виконується рівність, як зазначено вище

$$\mathbb{P}(A_1 \times A_2 \times \ldots \times A_N) = \mathbb{P}_1(A_1)\mathbb{P}_2(A_2)\ldots\mathbb{P}_N(A_N).$$

З теорії міри ми знаємо, що існує рівно одна така міра $\mathbb{P}$ на $\mathcal{F}_1 \otimes \mathcal{F}_2 \otimes \ldots \otimes \mathcal{F}_N$, і вона дорівнює $\mathbb{P}_1 \otimes \mathbb{P}_2 \otimes \ldots \otimes \mathbb{P}_N$ – добуткові мір $\mathbb{P}_1$, $\mathbb{P}_2$, ..., $\mathbb{P}_N$.

Аналогічні міркування можна провести у випадку нескінченної кількості експериментів, змодельованих імовірнісними просторами $(\Omega_i, \mathcal{F}_i, \mathbb{P}_i)$.

Приклад 3.7 (Схема випробувань Бернуллі)**.** Припустимо, що для кожного $i = 1, 2, \ldots, N$ ми маємо $\Omega_i = \{0,1\}$, $\mathcal{F}^i = 2^{\Omega_i}$ і $\mathbb{P}_i(\{1\}) = p$, де $p \in [0,1]$ – фіксований параметр. Ми бачимо, що кожен окремий простір $(\Omega_i, \mathcal{F}_i, \mathbb{P}_i)$ моделює експеримент (випробування Бернуллі), у якому є два можливі результати: 0 і 1, інтерпретовані як *невдача* і *успіх* (підкреслимо: ймовірність успіху дорівнює p і не залежить від номера випробування). За наведеною вище конструкцією ймовірнісний простір

$$(\{0,1\}^N, 2^{\Omega}, \mathbb{P}) = (\Omega_1 \times \Omega_2 \times \ldots \times \Omega_N, \mathcal{F}_1 \otimes \mathcal{F}_2 \otimes \ldots \otimes \mathcal{F}_N, \mathbb{P}_1 \otimes \mathbb{P}_2 \otimes \ldots \otimes \mathbb{P}_N)$$

моделює послідовність N незалежних випробувань Бернуллі. Ми називаємо цю послідовність *схемою випробувань Бернуллі.*

Звернімо увагу, що для будь-якого $\omega = (\omega_1, \omega_2, \ldots, \omega_N) \in \Omega$ ми маємо $\mathbb{P}(\{\omega\}) = p^k(1-p)^{N-k}$, де k — кількість одиниць у ω. Звідси випливає, що якщо $A_k = \{$кількість успіхів дорівнює $k\}$, то

$$\mathbb{P}(A_k) = \sum_{\omega \in A_k} \mathbb{P}(\{\omega\}) = |A_k| p^k (1-p)^{N-k} = \binom{N}{k} p^k (1-p)^{N-k}.$$

Нехай A_1, A_2, ... довільна послідовність випадкових подій. Тоді $\bigcap_{n=1}^{\infty} \bigcup_{m=n}^{\infty} A_m$ можна інтерпретувати як „відбулося нескінченно багато подій A_1, A_2, ...". Виявляється, що за певних припущень ця подія має ймовірність 0 або 1. А точніше, має місце такий факт.

Лема 3.1 (Бореля-Кантеллі)**.** *Нехай $(\Omega, \mathcal{F}, \mathbb{P})$ – ймовірнісний простір, A_1, A_2, ... $\in \mathcal{F}$.*

(i) Якщо $\sum_{n=1}^{\infty} \mathbb{P}(A_n) < \infty$, то

$$\mathbb{P}\Big(\bigcap_{n=1}^{\infty} \bigcup_{m=n}^{\infty} A_m\Big) = 0$$

(тому з імовірністю 1 *відбулося* скінченно багато *подій A_i).*

(ii) Якщо A_1, A_2, ... незалежні в сукупності і $\sum_{n=1}^{\infty} \mathbb{P}(A_n) = \infty$, то

$$\mathbb{P}\Big(\bigcap_{n=1}^{\infty} \bigcup_{m=n}^{\infty} A_m\Big) = 1.$$

Доведення. (i) Маємо

$$\mathbb{P}\Big(\bigcap_{n=1}^{\infty}\bigcup_{m=n}^{\infty} A_m\Big) \leqslant \mathbb{P}\Big(\bigcup_{m=n}^{\infty} A_m\Big) \leqslant \sum_{m=n}^{\infty} \mathbb{P}(A_m) \xrightarrow{n\to\infty} 0.$$

(ii) Доведемо, що протилежна подія $\bigcup_{n=1}^{\infty}\bigcap_{m=n}^{\infty} A'_m$ має ймовірність 0. Достатньо показати, що $\mathbb{P}(\bigcap_{m=n}^{\infty} A'_m) = 0$ для всіх n (тоді розглядувана протилежна подія буде лічильною сумою множин міри 0, а тому також матиме міру 0). Використовуючи теорему неперервності, для будь-яких n отримуємо

$$\mathbb{P}\Big(\bigcap_{m=n}^{\infty} A'_m\Big) = \mathbb{P}\Big(\bigcap_{k=n}^{\infty}\bigcap_{m=n}^{k} A'_m\ Big) = \lim_{k\to\infty} \mathbb{P}\Big(\bigcap_{m=n}^{k} A'_m\Big),$$

що укупі з незалежністю подій A_1, A_2, … дає

$$\lim_{k\to\infty}\prod_{m=n}^{k}\mathbb{P}(A'_m) = \lim_{k\to\infty}\prod_{m=n}^{k}(1-\mathbb{P}(A_m)) \leqslant \limsup_{k\to\infty} e^{-\sum_{m=n}^{k}\mathbb{P}(A_m)} = 0.$$ □

Насамкінець наведемо корисний факт, т. зв *Лему про $\pi-\lambda$ системи.* Щоб дати певну мотивацію, припустімо, що μ, ν є імовірнісними мірами в деякому вимірному просторі $(\Omega, \mathcal{F})$. Припустімо, ми хочемо показати, що ці міри рівні: хочемо перевірити, що для будь-якого $A \in \mathcal{F}$ виконується рівність

$$\mu(A) = \nu(A).$$

Виникає цілком природне запитання: чи достатньо перевірити цю тотожність для якогось конкретного класу подій A, наприклад, для генераторів σ-алгебри $\mathcal{F}$? Виявляється, що набір генераторів взагалі не є добрим вибором: а саме, ми повинні припустити, що цей клас додатково є π-системою.

Визначення 3.8. Припустимо, $\mathcal{K}$ є непорожнім класом підмножин Ω. Ми кажемо, що $\mathcal{K}$ є π-системою, якщо цей клас замкнуто відносно скінченних добутків: із A, $B \in \mathcal{K}$ випливає, що $A \cap B \in \mathcal{K}$.

Визначення 3.9. Припустимо, $\mathcal{L}$ є деяким класом підмножин Ω. Ми кажемо, що $\mathcal{L}$ є λ-системою, якщо виконуються такі умови:

(i) $\Omega \in \mathcal{L}$,

(ii) якщо A, $B \in \mathcal{L}$ і $A \subseteq B$, то $B \setminus A \in \mathcal{L}$,

(iii) якщо події A_1, A_2, … $\in \mathcal{L}$ не спадають, то $\bigcup_{n=1}^{\infty} A_n \in \mathcal{L}$.

Лема 3.2 (о π-λ системах)**.** *Якщо $\mathcal{L}$ є λ-системою, що містить π-систему $\mathcal{K}$, то $\mathcal{L}$ також містить σ-алгебру, породжену $\mathcal{K}$.*

Доведення. Розділимо доведення на три частини.

1) Зауважимо, що якщо $\mathcal{L}$ є λ-системою і A, $B \in \mathcal{L}$ несумісні, то як випливає з (i) і (ii)

$$A \cup B = (A' \setminus B)' \in \mathcal{L}.$$

2) Якщо $\mathcal{L}$ є і π-системою і λ-системою, то це σ-алгебра. Щоб довести це, зауважимо, що якщо A, $B \in \mathcal{L}$, то

$$A \cup B = A \cup (B \setminus (A \cap B)) \in \mathcal{L},$$

що випливає з 1) та визначень π-системи та λ-системи. Таким чином, за простою індукцією, $\mathcal{L}$ буде замкнуто через взяття скінченних сум. Тому якщо A_1, A_2, … є *будь-якою* послідовністю елементів з $\mathcal{L}$, то

$$\bigcup_{n=1}^{\infty} A_n = \bigcup_{n=1}^{\infty}\left(\bigcup_{k=1}^{n} A_k\right) \in \mathcal{L}.$$

На останньому кроці ми використовували те, що події A_1, A_2, … $\in \mathcal{L}$ не спадають.

3) Нехай Λ — клас усіх λ-систем, що містять $\mathcal{K}$. Покладемо $\mathcal{L}_0 = \bigcap_{\mathbf{L}\in\Lambda} \mathbf{L}$. Тоді $\mathcal{L}_0 \in \Lambda$ і $\mathcal{K} \subseteq \mathcal{L}_0 \subseteq \mathcal{L}$. Отже, достатньо довести, що $\mathcal{L}_0$ є σ-алгеброю: на підставі 2), достатньо показати, що $\mathcal{L}_0$ є π-системою. Візьмемо будь-яке $A \in \mathcal{K}$ і розглянемо клас

$$\mathcal{K}_1 = \{B : A \cap B \in \mathcal{L}_0\}.$$

Звичайно, $\mathcal{K}_1 \supseteq \mathcal{K}$, оскільки $\mathcal{K}$ є π-системою. Крім того, $\mathcal{K}_1$ є λ-системою:

(i) $\Omega \in \mathcal{K}_1$, тому що $A \cap \Omega = A \in \mathcal{K} \subseteq \mathcal{L}_0$;

(ii) якщо B_1, $B_2 \in \mathcal{K}_1$, $B_1 \subseteq B_2$, то

$$A \cap (B_2 \setminus B_1) = (A \cap B_2) \setminus (A \cap B_1) \in \mathcal{L}_0,$$

і тому $B_2 \setminus B_1 \in \mathcal{K}_1$;

(iii) якщо $B_1 \subseteq B_2 \subseteq \ldots \in \mathcal{K}_1$, то

$$A \cap \left(\bigcup_{n=1}^{\infty} B_n\right) = \bigcup_{n=1}^{\infty} (A \cap B_n) \in \mathcal{L}_0,$$

з чого випливає, що $\bigcup_{n=1}^{\infty} B_n \in \mathcal{K}_1$.

Таким чином, $\mathcal{K}_1$ містить $\mathcal{L}_0$, оскільки $\mathcal{L}_0$ є найменшою λ-системою, що містить $\mathcal{K}$. Отже, ми показали, що для будь-яких $A \in \mathcal{K}$ і будь-яких $B \in \mathcal{L}_0$, $A \cap B \in \mathcal{L}_0$. Далі ми повторюємо міркування: ми встановлюємо $B \in \mathcal{L}_0$ і визначаємо $\mathcal{K}_2 = \{A : A \cap B \in \mathcal{L}_0\}$. Маємо $\mathcal{K}_2 \supseteq \mathcal{K}$ і $\mathcal{K}_2$ є λ-системою, тому $\mathcal{K}_2 \supseteq \mathcal{L}_0$, одже для будь-яких A, $B \in \mathcal{L}_0$ виконується $A \cap B \in \mathcal{L}_0$. Доведення завершено. $\square$

У якості прикладу використання ми доведемо наступний факт.

Теорема 3.5. *Нехай $(\Omega, \mathcal{F}, \mathbb{P})$ є ймовірнісним простором. Припустимо, що родина $\{\mathcal{F}_\gamma\}_{\in\Gamma}$ незалежних σ-алгебр поділено на n підродин $\{\mathcal{F}_{\gamma^i}\}_{\gamma^i\in\Gamma^i}$, $i = 1, 2, \ldots, n$. Тоді σ-алгебри*

$$\sigma\left(\{\mathcal{F}_{\gamma^1}\}_{\gamma^1\in\Gamma^1}\right), \sigma\left(\{\mathcal{F}_{\gamma^2}\}_{\gamma^2\in\Gamma^2}\right), \ldots, \sigma\left(\{\mathcal{F}_{\gamma^n}\}_{\gamma^n\in\Gamma^n}\right),$$

породжені окремими підродинами також незалежні.

Доведення. Доведемо для $n = 2$, для більших n міркування аналогічні. Отже, ми маємо два сімейства $\{\mathcal{F}_\gamma\}_{\gamma\in\Gamma}$, $\{\mathcal{G}_\delta\}_{\delta\in\Delta}$ незалежних під-σ-алгебр $\mathcal{F}$, і ми повинні показати, що для будь-якого $A \in \sigma(\{\mathcal{F}_\gamma\}_{\gamma\in\Gamma})$, $B \in \sigma(\{\mathcal{G}_\delta\}_{\delta\in\Delta})$ виконується

$$(*) \qquad \mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B).$$

Завдяки незалежності σ-алгебр формула (*) справедлива для множин виду

$$A = A_{\gamma^1} \cap A_{\gamma^2} \cap \ldots \cap A_{\gamma^k}, \qquad A_{\gamma^i} \in \mathcal{F}_{\gamma^i},\, i = 1,\, 2,\, \ldots, k,$$

$$B = B_{\delta^1} \cap B_{\delta^2} \cap \ldots \cap B_{\delta^\ell}, \qquad B_{\delta^j} \in \mathcal{G}_{\delta^j},\, j = 1,\, 2,\, \ldots, \ell.$$

Встановимо A, як описано вище, і розглянемо клас $\mathcal{K} = \{B : \text{виконується (*)}\}$. Отже, ми маємо $\mathcal{K} \supseteq \{B_{\delta^1} \cap B_{\delta^2} \cap \ldots \cap B_{\delta^\ell} : B_{\delta^j} \in \mathcal{G}_{\delta^j}\}$, а останній клас, звичайно, є π-системою, що генерує $\sigma(\{\mathcal{G}_\delta\}_{\delta\in\Delta})$. Крім того, $\mathcal{K}$ є λ-системою:

(i) $\Omega \in \mathcal{K}$, тому що $\mathbb{P}(A \cap \Omega) = \mathbb{P}(A)\mathbb{P}(\Omega)$.

(ii) якщо $B_1,\, B_2 \in \mathcal{K}$ і $B_1 \subseteq B_2$, то $B_2 \setminus B_1 \in \mathcal{K}$:

$$\begin{aligned}\mathbb{P}(A \cap (B_2 \setminus B_1)) &= \mathbb{P}\big((A \cap B_2) \setminus (A \cap B_1)\big)\\ &= \mathbb{P}(A \cap B_2) - \mathbb{P}(A \cap B_1)\\ &= \mathbb{P}(A)\big(\mathbb{P}(B_2) - \mathbb{P}(B_1)\big) = \mathbb{P}(A)\mathbb{P}(B_2 \setminus B_1).\end{aligned}$$

(iii) Якщо $B_1 \subseteq B_2 \subseteq \ldots \in \mathcal{K}$, то $\bigcup_{n=1}^{\infty} B_n \in \mathcal{K}$, оскільки за теоремою неперервності,

$$\begin{aligned}\mathbb{P}\left(A \cap \left(\bigcup_{n=1}^{\infty} B_n\right)\right) &= \mathbb{P}\left(\bigcup_{n=1}^{\infty} (A \cap B_n)\right)\\ &= \lim_{n\to\infty} \mathbb{P}(A \cap B_n)\\ &= \lim_{n\to\infty} \mathbb{P}(A)\mathbb{P}(B_n) = \mathbb{P}(A)\mathbb{P}\left(\bigcup_{n=1}^{\infty} B_n\right).\end{aligned}$$

Таким чином, як випливає з Леми π-λ, $\mathcal{K}$ містить $\sigma(\{\mathcal{G}_\delta\}_{\delta\in\Delta})$, тобто (*) виконується для будь-якої множини A у вигляді $A_{\gamma^1} \cap A_{\gamma^2} \cap \ldots \cap A_{\gamma^k}$ і $B \in \sigma(\{\mathcal{G}_\delta\}_{\delta\in\Delta})$. Потім ми повторюємо міркування: ми встановлюємо $B \in \sigma(\{\mathcal{G}_\delta\}_{\delta\in\Delta})$ і визначаємо $\mathcal{L} = \{A : \text{виконується (*)}\}$. Клас $\mathcal{L}$ містить усі множини виду $A_{\gamma^1} \cap A_{\gamma^2} \cap \ldots \cap A_{\gamma^k}$, які складають π-систему, що генерує $\sigma(\{\mathcal{F}_\gamma\}_{\gamma\in\Gamma})$. Як і вище, ми перевіряємо, що $\mathcal{L}$ є λ-системою, тому з Леми π-λ випливає $\mathcal{L} \supseteq \sigma(\{\mathcal{F}_\gamma\}_{\gamma\in\Gamma})$. Отже (*) виконується для всіх $A \in \sigma(\{\mathcal{F}_\gamma\}_{\gamma\in\Gamma})$ і $B \in \sigma(\{\mathcal{G}_\delta\}_{\delta\in\Delta})$, звідки ми отримуємо бажану незалежність σ-алгебр. $\square$

4 Завдання

1. Група з n осіб ($n \geqslant 3$), яка включає осіб X, Y і Z, випадково ставлять у чергу. Яка ймовірність того, що

а) X стоїть безпосередньо перед Y, якщо Y стоїть безпосередньо перед Z?

б) X стоїть перед Y, якщо Y стоїть перед Z?

2. З колоди з 52 карт витягується 5 карт без заміни. Обчисліть ймовірність того, що у нас рівно 3 тузів, якщо ми знаємо, що

а) у нас є хоча б один туз;

б) маємо туза чорної масті;

в) маємо пікового туза;

г) перша витягнута карта – туз;

д) перша витягнута карта – чорний туз;

е) перша витягнута карта — піковий туз.

3. В урні три білі та чотири чорні кульки. Витягуємо кульку навмання, викидаємо її, не дивлячись на неї, а потім витягуємо з урни ще одну кульку.

а) Яка ймовірність того, що друга кулька біла?

б) Припустимо, друга кулька біла. Яка ймовірність того, що перша кулька чорна?

4. У популяції є 15% осіб хворих на дислексію. Якщо учень робить 6 і більше помилок на діагностичному тесті, він або вона вважається особою хворою на дислексію. Кожна людина з дислексією напевно зробить щонайменше 6 помилок у цьому тесті, але навіть людина без дислексії може зробити більше 5 помилок - це відбувається з імовірністю 0,1. Ян зробив 6 помилок у тесті. Яка ймовірність того, що він має дислексію? Яка ймовірність того, що він також зробить принаймні 6 помилок на наступному тесті?

5. На певній фабриці з вироблення телевізорів кожен телевізор може мати дефект з імовірністю p. На фабриці є три контрольні станції, і вироблений телевізор потрапляє на кожну зі станцій з однаковою ймовірністю. i-та станція виявляє несправний телевізор з імовірністю p_i ($i = 1, 2, 3$). Телевізори, не відбраковані на заводі, надходять до оптовика і там проходять додатковий контроль, який виявляє бракований телевізор з імовірністю p_0.

а) Обчислити ймовірність того, що даний телевізор нового виробництва буде у продажу (тобто пройде обидві перевірки).

б) Припустимо, телевізор вже є в магазині. Яка ймовірність того, що він несправний?

6. Кидаємо кубик двічі. Розглянемо події: A – перше число ділиться на 3; B – сума кинутих кубиків парна; C – отримано два рівних числа. Чи є події A, B незалежними? A, B, C незалежні?

7. n карток містить n різних дійсних чисел. Картки поміщали в коробку, ретельно перемішували, а потім випадковим чином витягували без заміни. Нехай A_k – k-те витягнуте число більше за попередні.

а) Доведіть, що $\mathbb{P}(A_k) = 1/k$, $k = 1, 2, \ldots, n$.

b) Доведіть, що події $A_1, A_2, \ldots, A_n$ незалежні.

8. Дано натуральні числа m, n і числа p, $q \in (0, 1)$, що задовольняють умову $p + q = 1$. Доведіть, що

$$(1 - p^n)^m + (1 - q^m)^n \geqslant 1.$$

9. Знайдіть найбільш ймовірну кількість успіхів у схемі n випробувань Бернуллі з імовірністю успіху p.

10. Гральний кубик кинули 10 разів. Яка ймовірність того, що шістка випаде під час першого кидка, якщо відомо, що

а) викинуто три шістки?

б) у наступних дев'яти випробуваннях було викинуто тільки шістки?

11. Кидаємо кубик, поки не викинемо п'ятірку або тричі шістку (всього, не обов'язково поспіль). Яка ймовірність того, що ми підкинемо рівно n разів?

12. Імовірність того, що в урні k кубів, становить $\frac{2^k}{k!}e^{-2}$, $k = 0, 1, 2, \ldots$. Ми випадковим чином витягуємо по колії всі кубики до останнього, і кидаємо кожен кубик. Яка ймовірність отримати I шісток?

13. Яка ймовірність того, що кількість успіхів у n випробуваннях Бернуллі з $p = \frac{1}{2}$ ділиться

а) на 3?

б) на 4?

14. Нехай $(\Omega, \mathcal{F}, \mathbb{P})$ — імовірнісний простір схеми n випробувань Бернуллі з імовірністю успіху p. Для будь-якого $0 \leqslant k \leqslant n$ нехай A_k одначає подію „мами рівно k успіхів". Доведіть, що для будь-яких $B \in \mathcal{F}$ і будь-яких k умовна ймовірність $\mathbb{P}(B|A_k)$ не залежить від p.

15. Кидаємо нескінченну кількість разів монету, для якої імовірність отримати голову становить $p \neq 1/2$. Для $n = 2, 4, \ldots$ розглянемо подію A_n - до n-го підкидання включно число голів дорівнює числу решек. Доведіть, що з імовірністю 1 відбудеться скінченна кількість подій A_1, A_2, $\ldots$.

16. Кидаємо нескінченну кількість разів монету, для якої імовірність отримати голову становить $p \in (0, 1]$. Доведіть, що з імовірністю 1 існує нескінченна кількість серій що складаються зі 100 голів поспіль.

17. Дано дві ймовірнісні міри μ, ν на $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$.

а) Припустимо, що для будь-якого числа $t > 0$ маємо $\nu([-t, t]) = \mu([-t, t])$. Доведіть, що якщо $A \in \mathcal{B}(\mathbb{R})$ є симетричним відносно 0, то $\mu(A) = \nu(A)$.

b) Припустимо, що $\mathcal{K}$ є деяким класом, який породжує $\mathcal{B}(\mathbb{R})$ (тобто задовольняє $\sigma(\mathcal{K}) = \mathcal{B}(\mathbb{R})$) . Чи з того що $\mu(A) = \nu(A)$ для кожного $A \in \mathcal{K}$, випливає рівність $\mu = \nu$?

5 Випадкові величини та їх розподіли

Ми підійшли до ключової концепції теорії ймовірностей.

Визначення 5.1. Нехай $(\Omega, \mathcal{F}, \mathbb{P})$ є ймовірнісним простором. d-вимірна випадкова величина (або d-вимірна випадкова змінна) — це будь-яка вимірна функція $X : \Omega \to \mathbb{R}^d$ (тобто X задовольняє умову: для будь-якого $B \in \mathcal{B}(\mathbb{R}^d)$ маємо $\{X \in B\} := X^{-1}(B) = \{\omega \in \Omega : X(\omega) \in B\} \in \mathcal{F}$). Зокрема, для $d = 1$ ми просто скажемо „випадкова змінна”.

Теорема 5.1. *Якщо $(\Omega, \mathcal{F}, \mathbb{P})$ є імовірнісним простором, X є d-вимірною випадковою змінною, а $f : \mathbb{R}^d \to \mathbb{R}^k$ є борелівською функцією, то $f(X)$ є k-вимірною випадковою змінною. Зокрема, теза справедлива для неперервних функцій f.*

Зокрема, ми бачимо, що якщо $X = (X_1, X_2, \ldots, X_d)$ – d-вимірна випадкова величина, то X_1, X_2, $\ldots$, X_d є одновимірними випадковими величинами.

Ще одна важлива концепція – розподіл випадкової величини. Нехай $(\Omega, \mathcal{F}, \mathbb{P})$ — імовірнісний простір, а X — d-вимірна випадкова величина. Для будь-якого $B \in \mathcal{B}(\mathbb{R}^d)$ визначимо

$$P_X(B) = \mathbb{P}(X \in B) := \mathbb{P}(\{\omega \in \Omega : X(\omega) \in B\})$$

(тобто P_X є образом міри $\mathbb{P}$ при перетворенні X). Легко перевірити, що $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d), P_X)$ є новим ймовірнісним простором.

Визначення 5.2. Міра P_X називається *розподілом* X.

Визначення 5.3. Нехай $(\Omega, \mathcal{F}, \mathbb{P})$ — імовірнісний простір, а X — d-вимірна випадкова величина. *Функцією розподілу ймовірностей* цієї випадкової величини є функція $F_X : \mathbb{R}^d \to [0, 1]$, визначена формулою

$$\begin{aligned} F_X(x_1, x_2, \ldots, x_d) &= \mathbb{P}(X_1 \leqslant x_1, X_2 \leqslant x_2, \ldots, X_d \leqslant x_d) \\ &= P_X((-\infty, x_1] \times (-\infty, x_2] \times \ldots \times (-\infty, x_d]). \end{aligned}$$

Приклади:

1. Один раз підкидаємо симетричну монету; $\Omega = \{O, R\}$. Нехай $X(O) = 1$, $X(R) = -1$. Тоді

$$F_X(t) = \mathbb{P}(X \leqslant t) = \begin{cases} 0 & \text{якщо } t < -1, \\ 1/2 & \text{якщо } -1 \leqslant t < 1, \\ 1 & \text{якщо } t \geqslant 1. \end{cases}$$

2. Вибираємо випадковим чином точку з кола радіусом R: маємо $(\Omega, \mathcal{F}, \mathbb{P}) = \left(K(0, R), \mathcal{B}(K(0, R))), \frac{|\cdot|}{\pi R^2}\right)$. Нехай $X(\omega) = \rho(\omega, 0)$ буде відстанню випадкової точки від центру. Маємо

$$F_X(t) = \mathbb{P}(X \leqslant t) = \begin{cases} 0 & \text{якщо } t < 0, \\ t^2/R^2 & \text{якщо } 0 \leqslant t < R, \\ 1 & \text{якщо } t \geqslant R. \end{cases}$$

Теорема 5.2 (Властивості функції розподілу)**.** *Нехай X є (одновимірною) випадковою величиною. Тоді*

а) F_X – неспадна.

б) $\lim_{t\to\infty} F_X(t) = 1$, $\lim_{t\to-\infty} F_X(t) = 0$.

в) F_X – неперервна зправа.

г) для будь-якого $t \in \mathbb{R}$ існує границя зліва $F_X(t-) = \lim_{s\uparrow t} F_X(s)$ і

$$F_X(t-) = \mathbb{P}(X < t) = \mathbb{P}(X \in (-\infty, t)).$$

д) F_X розривна в t_0 тоді і тільки тоді, коли $\mathbb{P}(X = t_0) > 0$ (це число t_0 тоді називаєтьса атомом розподілу). Точніше, для будь-якого $t_0 \in \mathbb{R}$ ми маємо $\mathbb{P}(X = t_0) = F_X(t_0) - F_X(t_0-)$.

е) для будь-яких $a < b$, маємо

$$\begin{aligned}
\mathbb{P}(a \leqslant X \leqslant b) &= \mathbb{P}(X \in [a, b]) = F_X(b) - F_X(a-),\\
\mathbb{P}(a < X \leqslant b) &= \mathbb{P}(X \in (a, b]) = F_X(b) - F_X(a),\\
\mathbb{P}(a \leqslant X < b) &= \mathbb{P}(X \in [a, b)) = F_X(b-) - F_X(a-),\\
\mathbb{P}(a < X < b) &= \mathbb{P}(X \in (a, b)) = F_X(b-) - F_X(a).
\end{aligned}$$

Доведення. а) Якщо $t_1 < t_2$, то $(-\infty, t_1] \subset (-\infty, t_2]$, а więc $F_X(t_1) = P_X((-\infty, t_1]) \leqslant P_X((-\infty, t_2]) = F_X(t_2)$.

б) Для будь-якої послідовності $(t_n)_{n\geqslant 1}$, що прямує до нескінченності, $\mathbb{R} = \bigcup_n(-\infty, t_n]$, а інтервали під сумою є зростаючими. Отже, використовуючи теорему неперервності,

$$1 = P_X(\mathbb{R}) = \lim_{n\to\infty} P_X((-\infty, t_n]) = \lim_{n\to\infty} F_X(t_n).$$

Другу частину доводимо аналогічно.

в) Візьмемо $t \in \mathbb{R}$ і послідовність $(t_n)_{n\geqslant 1}$, що збігається до t. Послідовність інтервалів $(t, t_n]$ спадає і $\bigcap_n(t, t_n] = \emptyset$, тому

$$\begin{aligned}
0 = P_X\left(\bigcap_n (t, t_n]\right) &= \lim_{n\to\infty} P_X((t, t_n]) = \lim_{n\to\infty} P_X((-\infty, t_n] \setminus (-\infty, t])\\
&= \lim_{n\to\infty}\Big[P_X((-\infty, t_n]) - P_X((-\infty, t])\Big] = \lim_{n\to\infty}(F_X(t_n) - F_X(t)).
\end{aligned}$$

г) Доведення подібне до в).

д) Безпосередньо випливає з г) і рівності $\mathbb{P}(X = t_0) = \mathbb{P}(X \leqslant t_0) - \mathbb{P}(X < t_0)$.

е) Доведемо лише першу рівність, інші перевіряємо аналогічно:

$$\mathbb{P}(a \leqslant X \leqslant b) = P_X([a, b]) = P_X((-\infty, b]) - P_X((-\infty, a)) = F_X(b) - F_X(a-).$$ □

Зауваження: Якщо функція F задовольняє умовам а), б) і в) наведеної вище теореми, то вона є функцією розподілу (доведення залишаємо як вправу).

Теорема 5.3 (Теорема єдиності)**.** *Функція розподілу d-вимірної випадкової величини однозначно визначає розподіл.*

Доведення. Припустимо, що X, Y — випадкові величини з однаковою функцією розподілу. Ми хочемо показати, що $P_X = P_Y$, тобто

$$(*) \qquad P_X(B) = P_Y(B)$$

для будь-якої борелівської підмножини $\mathbb{R}^d$. Клас усіх множин B, що задовольняють (*), утворює λ-систему. З іншого боку, рівність функцій розподілу дає

$$\begin{aligned} P_X((-\infty, x_1] \times (-\infty, x_2] \times \ldots \times (-\infty, x_d]) &= F_X(x) = F_Y(x) \\ &= P_Y((-\infty, x_1] \times (-\infty, x_2] \times \ldots \times (-\infty, x_d]), \end{aligned}$$

таким чином (*) виконується для множин виду $(-\infty, x_1] \times (-\infty, x_2] \times \ldots (-\infty, x_d]$, які складають π-систему, що генерує $\mathcal{B}(\mathbb{R}^d)$. Отже, теорема випливає безпосередньо з Леми π-λ. $\square$

Визначення 5.4. Припустимо, $X = (X_1, X_2, \ldots, X_d)$ є d-вимірною випадковою величиною і $1 \leqslant i_1 < i_2 < \ldots < i_k \leqslant d$ $(k < d)$. Тоді $(X_{i_1}, X_{i_2}, \ldots, X_{i_k})$ також є випадковою величиною, і її розподіл називається k-вимірним граничним розподілом розподілу X.

Зауваження: Граничні розподіли зазвичай не визначають спільний розподіл. Розглянемо такий приклад:

I - ми підкидаємо монету тричі і нехай X_1 буде кількістю решек до появи перших голів, а X_2 — загальна кількість голів.

II - ми робимо дві серії підкидань монет по три рази і нехай X_1' - кількість решек до появи перших голів в першій серії, а X_2' - загальна кількість голів у другій серії.

Очевидно, що X_1 і X_1' мають однаковий розподіл; так само X_2 і X_2' мають однаковий розподіл. З іншого боку, змінні (X_1, X_2) і (X_1', X_2') *не* мають однаковий розподіл: дійсно,

$$P_{(X_1, X_2)}(\{(3,3)\}) = 0 \quad \text{oraz} \quad P_{(X_1', X_2')}(\{(3,3)\}) = 2^{-6}.$$

Таким чином, граничні розподіли (X_1, X_2) і (X_1', X_2') ідентичні, але спільні розподіли різні.

У решті лекції ми будемо використовувати наступне позначення: якщо X є d-вимірною випадковою величиною, то

$$\sigma(X) := \{\{X \in B\} : B \in \mathcal{B}(\mathbb{R}^d)\}$$

це σ-алгебра подій, що генерує X.

Визначення 5.5. Нехай $(\Omega, \mathcal{F}, \mathbb{P})$ — імовірнісний простір, а $\{X_i\}_{i \in I}$ — родина випадкових величин, де X_i приймає значення в $\mathbb{R}^{d_i}$. Ми говоримо, що ці змінні незалежні, якщо σ-алгебри, породжені цими змінними, незалежні. Іншими словами, $\{X_i\}_{i \in I}$ незалежні тоді і тільки тоді коли для кожного n та будь-яких попарно різних $i_1, i_2, \ldots, i_n \in I$ і будь-яких $B_1 \in \mathcal{B}(\mathbb{R}^{d_{i_1}}), \ldots, B_n \in \mathcal{B}(\mathbb{R}^{d_{i_n}})$,

$$\mathbb{P}(X_{i_1} \in B_1, X_{i_2} \in B_2, \ldots, X_{i_n} \in B_n) = \mathbb{P}(X_{i_1} \in B_1)\mathbb{P}(X_{i_2} \in B_2) \ldots \mathbb{P}(X_{i_n} \in B_n).$$

Приклади:

1. Події $A_1, A_2, \ldots, A_n$ незалежні тоді і тільки тоді, коли $1_{A_1}, 1_{A_2}, \ldots, 1_{A_n}$ вони незалежні. Дійсно, це безпосередньо випливає з того факту, що для будь-якої події A і будь-якої борелівської підмножини прямої маємо

$$\{1_A \in B\} = \begin{cases} A & \text{якщо } 0 \notin B,\, 1 \in B, \\ A' & \text{якщо } 0 \in B,\, 1 \notin B, \\ \emptyset & \text{якщо } 0,\, 1 \notin B, \\ \Omega & \text{якщо } 0,\, 1 \in B. \end{cases}$$

2. У схемі n випробувань Бернуллі визначимо

$$X_i(\omega) = X_i(\omega_1, \omega_2, \ldots, \omega_n)$$
$$:= \omega_i = \begin{cases} 1 & \text{якщо в } i\text{-му випробуванні був успіх,} \\ 0 & \text{якщо в } i\text{-му випробуванні була невдача.} \end{cases}$$

Тоді X_1, X_2, $\ldots$, X_n є незалежними випадковими величинами. Це буде зрозуміло з фактів, які ми наведемо пізніше в лекції.

Обговоримо тепер деякі умови, які еквівалентні незалежності випадкових величин. Заради простоти ми зосередимося на одновимірному випадку, але наступна теорема також справедлива, коли змінні приймають векторні значення.

Теорема 5.4. *Припустимо, X_1, X_2, $\ldots$, X_n є випадковими величинами. Наступні умови еквівалентні.*

1) X_1, X_2, $\ldots$, X_n незалежні.

2) Для будь-яких B_1, B_2, $\ldots$, $B_n \in \mathcal{B}(\mathbb{R})$ $\{X_1 \in B_1\}$, $\{X_2 \in B_2\}$, $\ldots$, $\{X_n \in B_n\}$ незалежні.

3) $P_{(X_1,X_2,\ldots,X_n)} = P_{X_1} \otimes P_{X_2} \otimes \ldots \otimes P_{X_n}$,

4) Для будь-якого $x = (x_1, x_2, \ldots, x_n) \in \mathbb{R}^n$,

$$F_{(X_1,X_2,\ldots,X_n)}(x_1, x_2, \ldots, x_n) = F_{X_1}(x_1)F_{X_2}(x_2)\ldots F_{X_n}(x_n).$$

Доведення. Ми вже знаємо, що умови 1) і 2) еквівалентні.

2)⇒4). Маємо

$$F_{(X_1,X_2,\ldots,X_n)}(x_1, x_2, \ldots, x_n) = \mathbb{P}(X_1 \leqslant x_1, X_2 \leqslant x_2, \ldots, X_n \leqslant x_n)$$
$$= \mathbb{P}(X_1 \leqslant x_1)\mathbb{P}(X_2 \leqslant x_2)\ldots\mathbb{P}(X_n \leqslant x_n) = F_{X_1}(x_1)F_{X_2}(x_2)\ldots F_{X_n}(x_n).$$

4)⇒3) Нехай X' — n-вимірна випадкова величина, $P_{X_1} \otimes P_{X_2} \otimes \ldots \otimes P_{X_n}$. Для будь-якого $x = (x_1, x_2, \ldots, x_n)$ маємо

$$F_{X'}(x_1, x_2, \ldots, x_n) = P_{X_1} \otimes \ldots \otimes P_{X_n}\big((-\infty, x_1] \times (-\infty, x_2] \times \ldots \times (-\infty, x_n]\big)$$
$$= F_{X_1}(x_1)F_{X_2}(x_2)\ldots F_{X_n}(x_n) = F_{(X_1,X_2,\ldots,X_n)}(x_1, x_2, \ldots, x_n).$$

За теоремою єдиності з цього випливає $P_{X'} = P_{(X_1,X_2,\ldots,X_n)}$.

3)⇒1) Для будь-яких борелівських підмножин B_1, B_2, $\ldots$, B_n дійсної прямої маємо

$$\mathbb{P}(X_1 \in B_1, X_2 \in B_2, \ldots, X_n \in B_n) = P_{(X_1,X_2,\ldots,X_n)}(B_1 \times B_2 \times \ldots \times B_n)$$
$$= P_{X_1}(B_1)P_{X_2}(B_2)\ldots P_{X_n}(B_n) = \mathbb{P}(X_1 \in B_1)\mathbb{P}(X_2 \in B_2)\ldots\mathbb{P}(X_n \in B_n).$$

Доведення завершено. □

Перейдемо до більш детального обговорення типів і прикладів розподілів ймовірностей у $\mathbb{R}^d$.

Визначення 5.6. Ми кажемо, що d-вимірна випадкова величина X має дискретний (ступінчастий, атомарний) розподіл, якщо $\mathbb{P}(X \in S_X) = 1$, де

$$S_X = \{x \in \mathbb{R}^d : \mathbb{P}(X = x) > 0\}$$

це набір *атомів* розподілу.

Примітки:

1) Для будь-якої d-вимірної змінної X множина S_X не більше ніж зліченна, оскільки

$$S_X = \bigcup_{n\geqslant 1} \{x \in \mathbb{R}^d : \mathbb{P}(X = x) > 1/n\},$$

і кожна з множин, що фігурують під сумою, є скінченною.

2) Дискретний розподіл однозначно визначається щонайбільше зліченною множиною $S \subset \mathbb{R}^d$ і функцією $p : S \to [0,1]$ такою, що $\sum_{x\in S} p(x) = 1$. Дійсно, тоді

$$P(B) = \sum_{x\in S\cap B} p(x).$$

Відзначимо дуже простий факт.

Теорема 5.5. *Випадкові величини X_1, X_2, ..., X_n мають дискретний розподіл тоді і тільки тоді, коли змінна $(X_1, X_2, \ldots, X_n)$ має дискретний розподіл.*

У випадку скінченної родини дискретних змінних умова незалежності може бути перевірена наступним простим критерієм.

latwo **Теорема 5.6.** *Випадкові величини X_1, X_2, ..., X_n з дискретними розподілами незалежні тоді і тільки тоді, коли для будь-яких $x_1 \in S_{X_1}$, $x_2 \in S_{X_2}$, ..., $x_n \in S_{X_n}$ виконується*

$$\mathbb{P}(X_1 = x_1, X_2 = x_2, \ldots, X_n = x_n) = \mathbb{P}(X_1 = x_1)\mathbb{P}(X_2 = x_2)\ldots\mathbb{P}(X_n = x_n).$$

Доведення. $\Rightarrow$ Очевидно: постульована рівність є окремим випадком умови у визначенні незалежності.

$\Leftarrow$ Для зручності позначення ми доведемо лише для $n = 2$; випадок $n \geqslant 3$ розглядається аналогічно. Маємо

$$\begin{aligned}
&\mathbb{P}(X_1 \in B_1, X_2 \in B_2) = \mathbb{P}(X_1 \in S_{X_1} \cap B_1, X_2 \in S_{X_2} \cap B_2) \\
&= \sum_{x_1\in S_{X_1}\cap B_1, x_2\in S_{X_2}\cap B_2} \mathbb{P}(X_1 = x_1, X_2 = x_2) \\
&= \sum_{x_1\in S_{X_1}\cap B_1} \sum_{x_2\in S_{X_2}\cap B_2} \mathbb{P}(X_1 = x_1)\mathbb{P}(X_2 = x_2) = \mathbb{P}(X_1 \in B_1)\mathbb{P}(X_2 \in B_2). \square
\end{aligned}$$

Приклади:

1) *Розподіл зосереджений в $a \in \mathbb{R}^d$*, позначається δ_a. Змінна X має розподіл зосереджений в a, якщо $\mathbb{P}(X = a) = 1$; $S_X = \{a\}$.

2) *Двоточковий розподіл* зосереджений в a, $b \in \mathbb{R}^d$, $a \neq b$. X має двоточковий розподіл зосереджений в $\{a, b\}$, якщо $\mathbb{P}(X = a) = p$ і $\mathbb{P}(X = b) = 1 - p$ для деякого $p \in (0, 1)$.

3) *Розподіл Бернуллі (біноміальний розподіл)* з параметрами n, p ($n = 1, 2, \ldots, p \in (0,1)$), позначається $B(n, p)$. X ма розподіл $B(n, p)$, якщо $\mathbb{P}(X = k) = \binom{n}{k} p^k (1 - p)^{n-k}$, $k \in S_X = \{0, 1, 2, \ldots, n\}$. Іншими словами, якщо задана схема n випробувань Бернуллі з імовірністю успіху p, то загальна кількість успіхів має розподілу Бернуллі $B(n, p)$.

Відступ. Припустимо, що змінні X_1, X_2, ..., X_n задано формулою

$$(*) \qquad X_i = \begin{cases} 1 & \text{якщо в } i\text{-тому випробуванні бул успіх,} \\ 0 & \text{інакше.} \end{cases}$$

Тоді, як це легко перевірити за допомогою Теореми latwo 5.6, змінні $X_1, X_2, \ldots, X_n$ є незалежними. З іншого боку, $X_1 + X_2 + \ldots + X_n$ — це загальна кількість успіхів. Отже, ми отримали такий факт:

Нехай $X_1, X_2, \ldots, X_n$ – незалежні випадкові величини з однаковим розподілом (іноді ці дві властивості будуть скорочено називатися i.i.d.), заданими як $\mathbb{P}(X_i = 1) = p = 1-\mathbb{P}(X_i = 0)$. Тоді $X_1+X_2+\ldots+X_n$ має розподіл $B(n, p)$.

Ми використали тут просте спостереження:

Теорема 5.7. *Якщо d-вимірні змінні X, Y мають однаковий розподіл, а $f : \mathbb{R}^d \to \mathbb{R}^k$ є функцією Бореля, то $f(X)$ і $f(Y)$ теж мають однаковий розподіл.*

4) *Геометричний розподіл із параметром p ($0 < p < 1$)*, позначається $\mathrm{Geom}(p)$. Змінна X має розподіл $\mathrm{Geom}(p)$, якщо

$$\mathbb{P}(X = k) = (1-p)^k p, \qquad k \in S_X = \{0, 1, 2, \ldots\}.$$

Інтерпретація: припустимо, нам дано схему Бернуллі з нескінченною кількістю випробувань і ймовірністю успіху p. Нехай X буде кількістю невдач перед першим успіхом. Тоді X має геометричний розподіл з параметром p. Дійсно, ввівши змінні $X_1, X_2, \ldots$, як у попередньому прикладі, ми можемо написати

$$\begin{aligned}\mathbb{P}(X = k) &= \mathbb{P}(X_1 = 0, X_2 = 0, \ldots, X_k = 0, X_{k+1} = 1)\\ &= \mathbb{P}(X_1 = 0)\mathbb{P}(X_2 = 0)\ldots\mathbb{P}(X_k = 0)\mathbb{P}(X_{k+1} = 1) = (1-p)^k p.\end{aligned}$$

Іноді в літературі геометричний розподіл визначається трохи інакше: а саме $Y \sim$Geom(p), якщо для будь-якого $k \in S_Y = \{1, 2, \ldots\}$, ми маємо $\mathbb{P}(Y = k) = (1-p)^{k-1}p$. Тоді ця змінна інтерпретується як *час очікування першого успіху в нескінченній схемі Бернуллі з імовірністю p*. Це пов'язано з попередньою змінною таким чином: $Y = X + 1$.

5) *Розподіл Пуассона з параметром λ* ($\lambda > 0$), позначається $\mathrm{Pois}(\lambda)$. Змінна X має такий розподіл, якщо

$$\mathbb{P}(X = k) = \frac{\lambda^k}{k!}e^{-\lambda}, \qquad k \in S_X = \{0, 1, 2, \ldots\}.$$

Розподіл Пуассона виникає через відповідний перехів до границі для розподілів Бернуллі. Точніше, справедливий такий факт (без доведення).

Теорема 5.8 (Теорема Пуассон)**.** *Нехай $(p_n)_{n\geqslant 1}$ є послідовністю чисел в інтервалі $(0,1)$, що задовольняє умову $\lim_{n\to\infty} np_n = \lambda > 0$. Тоді для будь-якого числа $k \in \{0, 1, 2, \ldots\}$,*

$$\lim_{n\to\infty} \binom{n}{k} p_n^k (1-p_n)^{n-k} = \frac{\lambda^k}{k!}e^{-\lambda}.$$

Іншими словами, якщо X_n має розподіл $B(n,p)$, X має розподіл $\mathrm{Pois}(\lambda)$ і $np \approx \lambda$, то розподіли X_n і X близькі.

Переходимо до наступного важливого розділу розподілів.

Визначення 5.7. Ми кажемо, що d-вимірна випадкова величина X є неперервною, якщо існує борелівська функція $g : \mathbb{R}^d \to [0, \infty)$ така, що для будь-якої борелівської підмножини $B \subseteq\ R^d$,

$$\mathbb{P}(X \in B) = P_X(B) = \int_B g(x)\mathrm{d}x.$$

Функція g називається *щільністю розподілу.*

Зауваження:

1. Якщо розподіл P_X неперервний, то $S_X = \emptyset$, але не навпаки (необхідною і достатньою умовою є $\mathbb{P}(X \in B) = 0$ для будь-якого $B \subset \mathbb{R}^d$ з нульовою мірою Лебега).

2. Якщо $g : \mathbb{R}^d \to [0, \infty)$ — щільність деякого розподілу μ і $\tilde{g} : \mathbb{R}^d \to [0, \infty)$ є борелівською функцією, тоді $\tilde{g}$ є щільністю μ тоді і тільки тоді, коли $g = \tilde{g}$ м.в. Дійсно, імплікація $\Leftarrow$ очевидна:

$$\mu(A) = \int_A g(x)\mathrm{d}x = \int_A \tilde{g}(x)\mathrm{d}x.$$

Переходимо до імплікації $\Rightarrow$: маємо $\int_B g = \int_B \tilde{g}$ для будь-якої борелівської множини B. Припустимо, що теза не виконується; тоді одна з множин $\{g > \tilde{g}\}$, $\{g < \tilde{g}\}$ має додатну міру. Без втрати загальності припустимо, що це перша множина, і позначимо її через B. Це борелівська множина, і $\int_B g > \int_B \tilde{g}$, отримуємо суперечність.

3. Кожна борелівська функція $g : \mathbb{R}^d \to [0, \infty)$ така, що $\int_{\mathbb{R}^d} g(x)\mathrm{d}x = 1$ є щільністю деякого розподілу, що задано наступним чином: $\mu(B) = \int_B g(x)\mathrm{d}x$ для $B \in \mathcal{B}(\mathbb{R}^d)$.

4. Величина X неперервно розподілена тоді і тільки тоді, коли існує борелівська функція $g : \mathbb{R}^d \to [0, \infty)$ така, що

$$F_X(x_1, x_2, \ldots, x_d) = \int_{-\infty}^{x_1} \int_{-\infty}^{x_2} \cdots \int_{-\infty}^{x_d} g(y_1, y_2, \ldots, y_d)\mathrm{d}y_d\mathrm{d}y_{d-1} \ldots \mathrm{d}y_1.$$

Справді:

$\Rightarrow$ Маємо $F(x_1, \ldots, x_d) = \mathbb{P}((X_1, \ldots, X_d) \in (-\infty, x_1] \times \ldots \times (-\infty, x_d])$ і далі достатньо використати визначення щільності.

$\Leftarrow$ Легко перевірити, що $\int_{\mathbb{R}^d} g(x)\mathrm{d}x = 1$. Нехай μ — розподіл ймовірностей в $\mathbb{R}^d$, що задано щільністю g (див. Примітку 3 вище). Тоді, за попередньою імплікацією, ми маємо $F_X = F_\mu$, а отже, за теоремою єдиності, $P_X = \mu$.

5. Як висновок із попереднього зауваження отримуємо наступний факт.

Теорема 5.9. *Нехай X є d-вимірною випадковою величиною з функцією розподілу F, і нехай*

$$g(x_1, x_2, \ldots, x_d) = \frac{\partial^d F}{\partial x_1 \partial x_2 \ldots \partial x_d}(x_1, x_2, \ldots, x_d) \qquad \text{м.в.}$$

Тоді, якщо $\int_{\mathbb{R}^d} g = 1$, то X має неперервний розподіл і його щільність дорівнює g.

Приклад 5.1. Вибираємо навмання точку з кола радіуса R. Нехай X є відстанню точки від центра кола. Тоді, як ми вже знаємо,

$$F_X(t) = \begin{cases} 0 & \text{для } t < 0, \\ t^2/R^2 & \text{для } 0 \leqslant t < R, \\ 1 & \text{для } t \geqslant R. \end{cases}$$

Диференціюючи, отримуємо

$$g(t) = \begin{cases} 0 & \text{для } t < 0 \text{ lub } t \geqslant R, \\ 2t/R^2 & \text{для } 0 \leqslant t < R. \end{cases}$$

Легко перевірити, що $\int_{\mathbb{R}} g = 1$, тому g є щільністю розподілу X.

Теорема 5.10. *Якщо розподіл $X=(X_1,X_2,\ldots,X_d)$ неперервний, то його граничні розподіли також неперервні. Крім того*

$$g_{X_i}(x)=\int_{\mathbb{R}^{d-1}} g(x_1,x_2,\ldots,x_{i-1},x,x_{i+1},\ldots,x_d)dx_1\ldots dx_{i-1}dx_{i+1}\ldots dx_d.$$

Загалом, щоб отримати щільність вектора $(X_{i_1},X_{i_2},\ldots,X_{i_k})$, нам потрібно проінтегрувати щільність X за всіма x_i для $i\notin\{i_1,i_2,\ldots,i_k\}$.

Доведення. Маємо

$$\begin{aligned}\mathbb{P}(X_i\in B)=\mathbb{P}((X_1,X_2,\ldots,X_d)\in\mathbb{R}^{i-1}\times B\times\mathbb{R}^{d-i})&=\int_{\mathbb{R}^{i-1}\times B\times\mathbb{R}^{d-i}} g\\ &=\int_B\left(\int_{\mathbb{R}^{i-1}\times\mathbb{R}^{d-i}} g(x)\mathrm{d}x_1\ldots\mathrm{d}x_{i-1}\mathrm{d}x_{i+1}\ldots\mathrm{d}x_n\right)\mathrm{d}x_i.\end{aligned}$$

У випадку, коли граничний розподіл є багатовимірним, міркування аналогічні. □

Зауваження: Зворотна імплікація не виконується: достатньо взяти будь-яку неперервну одновимірну випадкову величину X і розглянути вектор (X,X) (який вже не є неперервним, оскільки він зосереджений на множині $\{(x,x):x\in\mathbb{R}\}$, яка має нульову міру Лебега).

Важливі приклади неперервних розподілів

1) *Рівномірний розподіл на множині* D, що позначається як $\mathcal{U}(D)$. Припустимо, що $D\in\mathcal{B}(\mathbb{R}^d)$ задовольняє умову $0<|D|<\infty$. Випадкова величина X рівномірно розподілена на D, якщо вона має щільність

$$g(x)=\frac{1}{|D|}1_D(x)=\begin{cases}1/|D| & \text{для } x\in D,\\ 0 & \text{для } x\notin D.\end{cases}$$

Для будь-якого $B\in\mathcal{B}(\mathbb{R}^d)$ ми маємо $\mathbb{P}(X\in B)=\int_B g(x)\mathrm{d}x=\frac{|B\cap D|}{|D|}$.

Зокрема, якщо $d=1$ і $D=[a,b]$, ми отримуємо розподіл о щільності $g(x)=\frac{1}{b-a}1_{[a,b]}(x)$ і функції розподілу

$$F_X(x)=\int_{-\infty}^x g(s)\mathrm{d}s=\begin{cases}0 & \text{для } x<a,\\ (x-a)/(b-a) & \text{для } a\leqslant x<b,\\ 1 & \text{для } x\geqslant b.\end{cases}$$

2) *Експоненціальний розподіл із параметром λ ($\lambda>0$)*, позначається $\mathrm{Exp}(\lambda)$. Випадкова величина X має такий розподіл, якщо вона має щільність

$$g(x)=\lambda e^{-\lambda x}1_{[0,\infty)}(x)=\begin{cases}0 & \text{для } x<0,\\ \lambda e^{-\lambda x} & \text{для } x\geqslant 0.\end{cases}$$

Як легко тоді порахувати

$$F_X(x)=\begin{cases}0 & \text{для } x<0,\\ 1-e^{-\lambda x} & \text{для } x\geqslant 0.\end{cases}$$

Експоненціальний розподіл використовується для моделювання часу очікування абсолютно випадкової події. Припустимо, що невід'ємна випадкова змінна X є таким часом очікування, і ми запишемо загальну випадковість як властивість *відсутності пам'яті*:

$$\mathbb{P}(X>t+s|X>s)=\mathbb{P}(X>t)\qquad \text{для всіх } s,\, t\geqslant 0.$$

Позначаючи $f(t) = \mathbb{P}(X > t)$, ми бачимо, що наведене вище рівняння еквівалентно $f(t+s) = f(t)f(s)$. Крім того, f не зростає, неперервна справа і задовольняє $f(0) = 1$, $\lim_{t\to\infty} f(t) = 0$, з чого вже випливає, що $f(t) = e^{-\lambda x}$ для деякого $\lambda > 0$, тому X має експоненціальний розподіл.

3) Розподіл Гауса (нормальний розподіл). Припустимо, m є фіксований вектор у $\mathbb{R}^d$, а A — симетрична позитивно визначена матриця $d \times d$. Випадкова величина X є нормально розподіленою (з параметрами m і A), якщо її щільність дорівнює

$$g(x) = \frac{\sqrt{\det A}}{(2\pi)^{d/2}} \exp\left(-\frac{1}{2}\langle A(x-m), x-m\rangle\right)$$

($\langle\cdot,\cdot\rangle$ означає скалярний добуток у $\mathbb{R}^d$). Перевіримо, що інтеграл від g справді дорівнює 1. З лінійної алгебри ми знаємо, що існує ізометрія (ортогональна матриця) $B : \mathbb{R}^d \to \mathbb{R}^d$ така, що B^tAB діагональна:

$$B^tAB = \begin{bmatrix} a_1 & 0 & \dots & 0 \\ 0 & a_2 & \dots & 0 \\ & & \dots & \\ 0 & 0 & \dots & a_d \end{bmatrix}.$$

Підставляючи $x - m = By$, отримуємо

$$\begin{aligned}
\int_{\mathbb{R}^d} g(x)\mathrm{d}x &= \frac{\sqrt{\det A}}{(2\pi)^{d/2}} \int_{\mathbb{R}^d} \exp\left(-\frac{1}{2}\langle ABy, By\rangle\right) |\det B| \mathrm{d}y \\
&= \frac{\sqrt{\det A}}{(2\pi)^{d/2}} \int_{\mathbb{R}^d} \exp\left(-\frac{1}{2}\langle B^{-1}ABy, y\rangle\right) \mathrm{d}y \\
&= \frac{\sqrt{\det A}}{(2\pi)^{d/2}} \int_{\mathbb{R}^d} \exp\left(-\frac{1}{2}\sum_{k=1}^d a_k y_k^2\right) \mathrm{d}y \\
&= \frac{\sqrt{\det A}}{(2\pi)^{d/2}} \prod_{k=1}^d \int_{\mathbb{R}} e^{-a_k y_k^2/2} \mathrm{d}y_k \\
&= \frac{\sqrt{\det A}}{\sqrt{a_1 a_2 \dots a_d}} \prod_{k=1}^d \left(\frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} e^{-z_k^2/2} \mathrm{d}z_k\right) = 1,
\end{aligned}$$

де на передостанньому кроці ми зробили заміну $z_k = \sqrt{a_k} y_k$, а в останньому використано рівність

$$\det A = a_1 a_2 \dots a_d \quad \text{та} \quad \frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} e^{-x^2/2} \mathrm{d}x = 1.$$

В особливому випадку $d = 1$, якщо $m \in \mathbb{R}$ і $\sigma > 0$ (і $[\sigma^{-2}]$ для матриці A), ми отримуємо щільність

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

Цей розподіл позначається $\mathcal{N}(m, \sigma^2)$. Зокрема, розподіл $\mathcal{N}(0, 1)$ називається стандартним нормальним розподілом (стандартним розподілом Гауса).

Для безперервно розподілених змінних ми маємо наступний критерій незалежності.

Теорема 5.11. *Припустимо, що g_1, g_2, …, g_n є щільностями. Тоді випадкові величини X_1, X_2, …, X_n із розподілами з щільностями g_1, g_2, …, g_n незалежні тоді і тільки тоді, коли змінна $(X_1, X_2, \dots, X_n)$ має щільність $g(x_1, x_2, \dots, x_n) = g_1(x_1)g_2(x_2)\dots g_n(x_n)$.*

Доведення. $\Rightarrow$ З незалежності змінних випливає

$$\begin{aligned}
&F_{(X_1,X_2,\ldots,X_n)}(x_1,x_2,\ldots,x_n)\\
&\qquad = F_{X_1}(x_1)F_{X_2}(x_2)\ldots F_{X_n}(x_n)\\
&\qquad = \int_{-\infty}^{x_1} g_1(y_1)\mathrm{d}y_1 \int_{-\infty}^{x_2} g_2(y_2)\mathrm{d}y_2 \ldots \int_{-\infty}^{x_n} g_n(y_n)\mathrm{d}y_n\\
&\qquad = \int_{(-\infty,x_1]\times\ldots\times(-\infty,x_n]} g_1(y_1)g_2(y_2)\ldots g_n(y_n)\mathrm{d}y_1\mathrm{d}y_2\ldots\mathrm{d}y_n,
\end{aligned}$$

з чого відразу випливає теза.

$\Leftarrow$ Записуємо таку ж послідовність рівностей, починаючи з кінця. $\square$

Приклад 5.2. Змінні X_1, X_2, ..., X_n, рівномірно розподілені на D_1, D_2, ..., D_n, незалежні тоді і тільки тоді, коли $(X_1, X_2, \ldots, X_n)$ рівномірно розподілено на $D_1 \times D_2 \times \ldots \times D_n$. Це безпосередньо випливає з того факту, що щільність рівномірного розподілу при $D_1 \times D_2 \times \ldots \times D_n$ дорівнює

$$\frac{1_{D_1\times D_2\times\ldots\times D_n}(x_1,x_2,\ldots,x_n)}{|D_1\times D_2\times\ldots\times D_n|} = \frac{1_{D_1}(x_1)}{|D_1|}\cdot\frac{1_{D_2}(x_2)}{|D_2|}\cdot\ldots\cdot\frac{1_{D_n}(x_n)}{|D_n|},$$

а множники праворуч — щільності рівномірного розподілу на D_1, D_2, ..., D_n відповідно.

Коли випадкові величини є незалежними та мають безперервний розподіл, існує ефективний метод обчислення розподілу їх суми. Справедливим є наступний факт.

Теорема 5.12. *Нехай X_1, X_2 незалежні одновимірні випадкових величин із розподілами з щільністями g_1 та g_2 відповідно. Тоді розподіл $X_1 + X_2$ має щільність*

$$g_1 * g_2(x) = \int_{\mathbb{R}} g_1(x-y)g_2(y)dy.$$

Отже, ми бачимо, що $g_1 * g_2$ є згорткою *щільностей g_1 і g_2.*

Доведення. Для будь-якого $B \in \mathcal{B}(\mathbb{R})$ маємо

$$\begin{aligned}
\mathbb{P}(X_1+X_2\in B) &= P_{(X_1,X_2)}(\{(x,y): x+y\in B\})\\
&= \iint_{\{(x,y):x+y\in B\}} g_{(X_1,X_2)}(x,y)\mathrm{d}x\mathrm{d}y\\
&= \iint_{\{(x,y):x+y\in B\}} g_1(x)g_2(y)\mathrm{d}x\mathrm{d}y\\
&= \iint_{\mathbb{R}^2} 1_B(x+y)g_1(x)g_2(y)\mathrm{d}x\mathrm{d}y\\
&= \int_{\mathbb{R}}\left(\int_{\mathbb{R}} 1_B(x+y)g_1(x)\mathrm{d}x\right)g_2(y)\mathrm{d}y\\
&= \int_{\mathbb{R}}\left(\int_{\mathbb{R}} 1_B(z)g_1(z-y)\mathrm{d}z\right)g_2(y)\mathrm{d}y\\
&= \int_{B}\left(\int_{\mathbb{R}} g_1(z-y)g_2(y)\mathrm{d}y\right)\mathrm{d}z\\
&= \int_{B} g_1 * g_2(z)\mathrm{d}z.\square
\end{aligned}$$

Приклади:

1) Якщо X_1, X_2 є незалежними випадковими величинами рівномірно розподілені на $[0,1]$, тоді $g_1(x) = g_2(x) = 1_{[0,1]}(x)$, тому $X_1 + X_2$ має щільність

$$
\begin{aligned}
g_1 * g_2(x) &= \int_{\mathbb{R}} 1_{[0,1]}(x-y)1_{[0,1]}(y)\mathrm{d}y \\
&= \int_{\mathbb{R}} 1_{[x-1,x]}(y)1_{[0,1]}(y)\mathrm{d}y \\
&= \big|[x-1,x] \cap [0,1]\big| = \begin{cases} 0 & \text{для } x < 0, \\ x & \text{для } 0 \leqslant x < 1, \\ 2-x & \text{для } 1 \leqslant x < 2, \\ 0 & \text{для } x \geqslant 2. \end{cases}
\end{aligned}
$$

2) Припустимо, X_1, X_2 — незалежні випадкові величини з розподілами $N(m_1, \sigma_1^2)$, $N(m_2, \sigma_2^2)$: тоді

$$
g_i(x) = \frac{1}{\sqrt{2\pi}\sigma_i} \exp\left(-\frac{(x-m_i)^2}{2\sigma_i^2}\right), \qquad i = 1,\, 2.
$$

$X_1 + X_2$ має щільність

$$
\begin{aligned}
g_1 * g_2(x) &= \frac{1}{2\pi\sigma_1\sigma_2} \int_{\mathbb{R}} \exp\left(-\frac{(x-y-m_1)^2}{2\sigma_1^2} - \frac{(y-m_2)^2}{2\sigma_2^2}\right)\mathrm{d}y \\
&= \frac{1}{\sqrt{2\pi(\sigma_1^2+\sigma_2^2)}} \exp\left(-\frac{(x-m_1-m_2)^2}{2(\sigma_1^2+\sigma_2^2)}\right)
\end{aligned}
$$

(Доведення останньої рівності залишаємо як вправу). Отже, $X_1 + X_2 \sim N(m_1 + m_2, \sigma_1^2 + \sigma_2^2)$. І загалом, за індукцією: якщо X_1, X_2, …, X_n є незалежними випадковими величинами з розподілами $N(m_1, \sigma_1^2)$, $N(m_2, \sigma_2)^2)$, …, $N(m_n, \sigma_n^2)$, то $X_1 + X_2 + \ldots + X_n$ має розподіл $N(m_1 + m_2 + \ldots + m_n, \sigma_1^2 + \sigma_2^2 + \ldots + \sigma_n^2)$.

6 Завдання

1. Кидаємо монету, для якої ймовірність випадіння голови становить $p \in (0, 1]$, поки не випаде k голів (загалом, не обов'язково поспіль). Нехай X — це кількість підкидань. Знайдіть розподіл X.

2. Кидаємо кубик двічі. Нехай X, Y позначають мінімальне та максимальне з отриманих чисел відповідно. Визначте розподіл X, Y і переконайтеся, що X і $7 - Y$ мають однаковий розподіл.

3. На перехресті вулиць певного напрямку на хвилину горить червоне світло, а на півхвилини зелене (вважаємо, що жовтого немає). Автомобіль прибуває на перехрестя (у заданому напрямку) у випадковий час. Нехай X — час перебування на перехресті; припускаємо, що пробки немає.

а) Визначте розподіл X.

б) Припустимо, що за 20 секунд автомобіль ще не проїхав перехрестя; яка ймовірність того, що він проїде протягом наступних 10 секунд?

4. Розподіл випадкової величини X задано формулою

$$F_X(t) = \begin{cases} 0 & \text{для } t < -1, \\ \frac{1}{2}(t+1) & \text{для } -1 \leqslant t < 0, \\ \frac{3}{4} & \text{для } 0 \leqslant t < 4, \\ 1 & \text{для } t \geqslant 4. \end{cases}$$

Обчислити $\mathbb{P}(X = -5)$, $\mathbb{P}(2 < X \leqslant 5)$, $\mathbb{P}(X = 4)$, $\mathbb{P}(-1 < X < 0)$.

5. Випадкова величина X має функцію розподілу

$$F(t) = \begin{cases} 0 & \text{для } t < 0, \\ t/2 & \text{для } 0 \leqslant t < 2, \\ 1 & \text{для } t \geqslant 2. \end{cases}.$$

Визначити функцію розподілу змінних $Y = \max(X, 1)$ і $Z = \min(X, X^2)$.

6. Нехай $F : \mathbb{R} \to [0, 1]$ – неперервна справа неспадна функція така, що $\lim_{t\to\infty} F(t) = 1$ і $\lim_{t\to-\infty} F(t) = 0$. Доведіть, що F є функцією розподілу деякої випадкової величини в деякому імовірнісному просторі.

7. З колоди 52 карт п'ять разів випадковим чином беремо по одній карті і кладемо назад. Нехай X — кількість витягнутих пік, Y — кількість витягнутих черв, а Z — кількість витягнутих валетів. Чи змінні X, Y є незалежними? Чи змінні X, Z є незалежними?

8. Випадкові величини $X_1, X_2, \ldots, X_n$ ($n \geqslant 6$) є незалежними та мають однаковий розподіл, визначений формулою $\mathbb{P}(X_i = -1) = \mathbb{P}(X_i = 1) = 1/2$, $i = 1, 2, \ldots, n$.

а) Чи змінні $X_1 + X_2$, X_1X_2 є незалежними?

б) Чи змінні $X_1 + X_2$, X_3, $X_4 + X_5X_6$ є незалежними?

в) Чи змінні X_1, X_1X_2, …, $X_1X_2 \ldots X_n$ є незалежними?

9. Випадкові величини X, Y незалежні, крім того для $n = 1, 2, \ldots$ маємо $\mathbb{P}(X = n) = (1-p)^{n-1}p$ і $\mathbb{P}(Y = n) = (1-q)^{n-1}q$. Обчислити $\mathbb{P}(X \leqslant Y)$.

10. Для будь-якого числа $\omega \in [0,1]$ нехай $X_n(\omega)$ буде n-м розрядом двійкового розкладання ω, $n = 1, 2, \ldots$ (якщо ω має два різних розкладання, ми беремо те, яке містить скінченну кількість одиниць). Доведіть, що $X_1, X_2, \ldots$ є незалежними випадковими величинами на ймовірнісному просторі $([0,1], \mathcal{B}([0,1]), |\cdot|)$.

11.X, Y незалежні, та X не має атомів. Доведіть, що $\mathbb{P}(X = Y) = 0$.

12. Випадкові величини $X_1, X_2, \ldots, X_n$ є незалежними і мають розподіли Пуассона з параметрами $\lambda_1, \lambda_2, \ldots, \lambda_n$. Доведіть, що $X_1 + X_2 + \ldots + X_n$ має розподіл Пуассона з параметром $\lambda_1 + \lambda_2 + \ldots + \lambda_n$.

13. Випадкова величина X не залежить від себе. Доведіть, що існує таке число c, що $\mathbb{P}(X = c) = 1$.

14. Випадкова величина X має експоненціальний розподіл з параметром 1.

a) Визначте розподіли $[X]$ і $\{X\}$.

b) Чи незалежні ці змінні?

Зауваження: $[x]$, $\{x\}$ позначають цілу частину та дробову частину числа $x \in \mathbb{R}$ відповідно.

15. Випадкова величина X рівномірно розподілена на інтервалі $[0,1]$. Визначити розподіл змінної $Y = -\ln X$.

16. Випадкова величина X має нормальний розподіл $\mathcal{N}(0,1)$. Визначити щільності змінних $Y = e^X$, $Z = X^2$.

17. Текст брошури містить $n = 100000$ символів. Під час введення (на комп'ютері) кожен символ може бути написаний з помилкою з ймовірністю $0,001$. У свою чергу редактор знаходить кожну з помилок з імовірністю $0,9$, а потім текст повертається автору, який знаходить кожну з помилок, що залишилися, з імовірністю $0,5$. Оцініть ймовірність того, що брошура міститиме не більше 3 помилок після обох редакцій.

18. Випадкові величини $X_1, X_2, \ldots, X_n$ є незалежними та мають експоненціальний розподіл з параметрами $\lambda_1, \lambda_2, \ldots, \lambda_n$ відповідно. Визначити розподіл змінної $Y = \max(X_1, X_2, \ldots, X_n)$.

19. Випадкова величина (X, Y) має щільність розподілу

$$g(x,y) = Cxy\mathbb{1}_{\{0 \leqslant x \leqslant y \leqslant 1\}}.$$

a) Обчислити C.

b) Обчислити $\mathbb{P}(X + Y \leqslant 1)$.

c) Визначити розподіл змінної X/Y.

г) Чи X, Y незалежні?

e) Чи X/Y, Y незалежні?

20. Змінні X, Y є незалежними та рівномірно розподіленими в інтервалі $[-1,1]$. Обчислити $\mathbb{P}(X^2 + Y^2 \leqslant 1)$.

21. Випадкова величина X має розподіл Коші, тобто ма щільність розподілу

$$g(x) = \frac{1}{\pi}\frac{1}{1+x^2}.$$

Доведіть, що змінні X, $1/X$ мають однаковий розподіл.

22. Нехай $\Gamma(r) = \int_0^\infty x^{r-1}e^{-x}dx$, $r > 0$. Ми кажемо, що X має розподіл *гамма з параметрами* λ, r (позначається $\Gamma(\lambda, r)$), якщо він має щільність

$$g_{\lambda,r}(x) = \frac{1}{\Gamma(r)}\lambda^r x^{r-1}e^{-\lambda x}\mathbb{1}_{[0,\infty)}(x).$$

а) Доведіть, що якщо X, Y є незалежними змінними $X \sim \Gamma(\lambda, r)$, $Y \sim \Gamma(\lambda, s)$, то $X+Y \sim \Gamma(\lambda, r+s)$.

б) Доведіть, що якщо X_1, X_2, …, X_n є незалежними випадковими величинами з розподілом $\mathrm{Exp}(\lambda)$, то $X_1 + X_2 + \ldots + X_n$ має розподіл $Gamma(\lambda, n)$.

в) Доведіть, що якщо X_1, X_2, …, X_n є незалежними випадковими величинами з розподілом $\mathcal{N}(0,1)$, то $X_1^2 + X_2^2 + \ldots + X_n^2$ має розподіл $\Gamma(1/2, n/2)$ (це називається *розподіл хі-квадрат із n ступенями свободи*).

23. Випадкові величини X, Y незалежні та експоненціально розподілені з параметром 1. Доведіть, що X/Y і $X+Y$ незалежні.

7 Параметри розподілу випадкової величини (математичне сподівання та дисперсія)

Визначення 7.1. (i) Припустимо, що X є одновимірною випадковою величиною в імовірнісному просторі $(\Omega, \mathcal{F}, \mathbb{P})$. Ми кажемо, що X має математичне сподівання, якщо інтеграл $\int_\Omega X(\omega) \mathrm{d}\mathbb{P}(\omega)$ є збіжним. Цей інтеграл і називається *математичним сподівання* X і позначається $\mathbb{E}X$.

(ii) Якщо $\mathbb{E}|X| < \infty$, то ми кажемо, що X інтегровна, і пишемо $X \in L^1(\Omega, \mathcal{F}, \mathbb{P})$.

(iii) Аналогічно, якщо $\mathbb{E}|X|^p < \infty$ для деякого $p > 0$, то ми говоримо, що X інтегровна до p-го степеня, і пишемо $X \in L^p(\Omega, \mathcal{F}, \mathbb{P})$.

(iv) Ми кажемо, що випадкова величина X є *обмеженою*, якщо існує таке число u, що $\mathbb{P}(|X| \geqslant u) = 0$. Позначення: $X \in L^\infty(\Omega, \mathcal{F}, \mathbb{P})$.

Зауваження: Наведені вище визначення також мають сенс, коли X є багатовимірною змінною, це буде обговорено пізніше в лекції.

Звернемо увагу на дві важливі нерівності:

1) *Нерівність Мінковського.* Якщо X, Y є випадковими величинами і $1 \leqslant p < \infty$, то

$$\operatorname{ess\,sup} |X + Y| \leqslant \operatorname{ess\,sup} |X| + \operatorname{ess\,sup} |Y|,$$

де $\operatorname{ess\,sup} \xi = \inf\{u : \mathbb{P}(\xi \geqslant u) = 0\}$ — т.зв. істотний супремум ξ.

2) *Нерівність Гельдера.* Припустимо, що X, Y є випадковими величинами і $p, q \in (1, \infty)$ є гармонічно спряженими числами, тобто $\frac{1}{p} + \frac{1}{q} = 1$. Тоді

$$\mathbb{E}|XY| \leqslant (\mathbb{E}|X|^p)^{1/p} (\mathbb{E}|Y|^q)^{1/q}.$$

Зауваження: Безпосередньо з визначення ми бачимо, що сподівання є лінійним оператором: точніше, якщо X_1, X_2, $\ldots$, X_n є інтегровними випадковими величинами, а a_1, a_2, $\ldots$, a_n $\mathbb{R}$, то $a_1 X_1 + a_2 X_2 + \ldots + a_n X_n$ є інтегровною, і

$$\mathbb{E}(a_1 X_1 + a_2 X_2 + \ldots + a_n X_n) = a_1 \mathbb{E}X_1 + a_2 \mathbb{E}X_2 + \ldots + a_n \mathbb{E}X_n.$$

З аналізу випливають наступні три теореми про перехід до границі під знаком інтеграла.

Теорема 7.1 (Теорема Леві про монотонну збіжність)**.** *Припустимо, що X_1, X_2, $\ldots$ є невід'ємними інтегровними випадковими величинами і $X_n(\omega) \leqslant X_{n+1}(\omega)$, $n = 1, 2, \ldots$. Тоді*

$$\mathbb{E}\big(\lim_{n\to\infty} X_n\big) = \lim_{n\to\infty} \mathbb{E}X_n.$$

Зокрема, змінна $\lim_{n\to\infty} X_n$ інтегровна тоді і тільки тоді, коли $\lim_{n\to\infty} \mathbb{E}X_n < \infty$.

Теорема 7.2 (Теорема Лебега про мажоровану збіжність)**.** *Припустимо, X_1, X_2, $\ldots$ є випадковими величинами, мажорованими деякою інтегровною змінною η: $|X_n| \leqslant \eta$ для $n = 1, 2, \ldots$. Якщо існує границя $X(\omega) = \lim_{n\to\infty} X_n(\omega)$ для майже всіх ω (майже всіх в сенсі $\mathbb{P}$), то $\lim_{n\to\infty} \mathbb{E}X_n = \mathbb{E}X$.*

Теорема 7.3 (Лема Фату)**.** *Припустимо, що X_1, X_2, $\ldots$ є невід'ємними випадковими величинами. Тоді $\mathbb{E} \liminf_{n\to\infty} X_n \leqslant \liminf_{n\to\infty} \mathbb{E}X_n$.*

Відступ: Припустимо, що X, Y є випадковими величинами. Ми кажемо, що X і Y рівні майже напевно (м.н.), якщо виконується $\mathbb{P}(X \neq Y) = 0$. Для $p \in [1, \infty]$ визначимо

$$||X||_p = \begin{cases} (\mathbb{E}|X|^p)^{1/p} & \text{для } 1 \leqslant p < \infty, \\ \operatorname{ess\,sup}|X| & \text{для } p = \infty. \end{cases}$$

Якщо ми ототожнюємо випадкові величини, що рівні м.н., то $||\cdot||_p$ є нормою на $L^p(\Omega, \mathcal{F}, \mathbb{P})$. Крім того, простір $L^p(\Omega, \mathcal{F}, \mathbb{P})$ з цією нормою є лінійним і повним (тобто є Банаховим простором).

Зауваження: Згідно зі старою нерівністю Гельдера ми маємо $||X||_p \leqslant ||X||_{p'}$, якщо $p < p'$. Звідси ми отримуємо включення $L^{p'}(\Omega, \mathcal{F}, \mathbb{P}) \subset L^p(\Omega, \mathcal{F}, \mathbb{P})$.

Визначення 7.2. Для $p \in (0, \infty)$ число $\mathbb{E}|X|^p$ називається p-м моментом змінної X.

Теорема 7.4 (Нерівність Чебишева)**.** *Припустимо, X — випадкова величина, а $f : [0, \infty) \to [0, \infty)$ — неспадна функція, така що $f(x) > 0$ для $x > 0$. Тоді для будь-якого числа $\lambda > 0$*

$$\mathbb{P}(|X| \geqslant \lambda) \leqslant \frac{\mathbb{E}f(|X|)}{f(\lambda)}.$$

Зокрема, приймаючи $f(x) = x^p$, $p > 0$, отримуємо

$$\mathbb{P}(|X| \geqslant \lambda) \leqslant \frac{\mathbb{E}|X|^p}{\lambda^p}.$$

Доведення. Маємо

$$\mathbb{E}f(|X|) \geqslant \mathbb{E}f(|X|)1_{\{|X| \geqslant \lambda\}} \geqslant \mathbb{E}f(\lambda)1_{\{|X| \geqslant \lambda\}} = f(\lambda)\mathbb{P}(|X| \geqslant \lambda). \qquad \square$$

Визначення 7.3. Припустімо, що X є одновимірною інтегровною з квадратом випадковою величиною (тобто $X \in L^2$). Число $\mathrm{Var}X = \mathbb{E}(X - \mathbb{E}X)^2$ називається дисперсією X.

Легко перевірити, що дисперсія має такі властивості. Припустивши $X \in L^2$, ми маємо:

(a) $\mathrm{Var}X = \mathbb{E}X^2 - (\mathbb{E}X)^2$,

(b) $\mathrm{Var}X \geqslant 0$, з $\mathrm{Var}X = 0$ тоді і тільки тоді, коли X має одноточковий розподіл.

(c) $\mathrm{Var}(aX + b) = a^2 \,\mathrm{Var}X$ для будь-яких дійсних чисел a, b.

(d) З нерівності Чебишева для будь-якого числа$\lambda > 0$, отримуємо

$$\mathbb{P}(|X - \mathbb{E}X| \geqslant \lambda) \leqslant \frac{\mathrm{Var}\,X}{\lambda^2}.$$

Відзначимо ще один загальний факт.

Теорема 7.5 (Про заміну змінних)**.** *Припустімо, що X є d-вимірною випадковою змінною на $(\Omega, \mathcal{F}, \mathbb{P})$ і $f : \mathbb{R}^d \to \mathbb{R}$ є борелівською функцією. Тоді*

$$\mathbb{E}f(X) = \int_{\mathbb{R}^d} f(x) P_X(dx),$$

якщо один із цих інтегралів існує.

Доведення. Використовуємо метод ускладнення функції.

(i) Спочатку припустимо, що f є характеристичною функцією деякої множини B: $f = 1_B$. Тоді доведена тотожність набуває вигляду $\mathbb{P}(X \in B) = P_X(B)$, що, звичайно, вірно.

(ii) Якщо f є простою функцією, тобто лінійною комбінацією характеристичних функцій, тоді досліджувана рівність виконується, оскільки обидві сторони залежать від f лінійним чином.

(iii) Припустимо $f \geqslant 0$. Тоді f є поточковою границею деякої невід'ємної послідовності $(f_n)_{n\geqslant 0}$ невід'ємних простих функцій. На підставі (ii) ми маємо

$$\mathbb{E}f_n(X) = \int_{\mathbb{R}^d} f_n(x)P_X(\mathrm{d}x), \qquad n = 1, 2, \dots,$$

тому достатньо взяти $n \to \infty$ і скористатися теоремою Лебега про монотонний перехід до границі.

(iv) Якщо f є довільною, ми розбиваємо її на різницю двох невід'ємних борелівських функцій: $f = f_+ - f_- = f1_{\{f\geqslant 0\}} + f1_{\{f<0\}}$, застосовуємо (iii) до функцій f_+ і f_-, і потім ми віднімаємо дві отримані тотожності. □

З наведеного факту випливають наступні

Висновки:

1) Якщо X є випадковою величиною, то $\mathbb{E}X = \int_{\mathbb{R}} xP_X(\mathrm{d}x)$ і

$$\operatorname{Var} X = \int_{\mathbb{R}} (x - \mathbb{E}X)^2 P_X(\mathrm{d}x) = \int_{\mathbb{R}} x^2 P_X(\mathrm{d}x) - (\mathbb{E}X)^2$$

(якщо ці інтеграли існують).

2) Математичне сподівання та дисперсія залежать лише від розподілу.

Як видно з наведеної вище теореми, якщо X є d-вимірною дискретною змінною, а $f : \mathbb{R}^d \to \mathbb{R}$ — борелівською функцією, то

$$\mathbb{E}f(X) = \int_{\mathbb{R}^d} f(x)P_X(\mathrm{d}x) = \sum_{x\in S_X} f(x)P_X(\{x\}) = \sum_{x\in S_X} f(x)\mathbb{P}(X = x),$$

якщо існує математичне сподівання. Так, зокрема, для $d = 1$ маємо

$$\mathbb{E}X = \sum_{x\in S_X} xP_X(\{x\}) = \sum_{x\in S_X} x\mathbb{P}(X = x),$$

$$\operatorname{Var} X = \mathbb{E}X^2 - (\mathbb{E}X)^2 = \sum_{x\in S_X} x^2 P_X(\{x\}) - (\mathbb{E}X)^2 = \sum_{x\in S_X} x^2\mathbb{P}(X = x) - (\mathbb{E}X)^2.$$

Якщо змінна має неперервний розподіл, ми визначаємо її параметри, використовуючи такий факт.

Теорема 7.6. *Припустимо, що d-вимірна випадкова величина X розподілена з щільністю g. Тоді для будь-якої борелівської функції $f : \mathbb{R}^d \to \mathbb{R}$ маємо*

$$\mathbb{E}f(X) = \int_{\mathbb{R}^d} f(x)P_X(dx) = \int_{\mathbb{R}^d} f(x)g(x)dx,$$

якщо існує математичне сподівання.

Dowód. Як і вище, ми використовуємо метод ускладнення функції. □

Отже, якщо X є одновимірною випадковою величиною з щільністю g, то

$$\mathbb{E}X = \int_{\mathbb{R}} xg(x)\mathrm{d}x, \quad \operatorname{Var} X = \int_{\mathbb{R}} x^2 g(x)\mathrm{d}x - (\mathbb{E}X)^2.$$

Приклади:

1) Нехай $P_X = \delta_a$, $a \in \mathbb{R}$. Тоді $\mathbb{E}X = a \cdot 1 = a$, $\operatorname{Var} X = \mathbb{E}X^2 - (\mathbb{E}X)^2 = a^2 - a^2 = 0$.

2) Припустимо, що P_X має двоточковий розподіл зосереджений в $\{a, b\}$, такий що $P_X(\{a\}) = p$, $P_X(\{b\}) = 1 - p$, $0 < p < 1$. Тоді

$$\mathbb{E}X = a \cdot p + b \cdot (1-p),$$

$$\operatorname{Var} X = a^2 \cdot p + b^2 \cdot (1-p) - (ap + b(1-p))^2 = (a-b)^2 p(1-p).$$

3) Тепер припустімо, що $P_X = B(n,p)$: $P_X(\{k\}) = \binom{n}{k} p^k (1-p)^{n-k}$, $k = 0, 1, \ldots, n$. Тоді, як вище

$$\mathbb{E}X = np \quad \text{і } \operatorname{Var} X = np(1-p).$$

Рахувати прямо за визначенням незручно: наприклад, маємо

$$\mathbb{E}X = \sum_{k=0}^{n} k \binom{n}{k} p^k (1-p)^{n-k}$$

і цю суму треба „згорнути”. Щоб уникнути рахунків, розглянемо незалежні випадкові величини X_1, X_2, …, X_n з однаковим двоточковим розподілом, заданим $\mathbb{P}(X_i = 1) = p = 1 - \mathbb{P}(X_i = 0)$. Тоді, як ми вже знаємо, $X_1 + X_2 + \ldots + X_n$ має розподіл $B(n,p)$, а отже, з лінійності математичного сподівання випливає

$$\mathbb{E}X = \mathbb{E}(X_1 + X_2 + \ldots + X_n) = \mathbb{E}X_1 + \mathbb{E}X_2 + \ldots + \mathbb{E}X_n = np.$$

Крім того,

$$\mathbb{E}X^2 = \mathbb{E}(X_1 + X_2 + \ldots + X_n)^2 = \sum_{k=1}^{n} \mathbb{E}X_k^2 + 2\sum_{k<\ell} \mathbb{E}X_k X_\ell.$$

Для будь-яких різних k, ℓ $X_k X_\ell$ має двоточковий розподіл зосереджений у $\{0, 1\}$, і як випливає з незалежністі X_k і X_ℓ,

$$\mathbb{P}(X_k X_\ell = 1) = \mathbb{P}(X_k = 1, X_\ell = 1) = \mathbb{P}(X_k = 1)\mathbb{P}(X_\ell = 1) = p^2.$$

Тому

$$\mathbb{E}X^2 = n \cdot p + 2\binom{n}{2} p^2 = np + n(n-1)p^2$$

і як наслідок,

$$\operatorname{Var} X = \mathbb{E}X^2 - (\mathbb{E}X)^2 = np - np^2 = np(1-p).$$

4) Нехай X має розподіл Пуассона з параметром $\lambda > 0$: $P_X(\{k\}) = e^{-\lambda}\frac{\lambda^k}{k!}$, $k = 0, 1, 2, \ldots$. Тоді

$$\mathbb{E}X = \sum_{k=0}^{\infty} k \cdot e^{-\lambda}\frac{\lambda^k}{k!} = \lambda e^{-\lambda} \sum_{k=1}^{\infty} \frac{\lambda^{k-1}}{(k-1)!} = \lambda.$$

Так само обчислюємо

$$\operatorname{Var} X = \sum_{k=0}^{\infty} k^2 \cdot e^{-\lambda}\frac{\lambda^k}{k!} - (\mathbb{E}X)^2 = \lambda.$$

5) Нехай $P_X = \mathcal{U}([a,b])$: $g(x) = \frac{1}{b-a} 1_{[a,b]}(x)$. Тоді

$$\mathbb{E}X = \int_{\mathbb{R}} x g(x) \mathrm{d}x = \frac{1}{b-a} \int_a^b x \mathrm{d}x = \frac{a+b}{2}$$

та

$$\operatorname{Var} X = \frac{1}{b-a} \int_a^b x^2 \mathrm{d}x - \left(\frac{a+b}{2}\right)^2 = \frac{(b-a)^2}{12}.$$

6) Далі припустимо, що $X \sim \operatorname{Exp}(\lambda)$, $\lambda > 0$, тобто X має розподіл з щільностю $g(x) = \lambda e^{-\lambda x} 1_{[0,\infty)}(x)$. Інтегруючи частинами, отримуємо

$$\mathbb{E}X = \int_0^\infty x \cdot \lambda e^{-\lambda x} \mathrm{d}x = -x e^{-\lambda x}\Big|_0^\infty + \int_0^\infty e^{-\lambda x} \mathrm{d}x = \frac{1}{\lambda}$$

і, виконавши аналогічні обчислення, отримуємо $\operatorname{Var} X = \frac{1}{\lambda^2}$.

7) Нарешті, припустимо що $P_X = N(m, \sigma^2)$, де $m \in \mathbb{R}$ і $\sigma > 0$. Тоді щільність X визначається формулою

$$g(x) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left[-\frac{(x-m)^2}{2\sigma^2}\right].$$

Підставляючи $y = (x-m)/\sigma$, ми отримуємо

$$\begin{aligned}
\mathbb{E}X &= \frac{1}{\sqrt{2\pi}\sigma} \int_{\mathbb{R}} x \cdot \exp\left[-\frac{(x-m)^2}{2\sigma^2}\right] \mathrm{d}x \\
&= \frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} (\sigma y + m) e^{-y^2/2} \mathrm{d}y \\
&= \frac{\sigma}{\sqrt{2\pi}} \int_{\mathbb{R}} y e^{-y^2/2} \mathrm{d}y + \frac{m}{\sqrt{2\pi}} \int_{\mathbb{R}} e^{-y^2/2} \mathrm{d}y = m.
\end{aligned}$$

Далі, знову застосовуючи наведену вище заміну, отримуємо

$$\begin{aligned}
\operatorname{Var} X &= \frac{1}{\sqrt{2\pi}\sigma} \int_{\mathbb{R}} (x-m)^2 e^{-(x-m)^2/(2\sigma^2)} \mathrm{d}x \\
&= \frac{\sigma^2}{\sqrt{2\pi}} \int_{\mathbb{R}} y^2 e^{-y^2/2} \mathrm{d}y \\
&= \frac{\sigma^2}{\sqrt{2\pi}} \big(-y e^{-y^2/2}\big)\Big|_{-\infty}^{\infty} + \frac{\sigma^2}{\sqrt{2\pi}} \int_{\mathbb{R}} e^{-y^2/2} \mathrm{d}y = \sigma^2.
\end{aligned}$$

Підкреслимо: отже, параметри m і σ^2, що входять до запису нормального розподілу, є його середнім і дисперсією відповідно.

8) Тут варто згадати ще один приклад. Припустимо, X має розподіл Коші, тобто розподіл з щільностю

$$g(x) = \frac{1}{\pi} \frac{1}{1+x^2}, \qquad x \in \mathbb{R}.$$

Тоді X не є інтегровною:

$$\mathbb{E}|X| = \int_{\mathbb{R}} |x| \cdot \frac{1}{\pi} \frac{1}{1+x^2} \mathrm{d}x = \infty.$$

Крім того, очікування X не існує:

$$\mathbb{E}X^+ = \int_{\mathbb{R}} x^+ g(x) \mathrm{d}x = \int_0^\infty \frac{1}{\pi} \frac{x}{1+x^2} \mathrm{d}x = \infty$$

і аналогічно $\mathbb{E}X^- = \infty$.

Тепер ми повернемося до зв'язків сподівання та дисперсії з незалежністю змінних.

Теорема 7.7. *Припустимо,* X_1, X_2, …, X_n *інтегровні та незалежні випадкові величини. Тоді* $X_1X_2\dots X_n$ *є інтегровними і виконується рівність*

$$\mathbb{E}X_1X_2\dots X_n = \mathbb{E}X_1\mathbb{E}X_2\dots\mathbb{E}X_n.$$

Доведення. Знаємо, що $P_{(X_1,X_2,\dots,X_n)} = P_{X_1}\otimes P_{X_2}\otimes\dots\otimes P_{X_n}$. Тому, використовуючи теорему про заміну змінних, отримуємо

$$\begin{aligned}\mathbb{E}|X_1X_2\dots X_n| &= \int_{\mathbb{R}^n} |x_1x_2\dots x_n| P_{(X_1,\dots,X_n)}(\mathrm{d}x_1\dots\mathrm{d}x_n)\\ &= \prod_{i=1}^{n}\int_{\mathbb{R}} |x_i| P_{X_i}(\mathrm{d}x_i) < \infty,\end{aligned}$$

тому $X_1X_2\dots X_n \in L^1(\Omega,\mathcal{F},\mathbb{P})$. Тепер достатньо повторити наведену вище послідовність рівнянь без модуля (що має сенс, оскільки, як ми щойно довели, усі сподівання існують). □

Зауваження: Зворотне не вірне. Для прикладу візьмемо інтегровні з квадратом змінні η_1, η_2 з однаковим розподілом, і покладемо $X_1 = \eta_1 + \eta_2$, $X_2 = \eta_1 - \eta_2$. Тоді $\mathbb{E}X_2 = 0$, тому $\mathbb{E}X_1\mathbb{E}X_2 = 0$; крім того, $\mathbb{E}X_1X_2 = \mathbb{E}\eta_1^2 - \mathbb{E}\eta_2^2 = 0$, в силу рівності розподілів. Однак загалом X_1 і X_2 не є незалежними: наприклад, розглянемо двічі кидання кубика і нехай η_i буде кількістю кубиків у i-му кидку, $i = 1, 2$. Тоді X_1, X_2 є залежними – вони мають однакову парність.

Переходимо до багатовимірної ситуації.

Визначення 7.4. Припустимо, що $X = (X_1, X_2, \dots, X_d)$ — це d-вимірна випадкова величина з інтегровними координатами (тобто $\mathbb{E}|X_i| < \infty$ для $i = 1,\ 2,\ ,l,\ n$). Математичним сподівнням X називається вектор ($\mathbb{E}X_1$, $\mathbb{E}X_2$, …, $\mathbb{E}X_d$).

Зауваження: 1) Якщо X, Y є d-вимірними інтегровними випадковими величинами і α, $\beta \in \mathbb{R}$, то $\alpha X + \beta Y$ також інтегровна.

2) d-вимірна змінна X інтегровна тоді і тільки тоді, коли $\mathbb{E}|X| < \infty$ (де $|\cdot|$ тут — евклідова норма). Це відразу випливає з оцінки

$$|X_j| \leqslant |X| \leqslant \sum_{i=1}^{d} |X_i|, \qquad j = 1, 2, \dots, d.$$

3) Якщо d-вимірна змінна X інтегровна, то $|\mathbb{E}X| \leqslant \mathbb{E}|X|$. Дійсно, для будь-якого вектора $a \in \mathbb{R}^d$ довжини 1 маємо

$$\langle \mathbb{E}X, a\rangle = \sum_{j=1}^{d} \mathbb{E}X_j \cdot a_j = \mathbb{E}\langle X, a\rangle \leqslant \mathbb{E}|X||a| = \mathbb{E}|X|$$

і беручи верхню суму по a (або, альтернативно, поклавши $a = \mathbb{E}X/|\mathbb{E}X|$), ми отримуємо бажану нерівність.

Визначення 7.5. Припустимо, що X_1, X_2 — інтегровні з квадратом випадкові величини. Число

$$\mathrm{Cov}\,(X,Y) = \mathbb{E}\big[(X-\mathbb{E}X)(Y-\mathbb{E}Y)\big]$$

ми називаємо *коваріацією* X і Y. Якщо $\mathrm{Cov}\,(X,Y) = 0$, ми кажемо, що X, Y *некорельовані.*

Як легко перевірити, коваріація має такі властивості:

(а) З нерівністі Шварца вмпливає, що коваріація добре визначена.

(б) $\mathrm{Cov}\,(X,Y) = \mathbb{E}XY - \mathbb{E}X\mathbb{E}Y$.

(в) Для будь-якої змінної $X \in L^2$ $\mathrm{Cov}\,(X,X) = \mathrm{Var}\,X$.

(г) $\mathrm{Cov}\,(X,Y) = \mathrm{Cov}\,(Y,X)$.

(д) Коваріація є білінійним оператором: якщо X, Y, $Z \in L^2$, то

$$\mathrm{Cov}\,(X+Y,Z) = \mathrm{Cov}\,(X,Z) + \mathrm{Cov}\,(Y,Z).$$

Крім того, якщо $X \in L^2$ і $a \in \mathbb{R}$, тоді $\mathrm{Cov}\,(X,a) = 0$.

Зауваження: Наведені вище міркування показують, що якщо X, $Y \in L^2$ незалежні, то вони некорельовані, але в другу сторону це невірно.

Теорема 7.8. *Нехай змінні X_1, X_2, …, X_n э інтегровними з квадратом. Тоді*

$$Var\,(X_1 + X_2 + \ldots + X_n) = \sum_{k=1}^{n} Var\,X_k + 2\sum_{k<\ell} Cov\,(X_k, X_\ell).$$

Зокрема, якщо X_1, X_2, …, X_n некорельовані, то

$$Var\,(X_1 + X_2 + \ldots + X_n) = Var\,X_1 + Var\,X_2 + \ldots + Var\,X_n.$$

Доведення. Перетворюємо:

$$\begin{aligned}
\mathrm{Var}\,(X_1 + X_2 + \ldots + X_n) &= \mathbb{E}\left[\sum_{j=1}^{n} X_j - \mathbb{E}\left(\sum_{j=1}^{n} X_j\right)\right]^2 \\
&= \mathbb{E}\left[\sum_{j=1}^{n}(X_j - \mathbb{E}X_j)\right]^2 \\
&= \sum_{j=1}^{n} \mathbb{E}(X_j - \mathbb{E}X_j)^2 + 2\sum_{i<j} \mathbb{E}(X_i - \mathbb{E}X_i)(X_j - \mathbb{E}X_j) \\
&= \sum_{j=1}^{n} \mathrm{Var}\,X_j + 2\sum_{i<j} \mathrm{Cov}\,(X_i, X_j).
\end{aligned}$$

Це доводить першу частину теореми. Якщо змінні тепер некорельовані, тоді всі коваріації, що з'являються у другій сумі, дорівнюють нулю. □

Порівнюючи одно- та багатовимірний випадки, ми бачимо, що сподівання одновимірної випадкової величини є числом, а сподівання багатовимірної змінної є вектором. Виникає закономірне питання про узагальнення дисперсії на багатовимірний випадок. Таким узагальнення є т. зв коваріаційна матриця.

Визначення 7.6. Припустимо, що $X = (X_1, X_2, \ldots, X_d)$ є d-вимірною випадковою змінною, що інтегрується з квадратом. Матриця

$$\Lambda = \begin{bmatrix} \mathrm{Cov}\,(X_1,X_1) & \mathrm{Cov}\,(X_1,X_2) & \ldots & \mathrm{Cov}\,(X_1,X_d) \\ \mathrm{Cov}\,(X_2,X_1) & \mathrm{Cov}\,(X_2,X_2) & \ldots & \mathrm{Cov}\,(X_2,X_d) \\ & & \ldots & \\ \mathrm{Cov}\,(X_d,X_1) & \mathrm{Cov}\,(X_d,X_2) & \ldots & \mathrm{Cov}\,(X_d,X_d) \end{bmatrix}$$

називається матрицею коваріацій змінної X.

Зауваження: Сподівання та матриця коваріацій d-вимірної випадкової величини залежать лише від розподілу.

Теорема 7.9 (Властивості матриці коваріацій)**.** *Матриця коваріацій змінної* $X = (X_1, X_2, \ldots, X_d)$ *симетрична і невід'ємно визначена.*

Доведення. Симетрія випливає безпосередньо з властивості (г) коваріації. Щоб довести невід'ємну визначеність, візьмемо будь-яку послідовність дійсних чисел t_1, t_2, …, t_d. Розглянемо одновимірну випадкову величину $\eta = t_1X_1 + t_2X_2 + \ldots + t_dX_d$, яка є інтегровною з квадратом (оскільки X_1, X_2, …, X_d також мають цю властивість). Ми маємо

$$\begin{aligned} 0 &\leqslant \operatorname{Var}\eta \\ &= \mathbb{E}\left(\sum_{j=1}^{d} t_j(X_j - \mathbb{E}X_j)\right)^2 \\ &= \sum_{i,j=1}^{d} \mathbb{E}\big[t_i(X_i - \mathbb{E}X_i)\cdot t_j(X_j - \mathbb{E}X_j)\big] \\ &= \sum_{i,j=1}^{d} t_it_j \operatorname{Cov}(X_i, X_j), \end{aligned}$$

звідки отримуємо невід'ємну визначеність.

Зробимо ще одне корисне спостереження. Припустимо, що коваріаційна матриця є *позитивно* визначеною, тобто для деяких t_1, t_2, …, t_d маємо

$$\sum_{i,j} t_it_j \operatorname{Cov}(X_i, X_j) = 0.$$

Це означає, що $\eta = t_1X_1 + t_2X_2 + \ldots + t_dX_d$ має одноточковий розподіл, тобто існує $c \in \mathbb{R}$ такий, що

$$\mathbb{P}(t_1X_1 + t_2X_2 + \ldots + t_dX_d = c) = 1,$$

тому з імовірністю 1 X приймає значення в деякому $d-1$-вимірному афінному підпросторі. □

Відзначимо корисний

Наслідок: Змінна $X = (X_1, X_2, \ldots, X_d)$ має попарно некорельовані координати тоді і тільки тоді, коли коваріаційна матриця є діагональною. Зокрема, якщо координати X_1, X_2, …, X_d незалежні, то X має діагональну коваріаційну матрицю (але не навпаки!).

Приклад 7.1. Розглянемо багатовимірний нормальний розподіл. Нехай $m \in \mathbb{R}^d$, нехай A — симетрична позитивно визначена матриця $d \times d$, і припустимо, що розподіл $X = (X_1, X_2, \ldots, X_d)$ має щільность

$$g(x) = \frac{\sqrt{\det A}}{(2\pi)^{d/2}} \exp\left[-\frac{1}{2}\langle A(x-m), (x-m)\rangle\right].$$

Теорема 7.10. *Маємо* $\mathbb{E}X = m$, *а коваріаційна матриця* X *дорівнює* A^{-1}.

Залишимо доведення цієї теореми як вправу.

Теорема 7.11. *Припустимо, $X = (X_1, X_2, \ldots, X_d)$ має d-вимірний нормальний розподіл. Тоді змінні X_1, X_2, …, X_d незалежні тоді і тільки тоді, коли вони некорельовані.*

Доведення. $\Rightarrow$ В цю сторонну імплікація виконується для будь-яких випадкових величин.

$\Leftarrow$ Якщо координати некорельовані, то, як відомо, коваріаційна матриця діагональна:

$$\Lambda = \begin{bmatrix} \sigma_1^2 & 0 & \ldots & 0 \\ 0 & \sigma_2^2 & \ldots & 0 \\ & \ldots & & \\ 0 & 0 & \ldots & \sigma_d^2 \end{bmatrix}.$$

Отже, $A = \Lambda^{-1}$ також є діагональною, і її елементи на головній діагоналі дорівнюють $1/\sigma_1^2$, $1/\sigma_2^2$, …, $1/\sigma_d^2$, тому

$$\begin{aligned} g(x) &= \frac{\sqrt{\sigma_1^{-2}\sigma_2^{-2}\ldots\sigma_d^{-2}}}{(2\pi)^{d/2}} \exp\left[-\frac{1}{2}\sum_{j=1}^{d}(x_j - m_j)\cdot\sigma_j^{-2}\right] \\ &= \prod_{j=1}^{d}\left(\frac{1}{\sqrt{2\pi}\sigma_j}e^{-(x_j-m_j)^2/(2\sigma_j^2)}\right) \\ &= g_1(x_1)g_2(x_2)\ldots g_d(x_d), \end{aligned}$$

де g_j — щільність розподілу $\mathcal{N}(m_j, \sigma_j^2)$. Звідси – незалежність. □

Розглянемо зараз т. зв проблему лінійної регресії, яка відіграє важливу роль у статистиці. Задачу можна сформулювати так. Припустимо, що у нас є випадкові величини X, Y і ми знаємо їх спільний розподіл. Крім того, припустімо, що ми спостерігаємо значення X, а Y складніше або неможливо виміряти. Це породжує цікаву проблему оптимального наближення Y на подставі X. Звичайно, ми повинні правильно поставити проблему; ми шукатимемо оптимальне *лінійне* наближення, тобто $aX+b$, a, $b \in \mathbb{R}$, і вимірюємо похибку в середньоквадратичному сенсі. Іншими словами, ми шукаємо константи a, $b \in \mathbb{R}$, для яких $f(a,b) = \mathbb{E}(Y - aX - b)^2$ приймає найменше значення.

Щоб розв'язати цю задачу, зауважимо, що при фіксованим a функція $b \mapsto f(a,b)$ є квадратним тричленом, який приймає найменше значення в $\mathbb{E}(Y - aX)$. Отже, достатньо визначити найменше значення функції

$$h(a) = f(a, \mathbb{E}(Y - aX)) = \mathbb{E}(Y - \mathbb{E}Y - a(X - \mathbb{E})X))^2 = a^2\mathrm{Var}X - 2a\mathrm{Cov}(X,Y) + \mathrm{Var}Y.$$

Якщо X є константою м.н. (тобто $\mathrm{Var}X$=0), то h є постійною функцією, і можна побачити, що оптимальна лінійна оцінка Y є її середнім: $aX + b = aX + (\mathbb{E}Y - a\mathbb{E}X) = \mathbb{E}Y$. Якщо $\mathrm{Var}X \neq 0$, то h є квадратичним тричленом відносно a, що приймає найменше значення в

$$a = \frac{\mathrm{Cov}(X,Y)}{\mathrm{Var}\ X}$$

і тому

$$b = \mathbb{E}Y - \mathbb{E}X \cdot \frac{\mathrm{Cov}(X,Y)}{\mathrm{Var}X}.$$

Зауваження:

1) Можна побачити, що для наведених вище розрахунків нам не потрібні були всі знання про спільний розподіл змінних (X, Y). Нам потрібно лише знати середні значення та дисперсії X, Y та їхні коваріації.

2) Припустимо, що дисперсії X і Y відмінні від нуля. Для наведених вище (оптимальних) a, b маємо

$$f(a,b) = \operatorname{Var} X(1 - \rho^2(X,Y)),$$

де

$$\rho(X,Y) = \frac{\operatorname{Cov}(X,Y)}{\sqrt{\operatorname{Var} X \operatorname{Var} Y}}$$

є т. зв *коефіцієнт кореляції*. Цей коефіцієнт має такі властивості:

(а) $-1 \leqslant \rho(X,Y) \leqslant 1$,

(б) $\rho(X,Y) = \rho(Y,X)$ і, для будь-яких a, $b \in \mathbb{R}$, $\rho(aX + b, Y) = \rho(X,Y)$.

(в) Якщо $|\rho(X,Y)| = 1$, то $X = \alpha Y + \beta$ для деяких α, $\beta \in \mathbb{R}$; іншими словами, між X і Y існує лінійна залежність.

(г) Рівність $\rho(X,Y) = 0$ виконується тоді і тільки тоді, коли X, Y некорельовані. Тоді найкращим наближенням Y є $\mathbb{E}Y$.

8 Завдання

1. Дано випадкову величину X таку, що $\mathbb{P}(X=0)=\mathbb{P}(X=1)=\frac{1}{4}$, $\mathbb{P}(X=-3)=\frac{1}{2}$. Обчисліть $\mathbb{E}X$, $\mathbb{E}\frac{1}{X+2}$, $\mathbb{E}\cos(\pi X)$ і $\operatorname{Var} X$.

2. Випадкова величина X ма розподіл Пуассона з параметром 2. Обчислити $\mathbb{E}6^X$.

3. Випадкова величина X має щільність розподілу

$$g(x) = \frac{3}{8}x^2 1_{[0,2]}.$$

Обчислити $\mathbb{E}X$, $\mathbb{E}\frac{1}{1+x^3}$ та $\operatorname{Var} X^2$.

4. Випадкова величина X має функцію розподілу

$$F(t) = \begin{cases} 0 & \text{jeśli } t < 0, \\ t/2 & \text{jeśli } 0 \leqslant t < 1, \\ 3/4 & \text{jeśli } 1 \leqslant t < 5, \\ 1 & \text{jeśli } t \geqslant 5. \end{cases}$$

Обчислити $\mathbb{E}(2X+1)$.

5. В урні 50 білих куль. Витягуємо і повертаємо по одній кулі, витягнуту кулю фарбуємо в червоний колір, якщо вона біла. Нехай X буде кількістю червоних кульок в урні після 20 розіграшів. Обчислити $\mathbb{E}X$ та $\operatorname{Var} X$.

6. Кожна сторона та кожна діагональ правильного шестикутника розфарбовано випадковим чином одним з трьох кольорів. Вибір кожного кольору однаково ймовірний, забарвлення різних розділів незалежні. Нехай X позначає кількість одноколірних трикутників, вершини яких є вершинами шестикутника. Обчислити $\mathbb{E}X$.

7. Ми кидаємо кубики, поки не отримаємо всі числа на гранях. Обчисліть математичне сподівання та дисперсію кількості підкидань.

8. Доведіть, що для будь-якої невід'ємної випадкової величини X і $p > 0$ справедлива формула

$$\mathbb{E}X^p = p\int_0^\infty t^{p-1}\mathbb{P}(X \geqslant t)\mathrm{d}t = p\int_0^\infty t^{p-1}\mathbb{P}(X > t)\mathrm{d}t.$$

Зробіть з цього висновок, що якщо X має дискретний розподіл зосереджений на невід'ємних цілих числах, то

$$\mathbb{E}X = \sum_{k=1}^{\infty}\mathbb{P}(X \geqslant k) = \sum_{k=0}^{\infty}\mathbb{P}(X > k).$$

9. Числа $1, 2, \ldots, n$ розставлено навмання в рядок $(a_1, a_2, \ldots, a_n)$. Нехай N — найбільше число таке, що $a_k > a_{k-1}$ для $k \leqslant N$. Обчислити $\mathbb{E}N$.

10. Дано послідовність незалежних випадкових величин $X_0, X_1, X_2, \ldots$ з однаковим розподілом і неперервною функцією розподілу. Нехай $\eta = \inf\{n : X_n > X_0\}$. Визначити розподіл η і обчислити $\mathbb{E}\eta$.

11. Паличку довжиною 1 ламають у навмання обраній точці з рівномірно розподіленою ймовірністю. Обчисліть сподівання відношення

а) довжини лівого шматка до довжини правого шматка.

б) довжини коротшого шматка до довжини довшого шматка.

12. Випадкові величини X, Y задовольняють умовам $\mathrm{Var}X = 3$, $\mathrm{Cov}(X, Y) = -1$, $\mathrm{Var}Y = 2$. Обчисліть $\mathrm{Var}(4X - 3Y)$ і $\mathrm{Cov}(5X - Y, 2X + Y)$.

13. Випадкова величина X має дисперсію $\sigma^2 < \infty$. Доведіть, що

$$\mathbb{P}(|X - \mathbb{E}X| > 3\sigma) \leqslant \frac{1}{9}.$$

14. Випадкові величини $\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_n$ є незалежними та мають однаковий розподіл $\mathbb{P}(\varepsilon_k = 1) = \mathbb{P}(\varepsilon_k = -1) = 1/2$, $k = 1, 2, \ldots, n$. Нехай $a_1, a_2, \ldots, a_n$ — послідовність дійсних чисел і $A = (\sum_{k=1}^n a_k^2)^{1/2}$. Доведіть, що

$$\mathbb{P}\left(\left|\sum_{k=1}^n a_k \varepsilon_k\right| > t\right) \leqslant 2\exp(-t^2/2A^2).$$

15. Змінні $\varepsilon_1, \varepsilon_2, \ldots$ є незалежними та мають однаковий розподіл $\mathbb{P}(\varepsilon_k = 1) = \mathbb{P}(\varepsilon_k = -1) = 1/2$, $k = 1, 2, \ldots$. Нехай $S_n = \varepsilon_1 + \varepsilon_2 + \ldots + \varepsilon_n$, $n = 1, 2, \ldots$. Доведіть, що

$$\mathbb{P}\left(\limsup_{n\to\infty} \frac{S_n}{\sqrt{2n\log n}} \leqslant 1\right) = 1$$

та

$$\mathbb{P}\left(\liminf_{n\to\infty} \frac{S_n}{\sqrt{2n\log n}} \geqslant -1\right) = 1.$$

16. Випадкова величина X має таку властивість: для $n = 1, 2, \ldots$ маємо

$$\mathbb{E}|X|^n \leqslant \binom{2n}{n}.$$

Доведіть, що $X \in L^\infty$ (тобто існує таке число M, що $\mathbb{P}(|X| \leqslant M) = 1$).

17. Випадкова величина X має нормальний розподіл в $\mathbb{R}^d$ із середнім m і коваріаційною матрицею Λ. Нехай T — афінне перетворення $\mathbb{R}^d$ на $\mathbb{R}^k$, $k \leqslant d$. Доведіть, що TX має нормальний розподіл в $\mathbb{R}^k$. Знайдіть її сподівання та коваріаційну матрицю.

18. Випадкова величина X має d-вимірний нормальний розподіл із щільністю

$$g(x) = \frac{\sqrt{\det A}}{(2\pi)^{d/2}} \exp\left[-\frac{1}{2}\langle A(x - m), (x - m)\rangle\right].$$

Доведіть, що $\mathbb{E}X = m$ та $\Lambda = A^{-1}$ (Λ є коваріаційною матрицею X).

19. Випадкова величина (X, Y) має двовимірний нормальний розподіл із середнім значенням $(0, 0)$ і коваріаційною матрицею

$$\Lambda = \begin{bmatrix} 2 & 1 \\ 1 & 1 \end{bmatrix}.$$

а) Визначте щільність (X, Y).

b) Визначте розподіл $X + 3Y$.

в) Знайдіть таке число $a \in \mathbb{R}$, щоб змінні $X + Y$, $X + aY$ були незалежними.

20. Передавач надсилає ξ, а приймач отримує $\eta = a\xi + \zeta$, де $a \in \mathbb{R}_+$ — коефіцієнт підсилення, а ζ - змінна перешкод. Ми припускаємо, що ξ і ζ є незалежними випадковими величинами, де $\mathbb{E}\xi = m$, $\mathrm{Var}\,\xi = 1$, $\mathbb{E}\zeta = 0$, $\mathrm{Var}\,,\zeta = \sigma^2$. Визначте коефіцієнт кореляції ξ і η і лінійну регресію ξ відносно η (тобто найкраще лінійне наближення ξ за допомогою η).

9 Різні види збіжності випадкових величин

Тепер перейдемо до граничної поведінки послідовностей випадкових величин. Почнемо з корисного факту.

Визначення 9.1. Припустимо, що $(\Omega, \mathcal{F}, \mathbb{P})$ — імовірнісний простір, а $\mathcal{F}_1$, $\mathcal{F}_2$, $\ldots \subseteq \mathcal{F}$ — послідовність σ-алгебр. Тоді σ-алгебру

$$\bigcap_{n=1}^{\infty} \sigma(\mathcal{F}_n, \mathcal{F}_{n+1}, \mathcal{F}_{n+2}, \ldots)$$

називаємо *залишковою σ-алгеброю.*

Приклад 9.1. Припустимо, що X_1, X_2, $\ldots$ — послідовність випадкових величин, а $\mathcal{F}_n = \sigma(X_n)$ — σ-алгебра, породжена змінною X_n, $n = 1$, 2, $\ldots$. Тоді кожна подія

$$\{(X_n)_{n\geqslant 1} \text{ сходиться}\},\ \{\sup_n |X_n| < \infty\},\ \left\{\sum_{n=1}^{\infty} X_n \text{ сходиться}\right\}$$

належить до залишкової σ-алгебри.

Теорема 9.1 (0 − 1 закон Колмогорова)**.** *Припустимо, що $(\Omega, \mathcal{F}, \mathbb{P})$ є ймовірнісним простором і σ-алгебри $\mathcal{F}_1$, $\mathcal{F}_2$, $\ldots \subseteq \mathcal{F}$ незалежні. Тоді для кожного $A \in \bigcap_{n=1}^{\infty} \sigma(\mathcal{F}_n, \mathcal{F}_{n+1}, \ldots)$ маємо $\mathbb{P}(A) = 0$ або $\mathbb{P}(A) = 1$.*

Лема 9.1. *Припустимо, $\mathcal{G}_1 \subseteq \mathcal{G}_2 \subseteq \ldots \subseteq \mathcal{F}$ є зростаючою послідовністю σ-алгебр, і нехай $\mathcal{G} = \sigma(\mathcal{G}_1, \mathcal{G}_2, \ldots)$. Тоді для будь-якого $A \in \mathcal{G}$ існує послідовність $(A_n)_{n\geqslant 1}$ така, що $A_n \in \mathcal{G}_n$ для кожного n і $lim_{n\to\infty} \mathbb{P}(A \Delta A_n) = 0$ (Δ тут означає симетричну різницю множин).*

Доведення. Введемо клас множини

$$\mathcal{K} = \{A \in \mathcal{G} : \text{є послідовність } (A_n)_{n\geqslant 1} \text{що задовольняє умови леми}\}.$$

Очевидно, що $\bigcup_{n\geqslant 1} \mathcal{G}_n \subset \mathcal{K}$, і в лівій частині мами π-систему. Отже, достатньо показати, що $\mathcal{K}$ є λ-системою. Перевіряємо:

(i) $\Omega \in \mathcal{K}$ - це очевидно.

(ii) Припустимо, що A, $B \in \mathcal{K}$ задовольняють умову $A \subseteq B$ і нехай $(A_n)_{n\geqslant 1}$, $(B_n)_{n\geqslant 1}$ є відповідними наближеннями. Ми покажемо, що послідовність $(B_n \setminus A_n)_{n\geqslant 1}$ апроксимує $B \setminus A$. Очевидно $B_n \setminus A_n \in \mathcal{G}_n$ для кожного n. Крім того, використовуючи тотожність $\mathbb{P}(A\Delta B) = \mathbb{E}|1_A - 1_B|$, маємо

$$\begin{aligned} \mathbb{P}\big((B_n \setminus A_n)\Delta(B \setminus A)\big) &= \mathbb{E}\big|1_{B_n\setminus A_n} - 1_{B\setminus A}\big| \\ &= \mathbb{E}\big|1_{B_n} - 1_{A_n\cap B_n} - 1_B + 1_A\big| \\ &\leqslant \mathbb{E}|1_{B_n} - 1_B| + \mathbb{E}|1_A - 1_{A_n}1_{B_n}|. \end{aligned}$$

Перший член дорівнює $\mathbb{P}(B_n\Delta B)$, тому він збігається до 0, коли $n \to \infty$. Крім того,

$$\begin{aligned} \mathbb{E}|1_A - 1_{A_n}1_{B_n}| &= \mathbb{E}\big|1_A - (1_{A_n} - 1_A + 1_A)(1_{B_n} - 1_B + 1_B)\big| \\ &\leqslant \mathbb{E}|1_A - 1_A 1_B| + \mathbb{E}1_A|1_{B_n} - 1_B| \\ &\quad + \mathbb{E}1_B|1_{A_n} - 1_A| + \mathbb{E}|1_{A_n} - 1_A||1_{B_n} - 1_B| \\ &\leqslant 0 + \mathbb{P}(B\Delta B_n) + \mathbb{P}(A\Delta A_n) + 2\mathbb{P}(A\Delta A_n) \end{aligned}$$

також прямує до нуля, коли n прямує до нескінченності. Отже ми довели, що $B \setminus A \in \mathcal{K}$.

(iii) Якщо $A_1, A_2, \ldots$ є зростаючою послідовністю елементів із $\mathcal{K}$, то $\bigcup_{n=1}^{\infty} A_n$ також лежи в $\mathcal{K}$. Дійсно, сума $\bigcup_{n=1}^{\infty} A_n$ може бути як завгодно точно апроксимована $\bigcup_{n=1}^{N} A_n = A_N$, і тоді множина A_N апроксимується за допомогою відповідної послідовністі подій з σ-алгебр $\mathcal{G}_n$. □

Доведення теореми Колмогорова. Розглянемо висхідну послідовність σ-алгебр

$$\mathcal{G}_n = \sigma(\mathcal{F}_1, \mathcal{F}_2, \ldots, \mathcal{F}_n), \qquad n = 1, 2, \ldots.$$

Візьмемо подію A, що належить залишковій σ-алгебрі. Звичайно, вона також належить до σ-алгебри

$$\sigma(\mathcal{F}_1, \mathcal{F}_2, \ldots) = \sigma(\mathcal{G}_1, \mathcal{G}_2, \ldots),$$

отже, за наведеною вище лемою ми можемо вказати послідовність $(A_n)_{n\geqslant 1}$ таку, що $A_n \in \mathcal{G}_n$ і $\mathbb{P}(A\Delta A_n) \to 0$, коли $n \to \infty$. Оскільки $\mathbb{P}(A\Delta A_n) = \mathbb{P}(A \setminus A_n) + \mathbb{P}(A_n \setminus A)$, то кожна з цих двох ймовірностей також збігається до 0. Але

$$\mathbb{P}(A \setminus A_n) = \mathbb{P}(A) - \mathbb{P}(A \cap A_n),$$

тому $\mathbb{P}(A \cap A_n) \to \mathbb{P}(A)$. На додаток,

$$\mathbb{P}(A_n \setminus A) = \mathbb{P}(A_n) - \mathbb{P}(A \cap A_n),$$

тому в поєднанні з попередньою збіжністю $\mathbb{P}(A_n) \to \mathbb{P}(A)$. Нарешті, для будь-яких n маємо

$$A \in \sigma(\mathcal{F}_{n+1}, \mathcal{F}_{n+2}, \ldots), \qquad A_n \in \mathcal{G}_n,$$

і ці σ-алгебри незалежні. Отже, A і A_n також незалежні, $\mathbb{P}(A \cap A_n) = \mathbb{P}(A)\mathbb{P}(A_n)$, ліва сторона йде до $\mathbb{P}(A)$, а права сторона до $\mathbb{P}(A)^2$. Отже, $\mathbb{P}(A) \in \{0, 1\}$. □

Тепер перейдемо до збіжності випадкових величин.

Визначення 9.2. Нехай $(X_n)_{n\geqslant 1}$ є послідовністю випадкових величин зі значеннями в $\mathbb{R}^d$. Ми говоримо, що

(i) X_n збігається до X майже напевно, якщо $\mathbb{P}(\lim_{n\to\infty} X_n = X) = 1$. Позначення: $X_n \to X$ п.н.

(ii) ($p \geqslant 1$, $d = 1$) X_n збігається до X в L^p, якщо $X_1, X_2, \ldots \in L^p$ і $\lim_{n\to\infty} ||X_n - X||_p = 0$ (нагадаємо: $||\xi||_p = (\mathbb{E}|\xi|^p)^{1/p}$ для $p < \infty$, $||\xi||_\infty = \operatorname{esssup}|\xi|$). Позначення: $X_n \to X$ в L^p.

(iii) X_n збігається до X за ймовірністю, якщо для кожного $\varepsilon > 0$ $\lim_{n\to\infty} \mathbb{P}(|X_n - X| > \varepsilon) = 0$. Позначення: $X_n \xrightarrow{\mathbb{P}} X$.

Теорема 9.2. *Якщо $X_n \to X$ п.н., то $X_n \xrightarrow{\mathbb{P}} X$. Протилежна імплікація не виконується.*

Доведення. За визначенням, послідовність $(X_n)_{n\geqslant 1}$ майже напевно сходиться до X, якщо

$$\mathbb{P}\left(\bigcap_{\varepsilon>0} \bigcup_{N=1}^{\infty} \bigcap_{n\geqslant N} \{\omega : |X_n(\omega) - X(\omega)| < \varepsilon\}\right) = 1.$$

Це еквівалентно умові, що для кожного $\varepsilon > 0$

$$\mathbb{P}\left(\bigcup_{N=1}^{\infty} \bigcap_{n\geqslant N} \{\omega : |X_n(\omega) - X(\omega)| < \varepsilon\}\right) = 1.$$

Але послідовність подій $\left(\bigcap_{n\geqslant N}\{|X_n - X| < \varepsilon\}\right)_{N\geqslant 1}$ є зростаючою; згідно з теоремою про неперервність вищенаведена рівність означає, що для будь-якого $\varepsilon > 0$,

$$\lim_{N\to\infty} \mathbb{P}\left(\bigcap_{n\geqslant N}\{\omega : |X_n(\omega) - X(\omega)| < \varepsilon\}\right) = 1.$$

Тому тим більше

$$\lim_{N\to\infty} \mathbb{P}\left(\{\omega : |X_N(\omega) - X(\omega)| < \varepsilon\}\right) = 1,$$

тобто після перемикання на протилежну подію $\mathbb{P}(|X_N - X| \geqslant \varepsilon) \to 0$.

Щоб довести, що протилежна імплікація не виконується, розглянемо наступний приклад. Припустимо, що ймовірнісний простір — це інтервал $[0,1]$ з його борелівськими підмножинами та мірою Лебега. Нехай

$$\begin{aligned}
&X_1 = 1_{[0,1)},\\
&X_2 = 1_{[0,1/2)}, \quad X_3 = 1_{[1/2,1)},\\
&X_4 = 1_{[0,1/4)}, \quad X_5 = 1_{[1/4,1/2)}, \quad X_6 = 1_{[1/2,3/4)}, \quad X_7 = 1_{[3/4,1)},\\
&\quad \ldots
\end{aligned}$$

Тоді послідовність $(X_n)_{n\geqslant 0}$ збігається до 0 за ймовірністю: для будь-якого ε $\mathbb{P}(|X_n - 0| > \varepsilon)$ є степенем двійки, що має за кожним разом менший цілий показник. З іншого боку, для будь-якого $\omega \in [0,1)$ числова послідовність $(X_n(\omega))_{n\geqslant 1}$ не збіжна; це послідовність, що містить нескінченну кількість нулів і нескінченну кількість одиниць. $\square$

Теорема 9.3. *Якщо $X_n \to X$ в L^p, то $X_n \xrightarrow{\mathbb{P}} X$. Протилежна імплікація не виконується.*

Доведення. Згідно з нерівністю Чебишева, для будь-якого ε маємо

$$\mathbb{P}(|X_n - X| \geqslant \varepsilon) \leqslant \frac{\mathbb{E}|X_n - X|^p}{\varepsilon^p} \xrightarrow{n\to\infty} 0.$$

Покажемо, що імплікація в інший бік не виконується. Ми розглянемо лише $p < \infty$, залишивши випадок $p = \infty$ Читачеві. Розглянемо ймовірнісний простір $([0,1], \mathcal{B}(0,1), |\cdot|)$ і послідовність $(X_n)_{n\geqslant 1}$ змінних, заданих формулою

$$X_n(\omega) = n^{1/p} 1_{[0,1/n]}(\omega).$$

Тоді $X_n \xrightarrow{\mathbb{P}} 0$: для будь-якого ε маємо

$$\mathbb{P}(|X_n - 0| > \varepsilon) \leqslant 1/n \to 0.$$

Таким чином, якби послідовність $(X_n)_{n\geqslant 1}$ сходилася в L^p, вона сходилася б до змінної з нульовим носієм (за наслідком, який ми щойно довели). Але

$$\mathbb{E}|X_n - 0|^p = \mathbb{E}|X_n|^p = 1 \not\to 0. \qquad \square$$

Теорема 9.4. *Якщо $p < p'$ і $X_n \to X$ в $L^{p'}$, то $X_n \to X$ в L^p.*

Доведення. Це безпосередньо випливає з нерівності Гельдера: маємо

$$||X_n - X||_p \leqslant ||X_n - X||_{p'} \xrightarrow{n\to\infty} 0. \qquad \square$$

Теорема 9.5. *a) Послідовність $(X_n)_{n\geqslant 1}$ збігається за ймовірністю тоді і тільки тоді, коли вона задовольняє умову Коші за ймовірністю:*

$$\forall_{\delta>0}\forall_{\varepsilon>0}\exists_N\forall_{m,n\geqslant N}\mathbb{P}(|X_n - X_m| > \varepsilon) < \delta.$$

b) Якщо X_n збігається до X за ймовірністю, то існує підпослідовність $(n_k)_{k\geqslant 1}$ така, що $(X_{n_k})_{k\geqslant 1}$ збігається до X п.н.

Визначення 9.3. Припустимо, що $\{X_i\}_{i\in\mathcal{I}}$ є родиною інтегровних випадкових величин. Ми говоримо, що ця родина є рівномірно інтегровною, якщо

$$\sup_{i\in\mathcal{I}}\int_{\{|X_i|\geqslant r\}}|X_i|\mathrm{d}\mathbb{P}\xrightarrow{r\to\infty}0.$$

Приклади:

1) Припустимо, що існує невід'ємна інтегровна змінна η така, що $|X_i|\leqslant\eta$ для всіх $i\in\mathcal{I}$. Тоді $\{X_i\}_{i\in\mathcal{I}}$ є рівномірно інтегровною родиною. Дійсно,

$$\sup_{i\in\mathcal{I}}\int_{\{|X_i|\geqslant r\}}|X_i|\mathrm{d}\mathbb{P}\leqslant\int_{\{\eta\geqslant r\}}|\eta|\mathrm{d}\mathbb{P}\xrightarrow{r\to\infty}0$$

за теоремою Лебега про мажоровану збіжність.

2) Кожна скінченна родина інтегровних змінних є рівномірно інтегровною: просто скористайтеся попереднім прикладом, взявши $\eta=\sum_{i\in\mathcal{I}}|X_i|$.

3) Будь-яка рівномірно інтегровна родина випадкових величин при додаванні до неї скінченної кількості інтегровних змінних залишається рівномірно інтегровною.

4) Розглянемо ймовірнісний простір $([0,1],\mathcal{B}([0,1]),|\cdot|)$ і послідовність X_1, X_2, …, задану $X_n(\omega)=n^2 1_{[0,1/n]}(\omega)$. Ця родина не є рівномірно інтегровною: маємо

$$\{|X_n|\geqslant m^2\}=\{n^2 1_{[0,1/n]}\geqslant m^2\}=\begin{cases}\emptyset & \text{для } n<m,\\ [0,1/n] & \text{для } n\geqslant m,\end{cases}$$

отже

$$\int_{\{|X_n|\geqslant m^2\}}|X_n|\mathrm{d}\mathbb{P}=\begin{cases}0 & \text{для } n<m,\\ n^2\cdot 1/n & \text{для } n\geqslant m,\end{cases}$$

і для кожного r, $\sup_n\int_{\{|X_n|\geqslant r\}}|X_n|\mathrm{d}\mathbb{P}=\infty$.

З іншого боку, родина змінних $\{Y_n\}_{n\geqslant 1}=\{\sqrt{n}1_{[0,1/n]}\}_{n\geqslant 1}$ є рівномірно інтегровною. Повторюючи наведені міркування, ми бачимо, що

$$\int_{\{|X_n|\geqslant r\}}|X_n|\mathrm{d}\mathbb{P}=\begin{cases}0 & \text{для } \sqrt{n}<r,\\ \sqrt{n}\cdot 1/n & \text{для } \sqrt{n}\geqslant r.\end{cases}$$

Отже, для фіксованого r,

$$\sup_n\int_{\{|X_n|\geqslant r\}}|X_n|\mathrm{d}\mathbb{P}\leqslant 1/r,$$

що збігається до 0, коли $r\to\infty$.

Отримаємо тепер умову еквівалентну рівномірної інтегровності.

Теорема 9.6. *Родина $\{X_i\}_{i\in\mathcal{I}}$ рівномірно інтегровна тоді і тільки тоді, коли виконуються такі дві умови:*

1° $\sup_{i\in\mathcal{I}}\mathbb{E}|X_i|<\infty$,

2° Для кожного $\varepsilon>0$ існує такий $\delta>0$, що якщо подія A задовольняє $\mathbb{P}(A)<\delta$, то

$$\int_A |X_i| d\mathbb{P} < \varepsilon, \qquad i\in\mathcal{I}.$$

Доведення. $\Rightarrow$ Почнемо з умови 2°. Для кожного $A\in\mathcal{F}$ і $i\in\mathcal{I}$ маємо

$$\int_A |X_i|\mathrm{d}\mathbb{P} = \int_{A\cap\{|X_i|\geqslant r\}}|X_i|\mathrm{d}\mathbb{P} + \int_{A\cap\{|X_i|<r\}}|X_i|\mathrm{d}\mathbb{P} \leqslant \sup_{i\in\mathcal{I}}\int_{\{|X_i|\geqslant r\}}|X_i|\mathrm{d}\mathbb{P} + r\mathbb{P}(A).$$

Отже, при $\varepsilon>0$ ми беремо r так, щоб перший член був меншим за $\varepsilon/2$ (це можливо за визначенням рівномірної інтегровності); далі беремо $\delta=\varepsilon/(2r)$: тоді другий член також менше $\varepsilon/2$. Крім того, взявши $A=\Omega$, ми отримаємо, що для кожного r,

$$\sup_{i\in\mathcal{I}}\mathbb{E}|X_i| \leqslant \sup_{i\in\mathcal{I}}\int_{\{|X_i|\geqslant r\}}|X_i|\mathrm{d}\mathbb{P} + r < \infty,$$

що є бажаною умовою 1°.

$\Leftarrow$ Для будь-якого $i\in\mathcal{I}$ з нерівності Чебишева та 1° випливає

$$\mathbb{P}(|X_i|\geqslant r) \leqslant \frac{\mathbb{E}|X_i|}{r} \leqslant \sup_{i\in\mathcal{I}}\mathbb{E}|X_i|/r < \infty.$$

Тоді для будь-якого $\varepsilon>0$ виберемо δ з умови 2°. Наведений вище розрахунок показує, що для достатньо великих r ми маємо $\sup_{i\in\mathcal{I}}\mathbb{P}(|X_i|\geqslant r)<\delta$, тому з 2°,

$$\sup_{i\in\mathcal{I}}\int_{\{|X_i|\geqslant r\}}|X_i|\mathrm{d}\mathbb{P} < \varepsilon.$$

Це означає, що виконується умова рівномірної інтегровності. □

Теорема 9.7. *Нехай $p\geqslant 1$ — фіксоване число. Послідовність $(X_n)_{n\geqslant 1}$ збігається в L^p тоді і тільки тоді, коли вона збігається за ймовірністю та родина $\{|X_n|^p\}_{n\geqslant 1}$ є рівномірно інтегровною.*

Доведення. $\Rightarrow$ Збіжність за ймовірністю мами автоматично, залишається показати рівномірну інтегровність. Для будь-якого $A\in\mathcal{F}$,

$$(*)\qquad \left(\int_A |X_n|^p\mathrm{d}\mathbb{P}\right)^{1/p} = ||X_n 1_A||_p \leqslant ||X1_A||_p + ||(X-X_n)1_A||_p.$$

Для $A=\Omega$ нерівність $(*)$ дає $||X_n||_p\leqslant ||X||_p+||X_n-X||_p$, тому $\sup_n||X_n||_p<\infty$, що дає умову 1° з попередньої теореми. Щоб довести 2°, виберемо $\varepsilon>0$. З визначення збіжності в L^p існує N таке, що $||X_n-X||_p<\varepsilon/2$ для $n\geqslant N$. Родина $\{|X|^p,|X_1-X|^p,\ldots,|X_N-X|^p\}$ є скінченною і містить інтегровні випадкові змінні, тому вона рівномірно інтегровна (див. приклад 2 вище) і задовольняє умову 2°: існує такий δ, що якщо $\mathbb{P}(A)<\delta$, то

$$\left(\int_A |X|^p\mathrm{d}\mathbb{P}\right)^{1/p} < \varepsilon/2, \qquad \left(\int_A |X_i-X|^p\mathrm{d}\mathbb{P}\right)^{1/p} < \varepsilon/2, \quad i=1,2,\ldots,N.$$

Тепер достатньо поєднати всі наведені міркування та (*): якщо $\mathbb{P}(A)<\delta$, то

$$\sup_n\int_A |X_n|^p\mathrm{d}\mathbb{P} \leqslant \varepsilon^p.$$

$\Leftarrow$ Оскільки $X_n \to X$ за ймовірністю, ми можемо вибрати підпослідовність $(X_{n_k})_{k\geqslant 1}$, що збігається до X майже напевно. З Леми Фату, $X \in L^p$:

$$\mathbb{E}|X|^p = \mathbb{E}\lim_{k\to\infty}|X_{n_k}|^p \leqslant \liminf_{k\to\infty}\mathbb{E}|X_{n_k}|^p \leqslant \sup_n \mathbb{E}|X_n|^p < \infty.$$

Далі маємо

$$\begin{aligned} ||X_n - X||_p &\leqslant ||(X_n - X)1_{\{|X_n - X|\geqslant\alpha\}}||_p + ||(X_n - X)1_{\{|X_n - X|<\alpha\}}||_p \\ &\leqslant \left(\int_{\{|X_n - X|\geqslant\alpha\}}|X_n|^p \mathrm{d}\mathbb{P}\right)^{1/p} + \left(\int_{\{|X_n - X|\geqslant\alpha\}}|X|^p \mathrm{d}\mathbb{P}\right)^{1/p} \\ &\quad + \left(\int_{\{|X_n - X|<\alpha\}}|X_n - X|^p \mathrm{d}\mathbb{P}\right)^{1/p}. \end{aligned}$$

Отже виберемо будь-який $\varepsilon > 0$ і положимо $\alpha = \varepsilon/3$. З умови 2° ми отримуємо існування δ такого, що якщо $\mathbb{P}(A) < \delta$, то

$$\sup_n \int_A |X_n|^p \mathrm{d}\mathbb{P} < (\varepsilon/3)^p\,, \qquad \int_A |X|^p < (\varepsilon/3)^p.$$

Крім того, за визначенням збіжності за ймовірністю існує N таке, що для $n \geqslant N$, $\mathbb{P}(|X_n - X| \geqslant \alpha) < \delta$. Звідси теза, оскільки перші два доданки у наведеній вище оцінці менші за $\varepsilon/3$ і

$$\left(\int_{\{|X_n - X|<\alpha\}}|X_n - X|^p \mathrm{d}\mathbb{P}\right)^{1/p} \leqslant \left(\int_{\{|X_n - X|<\alpha\}}\alpha^p \mathrm{d}\mathbb{P}\right)^{1/p} \leqslant \alpha = \varepsilon/3. \qquad \square$$

10 Завдання

1. Змінні $(X_n)_{n\geqslant 1}$ є незалежними змінними Радемахера. Доведіть, що $(X_n)_{n\geqslant 1}$ не збігається п.н. Чи збігається $(X_n)_{n\geqslant 1}$ за ймовірністю?

2. Дано послідовність $(X_n)_{n\geqslant 1}$ як вище. Доведіть, що ряд $\sum_{n=1}^{\infty} 2^{-n} X_n$ збігається п.н. і визначте граничний розподіл.

3. Послідовності $(X_n)_{n\geqslant 1}$, $(Y_n)_{n\geqslant 1}$ збігаються за ймовірністю до X, Y відповідно. Доведіть, що

a) $(X_n + Y_n)_{n\geqslant 1}$ збігається за ймовірністю до $X + Y$.

b) $(X_n Y_n)_{n\geqslant 1}$ збігається за ймовірністю до XY.

4. Дано інтегровну випадкову величину X. Нехай для $n \geqslant 1$,

$$X_n(\omega) = \begin{cases} -n & \text{jeśli } X(\omega) < -n, \\ X(\omega) & \text{jeśli } |X(\omega)| \leqslant n, \\ n & \text{jeśli } X(\omega) > n. \end{cases}$$

Чи збігається $(X_n)_{n\geqslant 1}$ до X п.н.? Чи збігається вона в L^1?

5. Дано послідовності $(X_n)_{n\geqslant 1}$, $(Y_n)_{n\geqslant 1}$ збіжні п.н. до змінних X, Y. Доведіть, що якщо для кожного n змінні X_n і Y_n мають однаковий розподіл, то X і Y також мають однаковий розподіл.

6. Змінні X_1, X_2, $\ldots$ є незалежними випадковими величинами з експоненціальним розподілом з параметром λ.

(a) Доведіть, що якщо $\lambda > 1$, то з ймовірністю 1 маємо $\{X_n < \log n\}$ для достатньо великих n, а якщо $\lambda \geqslant 1$, то з імовірність 1 маємо $X_n \geqslant \log n$ для нескінченної кількості n.

(б) Дослідіть збіжність п.н. $(X_n/\log n)_{n\geqslant 2}$.

7. Змінні X_1, X_2, $\ldots$ незалежні, невід'ємні та мають однаковий розподіл (не рівний δ_0). Доведіть, що $\sum_{n=1}^{\infty} X_n = \infty$ з імовірністю 1.

8. Випадкові величини X_1, X_2, $\ldots$ незалежні, мають однаковий розподіл і задовольняють умову $\mathbb{P}(|X_i| < 1) = 1$. Доведіть, що $\lim_{n\to\infty} X_1 X_2 \ldots X_n = 0$ п.н.

9. X_1, X_2, $\ldots$ є незалежними та мають однаковий розподіл.

(a) Доведіть, що послідовність середніх

$$\frac{X_1 + X_2 + \ldots + X_n}{n}, \qquad n = 1, 2, \ldots$$

є або збіжною п.н., або розбіжною з імовірністю 1.

(b) Доведіть, що якщо ця послідовність сходиться п.н., то її границя має одноточковий розподіл.

10. Дано послідовність $(X_n)_{n\geqslant 1}$ незалежних випадкових величин таких, що для $n \geqslant 1$ X_n має розподіл Пуассона з параметром $1/n$. Чи збігається $(X_n)_{n\geqslant 1}$ за ймовірністю? Чи збігається п.н.? Чи збігається в L^2? Чи збігається в $L^{3/2}$?

11. Дано послідовності змінних $(X_n)_{n\geqslant 1}$, $(Y_n)_{n\geqslant 1}$, де $X_n \to X$ в L^p і $Y_n \to Y$ в L^q, де $p, q > 1$ задовольняють умову $1/p + 1/q = 1$. Доведіть, що $(X_n Y_n)_{n\geqslant 1}$ збігається в L^1 до XY.

12. Яким умовам має задовольняти непорожня множина $\Lambda \subseteq (0, \infty)$ для родини випадкових величин $(X_\lambda)_{\lambda\in\Lambda}$, де

(a) $X_\lambda \sim \mathcal{U}([0,\lambda])$,

(b) $X_\lambda \sim \operatorname{Exp}(\lambda)$,

аби вона була рівномірно інтегровною?

13. Функція $G : [0,\infty) \to [0,\infty)$ задана так, що $\lim_{t\to\infty} \frac{G(t)}{t} = \infty$. Нехай $(X_i)_{i\in I}$ — це родина випадкових величин, таких що $\sup_{i\in I} \mathbb{E}G(|X_i|) < \infty$. Доведіть, що ця родина рівномірно інтегровна.

11 Закони великих чисел

Закони великих чисел говорять нам про граничну поведінку послідовності середніх арифметичних
$$\frac{X_1 + X_2 + \ldots + X_n}{n}, \qquad n = 1, 2, \ldots,$$
з різними припущеннями щодо змінних. Почнемо зі слабкого закону великих чисел (СЗВЧ): термін „слабкий" походить від того факту, що тут йдеться про збіжність за ймовірністю.

Теорема 11.1. *Нехай $X_1, X_2, \ldots$ – випадкові величини інтегровні з квадратом. Якщо ці змінні некорельовані та мають рівномірно обмежену дисперсію, тоді*
$$\frac{X_1 + X_2 + \ldots + X_n - \mathbb{E}(X_1 + X_2 + \ldots + X_n)}{n} \to 0$$
за ймовірністю. Зокрема, якщо змінні X_i мають однакове математичне сподівання, то
$$\frac{X_1 + X_2 + \ldots + X_n}{n} \xrightarrow{\mathbb{P}} \mathbb{E}X_1.$$

Доведення. Зазначимо, що
$$\begin{aligned}
&\mathbb{E}\left|\frac{X_1 + X_2 + \ldots + X_n - \mathbb{E}(X_1 + X_2 + \ldots + X_n)}{n}\right|^2 \\
&= \operatorname{Var}\left(\frac{X_1 + X_2 + \ldots + X_n - \mathbb{E}(X_1 + X_2 + \ldots + X_n)}{n}\right) \\
&= \frac{1}{n^2}\operatorname{Var}\left(\sum_{k=1}^{n} X_k\right) \\
&= \frac{1}{n^2}\sum_{k=1}^{n}\operatorname{Var} X_k \leqslant \frac{\sup_{k\geqslant 1}\operatorname{Var} X_k}{n}.
\end{aligned}$$
Отже, для фіксованого $\varepsilon > 0$ з нерівності Чебишева випливає
$$\begin{aligned}
&\mathbb{P}\left(\left|\frac{X_1 + X_2 + \ldots + X_n - \mathbb{E}(X_1 + X_2 + \ldots + X_n)}{n}\right| \geqslant \varepsilon\right) \\
&\leqslant \frac{\sup_{k\geqslant 1}\operatorname{Var} X_k}{n\varepsilon^2}.
\end{aligned}$$
Залишилось зауважити, що останній вираз збігається до 0, коли $n \to \infty$. Звідси отримуємо бажану збіжність за ймовірністю. ☐

Як окремий випадок ми отримуємо т.зв. слабкий закон великих чисел Бернуллі. А саме, розглянемо послідовність $(\xi_n)_{n\geqslant 1}$ випадкових величин (не обов'язково незалежних), при чому для $n \geqslant 1$ змінна ξ_n має розподіл $B(n, p)$ де $p \in (0, 1)$ — фіксований параметр. Тоді
$$\frac{\xi_n}{n} \xrightarrow{\mathbb{P}} p.$$
Дійсно, достатньо взяти послідовність $(X_n)_{n\geqslant 1}$ незалежних (і, отже, некорельованих) випадкових величин з однаковим двоточковим розподілом $\mathbb{P}(X_i = 1) = p = 1 - \mathbb{P}(X_i = 0)$. Тоді $\xi_n \sim X_1 + X_2 + \ldots + X_n$, тому для $\varepsilon > 0$,
$$\lim_{n\to\infty} \mathbb{P}\left(\left|\frac{\xi_n}{n} - p\right| > \varepsilon\right) = \lim_{n\to\infty} \mathbb{P}\left(\left|\frac{X_1 + X_2 + \ldots + X_n}{n} - p\right| > \varepsilon\right) = 0.$$

Головним результатом цієї глави є т. зв. сильний закон великих чисел (ЗВЛ) (Теорема 11.4 нижче), який говорить про збіжність майже напевно. Почнемо з деяких підготовчих фактів.

Теорема 11.2 (Нерівність Колмогорова)**.** *Припустимо, X_1, X_2, …, X_n незалежні та центровані інтегровні з квадратом випадкові величини. Тоді для будь-якого $\alpha > 0$,*

$$\mathbb{P}(\max_{1\leqslant k\leqslant n} |X_1 + X_2 + \ldots + X_k| \geqslant \alpha) \leqslant \frac{1}{\alpha^2}\, Var(X_1 + X_2 + \ldots + X_n).$$

Доведення. Введемо позначення $S_0 = 0$ та $S_k = X_1 + X_2 + \ldots + X_k$ для $k = 1, 2, \ldots, n$. Розглянемо події

$$A_k = \{|S_j| < \alpha \text{ для } j < k \ \text{ та } \ |S_k| \geqslant \alpha\},$$

$k = 1, 2, \ldots, n$. Очевидно для будь-яких k маємо $A_k \in \sigma(X_1, X_2, \ldots, X_k)$. Крім того, події A_1, A_2, …, A_n несумісні парами і дають $B := \{\max_{1\leqslant k\leqslant n} |S_k| \geqslant \alpha\}$. Далі,

$$\begin{aligned}
\operatorname{Var} S_n &= \mathbb{E}S_n^2 \\
&= \int_B S_n^2 \mathrm{d}\mathbb{P} + \int_{B'} S_n^2 \mathrm{d}\mathbb{P} \\
&\geqslant \int_B S_n^2 \mathrm{d}\mathbb{P} \\
&= \sum_{k=1}^n \int_{A_k} S_n^2 \mathrm{d}\mathbb{P} \\
&= \sum_{k=1}^n \int_{A_k} (S_k + S_n - S_k)^2 \mathrm{d}\mathbb{P} \\
&= \sum_{k=1}^n \left[\int_{A_k} S_k^2 \mathrm{d}\mathbb{P} + 2\int_{A_k} (S_n - S_k) S_k \mathrm{d}\mathbb{P} + \int_{A_k} (S_n - S_k)^2 \mathrm{d}\mathbb{P}\right] \\
&\geqslant \sum_{k=1}^n \left[\int_{A_k} S_k^2 \mathrm{d}\mathbb{P} + 2\int_\Omega (S_n - S_k) S_k 1_{A_k} \mathrm{d}\mathbb{P}\right].
\end{aligned}$$

Але для будь-яких k змінні $S_n - S_k$ і $S_k 1_{A_k}$ є незалежними (перша це $X_{k+1} + X_{k+2} + \ldots + X_n$, а друга залежить лише від X_1, X_2, …, X_k).

$$\int_\Omega (S_n - S_k) S_k 1_{A_k} \mathrm{d}\mathbb{P} = \mathbb{E}(S_n - S_k) S_k 1_{A_k} = \mathbb{E}(S_n - S_k)\mathbb{E}S_k 1_{A_k} = 0.$$

Отже, продовжуючи

$$\operatorname{Var} S_n \geqslant \sum_{k=1}^n \int_{A_k} S_k^2 \mathrm{d}\mathbb{P} \geqslant \sum_{k=1}^n \int_{A_k} \alpha^2 \mathrm{d}\mathbb{P} = \alpha^2 \sum_{k=1}^n \mathbb{P}(A_k) = \alpha^2 \mathbb{P}(B).$$

Доведення завершено. □

L2 **Теорема 11.3.** *Нехай X_1, X_2, … є послідовністю незалежних центрованих інтегровних з квадратом випадкових величин. Якщо $\sum_{n=1}^\infty Var\, X_n < \infty$, то $\sum_{n=1}^\infty X_n$ збігається п.н.*

Доведення. Як легко перевірити, маємо

$$\begin{aligned}
&\mathbb{P}\left(\sum_{n=1}^\infty X_n \text{ jest rozbieżny}\right) \\
&= \mathbb{P}\left(\exists_{\gamma\in\mathbb{N}_+} \forall_n \sup_{k\geqslant 0} |X_n + X_{n+1} + \ldots + X_{n+k}| > \frac{1}{\gamma}\right) \\
&= \mathbb{P}\left(\bigcup_{\gamma\in\mathbb{N}_+} \forall_n \sup_{k\geqslant 0} |X_n + X_{n+1} + \ldots + X_{n+k}| > \frac{1}{\gamma}\right).
\end{aligned}$$

Тож достатньо показати, що для кожного $\gamma \in \mathbb{N}_+$,

$$(*) \qquad \mathbb{P}\left(\forall_n \sup_{k \geqslant 0} |X_n + X_{n+1} + \ldots + X_{n+k}| > \frac{1}{\gamma}\right) = 0.$$

Але для кожного n зазначену вище ймовірність можна заздалегідь оцінити так

$$\begin{aligned}
&\mathbb{P}\left(\sup_{k \geqslant 0} |X_n + X_{n+1} + \ldots + X_{n+k}| > \frac{1}{\gamma}\right) \\
&= \lim_{m \to \infty} \mathbb{P}\left(\sup_{0 \leqslant k \leqslant m} |X_n + X_{n+1} + \ldots + X_{n+k}| > \frac{1}{\gamma}\right) \\
&\leqslant \limsup_{m \to \infty} \gamma^2 \sum_{k=0}^{m} \operatorname{Var} X_{n+k} = \gamma^2 \sum_{k=n}^{\infty} \operatorname{Var} X_k.
\end{aligned}$$

Якщо тепер взяти $n \to \infty$, то останній вираз збігається до 0. Отже, ймовірність (*) має бути 0. $\square$

Лема 11.1 (Кронекера)**.** *Припустимо, що $(a_n)_{n \geqslant 1}$ — це така послідовність чисел, що ряд $\sum_{n=1}^{\infty} a_n/n$ збігається. Тоді $(a_1 + a_2 + \ldots + a_n)/n \to 0$.*

Доведення. Позначимо $S_n = \sum_{k=1}^{n} a_k/k$. Маємо $a_n = n(S_n - S_{n-1})$ для всіх n і

$$\begin{aligned}
\frac{a_1 + a_2 + \ldots + a_n}{n} &= \frac{S_1 + 2(S_2 - S_1) + \ldots + n(S_n - S_{n-1})}{n} \\
&= \frac{nS_n - S_1 - S_2 - \ldots - S_{n-1}}{n} \to 0. \square
\end{aligned}$$

Переходимо до основної теореми.

РЧРЎРЊ **Теорема 11.4** (ЗВЛ Колмогорова)**.** *Нехай X_1, X_2, … є послідовністю незалежних випадкових величин з однаковим розподілом.*

(а) Якщо $X_n \in L^1$ і $m = \mathbb{E}X_1$, то

$$\frac{X_1 + X_2 + \ldots + X_n}{n} \to m \quad \text{п.н.}$$

(б) Якщо $X_n \notin L^1$, то

$$\mathbb{P}\left(\limsup_{n \to \infty} \left|\frac{X_1 + X_2 + \ldots + X_n}{n}\right| = \infty\right) = 1.$$

Доведення. (а) Припустимо спочатку, що змінні X_1, X_2, … інтегровні з квадратом. Тоді теза випливає з наведених двох допоміжних фактів. Дійсно, використовуючи Теорему 11.3, ряд $\sum_{n=1}^{\infty} \frac{X_n - m}{n}$ збігається п.н., оскільки

$$\operatorname{Var} \quad \frac{X_n - m}{n} = \frac{\operatorname{Var} X_n}{n^2},$$

і далі використуєми Лему Кронекера.

Тепер розглянемо загальний випадок. Введемо нову послідовність $(X'_n)_{n \geqslant 1}$ випадкових величин, заданих як

$$X'_n(\omega) = X_n(\omega) 1_{(-n,n)}(X_n(\omega)) = \begin{cases} X_n(\omega) & \text{якщо } |X_n(\omega)| < n, \\ 0 & \text{якщо } |X_n(\omega)| \geqslant n. \end{cases}$$

Тоді X'_1, X'_2, … є незалежними випадковими величинами інтегровними з квадратом. Можемо написати

$$\frac{X_1 + X_2 + \ldots + X_n}{n} - m = I_n + II_n + III_n,$$

де

$$
\begin{aligned}
I_n &= \frac{X_1 + X_2 + \ldots + X_n - (X_1' + X_2' + \ldots + X_n')}{n}, \\
II_n &= \frac{X_1' + X_2' + \ldots + X_n' - (\mathbb{E}X_1' + \mathbb{E}X_2' + \ldots + \mathbb{E}X_n')}{n}, \\
III_n &= \frac{\mathbb{E}X_1' + \mathbb{E}X_2' + \ldots + \mathbb{E}X_n'}{n} - m.
\end{aligned}
$$

Дослідимо поведінку кожного з компонентів I_n, II_n, III_n при $n \to \infty$. Відповідно до теореми Лебега про мажоровану збіжність,

$$
\mathbb{E}X_n' = \mathbb{E}X_n 1_{\{|X_n|<n\}} = \mathbb{E}X_1 1_{\{|X_1|<n\}} \xrightarrow{n\to\infty} \mathbb{E}X_1 = m,
$$

з чого випливає, що $III_n \to 0$. Далі зауважимо, що

$$
\begin{aligned}
\sum_{n=1}^{\infty} \mathbb{P}(X_n \neq X_n') &= \sum_{n=1}^{\infty} \mathbb{P}(|X_n| \geqslant n) \\
&= \sum_{n=1}^{\infty} \mathbb{P}(|X_1| \geqslant n) \\
&\leqslant \int_0^{\infty} \mathbb{P}(|X_1| \geqslant t)\mathrm{d}t = \mathbb{E}|X_1| < \infty.
\end{aligned}
$$

Таким чином, згідно з лемою Бореля-Кантеллі, з імовірністю 1 відбудеться лише скінченна кількість подій $\{X_n \neq X_n'\}$. Іншими словами, майже для всіх ω послідовності $(X_n(\omega))_{n\geqslant 1}$ і $(X_n'(\omega))_{n\geqslant 1}$ покриваються починаючи з певного моменту. Отже $I_n \to 0$ п.н.

Залишається лише показати, що $II_n \to 0$ п.н. Відповідно до Леми Кронекера достатньо довести, що ряд $\sum_{n=1}^{\infty}(X_n' - \mathbb{E}X_n')/n$ майже напевно збігається. Ми будемо використовувати Теорему 11.3. Маємо

$$
\begin{aligned}
\mathrm{Var}\left(\frac{X_n' - \mathbb{E}X_n'}{n}\right) &= \frac{1}{n^2}(\mathbb{E}(X_n')^2 - (\mathbb{E}X_n')^2) \\
&\leqslant \frac{1}{n^2}\mathbb{E}(X_n')^2 \\
&= \frac{1}{n^2}\sum_{k=1}^{\infty}\int_{\{k-1\leqslant|X_n'|<k\}} |X_n'|^2 \mathrm{d}\mathbb{P} \\
&= \frac{1}{n^2}\sum_{k=1}^{n}\int_{\{k-1\leqslant|X_n'|<k\}} |X_n'|^2 \mathrm{d}\mathbb{P} \\
&= \frac{1}{n^2}\sum_{k=1}^{n}\int_{\{k-1\leqslant|X_1|<k\}} |X_1|^2 \mathrm{d}\mathbb{P} \\
&\leqslant \frac{1}{n^2}\sum_{k=1}^{n} k \int_{\{k-1\leqslant|X_1|<k\}} |X_1| \mathrm{d}\mathbb{P}.
\end{aligned}
$$

Тому

$$
\begin{aligned}
\sum_{n=1}^{\infty}\mathrm{Var}\left(\frac{X_n' - \mathbb{E}X_n'}{n}\right) &\leqslant \sum_{n=1}^{\infty}\sum_{k=1}^{n}\frac{k}{n^2}\mathbb{E}|X_1| 1_{\{k-1\leqslant|X_1|<k\}} \\
&= \sum_{k=1}^{\infty}\sum_{n=k}^{\infty}\frac{k}{n^2}\mathbb{E}|X_1| 1_{\{k-1\leqslant|X_1|<k\}} \\
&= \sum_{k=1}^{\infty} k\mathbb{E}|X_1| 1_{\{k-1\leqslant|X_1|<k\}} \cdot \sum_{n=k}^{\infty}\frac{1}{n^2}.
\end{aligned}
$$

Але

$$\sum_{n=k}^{\infty}\frac{1}{n^2}=\frac{1}{k^2}+\sum_{n=k+1}^{\infty}\frac{1}{n^2}\leqslant\frac{1}{k}+\int_k^{\infty}\frac{1}{x^2}\mathrm{d}x=\frac{2}{k}.$$

Отже, беручи до уваги все вище, ми отримуємо

$$\sum_{n=1}^{\infty}\mathrm{Var}\left(\frac{X_n'-\mathbb{E}X_n'}{n}\right)\leqslant 2\sum_{k=1}^{\infty}\mathbb{E}|X_1|1_{\{k-1\leqslant|X_1|<k\}}=2\mathbb{E}|X_1|<\infty.$$

Звідси теза (а).

(б) Маємо

$$\frac{X_n}{n}=\frac{X_1+X_2+\ldots+X_n}{n}-\frac{n-1}{n}\frac{X_1+X_2+\ldots+X_{n-1}}{n-1}.$$

З цього випливає, що якщо послідовність $((X_1(\omega)+X_2(\omega)+\ldots+X_n(\omega))/n)_{n\geqslant 1}$ обмежена для деякого ω, то $(X_n(\omega)/n)_{n\geqslant 1}$ також має цю властивість. Тому достатньо показати, що

$$\mathbb{P}\left(\text{рядок }\left(\frac{X_n}{n}\right)_{n\geqslant 1}\text{ не обмежено}\right)=1.$$

Маємо

$$\mathbb{P}\left(\left(\frac{X_n}{n}\right)\text{ не обмежено}\right)=\mathbb{P}\left(\bigcap_{M\in\mathbb{N}}\left\{\frac{|X_n|}{n}>M\text{ для нескінченно багатьох }n\right\}\right),$$

тож теза буде вірною, якщо ми доведемо це для всіх $M\in\mathbb{N}$,

$$\mathbb{P}\left(\frac{|X_n|}{n}>M\text{ для нескінченно багатьох }n\right)=1.$$

Зверніть увагу, що події $\{|X_n|/n>M\}$, $n=1,2,\ldots$, незалежні. Крім того

$$\begin{aligned}\sum_{n=1}^{\infty}\mathbb{P}(|X_n|/n>M)&=\sum_{n=1}^{\infty}\mathbb{P}(|X_1|>nM)\\&=\sum_{n=1}^{\infty}\sum_{k=n}^{\infty}\mathbb{P}\left(kM<|X_1|\leqslant(k+1)M\right)\\&=\sum_{k=1}^{\infty}\sum_{n=1}^{k}\mathbb{P}\left(kM<|X_1|\leqslant(k+1)M\right)\\&=\sum_{k=1}^{\infty}k\mathbb{P}(kM<|X_1|\leqslant(k+1)M)\\&\geqslant-1+\frac{1}{M}\sum_{k=1}^{\infty}(k+1)M\mathbb{P}\left(kM<|X_1|\leqslant(k+1)M\right)\\&\geqslant-1+\frac{1}{M}\sum_{k=1}^{\infty}\int_{\{kM<|X_1|\leqslant(k+1)M\}}|X_1|\mathrm{d}\mathbb{P}\\&=-1+\frac{1}{M}\mathbb{E}|X_1|1_{\{|X_1|>M\}}=\infty.\end{aligned}$$

Таким чином, з леми Бореля-Кантеллі випливає теорема. □

Зараз ми обговоримо одне із застосувань сильного закону великих чисел, т. зв. емпіричну функція розподілу.

Визначення 11.1. Нехай $X_1, X_2, \ldots$ — незалежні змінні з однаковим розподілом із функцією розподілу F. Тоді ми називаємо емпіричною функцією розподілу функцію

$$\begin{aligned}\mathbb{F}_n(t) &= \frac{1_{\{X_1 \leqslant t\}} + 1_{\{X_2 \leqslant t\}} + \ldots + 1_{\{X_n \leqslant t\}}}{n} \\ &= \frac{1_{(-\infty,t]}(X_1) + 1_{(-\infty,t]}(X_2) + \ldots + 1_{(-\infty,t]}(X_n)}{n}.\end{aligned}$$

Зверніть увагу, що для кожного $\omega \in \Omega$ функція $\mathbb{F}_n$ є функцією розподілу (як функція змінної t). Одним із основних результатів математичної статистики є наступна теорема.

Теорема 11.5 (Глівенко-Кантеллі)**.** *Якщо X_1, X_2, …, F, $\mathbb{F}_n$ такі, як вище, то*

$$\sup_{t \in \mathbb{R}} |\mathbb{F}_n(t) - F(t)| \xrightarrow{n\to\infty} 0$$

майже напевно.

У доказі ми використаємо наступну лему (без доведення: залишимо це як вправу).

Лема 11.2. *Нехай F, F_1, F_2, … — розподіли, а S — множина точок розриву функції F. Припустимо, що Q — щільна зліченна підмножина $\mathbb{R}$ така, що $\lim_{n\to\infty} F_n(t) = F(t)$ для кожного $t \in Q$. Тоді, якщо для кожного $t \in S$ маємо $F_n(t) - F_n(t-) \to F(t) - F(t-)$, то*

$$\sup_{t \in \mathbb{R}} |F_n(t) - F(t)| \xrightarrow{n\to\infty} 0.$$

Доведення теореми Глівенко-Кантеллі. Зафіксуємо будь-яку щільну зліченну підмножину $Q \subset \mathbb{R}$ і нехай S буде множиною точок розриву F. Із ЗВЛ випливає, що для кожного $t \in Q$ маємо

$$\mathbb{F}_n(t) \xrightarrow{n\to\infty} F(t) \qquad \text{п.н.}$$

оскільки $\mathbb{E}1_{(-\infty,t]}(X_1) = \mathbb{P}(X_1 \leqslant t) = F(t)$. Так само для будь-якого $t \in S$,

$$\begin{aligned}\mathbb{F}_n(t) - \mathbb{F}_n(t-) &= \frac{1_{\{t\}}(X_1) + 1_{\{t\}}(X_2) + \ldots + 1_{\{t\}}(X_n)}{n} \\ &\xrightarrow{n\to\infty} \mathbb{E}1_{\{t\}}(X_1) = \mathbb{P}(X_1 = t) = F(t) - F(t-).\end{aligned}$$

Отже множина

$$\Omega_0 = \bigcap_{t \in Q} \{\mathbb{F}_n(t) \to F(t)\} \cap \bigcap_{t \in S} \{\mathbb{F}_n(t) - \mathbb{F}_n(t-) \to F(t) - F(t-)\}$$

має повну міру, як лічильний перетин множин повної міри. Отже, за лемою вище для кожного $\omega \in \Omega_0$ маємо рівномірну збіжність $\mathbb{F}_n \to F$. □

Наприкінці цього розділу ми обговоримо попередні результати, пов'язані зі збіжністю рядів незалежних випадкових величин. Почнемо з наступного факту.

tw22 **Теорема 11.6.** *Припустимо, X_1, X_2, … є незалежними, спільно обмеженими випадковими величинами (тобто існує $a > 0$ таке, що $|X_n| \leqslant a$ з імовірністю 1 для $n = 1, \ ,2, \ldots$). Якщо ряд $\sum_{n=1}^{\infty} X_n$ збігається майже напевно, то він також збігається в L^p для будь-якого $p \geqslant 1$.*

Лема 11.3 (Нерівність Гофмана-Йоргенсена)**.** *Нехай X_1, X_2, … є незалежними випадковими величинами (можливо в $\mathbb{R}^d$ або, загалом, в певному Банаховому просторі). Визначимо $S_0 = 0$, $S_n = X_1 + X_2 + \ldots + X_n$, $n \geqslant 1$. Тоді для будь-якого s, t, $a \geqslant 0$,*

$$\begin{aligned}
&\mathbb{P}(\max_{1\leqslant k\leqslant n} |S_k| > s+t+a) \\
&\qquad \leqslant \mathbb{P}(\max_{1\leqslant k\leqslant n} |X_k| > a) + \mathbb{P}(\max_{1\leqslant k\leqslant n} |S_k| > s)\mathbb{P}(\max_{1\leqslant k\leqslant n} |S_n - S_k| > t/2).
\end{aligned}$$

Доведення. Нехай $\tau = \inf\{k \geqslant 1 : |S_k| > s\}$ (ми припускаємо $\inf \emptyset = \infty$). Зазначимо, що

$$\begin{aligned}
&\{\max_{1\leqslant k\leqslant n} |S_k| > s+t+a\} \\
&\subseteq \{\max_{1\leqslant k\leqslant n} |X_k| > a\} \cup \bigcup_{j=1}^{n} \Big(\{\max_{1\leqslant k\leqslant n} |X_k| \leqslant a\} \cap \{\tau = j\} \cap \{\max_{1\leqslant k\leqslant n} |S_k| > s+t+a\}\Big) \\
&= \{\max_{1\leqslant k\leqslant n} |X_k| > a\} \cup \bigcup_{j=1}^{n} A_j.
\end{aligned}$$

Встановимо $\omega \in A_j$. Ми маємо $S_j(\omega) > s$; крім того, $|S_{j-1}(\omega)| \leqslant s$ і $|X_j(\omega)| \leqslant a$, звідки через нерівність трикутника випливає, що $|S_j| \leqslant s + a$. Отже $|S_\ell(\omega)| > s + t + a$ для деякого $\ell > j$. Тож ми можемо писати

$$\begin{aligned}
s+t+a &< |S_\ell| \\
&\leqslant |S_j| + |S_\ell - S_j| \\
&\leqslant s + a + |S_n - S_\ell| + |S_n - S_j| \\
&\leqslant s + a + 2\max_{j\leqslant k\leqslant n} |S_n - S_k|.
\end{aligned}$$

Ми робимо висновок, що $\max_{j\leqslant k\leqslant n} |S_n - S_k| > t/2$. Тому

$$A_j \subset \{\tau = j\} \cap \{\max_{j\leqslant k\leqslant n} |S_n - S_k| > t/2\}$$

і події, що перетинаються, є незалежними: справді, перша залежить лише від змінних X_1, X_2, …, X_j, тоді як остання записується в термінах інших змінних. Зібравши всі вищенаведені факти, маємо

$$\mathbb{P}(\max_{1\leqslant k\leqslant n} |S_k| > s+t+a) \leqslant \mathbb{P}(\max_{1\leqslant k\leqslant n} |X_k| > a) + \sum_{j=1}^{n} \mathbb{P}(\tau = j)\mathbb{P}(\max_{1\leqslant k\leqslant n} |S_n - S_k| > t/2)$$

і залишається помітити, що

$$\sum_{j=1}^{n} \mathbb{P}(\tau = j) = \mathbb{P}(\tau < \infty) = \mathbb{P}(\max_{1\leqslant k\leqslant n} |S_k| > s). \qquad \square$$

Доведення теореми tw22 *11.6.* Як і вище, нехай $S_n = X_1 + X_2 + \ldots + X_n$, $n = 1, 2, \ldots$. За припущеннями, $(S_n)_{n\geqslant 1}$ збігається п.н., тому для кожного $\varepsilon \in (0, 1)$ знайдеться m таке, що

$$\mathbb{P}(\max_{m\leqslant k\leqslant n} |S_n - S_k| > \varepsilon/2) < \varepsilon$$

поки $n > m$. Маємо

$$\begin{aligned}
\mathbb{E}|S_n - S_m|^p &= p\int_0^\infty \alpha^{p-1}\mathbb{P}(|S_n - S_m| > \alpha)\mathrm{d}\alpha \\
&= \sum_{r=0}^{\infty} p\int_{(r+1)(\varepsilon+a)>\alpha>r(\varepsilon+a)} \alpha^{p-1}\mathbb{P}(|S_n - S_m| > \alpha)\mathrm{d}\alpha \\
&= p\int_0^{\varepsilon+a} \alpha^{p-1}\mathbb{P}(|S_n - S_m| > \alpha)\mathrm{d}\alpha \\
&\quad + \sum_{r=1}^{\infty} p\int_{(r+1)(\varepsilon+a)>\alpha>r(\varepsilon+a)} \alpha^{p-1}\mathbb{P}(|S_n - S_m| > r(\varepsilon+a))\mathrm{d}\alpha.
\end{aligned}$$

Але $S_n - S_m = X_{m+1} + X_{m+2} + \ldots + X_n$, тож застосовуючи до цих змінних нерівність Гофмана-Йоергенсена з параметрами $s = (r-1)(\varepsilon + a)$, $t = \varepsilon$ і $a = a$, отримуємо

$$\begin{aligned}
\mathbb{P}(\max_{m<k\leqslant n}|S_k - S_m| > r(\varepsilon+a)) \leqslant 0 + \mathbb{P}(\max_{m<k\leqslant n}|S_k - S_m| > (r-1)(\varepsilon+a))\times \\
\times\mathbb{P}(\max_{m<k\leqslant n}|S_n - S_k| > \varepsilon/2).
\end{aligned}$$

Другий множник оцінюється ε, тому за допомогою простої індукції ми отримуємо, що

$$\mathbb{P}(|S_n - S_m| > r(\varepsilon+a)) \leqslant \mathbb{P}(\max_{m<k\leqslant n}|S_n - S_m| > r(\varepsilon+a)) \leqslant \varepsilon^r.$$

Тому

$$\begin{aligned}
\mathbb{E}|S_n - S_m|^p &\leqslant p\int_0^{\varepsilon}\alpha^{p-1}\mathrm{d}\alpha + p\int_{\varepsilon}^{\varepsilon+a}\alpha^{p-1}\mathbb{P}(|S_n - S_m| > \varepsilon)\mathrm{d}\alpha \\
&\quad + \sum_{r=1}^{\infty} p\int_{r(\varepsilon+a)}^{(r+1)(\varepsilon+a)}\alpha^{p-1}\varepsilon^r\mathrm{d}\alpha \\
&\leqslant \varepsilon^p + (\varepsilon+a)^p\cdot\varepsilon + \varepsilon\sum_{r=1}^{\infty}\varepsilon^{r-1}((r+1)^p - r^p)(\varepsilon+a)^p \\
&\leqslant \varepsilon\cdot C,
\end{aligned}$$

де C — деяка константа, що залежить лише від p і a. Отже, $(S_n)_{m\geqslant 0}$ задовольняє умову Коші в L^p, тому він збігається в L^p. □

Введемо такі позначення: для випадкової величини X і $a > 0$ нехай

$$X^a(\omega) = \begin{cases} a & \text{для } X(\omega) > a, \\ X(\omega) & \text{для } |X(\omega)| \leqslant a, \\ -a & \text{для } X(\omega) < -a. \end{cases}$$

Теорема 11.7 (Колмогорова про три ряди)**.** *Припустимо, X_1, X_2, … — незалежні випадкові величини, а a — фіксоване позитивне число. Тоді ряд $\sum_{n=0}^{\infty} X_n$ збігається п.н. тоді і тільки тоді, коли збігаються ряди*

$$\sum_{n=1}^{\infty}\mathbb{E}X_n^a, \quad \sum_{n=1}^{\infty} Var\, X_n^a, \quad \sum_{n=1}^{\infty}\mathbb{P}(|X_n| > a).$$

У доказі використаємо такий простий факт.

Лема 11.4. *Нехай* X_1*,* X_2*,* $\ldots$ *— випадкові величини, що задовольняють умову* $\sum_{n=1}^{\infty}\mathbb{P}(|X_n|>a)<\infty$ *для деякого* $a>0$*. Тоді ряд* $\sum_{n=1}^{\infty}X_n$ *збігається п.н. тоді і тільки тоді, коли ряд* $\sum_{n=1}^{\infty}X_n^a$ *збігається п.н.*

Доведення. Згідно з лемою Бореля-Кантеллі, для майже всіх ω послідовності $X_n(\omega)$, $X_n^a(\omega)$ співпадають з певного моменту. Звідси відразу випливає теза. □

Доведення теореми про три ряди. $\Leftarrow$ Згідно з лемою вище достатньо показати, що ряд $\sum_{n=1}^{\infty}X_n^a$ збігається п.н. Змінні $X_n^a-\mathbb{E}X_n^a$ центровані, незалежні, обмежені та

$$\sum_{n=1}^{\infty}\operatorname{Var}(X_n^a-\mathbb{E}X_n^a)=\sum_{n=1}^{\infty}\operatorname{Var}X_n^a<\infty,$$

тому за теоремою 11.3 ряд $\sum_{n=1}^{\infty}(X_n^a-\mathbb{E}X_n^a)$ збігається п.н. Оскільки числовий ряд $\sum_{n=1}^{\infty}\mathbb{E}X_n^a$ також збігається, маємо тезу.

$\Rightarrow$ Припустимо, всупереч тезі, що $\sum_{n=1}^{\infty}\mathbb{P}(|X_n|>a)=\infty$. Тоді за лемою Бореля-Кантеллі існує нескінченна кількість нерівностей $|X_n|>a$ з імовірністю 1, що виключає збіжність ряду $\sum_{n=1}^{\infty}X_n$, отримуємо протиріччя. Отже, $\sum_{n=1}^{\infty}\mathbb{P}(|X_n|>a)<\infty$, тому з наведеної вище леми випливає, що ряд $\sum_{n=1}^{\infty}X_n^a$ збігається п.н. Використовуючи теорему 11.6, ми отримуємо, що ряд $\sum_{n=1}^{\infty}X_n^a$ збігається в L^1 і L^2. Так, зокрема, числова послідовність

$$\left(\mathbb{E}\sum_{n=1}^{N}X_n^a\right)_{N\geqslant 1}=\left(\sum_{n=1}^{N}\mathbb{E}X_n^a\right)_{N\geqslant 1}$$

є збіжними. Крім того, те саме вірно для

$$\left(\mathbb{E}\left(\sum_{n=1}^{N}(X_n^a-\mathbb{E}X_n^a)\right)^2\right)_{N\geqslant 1}=\left(\operatorname{Var}\left(\sum_{n=1}^{N}X_n^a\right)\right)_{N\geqslant 1}=\left(\sum_{n=1}^{N}\operatorname{Var}X_n^a\right)_{N\geqslant 1}.$$

Отримуємо протиріччя. □

Приклад 11.1. Нехай X_1, X_2, $\ldots$ — послідовність незалежних випадкових величин, така що X_n має експоненціальний розподіл із параметром λ_n. Знайдемо умову на послідовність $(\lambda_n)_{n\geqslant 0}$, яка буде еквівалентна збіжності ряду $\sum_{n=1}^{\infty}X_n$ майже напевно.

За теоремою Колмогорова збіжність п.н. має місце тоді і тільки тоді

$$(1)\qquad \sum_{n=1}^{\infty}\mathbb{P}(|X_n^1|>1)<\infty,\quad \text{czyli}\quad \sum_{n=1}^{\infty}e^{-\lambda_n}<\infty,$$

$$(2)\qquad \sum_{n=1}^{\infty}\mathbb{E}X_n^1<\infty,\quad \text{czyli}\quad \sum_{n=1}^{\infty}\left(\frac{1-e^{-\lambda_n}}{\lambda_n}-e^{-\lambda_n}\right)<\infty,$$

$$(3)\qquad \sum_{n=1}^{\infty}\operatorname{Var}X_n^1<\infty,\quad \text{czyli}\quad \sum_{n=1}^{\infty}\left(\frac{1}{\lambda_n^2}-e^{-\lambda_n}-\frac{(1+\lambda_n)^2}{\lambda_n^2}e^{-2\lambda_n}\right)<\infty.$$

Тепер припустимо, що ряд $\sum_{n=1}^{\infty}X_n$ збігається п.н. Тоді з (1) маємо $\lambda_n\to\infty$, що в поєднанні з (2) призводить до висновку, що $\sum_{n=1}^{\infty}\frac{1}{\lambda_n}<\infty$.

Ми покажемо, що умова $\sum_{n=1}^{\infty}\frac{1}{\lambda_n}<\infty$ є достатньою. Дійсно, з цієї умови випливає, що $\lambda_n\to\infty$, а отже $\sum_{n=1}^{\infty}\frac{1}{\lambda_n^2}<\infty$ (оскільки для великих n маємо $\frac{1}{\lambda_n^2}\leqslant\frac{1}{\lambda_n}$). Далі маємо $e^{-\lambda_n}\leqslant\frac{1}{1+\lambda_n}\leqslant\frac{1}{\lambda_n}$ і $e^{-2\lambda_n}\leqslant\frac{1}{\lambda_n^2}$, отже, маємо збіжність усіх трьох рядів (1), (2) і (3).

12 Завдання

1. Дано послідовність $(X_n)_{n\geqslant 1}$ незалежних пуассонівських випадкових величин з параметром 2. Доведіть, що послідовність

$$\frac{X_1X_2 + X_2X_3 + \ldots + X_nX_{n+1}}{n+2009}, \qquad n = 1,\, 2,\, \ldots,$$

є збіжна п.н. і знайдіть її границю.

2. Дано послідовність $(X_n)_{n\geqslant 1}$ незалежних випадкових величин, де для $n \geqslant 1$ змінна X_n рівномірно розподілена на інтервалі $(1/n, 1]$. Доведіть, що послідовність

$$\frac{X_1 + X_2 + \ldots + X_n}{n}, \qquad n = 1,\, 2,\, \ldots,$$

є збіжна п.н. і знайдіть її границю.

3. Дано послідовність $(X_n)_{n\geqslant 1}$ незалежних невід'ємних випадкових величин з однаковим розподілом. Доведіть, що якщо $\mathbb{E}X_1 = \infty$, то п.н.

$$\frac{X_1 + X_2 + \ldots + X_n}{n} \to \infty.$$

4. Дано послідовність $(A_n)_{n\geqslant 1}$ незалежних подій, $p_n = \mathbb{P}(A_n)$. Доведіть, що

$$\frac{1_{A_1} + 1_{A_2} + \ldots + 1_{A_n}}{n} - \frac{p_1 + p_2 + \ldots + p_n}{n} \to 0$$

за ймовірністю.

5. Дано послідовність $(X_n)_{n\geqslant 1}$ незалежних інтегровних випадкових величин з однаковим розподілом. Доведіть, що послідовність

$$\frac{X_1 + X_2 + \ldots + X_n}{n}, \qquad n \geqslant 1,$$

збігається в L^1 до $\mathbb{E}X_1$.

6. Дано послідовність $(N_n)_{n\geqslant 1}$ випадкових величин (не обов'язково незалежних), де для $n \geqslant 1$ змінна N_n має розподіл Пуассона з параметром n. Доведіть, що $N_n/n \to 1$ в L^1.

7. $X_1,\, X_2,\, \ldots$ незалежні та рівномірно розподілені на $[-1, 1]$. Чи послідовність

$$\frac{X_1 + X_2^2 + \ldots + X_n^n}{n}, \qquad n = 1,\, 2,\, \ldots,$$

сходиться п.н.?

8. Обчисліть границі

$$\lim_{n\to\infty} \frac{1}{2^n} \int_{-1}^{1} \int_{-1}^{1} \cdots \int_{-1}^{1} \frac{x_1^2 + x_2^2 + x_3^3 + x_4^4 + \ldots + x_n^n}{x_1^2 + x_2^2 + \ldots + x_n^2} \mathrm{d}x_1 \mathrm{d}x_2 \ldots \mathrm{d}x_n$$

та

$$\lim_{n\to\infty} \int_0^1 \int_0^1 \cdots \int_0^1 f(\sqrt[n]{x_1 x_2 \ldots x_n}) \mathrm{d}x_1 \mathrm{d}x_2 \ldots \mathrm{d}x_n,$$

де $f : [0, 1] \to \mathbb{R}$ є фіксованою неперервною функцією.

9. Випадкові величини $X_1,\ X_2,\ \ldots$ незалежні, та для $n \geqslant 1$ розподіл X_n задається таким чином:

$$\mathbb{P}(X_n = 0) = 1/2,\ \mathbb{P}(X_n = 1) = 1/2 - \frac{1}{4n^2},\ \mathbb{P}(X_n = n) = \frac{1}{4n^2}.$$

Доведіть, що послідовність

$$\frac{X_1 + X_2 + \ldots + X_n}{n}$$

є збіжна п.н. і знайдіть її границю.

10. Змінні $\varepsilon_1, \varepsilon_2, \ldots$ незалежні і мають розподіл Радемахера. Доведіть, що для $\alpha > 1/2$, послідовність

$$\frac{\varepsilon_1 + \varepsilon_2 + \ldots + \varepsilon_n}{n^\alpha}, \qquad n = 1,\ 2,\ \ldots$$

є збіжна п.н.

11. Доведіть наступну теорему про два ряди: якщо $(X_n)_{n\geqslant 1}$ є послідовністю незалежних випадкових величин, що інтегруються з квадратом, так що числові ряди

$$\sum_{n=1}^{\infty} \mathbb{E}X_n, \qquad \sum_{n=1}^{\infty} \operatorname{Var} X_n$$

є збіжними, то ряд $\sum_{n=1}^{\infty} X_n$ є збіжним п.н.

12. Дано послідовність $(X_n)_{n\geqslant 1}$ незалежних випадкових величин, таких що

$$\mathbb{P}(X_n = -n) = \mathbb{P}(X_n = n) = \frac{1}{n^3},\ \mathbb{P}(X_n = 0) = 1 - \frac{2}{n^3}.$$

Доведіть, що $\sum_{n=1}^{\infty} X_n$ є збіжним п.н.

13. Дано послідовність $(\varepsilon_n)_{n\geqslant 1}$ незалежних змінних Радемахера. Якій умові має задовольняти послідовність $(a_n)_{n\geqslant 1}$, щоб ряд $\sum_{n=1}^{\infty} a_n\varepsilon_n$ був збіжним п.н.?

14. Дано послідовність (X_n) незалежних випадкових величин, таких що для $n \geqslant 1$ змінна X_n рівномірно розподілена на відрізку $[-n, n]$. При яких значеннях параметра $p > 0$ ряд

$$\sum_{n=1}^{\infty} \frac{X_n}{n^p}$$

є збіжним п.н.?

13 Теорема де Муавра-Лапласа

Тепер ми звернемося до надзвичайно важливого і корисного факту, який дозволяє нам апроксимувати розподіл Бернуллі $B(n,p)$ нормальним розподілом. Суттєве узагальнення наступних результатів буде дано на лекціях з теорії ймовірності II, з нагоди т.зв. Центральної граничної теореми.

Припустимо, $\Phi : \mathbb{R} \to [0,1]$ — функція розподілу, а g — щільність стандартного нормального розподілу:

$$g(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}, \quad \Phi(t) = \int_{-\infty}^{t} g(x)\mathrm{d}x.$$

Нехай p — фіксоване число в діапазоні $(0,1)$, $q = 1-p$ і S_n — випадкова величина з розподілом $B(n,p)$.

Теорема 13.1. *Припустимо, що k — таке ціле число, що*

$$(*) \qquad |k-np| \cdot \frac{\max(p,q)}{npq} \leqslant 1/2.$$

Тоді

$$\mathbb{P}(S_n = k) = \frac{1}{\sqrt{2\pi npq}} \exp\left(-\frac{(k-np)^2}{2npq} + R(n,k)\right),$$

де

$$R(n,k) \leqslant \frac{3|k-np|}{4npq} + \frac{|k-np|^3}{3n^2p^2q^2} + \frac{1}{3npq}.$$

Доведення. Застосовуючи формулу Стірлінга

$$n! = \sqrt{2\pi n} n^n e^{-n+\theta_n/(12n)}, \qquad 0 < \theta_n < 1,$$

отримуємо

$$\begin{aligned} \mathbb{P}(S_n = k) &= \sqrt{\frac{n}{2k\pi(n-k)}} \cdot \left(\frac{np}{k}\right)^k \left(\frac{nq}{n-k}\right)^{n-k} \cdot \exp\left(\frac{\theta_n}{12n} - \frac{\theta_k}{12k} - \frac{\theta_{n-k}}{12(n-k)}\right) \\ &= I \cdot II \cdot III. \end{aligned}$$

Розглянемо по черзі множники I, II і III. Маємо

$$I = \frac{1}{\sqrt{2\pi npq}} \cdot \left(1 + \frac{k-np}{npq} \cdot q\right)^{-1/2} \left(1 - \frac{k-np}{npq} \cdot p\right)^{-1/2} = \frac{1}{\sqrt{2\pi npq}} \cdot e^{R_1(n,k)}.$$

Для будь-якого $x \geqslant -1/2$ справедлива оцінка

$$|\log(1+x) - x| \leqslant x^2,$$

звідки випливає, що

$$\begin{aligned} R_1(n,k) &= -\frac{1}{2}\left[\log\left(1 + \frac{k-np}{npq} \cdot q\right) + \log\left(1 - \frac{k-np}{npq} \cdot p\right)\right] \\ &= \frac{1}{2}\frac{(p-q)(k-np)}{npq} + R_1'(n,k), \end{aligned}$$

де $|R_1'(n,k)| \leqslant \frac{1}{2}(p^2+q^2)(k-np)^2/(n^2p^2q^2) \leqslant |k-np|/(4npq)$, за припущенням (*). Отже,

$$|R_1(n,k)| \leqslant \frac{3|k-np|}{4npq}.$$

Далі маємо

$$\begin{aligned}\log II &= k\log\left(\frac{np}{k}\right) + (n-k)\log\left(\frac{nq}{n-k}\right)\\ &= -np\cdot\frac{k}{np}\log\left(\frac{k}{np}\right) - nq\cdot\frac{n-k}{nq}\log\left(\frac{n-k}{nq}\right)\\ &= -np\cdot\left(1+\frac{k-np}{np}\right)\log\left(1+\frac{k-np}{np}\right)\\ &\quad - nq\cdot\left(1-\frac{k-np}{nq}\right)\log\left(1-\frac{k-np}{nq}\right).\end{aligned}$$

Тепер скористаємося з нерівності

$$\left|(1+x)\log(1+x) - x - \frac{x^2}{2}\right| \leqslant \frac{1}{3}|x|^3,$$

справедливої для $x \geqslant -1/2$: якщо (*) виконується, то

$$\begin{aligned}\log II &= -np\left(\frac{k-np}{np} + \frac{1}{2}\left(\frac{k-np}{np}\right)^2\right)\\ &\quad - nq\left(-\frac{k-np}{nq} + \frac{1}{2}\left(\frac{k-np}{nq}\right)^2\right) + R_2'(n,k)\\ &= -\frac{1}{2}\frac{(k-np)^2}{npq} + R_2'(n,k),\end{aligned}$$

де

$$|R_2'(n,k)| \leqslant \frac{1}{3}\left(np\left|\frac{k-np}{np}\right|^3 + nq\left|\frac{k-np}{nq}\right|^3\right) = \frac{|k-np|^3}{3n^2p^2q^2}(p^2+q^2) \leqslant \frac{|k-np|^3}{3n^2p^2q^2}.$$

Нарешті, ми маємо $III = e^{R_3(n,k)}$, де

$$-\left(\frac{1}{12k} + \frac{1}{12(n-k)}\right) < R_3(n,k) < \frac{1}{12n}.$$

Рівнозначно,

$$-\frac{1}{12npq}\left(1+\frac{k-np}{npq}\cdot q\right)^{-1}\left(1-\frac{k-np}{npq}\cdot p\right)^{-1} < R_3(n,k) < \frac{1}{12n},$$

з якого в силу (*) випливає оцінка

$$|R_3(n,k)| \leqslant \frac{1}{3npq}.$$

Поєднуючи наведені вище нерівності для $R_i(n,k)$, ми отримуємо тезу. $\square$

Наступна теорема, т.зв. інтегральна теорема де Муавра-Лапласа дозволяє нам приблизно визначити ймовірність того, що кількість успіхів потрапляє в певний діапазон.

Теорема 13.2. *Припустимо, що a, $b \geqslant 0$ задовольняють умову*

$$(*) \qquad |a-np|\cdot\frac{\max(p,q)}{npq} \leqslant 1/2, \qquad |b-np|\cdot\frac{\max(p,q)}{npq} \leqslant 1/2.$$

Тоді

$$\mathbb{P}\left(a \leqslant S_n \leqslant b\right) = \left[\Phi\left(\frac{b-np+\frac{1}{2}}{\sqrt{npq}}\right) - \Phi\left(\frac{a-np-\frac{1}{2}}{\sqrt{npq}}\right)\right]e^{D(n,a,b)},$$

де

$$|D(n,a,b)| \leqslant \max_{k\in\{a,b\}}\left[\frac{5}{4}\frac{|k-np|}{npq} + \frac{1}{3}\frac{|k-np|^3}{n^2p^2q^2}\right] + \frac{1}{3npq} + \frac{1}{8npq}.$$

Доведення. Нагадаємо, що g — це щільність стандартного нормального розподілу. Позначимо

$$x_k = \frac{k-np}{\sqrt{npq}}, \qquad h = \frac{1}{\sqrt{npq}}.$$

З теореми про середнє значення маємо

$$\Phi\left(x_k + h/2\right) - \Phi\left(x_k - h/2\right) = hg(\xi_k),$$

де $\xi_k \in (x_k - h/2, x_k + h/2)$. Іншими словами, маємо

$$\begin{aligned} &hg\left(x_k\right) \\ &= \exp\left(\frac{1}{2}(\xi_k^2 - x_k^2)\right)\Phi\left(x_k + h/2\right) - \Phi\left(x_k - h/2\right). \end{aligned}$$

Далі, $|\xi_k^2 - x_k^2| = |\xi_k + x_k| \cdot |\xi_k - x_k| \leqslant \frac{1}{2}h\left(2|x_k| + \frac{1}{2}h\right) = h|x_k| + \frac{1}{4}h^2$, тому

$$hg(x_k) = e^{r_k}\left[\Phi(x_k + h/2) - \Phi(x_k - h/2)\right],$$

де $|r_k| \leqslant \frac{1}{2}h|x_k| + \frac{1}{8}h^2$. В поєднанні з попередньою теоремою ми отримуємо

$$\mathbb{P}(S_n = k) = e^{r_k + R(n,k)}\left[\Phi(x_k + h/2) - \Phi(x_k - h/2)\right].$$

Нехай $d = \max_{k\in\{a,a+1,\ldots,b\}} |r_k + R(n,k)|$; отримуємо

$$e^{-d}\left[\Phi(x_k + h/2) - \Phi(x_k - h/2)\right] \leqslant \mathbb{P}(S_n = k) \leqslant e^{d}\left[\Phi(x_k + h/2) - \Phi(x_k - h/2)\right].$$

Записуючи ці нерівності для $k = a,\ a+1,\ \ldots,\ b$ і додаючи їх, отримуємо тезу. □

Нарешті, сформулюємо (без доказів) факт, який містить зручну оцінку похибки апроксимації в теоремі де Муавра-Лапласа.

Теорема 13.3. *З такими позначеннями, як вище, маємо*

$$\sup_{t\in\mathbb{R}}\left|\mathbb{P}\left(\frac{S_n - np}{\sqrt{npq}} \leqslant t\right) - \Phi(t)\right| \leqslant \frac{p^2+q^2}{\sqrt{npq}}.$$

14 Завдання

1. Ймовірність народження хлопчика становить $0,517$. Яка ймовірність того, що серед $n = 10000$ новонароджених число хлопчиків не перевищить кількість дівчаток?

2. Кидаємо симетричну монету, поки не отримаємо 200 голів (не обов'язково поспіль). Яка приблизна ймовірність того, що ми кинемо більше ніж 440 разів?

3. У меблевий магазин привезли 150 парт I типу, і 754 парт II типу. Відомо, що парти I типу вдвічі популярніші (тобто ймовірність того, що клієнт, купуючи стіл, вибере парту I типу, становить $2/3$). Яка приблизна ймовірність того, що один із перших 200 клієнтів, які купують столи, не отримає бажану модель?

4. Виявлено, що в середньому 30% із загальної кількості прийнятих студентів закінчують коледж вчасно. Скільки осіб потрібно прийняти на перший курс, щоб з імовірністю не менше 0,9 вчасно закінчили не менше 50 осіб?

5. В одному експерименті ймовірність A становить $0,7$. Скільки разів потрібно повторити цей експеримент, щоб з імовірністю $0,9$ частота A не відрізнялася від $0,7$ більш ніж на $0,1$? Чи можемо ми щось сказати про кількість необхідних ітерацій, якщо не знаємо ймовірності A?

6. а) Ми кидаємо кубик 4500 разів, імовірність випадіння шістки становить $1/6$. Обчисліть приблизну ймовірність того, що кількість викинутих шісток перевищить 450.

б) Припустимо, що ймовірність отримати шістку становить $1/1000$. Яка приблизна ймовірність того, що кількість викинутих шісток перевищить 2?

7. Дано послідовність $(\varepsilon_n)_{n\geqslant 1}$ незалежних випадкових величин Радемахера. Доведіть, що послідовність

$$\frac{\varepsilon_1 + \varepsilon_2 + \ldots + \varepsilon_n}{\sqrt{n}}, \qquad n = 1,\, 2,\, \ldots,$$

майже напевно не збігається.

15 Умовне сподівання

Умовне сподівання є одним із ключових понять теорії ймовірностей. Почнемо з ситуації, коли умовою є певна подія.

Визначення 15.1. Припустимо, що $(\Omega, \mathcal{F}, \mathbb{P})$ — імовірнісний простір, а B — позитивна ймовірна подія. Нехай X — інтегровна випадкова величина. Умовне сподівання X за умови B, є числом

$$\mathbb{E}(X|B) = \int_\Omega X(\omega)\mathbb{P}(d\omega|B).$$

Теорема 15.1. *З припущеннями, як вище,*

$$(*) \qquad \mathbb{E}(X|B) = \frac{1}{\mathbb{P}(B)} \int_B X d\mathbb{P}.$$

Доведення. Ми використовуємо стандартний метод ускладнення змінної X.

1.Припустимо спочатку, що $X = 1_A$, де $A \in \mathcal{F}$. Маємо

$$\mathbb{E}(X|B) = \mathbb{P}(A|B) = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)} = \frac{1}{\mathbb{P}(B)} \int_B 1_A d\mathbb{P}.$$

2.З лінійності, доведена рівність справедлива також для простих змінних (лінійних комбінацій індикаторів подій).

3. Тепер, якщо X є невід'ємною випадковою змінною, ми беремо неспадну послідовність $(X_n)_{n\geqslant 1}$ простих змінних, що майже напевно збігається до X. Записуючи (*) для X_n і переходячи до границі $n \to \infty$, отримуємо (*) для X за теоремою Лебега про монотонний перехід до границі під знаком інтеграла.

4. Якщо X є будь-якою випадковою величиною, ми представляємо X у вигляді $X = X_+ - X_-$, використовуємо (*) для X_+ і X_-, і після віднімання сторін отримуємо (*) для X. $\square$

Тепер розглянемо наступний приклад. Припустімо, що $\{B_i\}_{i=1,2,\ldots,n}$ є розбиттям Ω на події позитивної міри. Нехай X — випадкова інтегрувана величина. Визначте η як $\eta(\omega) = \mathbb{E}(X|B_i)$, якщо $\omega \in B_i$, $i = 1, 2, \ldots, n$. Ми інтерпретуємо η як середнє значення X, якщо знаємо все про події в σ-алгебрі, що створено розбиттям $\{B_i\}$. Змінна η має такі властивості:

1) η вимірна відносно $\sigma(B_1, B_2, \ldots, B_n)$, тому що є постійною на будь-якій події B_i,

2) Для кожного $i = 1, 2, \ldots, n$ ми маємо

$$\int_{B_i} \eta \mathrm{d}\mathbb{P} = \mathbb{E}(X|B_i) \cdot \mathbb{P}(B_i) = \int_{B_i} X \mathrm{d}\mathbb{P},$$

з чого випливає, що

$$\int_B \eta \mathrm{d}\mathbb{P} = \int_B X \mathrm{d}\mathbb{P}$$

для будь-якого $B \in \sigma(B_1, B_2, \ldots, B_n)$.

Це призводить до визначення умовного сподівання відносно σ-алгебри.

Визначення 15.2. Припустімо, що $(\Omega, \mathcal{F}, \mathbb{P})$ — імовірнісний простір, $\mathcal{M}$ — під-σ-алгебра $\mathcal{F}$, а X — випадкова інтегровна змінна. Умовне сподівання X за умови $\mathcal{M}$ є випадковою величиною η, що задовольняє наступні дві умови:

1) η вимірна відносно $\mathcal{M}$.

2) Для кожного $B \in \mathcal{M}$,

$$\int_B \eta \mathrm{d}\mathbb{P} = \int_B X \mathrm{d}\mathbb{P}.$$

Позначення: $\mathbb{E}(X|\mathcal{M})$.

Зокрема, якщо $X = 1_A$, $A \in \mathcal{F}$, ми визначаємо умовну ймовірність A за умови $\mathcal{M}$ через $\mathbb{P}(A|\mathcal{M}) = \mathbb{E}(1_A|\mathcal{M})$.

Теорема 15.2. *Припустимо, що X є інтегровною випадковою величиною, а $\mathcal{M}$ є під-σ-алгеброю $\mathcal{F}$. Тоді умовне математичне сподівання існує і визначається однозначно з точністю до рівності п.н.*

Доведення. Для будь-якого $B \in \mathcal{M}$ ми визначаємо $\nu(B) = \int_B X \mathrm{d}\mathbb{P}$. Функція $\nu : \mathcal{M} \to \mathbb{R}$ є лічильно-адитивною функцією множини. Крім того, якщо $\mathbb{P}(B) = 0$, то $\nu(B) = 0$ (це називається абсолютною неперервністю ν відносно $\mathbb{P}$). За теоремою Радона-Нікодима існує $\mathcal{M}$-вимірна випадкова величина η, яка є щільністю ν відносно $\mathbb{P}$, тобто така, що для всіх $B \in \mathcal{M}$,

$$\int_B X \mathrm{d}\mathbb{P} = \nu(B) = \int_B \eta \mathrm{d}\mathbb{P}.$$

Єдиність очевидна: якщо η_1, η_2 є випадковими величинами, що задовольняють 1) і 2), то, зокрема, для кожного $B \in \mathcal{M}$, $\int_B \eta_1 \mathrm{d}\mathbb{P} = \int_B \eta_2 \mathrm{d}\mathbb{P}$, отже $\eta_1 = \eta_2$ п.н. $\square$

Приходимо до поняття умовного сподівання випадкової величини. Нам знадобиться наступний допоміжний факт.

Лема 15.1. *Припустимо, що Y — випадкова величина. Тоді кожна випадкова величина X, вимірна відносно $\sigma(Y)$, має вигляд $f(Y)$ для деякої борелівської функції f.*

Доведення. Знову використовуємо метод змінного ускладнення.

1. Припустимо, що $X = 1_A$, де $A \in \sigma(Y)$. Тоді $A = \{Y \in B\}$ для деякого B, звідки $X = 1_B(Y)$, тобто як f можна взяти показник 1_B.

2. Якщо X — проста змінна, то за f беремо лінійну комбінацію відповідних показників (див. попередній пункт).

3. Припустимо, що X є невід'ємною випадковою величиною. Існує неспадна послідовність (X_n) простих $\sigma(Y)$-вимірних випадкових величин, що збігається до X. На підставі 2) ми маємо $X_n = f_n(Y)$ для деякої послідовності функцій (f_n). Аби перевірити легко, достатньо взяти

$$f(x) = \begin{cases} \lim_{n\to\infty} f_n(x) & \text{якщо границя існує,} \\ 0 & \text{якщо границя не існує.} \end{cases}$$

4. Тепер, якщо X є будь-якою випадковою величиною, то ми маємо $X = X_+ - X_- = f_+(Y) - f_-(Y) = f(Y)$, де f_+, f_- це борелевські функції, що відповідають $\sigma(Y)$-вимірним X_+ і X_-. $\square$

Визначення 15.3. Нехай X, Y — випадкові величини, X інтегровна. Ми визначаємо умовне сподівання X за умови Y як

$$\mathbb{E}(X|Y) = \mathbb{E}(X|\sigma(Y)).$$

Зауваження: За лемою маємо $\mathbb{E}(X|Y) = f(Y)$ для деякої борелевської функції f. Число $f(y)$ можна інтерпретувати як $\mathbb{E}(X|Y = y)$.

Приклади:

1. Припустимо, що X, Y мають дискретні розподіли. Позначимо

$$P_Y(y) = \mathbb{P}(Y = y) \ \text{ oraz } \ P_{(X,Y)}(x, y) = \mathbb{P}(X = x, Y = y).$$

Якщо h — будь-яка борелівська функція така, що $h(X) \in L^1$, тоді

$$\mathbb{E}(h(X)|Y) = \sum_{x \in S_X} h(x) \frac{P_{(X,Y)}(x, Y)}{P_Y(Y)}.$$

Щоб довести це, необхідно перевірити, що права частина (надалі η) задовольняє властивості 1) і 2) з визначення $\mathbb{E}(h(X)|\sigma(Y)))$. Перша умова зрозуміла - η, як функція Y, є $\sigma(Y)$-вимірною. Отже, розберемося з другою умовою. Нехай $B \in \sigma(Y)$. Оскільки Y має дискретний розподіл, B є щонайбільше рахунковою сумою подій виду $\{Y = y\}$ і подій з імовірністю 0. Тож достатньо перевірити 2) для множин B виду $\{Y = y\}$. Маємо

$$\int_{\{Y=y\}} \eta \mathrm{d}\mathbb{P} = \int_{\{Y=y\}} \sum_{x \in S_X} h(x) \frac{P_{X,Y}(x, y)}{P_Y(y)} \mathrm{d}\mathbb{P} = \sum_{x \in S_X} h(x) P_{X,Y}(x, y)$$

та

$$\int_{\{Y=y\}} h(X) \mathrm{d}\mathbb{P} = \sum_{x \in S_X} h(x) \int_{\{Y=y\}} 1_{\{X=x\}} \mathrm{d}\mathbb{P} = \sum_{x \in S_X} h(x) P_{X,Y}(x, y).$$

2. Приклад. Припустимо, X, Y незалежні Пуассонівські випадкові величини з параметрами λ, μ відповідно. Ми знайдемо $\mathbb{E}(X|X + Y)$.

Відомо, що $X + Y$ має розподіл Пуассона з параметром $\lambda + \mu$. Отже

$$P_{X+Y}(k) = \frac{(\lambda + \mu)^k}{k!} e^{-(\lambda+\mu)}, \qquad k = 0, 1, 2, \ldots.$$

Крім того, якщо $k \geqslant \ell \geqslant 0$, то

$$\begin{aligned} P_{X,X+Y}(\ell, k) &= \mathbb{P}(X = \ell, X + Y = k) = \mathbb{P}(X = \ell)\mathbb{P}(Y = k - \ell) \\ &= \frac{\lambda^\ell}{\ell!} e^{-\lambda} \cdot \frac{\mu^{k-\ell}}{(k - \ell)!} e^{-\mu} \end{aligned}$$

і

$$\frac{P_{X,X+Y}(\ell, k)}{P_{X+Y}(k)} = \frac{k! \lambda^\ell \mu^{k-\ell}}{\ell!(k - \ell)!(\lambda + \mu)^k} = \binom{k}{\ell} \left(\frac{\lambda}{\lambda + \mu}\right)^\ell \left(1 - \frac{\lambda}{\lambda + \mu}\right)^{k-\ell}.$$

Звідки

$$\mathbb{E}(X|X + Y) = \frac{\lambda}{\lambda + \mu}(X + Y).$$

3. Припустимо, що (X, Y) має розподіл зі щільністю g, і нехай $g_Y(y) = \int_{\mathbb{R}} g(x, y)dx$ — це щільність Y. Визначимо *умовну щільність* за формулою

$$g_{X|Y}(x|y) = \begin{cases} \frac{g(x,y)}{g_Y(y)} & \text{якщо } g_Y(y) \neq 0, \\ 0 & \text{якщо } g_Y(y) = 0. \end{cases}$$

Тоді для будь-якої борелівської функції $h : \mathbb{R} \to \mathbb{R}$ такої, що $h(X) \in L^1$, маємо

$$(*) \qquad \mathbb{E}(h(X)|Y) = \int_{\mathbb{R}} h(x) g_{X|Y}(x|Y) dx.$$

Дійсно, перевіримо, що права частина задовольняє умови 1) і 2) з визначення $\mathbb{E}(h(X)|Y)$. Очевидно, умова 1) виконується - права частина є функцією з Y. Переходимо до 2). Для будь-якого $B \in \sigma(Y)$ ми маємо, що $B = \{Y \in A\}$ для деякого $A \in \mathbb{R}$ і

$$\begin{aligned}\int_B h(X)d\mathbb{P} &= \int_\Omega 1_{\{Y\in A\}} h(X)d\mathbb{P} = \int_{\mathbb{R}^2} 1_{\{y\in A\}} h(x)g(x,y)dxdy \\ &= \int_{\mathbb{R}} 1_{\{y\in A\}} g_Y(y)\int_{\mathbb{R}} h(x)g_{X|Y}(x|y)dxdy = \int_B \int_{\mathbb{R}} h(x)g_{X|Y}(x|Y)dxd\mathbb{P}.\end{aligned}$$

4. Приклад. Припустимо, що (X, Y) рівномірно розподілено в трикутнику

$$D = \{(x,y) : 0 \leqslant x \leqslant y \leqslant 1\}.$$

Облічимо $\mathbb{E}(X|Y)$ та $\mathbb{P}(X \leqslant 1/2|Y)$.

Маємо $g(x,y) = 21_{\{0\leqslant x\leqslant y\leqslant 1\}}$ та

$$g_Y(y) = \int_{\mathbb{R}} g(x,y)\mathrm{d}x = 2y1_{[0,1]}(y).$$

Тому умовна щільність $g_{X|Y}$ визначається формулою

$$g_{X|Y}(x|y) = \begin{cases} \frac{1}{y}1_{[0,y]}(x) & \text{якщо } y \in (0,1], \\ 0 & \text{для інших } y. \end{cases}$$

Звідки

$$\mathbb{E}(X|Y) = \int_{\mathbb{R}} x g_{X|Y}(x|Y)\mathrm{d}x = \frac{1}{Y}\int_0^Y x\mathrm{d}x = \frac{Y}{2}$$

та

$$\begin{aligned}\mathbb{P}(X \leqslant 1/2|Y) &= \mathbb{E}[1_{(-\infty,1/2]}(X)|Y] \\ &= \int_{\mathbb{R}} 1_{(-\infty,1/2]}(x)g_{X|Y}(x|Y)\mathrm{d}x \\ &= \frac{1}{Y}\int_0^Y 1_{(-\infty,1/2]}(x)1_{[0,Y]}(x)\mathrm{d}x \\ &= \begin{cases} 1 & \text{якщо } Y \leqslant 1/2, \\ 1/(2Y) & \text{якщо } Y > 1/2. \end{cases}\end{aligned}$$

Властивості умовного сподівання

Припустимо, що $(\Omega, \mathcal{F}, \mathbb{P})$ є фіксованим імовірнісним простором, і нехай $\mathcal{M}$ є деякою під-σ-алгеброю $\mathcal{F}$. Крім того, ми припускаємо, що всі умовні випадкові величини інтегровні.

0. Маємо $\mathbb{E}(\mathbb{E}(X|\mathcal{M})) = \mathbb{E}X$. Це відразу випливає з 2), якщо взяти $B = \Omega$.

Приклад 15.1. Кількість аварій у певний день у місті має розподіл Пуассона з параметром 5. Сума збитків, спричинених аварією, рівномірно розподілена в діапазоні $[2, 10]$. Нехай X буде загальним збитком за день. Визначити $\mathbb{E}X$.

Рішення: Введемо випадкову змінну Y, задану як кількість аварій за певний день. Змінна Y має розподіл Пуассона з параметром 5, причому з умов задачі $\mathbb{E}(X|Y) = 6Y$. Дійсно, середня сума збитку, заподіяного в одній аварії, становить 6, отже, якщо було Y аварій, середня сума збитків становить $6Y$. Отже, користуючись з властивості 0., отримуємо

$$\mathbb{E}X = \mathbb{E}(\mathbb{E}(X|Y)) = \mathbb{E}6Y = 30.$$

1. Нехай α, $\beta \in \mathbb{R}$. Тоді

$$\mathbb{E}(\alpha X_1 + \beta X_2|\mathcal{M}) = \alpha\mathbb{E}(X_1|\mathcal{M}) + \beta\mathbb{E}(X_2|\mathcal{M}).$$

Дійсно: перевіримо, що права частина (надалі позначається R) задовольняє умови 1) і 2) з визначення $\mathbb{E}(\alpha X_1 + \beta X_2|\mathcal{M})$. Перша умова очевидна. Щоб перевірити другу, зауважимо, що для довільної $B \in \mathcal{M}$,

$$\begin{aligned}\int_B R d\mathbb{P} &= \alpha \int_B \mathbb{E}(X_1|\mathcal{M})\mathrm{d}\mathbb{P} + \beta \int_B \mathbb{E}(X_2|\mathcal{M})\mathrm{d}\mathbb{P} = \alpha \int_B X_1 \mathrm{d}\mathbb{P} + \beta \int_B X_2 \mathrm{d}\mathbb{P} \\ &= \int_B \alpha X_1 + \beta X_2 d\mathbb{P}.\end{aligned}$$

2. Якщо X є невід'ємною випадковою величиною, то $\mathbb{E}(X|\mathcal{M}) \geqslant 0$ п.н. Дійсно, нехай $B = \{\mathbb{E}(X|\mathcal{M}) < 0\}$. Тоді $B \in \mathcal{M}$ і

$$\int_B \mathbb{E}(X|\mathcal{M}) d\mathbb{P} = \int_B X d\mathbb{P}.$$

Ми бачимо, що якби B мала позитивну ймовірність, ліва частина була б від'ємною, а права – невід'ємною.

3. Маємо

$$(*) \qquad |\mathbb{E}(X|\mathcal{M})| \leqslant \mathbb{E}(|X||\mathcal{M}) \quad \text{п.н.}$$

Дійсно, як випливає з 1. і 2., нерівність $X \leqslant Y$ п.н. тягне за собою $\mathbb{E}(X|\mathcal{M}) \leqslant \mathbb{E}(Y|\mathcal{M})$. Отже, з імовірністю 1,

$$\mathbb{E}(X_1|\mathcal{M}) \leqslant \mathbb{E}(|X_1||\mathcal{M})$$

і

$$-\mathbb{E}(X_1|\mathcal{M}) \leqslant \mathbb{E}(|X_1||\mathcal{M}).$$

Зауваження: Беручи у (*) сподівання обох сторін, ми отримуємо відповідно до 0.,

$$\mathbb{E}(|\mathbb{E}(X|\mathcal{M})|) \leqslant \mathbb{E}|X|.$$

Іншими словами, лінійний оператор $\mathbb{E}(\cdot|\mathcal{M}) : L^1(\Omega, \mathcal{F}, \mathbb{P}) \to L^1(\Omega, \mathcal{F}, \mathbb{P})$ є стискальним.

4. Умовний варіант теореми Лебега про монотонний перехід до границі. Нехай X_n — неспадна послідовність невід'ємних випадкових величин, збіжна п.н. до $X \in L^1$. Тоді $\mathbb{E}(X_n|\mathcal{M}) \uparrow \mathbb{E}(X|\mathcal{M})$ п.н.

Щоб показати це, зауважимо, що в силу 1 і 2 послідовність $(\mathbb{E}(X_n|\mathcal{M}))$ не спадає з імовірністю 1, і зокрема збігається. Позначимо його границю η, $\mathbb{E}(X_1|\mathcal{M}) \leqslant \eta \leqslant \infty$. Тепер нехай $B \in \mathcal{M}$. Відповідно до 2) і безумовній теореми Лебега маємо

$$\int_B X = \lim_{n\to\infty} \int_B X_n = \lim_{n\to\infty} \int_B \mathbb{E}(X_n|\mathcal{M}) = \int_B \eta.$$

Оскільки η $\mathcal{M}$-вимірна, наведена рівність означає, що $\eta = \mathbb{E}(X|\mathcal{M})$.

5. Аналогічно можно довести умовний варіант теореми Лебега про мажорований перехід до границі під знаком інтеграла та умовний варіант леми Фату.

6. Припустимо, що X_1 є вимірною відносно $\mathcal{M}$. Тоді

$$\mathbb{E}(X_1X_2|\mathcal{M}) = X_1\mathbb{E}(X_2|\mathcal{M}) \quad \text{п.н.} \tag{+}$$

Зокрема, беручи $X_2 \equiv 1$, ми отримуємо, що $\mathbb{E}(X_1|\mathcal{M}) = X_1$.

Перевіримо, що права частина задовольняє умовам 1) і 2) з визначення $\mathbb{E}(X_1X_2|\mathcal{M})$. Умова 1) очевидна, тому залишається перевірити 2). Ми будемо використовувати метод ускладнення X_1. а) Якщо $X_1 = 1_A$, де $A \in \mathcal{M}$, то для будь-якого $B \in \mathcal{M}$,

$$\int_B X_1\mathbb{E}(X_2|\mathcal{M})\mathrm{d}\mathbb{P} = \int_{A\cap B} \mathbb{E}(X_2|\mathcal{M})\mathrm{d}\mathbb{P} = \int_{A\cap B} X_2\mathrm{d}\mathbb{P} = \int_B X_1X_2\mathrm{d}\mathbb{P}.$$

б) Якщо X_1 є простою змінною, то формулу (+) отримуємо з а) і лінійності умовних сподівань.

в) Якщо X_1 є невід'ємною випадковою величиною, то існує неспадна послідовність (Y_n) $\mathcal{M}$-вимірних простих величин, збіжних п.н. до X_1. Розділимо $X_2 = X_2^+ - X_2^-$ і застосуємо b) до змінних Y_n і X_2^+:

$$\mathbb{E}(Y_nX_2^+|\mathcal{M}) = Y_n\mathbb{E}(X_2^+|\mathcal{M}).$$

Переходячи до границі $n \to \infty$ і використовуючи умовне версію Теореми Лебега (властивість 4) отримуємо

$$\mathbb{E}(X_1X_2^+|\mathcal{M}) = X_1\mathbb{E}(X_2^+|\mathcal{M}).$$

Замінивши X_2^+ на X_2^- і повторивши міркування, отримуємо

$$\mathbb{E}(X_1X_2^-|\mathcal{M}) = X_1\mathbb{E}(X_2^-|\mathcal{M})$$

і після віднімання обох сторін маємо (+).

d) Якщо X_1 є будь-якою випадковою величиною, то ми розбиваємо її на різницю $X_1^+ - X_1^-$, застосовуємо c) до змінних X_1^+, X_2 і X_1^-, X_2, і віднімаємо отримані рівності.

7. Якщо $\mathcal{M}_1 \subset \mathcal{M}_2$ є під-σ-алгебрами $\mathcal{F}$, то

$$\mathbb{E}(X|\mathcal{M}_1) = \mathbb{E}(\mathbb{E}(X|\mathcal{M}_2)|\mathcal{M}_1) = \mathbb{E}(\mathbb{E}(X|\mathcal{M}_1)|\mathcal{M}_2). \tag{=}$$

Почнемо з того, що вирази обох сторонах рівні. Це безпосередньо випливає з попередньої властивості: випадкова величина $\mathbb{E}(X|\mathcal{M}_1)$ вимірна відносно $\mathcal{M}_2$. Отже, достатньо довести, що перші два доданки в (=) рівні. Візьмемо $B \in \mathcal{M}_1$. Ми маємо $B \in \mathcal{M}_2$, отже

$$\int_B \mathbb{E}(X|\mathcal{M}_1) = \int_B X = \int_B \mathbb{E}(X|\mathcal{M}_2) = \int_B \mathbb{E}(\mathbb{E}(X|\mathcal{M}_2)|\mathcal{M}_1),$$

звідки випливає теза.

8. Припустимо, X не залежить від $\mathcal{M}$. Тому $\mathbb{E}(X|\mathcal{M}) = \mathbb{E}X$. Дійсно, перевіримо, що $\mathbb{E}X$ задовольняє умови 1) і 2) у визначенні $\mathbb{E}(X|\mathcal{M})$. Умова 1) очевидна: $\mathbb{E}X$ є постійною випадковою величиною, а отже, вимірною відносно кожної σ-алгебри. Тепер нехай $B \in \mathcal{M}$. З незалежністі 1_B і X випливає,

$$\int_B \mathbb{E}X\mathrm{d}\mathbb{P} = \mathbb{E}1_B\mathbb{E}X = \mathbb{E}(1_BX) = \int_B X\mathrm{d}\mathbb{P}.$$

9. Нерівність Йенсена. Якщо $f : \mathbb{R} \to \mathbb{R}$ є функцією опуклою такою, що $f(X)$ інтегровна, то

$$\mathbb{E}(f(X)|\mathcal{M}) \geqslant f(\mathbb{E}(X|\mathcal{M})).$$

Нам знадобиться наступний простий факт. Ми залишимо доведення як просту вправу.

Лема 15.2. *Нехай $f : \mathbb{R} \to \mathbb{R}$ — опукла функція. Тоді існують такі послідовності (a_n), (b_n), що для будь-якого $x \in \mathbb{R}$,*

$$f(x) = \sup_n (a_n x + b_n).$$

Повернемося до доказу 9. Для послідовностей (a_n), (b_n), існування яких гарантує наведена вище лема, ми маємо $f(X) \geqslant a_n X + b_n$ для кожного n. Отже, в силу 1 і 2 з імовірністю 1,

$$\mathbb{E}(f(X)|\mathcal{M}) \geqslant a_n \mathbb{E}(X|\mathcal{M}) + b_n.$$

Оскільки послідовності (a_n), (b_n) є зліченними, ми можемо взяти верхню суму n у правій частині, і тоді з імовірністю 1 виконуєтсья нерівність:

$$\mathbb{E}(f(X)|\mathcal{M}) \geqslant \sup_n (a_n \mathbb{E}(X|\mathcal{M}) + b_n) = f(\mathbb{E}(X|\mathcal{M})).$$

Зауваження: Як висновок отримуємо, що для $p \geqslant 1$ і $X \in L^p(\Omega, \mathcal{F}, \mathbb{P})$,

$$\mathbb{E}(|X|^p|\mathcal{M}) \geqslant [\mathbb{E}(|X||\mathcal{M})]^p.$$

Отже, після взяття сподівання обох сторін, $\mathbb{E}(|\mathbb{E}(X|\mathcal{M})|^p) \leqslant \mathbb{E}|X|^p$, або

$$||\mathbb{E}(X|\mathcal{M})||_p \leqslant ||X||_p.$$

Отже, умовне сподівання $\mathbb{E}(\cdot|\mathcal{M})$ є стискальним відображенням у L^p.

Нарешті, розберемося з питанням нелінійної регресії. Припустимо, X, Y — інтегровні з квадратом випадкові величини. Ми спостерігаємо за змінною Y і за допомогою цих даних хочемо найкраще наблизити X (у середньоквадратичному значенні) змінною у вигляді $h(Y)$. Точніше, ми шукаємо таку борелівську функцію f, що

$$\mathbb{E}(X - f(Y))^2 = \min_h \mathbb{E}(X - h(Y))^2.$$

Якщо ми звузимося до класу лінійних функцій, це приведе до задачі лінійної регресії, розглянутої раніше.

Теорема 15.3. *Розв'язанням вищезазначеної проблеми є $f(Y) = \mathbb{E}(X|Y)$.*

Доведення. Візьмемо будь-яку борелівську функцію h. Маємо

$$\begin{aligned}\mathbb{E}(X - h(Y))^2 &= \mathbb{E}(X - f(Y) + f(Y) - h(Y))^2 \\ &= \mathbb{E}(X - f(Y))^2 + 2\mathbb{E}(X - f(Y))(f(Y) - h(Y)) + \mathbb{E}(f(Y) - h(Y))^2.\end{aligned}$$

Але $f(Y) - h(Y)$ є вимірна відносно $\sigma(Y)$. Отже, використовуючи властивості 0., 1. і 6., отримуємо

$$\begin{aligned}\mathbb{E}(X - f(Y))(f(Y) - h(Y)) &= \mathbb{E}\Big\{\mathbb{E}\big[(X - f(Y))(f(Y) - h(Y))|Y\big]\Big\} \\ &= \mathbb{E}\Big\{(f(Y) - h(Y))\mathbb{E}(X - f(Y)|Y)\Big\} = 0.\end{aligned}$$

Отже, середній член у попередній послідовності рівнянь дорівнює нулю, і ми отримуємо

$$\mathbb{E}(X - h(Y))^2 \geqslant \mathbb{E}(X - f(Y))^2.$$

Звідси теза. $\square$

16 Завдання

1. Випадкові величини ε_1, ε_2, ε_3 незалежні і мають однаковий розподіл $\mathbb{P}(\varepsilon_i = -1) = \mathbb{P}(\varepsilon_i = 1) = 1/2$, $i = 1, 2, 3$. Обчисліть $\mathbb{E}(\varepsilon_1 | \varepsilon_1 + \varepsilon_2 + \varepsilon_3)$ і $\mathbb{E}(\varepsilon_1\varepsilon_2 | e_1 + e_2 e_3)$.

2. Випадкові величини X, Y є незалежними, де X має розподіл Бернуллі $B(n, p)$, а Y має розподіл Бернуллі $B(m, p)$. Визначте $\mathbb{E}(X + Y | X)$ і $\mathbb{E}(X | X + Y)$.

3. Гральний кубик кидається, а потім кидається стільки разів, скільки випало в перший раз. Обчисліть сподівання кількості отриманих трійок.

4. Урна містить a білих куль, b чорних куль і c червоних куль (a, b, c — натуральні числа). Беремо навмання по одній кульці за раз (з повертанням), доки не буде витягнуто червону кульку. Знайти сподівання кількості розіграшів, у яких витягнуто білу кульку.

5. Відомо, що p відсотків монет є підробками, з орлом по обидва боки. Ми випадково беремо n монет і кидаємо кожну один раз. Нехай F позначає кількість кинутих фальшивих монет, а O - кількість отриманих голів. Доведіть, що $\mathbb{E}(F|O) = \frac{2p}{100+p} O$.

6. Випадкова величина (X, Y) має щільність

$$g(x, y) = \frac{x^3}{2} e^{-x(y+1)} 1_{\{x>0,\, y>0\}}.$$

Визначити $\mathbb{E}(Y|X)$, $\mathbb{E}(Y^2|X^2)$ і $\mathbb{P}(Y > 1|X^3 + 1)$.

7. Випадкові величини X, Y незалежні та експоненціально розподілені з параметром 1. Обчислити $\mathbb{P}(X \in B|X + Y)$ (для $B \in \mathcal{B}(\mathbb{R})$) і $\mathbb{E}(\sin X|X + Y)$.

8. Випадкова величина X має експоненціальний розподіл з параметром 1, а Y є випадковою величиною, такою що якщо $X = x$, то Y має експоненціальний розподіл з параметром x.

a) Знайдіть розподіл Y.

b) Обчисліть $\mathbb{P}(X > r|Y)$.

9. Випадкова величина (X, Y) нормально розподілена з середнім 0, $\mathrm{Var}X = \sigma_1^2$, $\mathrm{Var}Y = \sigma_2^2$, $\mathrm{Cov}(X, Y) = c$. Обчислити $\mathbb{P}(Y \in B|X)$ (для $B \in \mathcal{B}(\mathbb{R})$) і $\mathbb{E}(Y|X)$.

10. X_1, X_2, …, X_n є незалежними та мають однаковий розподіл зі скінченним середнім. Обчислити $\mathbb{E}(X_1|X_1 + X_2 + \ldots + X_n)$.

11. Припустимо, що X, Y — випадкові величини, а $\mathcal{G}$ — σ-алгебра така, що X є вимірним відносно $\mathcal{G}$, а Y є незалежною від $\mathcal{G}$. Нехай $\phi : \mathbb{R}^2 \to \mathbb{R}$ — така борелівська функція, що $\phi(X, Y)$ — інтегровна випадкова величина. Доведіть, що

$$\mathbb{E}[\phi(X, Y)|\mathcal{G}] = \Phi(X),$$

де $\Phi(x) = \mathbb{E}\phi(x, Y)$.

12. Припустимо, що X є інтегровною випадковою величиною, а σ-алгебра $\mathcal{G}$ не залежить від X і σ-алгебри $\mathcal{M}$. Доведіть, що

$$\mathbb{E}(X|\sigma(\mathcal{G}, \mathcal{M})) = \mathbb{E}(X|\mathcal{M}).$$

13. X, Y, Z є незалежними, при чому X ма стандартніий нормальний розклад, Y є невід'ємною обмеженою змінною, а Z є змінною Радемахера. Обчислити $\mathbb{E}(e^{XY}|Y)$ та $\mathbb{E}(e^{XY}|YZ)$.

14. N, X_1, X_2, … є незалежними, при чому N має розподіл Пуассона з параметром 3, а X_n має рівномірний розподіл на $[0, 1]$, $n = 1, 2, \ldots$. Обчислити $\mathbb{E}(X_1 + X_2 + \ldots + X_{N+1})$.