Лекція 3. Статистичне виведення в R. Загальні принципи

Данило Тавров

2023-02-22

План лекції

1. Поняття про статистичне виведення
2. Основні поняття теорії ймовірностей
3. Огляд деяких найважливіших розподілів

Вступні зауваги

- Найближчі дві лекції ми присвятимо пригадуванню основних відомостей із теорії ймовірностей та статистики
- Сьогодні ми згадаємо основні поняття з теорії ймовірностей та подивимося, як їх реалізовано в R
- Корисними матеріалами є:
  - Мій конспект лекцій із «Теорії ймовірностей» (викладено на диску в каталозі з Лекцією 3)
  - Фундаментальна книжка *All of Statistics*, розділи 1–4 (викладено на диску в загальному каталозі з літературою)

Базові поняття

- Під **статистичним виведенням** (statistical inference) будемо розуміти процес формування висновків про **популяцію**[1] (population) на основі деякої **вибірки** (sample)
  - Популяція містить повний набір даних про всі можливі спостереження
  - Вибірка є підмножиною популяції
- Маючи на руках дані певної вибірки, ми намагаємося з'ясувати, яка **ймовірнісна модель** (probability model, або ще кажуть **data generating process**, DGP) ці дані породила
  - У термінах теорії ймовірностей, нас цікавить **спільний розподіл** (joint probability distribution) усіх випадкових величин із вибірки
- **Параметром** (parameter) модели є деяка її числова характеристика
- **Статистикою** (statistic) є функція від даних, яка описує вибірку
- Ми хочемо використати підраховані статистики для **виведення** (to infer) знання про параметри популяції

[1] Також побутує поняття «генеральна сукупність», але ми ним користуватися не будемо

Базові поняття

- Зв'язок між теорією ймовірностей і статистичним виведенням:

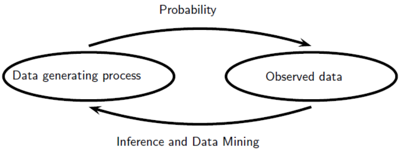


Рис. 1.1: Імовірність і виведення (Wasserman, p. ix)

- Теорія ймовірностей є формальною мовою невизначености
- У теорії ймовірностей ми намагаємося відповісти на питання: «Якщо ми знаємо DGP, які є властивості згенерованих даних?»
- У статистичному виведенні ми намагаємося відповісти на питання: «Якщо ми знаємо дані, що можна сказати про DGP, який їх згенерував?»
- Фактично статистичним виведенням значною мірою є й те, що називають data analysis, machine learning і data mining
- До статистичного виведення в широкому сенсі можна віднести як інференційний аналіз, так і прогнозний, і причиново-наслідковий
- Адже як для прогнозування, так і для встановлення причиново-наслідкових зв'язків потрібно знати характеристики DGP

Зв’язок між статистикою і машинним навчанням

І статистика, і data mining, і machine learning пов’язані зі збором та аналізом даних. Протягом певного часу статистичні дослідження виконувалися на кафедрах статистики, а data mining і machine learning досліджувалися на кафедрах комп’ютерних наук. Статистики вважали, що комп’ютерники винаходять велосипед. Комп’ютерники вважали, що статистична теорія не розв’язує їхніх задач. (Wasserman, p. vii)

Приклади

- Як рівень мінімальної зарплати впливає на безробіття?
- Який зв'язок між освітою та доходами?
- Чи зменшується розмір мозку від недосипання?
- Чи є собаки, яких вигулюють чоловіки, агресивніші?
- Які висновки можна зробити на основі соціологічного опитування?
- Які ділянки мозку відповідають за певні активності?
- Чи є ліки ефективні?

Виведення на основі даних спостережень

- В ідеалі ми воліли би провести **експеримент** (experiment) і зібрати всі потрібні дані
- Проте на практиці це можливо тільки в окремих випадках і в окремих галузях (медицині, біостатистиці, дуже обмежено в економіці тощо)
- У нашому курсі ми **не будемо** займатися питання розроблення експериментів
- Ми працюватимемо з **даними спостережень** (observational data)
- Вибірка, із якою ми працюємо, може викликати питання:
  - Чи є вибірка **репрезентативною** (representative) для популяції?
  - Чи наявні в даних спостереження відібрано з однаковою ймовірністю?
  - Які існують додаткові змінні (відомі і наявні; відомі, але **не** наявні; **не** відомі і **не** наявні), які можуть спотворити наші висновки?
  - Який вплив мають пропущені дані?
  - І т.д.
- Виконуючи статистичне виведення, потрібно завжди робити певні **припущення** (assumptions) і правильно їх мотивувати

Два різні підходи до статистичного виведення

- Згідно з **частотним підходом** (frequentist approach), імовірність деякої події — це гранична частота настання події за умови нескінченного повторювання одного й того ж експерименту
- Відтак **частотне виведення** (frequentist inference) слідує цій інтерпретації для оцінювання параметрів та керування похибками
- Згідно з **Беєсівським підходом** (Bayesian approach), імовірність являє собою **ступінь упевнености** (belief) у настанні деякої події
- Відтак **Беєсівське виведення** (Bayesian inference) слідує цій інтерпретації та розглядає всі невідомі параметри моделей як випадкові
- У нашому курсі розглянемо обидва підходи

План лекції

Дисклеймер

- Ми коротко згадаємо основні визначення та властивості
- Ви це вже все знаєте з попередніх курсів
- Детальнішу інформацію можна почитати в конспекті лекцій з «Теорії ймовірностей» на диску (КЛТЙ)
- Якщо щось незрозуміло чи не пам'ятаєте — відразу заповнюйте прогалини

Вимірний простір

Визначення 2.1

- **Простір елементарних подій** (sample space) Ω для деякого випадкового експерименту — множина всіх можливих його результатів. Може бути:
  - Скінченний (finite): $|\Omega| < \infty$
  - Зліченний (countable): $|\Omega| = \aleph_0$
  - Незліченний (uncountable): $|\Omega| = \mathfrak{c}$
- Елемент $\omega \in \Omega$ — **елементарна подія** (sample)
- **Подія** (event) A — це деяка підмножина (subset) простору Ω: $A \subseteq \Omega$

□

- На просторі Ω ми визначаємо σ**-алгебру** $\mathcal{A}$ — множину подій, які нас цікавлять[2]
  - На скінченному й зліченному просторах ми беремо множину всіх підмножин: $\mathcal{A} = 2^\Omega$
  - На незліченному просторі (як правило, $\mathbb{R}$ або $\mathbb{R}^k$) ми беремо **Борелеву** σ**-алгебру** (Borel σ-algebra)[3] $\mathcal{B}$ або $\mathcal{B}^k$

Визначення 2.2

- $(\Omega, \mathcal{A})$ — **вимірний простір** (measurable space)
- Множини в σ-алгебрі $\mathcal{A}$ — **вимірні множини** (measurable sets)

□

[2] Деталі можна почитати в КЛТЙ 1.5

[3] Еміль Борель (Félix Édouard Justin Émile Borel, 1871–1956) — французький математик

Імовірнісна міра

Визначення 2.3

- Функція $\mu : \mathcal{A} \to \mathbb{R}$ — **міра** (measure), якщо:
  - $\mu(A) \geq 0$ для всіх $A \in \mathcal{A}$
  - μ є σ**-адитивною** (σ-additive):

$$\mu\left(\bigcup_{j=1}^{\infty} A_j\right) = \sum_{j=1}^{\infty} \mu(A_j)\ , \quad A_j \in \mathcal{A}\ ,\ j = 1, 2, \ldots\ , \quad A_i \cap A_j = \emptyset\ ,\ i \neq j \tag{2.1}$$

□

- Якщо $\mu(\Omega) = 1$, то μ називають **імовірнісною мірою** (probability measure) і позначають через $\mathbb{P}$
- Трійку $(\Omega, \mathcal{A}, \mathbb{P})$ називають **імовірнісним простором** (probability space)
- Деякі властивості ймовірнісної міри:
  - $\mathbb{P}(\emptyset) = 0$
  - $\mathbb{P}(A) \leq 1$
  - $\mathbb{P}(A^c) = 1 - \mathbb{P}(A)$
  - $\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B) - \mathbb{P}(A \cap B)$
  - Якщо $A_1 \subseteq A_2$, то $\mathbb{P}(A_1) \leq \mathbb{P}(A_2)$

Випадкові величини

Визначення 2.4

Дійснозначну[4] функцію $X : \Omega \to \mathbb{R}$ називають **випадковою величиною** (random variable) □

- Наприклад, Ω може містити всіх можливих людей
- Тоді $X : \Omega \to \mathbb{R}^+$ відповідає зросту кожної людини
- Випадкові величини, як правило, позначаємо великими літерами латинського алфавіту (X, Y, Z тощо)
- Значення, яких вони набувають — маленькими (x, y, z тощо)

[4] Строго формально, ще й вимірну; див. КЛТЙ 4.1

Розподіл випадкової величини

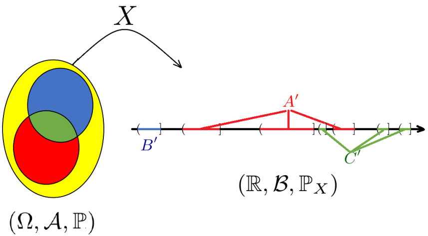
Визначення 2.5

- Випадкова величина X задає на вимірному просторі $(\mathbb{R}, \mathcal{B})$ ймовірнісну міру

$$\mathbb{P}_X(A) \equiv \mathbb{P}_X(X \in A) = \mathbb{P}\left(X^{-1}(A)\right), \qquad A \in \mathcal{B} \tag{2.2}$$

- Таку міру називають **розподілом** (distribution) X
- Випадкові величини X і Y **рівні за розподілом** (equal in distribution), $X \overset{d}{=} Y$, якщо $\mathbb{P}_X(A) = \mathbb{P}_Y(A)$ для всіх $A \in \mathcal{B}$

□



Як рахувати ймовірності з випадковими величинами?

- Нехай маємо випадкову величину $X : (\Omega, \mathcal{A}, \mathbb{P}) \to (\mathbb{R}, \mathcal{B})$ з розподілом $\mathbb{P}_X$
- Наприклад, це може бути «вага людини, кг»
- Ми не маємо жодного уявлення про простір $(\Omega, \mathcal{A}, \mathbb{P})$, на якому вона визначена
- Ми тільки спостерігаємо значення X
- Нас цікавить обчислити ймовірність, наприклад, події $X > 100$
- За (2.2) ми повинні порахувати $\mathbb{P}_X\left(X > 100\right) = \mathbb{P}\left(X^{-1}(100; \infty)\right)$
- Без знання простору $(\Omega, \mathcal{A}, \mathbb{P})$ це зробити просто неможливо
- Тому на практиці використовують такі випадкові величини, розподіл яких можна **легко** характеризувати

Абсолютно неперервні міри

Визначення 2.6

- Нехай на вимірному просторі $(\Omega, \mathcal{A})$ визначено дві міри, μ і ν
- Міра ν є **абсолютно неперервною відносно** μ (absolutely continuous with respect to μ), що позначають як $\nu \ll \mu$, якщо

$$\forall A \in \mathcal{A}\ ,\ \ \mu(A) = 0 \Rightarrow \nu(A) = 0 \qquad \square$$

Теорема 2.7 (Теорема Радона-Никодима (Radon–Nikodym theorem)[5])

- Нехай на вимірному просторі $(\Omega, \mathcal{A})$ визначено дві (σ-скінченні) міри μ і ν, $\nu \ll \mu$
- Тоді існує невід'ємна (вимірна) функція $f : \Omega \to [0; \infty)$ така, що

$$\nu(A) = \int_A f\, d\mu\ , \qquad \forall A \in \mathcal{A} \tag{2.3}$$

[5] Йоганн Карл Август Радон (Johann Karl August Radon, 1887–1956) — австрійський математик. Отто Марчин Никодим (Otto Marcin Nikodym, 1887–1974) — польський математик

Два основні типи випадкових величин

- На практиці нас цікавлять такі випадкові величини X, розподіл яких $\mathbb{P}_X$ абсолютно неперервний відносно деякої міри
- Тоді ймовірність будь-якої події $X \in A$, $A \in \mathcal{B}$, можна обчислити за допомогою (2.3) як $\mathbb{P}_X(X \in A) = \int_A f\,d\mu$

1. Випадкова величина X **дискретна**, якщо її образ зліченний, тобто $X \in X_0 = \{x_1, x_2, \dots\}$
   - Тоді можна показати, що $\mathbb{P}_X \ll \#$
   - Тут # — це **лічна міра** (counting measure): $\#(A) = |X_0 \cap A|$, $A \in \mathcal{B}$
   - Тоді $\mathbb{P}(X \in A) = \int_A f\,d\# = \sum_{x \in X_0 \cap A} f(x)$
2. Випадкова величина X **неперервна**, якщо її образ незліченний і $\mathbb{P}_X \ll \lambda$
   - Тут λ — це **міра Лебега** (Lebesgue measure)[6]
   - Міра Лебега — **єдина (σ-скінченна[7]) міра, яка має інтерпретацію довжини інтервалу**
   - Тобто $\lambda((a;b]) = b - a$
   - Тоді[8] $\mathbb{P}(X \in A) = \int_A f\,d\lambda = \int_A f(x)\,dx$

[6] Анрі Лебег (Henri Lebesgue, 1875–1941) — французький математик

[7] Це для нашого курсу непринципово; дис. КЛТЙ 2.4

[8] На практиці зустрічаються тільки такі величини, де перехід до інтегралу Рімана справді можливий

Функція ймовірности дискретної випадкової величини

- Отже якщо X дискретна, то $\mathbb{P}_X(X \in A) = \int_A f\, d\mu = \sum_{x \in X_0 \cap A} f(x)$
- **Носієм** (support) дискретної величини є її образ: $\text{supp}(X) = X_0$
  - Наприклад, X може відповідати числу дітей у сім'ї, і тоді $\text{supp}(X) = \{0, 1, 2, \dots\}$
  - Або числу гербів після підкидання двох монеток, і тоді $\text{supp}(X) = \{0, 1, 2\}$

Визначення 2.8

Таку функцію f називають **функцією ймовірности** (probability mass function) і частіше позначають через $p_X(x) \equiv \mathbb{P}_X(X = x)$ □

- Ця функція дає ймовірність будь-якого елемента носія
- За межами носія $p_X(x) = 0$, $x \notin \text{supp}(X)$
- Також повинно бути справедливим $\sum_{x \in \text{supp}(X)} p_X(x) = 1$

Щільність розподілу неперервної випадкової величини

- Якщо ж X неперервна, то $\mathbb{P}_X(X \in A) = \int_A f_X \, d\lambda = \int_A f_X(x)\, dx$
- **Носієм** неперервної величини є $\operatorname{supp}(x) = \{x \in \mathbb{R} : f(x) > 0\}$
  - Наприклад, X може відповідати зросту людини, і тоді $\operatorname{supp}(X) = (0; \infty)$
  - Або частці пацієнтів, що одужали після вживання ліків, і тоді $\operatorname{supp}(X) = [0; 1]$

Визначення 2.9

Таку функцію f_X називають **щільністю розподілу** (probability density function, PDF) □

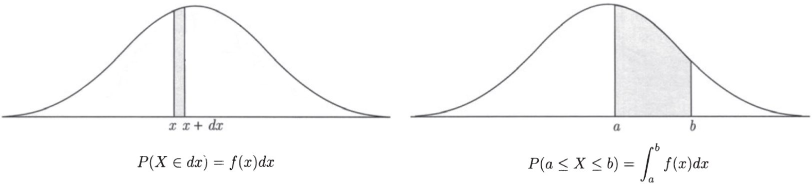
Теорема 2.10

Будь-яка функція $f : \mathbb{R} \to [0; \infty)$ є щільністю для деякого розподілу тоді й тільки тоді, коли:

- $f(x) \geq 0$ для всіх $x \in \mathbb{R}$
- $\int_{-\infty}^{\infty} f(t)\, dt = 1$

Інтерпретація щільности розподілу

- Щільність розподілу дуже подібна до функції ймовірности, проте не зовсім
- Для розподілів із неперервною функцією розподілу $\mathbb{P}_X\left(X = x\right) = 0$
- А щільність розподілу — це ймовірність **нескінченно малого інтервалу**: $\mathbb{P}_X\left(X \in (x; x + dx)\right) = f(x)dx$



- Щільність $f(x)$ сама по собі **не є ймовірністю**
- Вона навіть може перевищувати 1

Події міри нуль

- Подію A, імовірність якої $\mathbb{P}(A) = 0$, називають **подією міри нуль** (measure zero event, null set)
  - Множинами міри нуль є множини з одного елемента $\{x\}$, $x \in \mathbb{R}$
  - Також будь-які зліченні множини ($\mathbb{N}$, $\mathbb{Z}$, парні числа тощо)
  - В $\mathbb{R}^k$ це є будь-які множини розмірности $\mathbb{R}^{k-1}$ (лінії на площині і т.п.)
- Кажуть, що деяке твердження істинне **майже напевно** (almost surely, a.s.) або **з імовірністю 1** (with probability 1, w.p.1), якщо воно **не виконується** на множині міри нуль
- Наприклад, можемо казати, що випадкова величина $X > 0$ майже напевно
  - Це означає, що вона **може** набувати від'ємних значень, але такі події є нульовими
  - Тобто на практиці вони ніколи не виринатимуть
- Доведення низки теоретичних властивостей прекрасно працює, якщо якась властивість виконується тільки «майже» напевно
  - Наприклад, може вимагатися функція, яка є неперервною майже напевно
  - Це означає, що вона в принципі може мати зліченну кількість точок розриву
  - Наприклад, $f(x) = 1/x$ є неперервною майже напевно
- Також будь-яка щільність унікальна **з точністю до міри нуль**
  - Тобто якщо в одній точці поміняти значення щільности, суть не зміниться
  - Адже інтеграл від значення в одній точці не зміниться

Функції розподілу (1)

- Окрім дискретних і неперервних величин, існують величини мішані
- Обчислювати ймовірності з такими величинами через відповідні інтеграли неочевидно
- Існує апарат, який працює **для повністю довільних** випадкових величин

Визначення 2.11

Функцією розподілу (distribution function, інколи cumulative distribution function, CDF) випадкової величини X є функція

$$F_X(x) \equiv \mathbb{P}_X\left(X \leq x\right) = \mathbb{P}_X\left((-\infty; x]\right), \qquad x \in \mathbb{R} \tag{2.4}$$

□

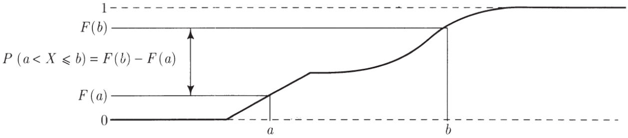
- Можна довести, що між розподілом і функцією розподілу існує взаємно однозначне відображення
- Тобто кожна випадкова величина має функцію розподілу, навіть якщо вона не має функції ймовірности чи щільности

Функції розподілу (2)

- Функція розподілу $F_X(x)$ має такі властивості:
  - Є неспадною: $F_X(x) \leq F_X(y)$, якщо $x \leq y$
  - Є неперервною справа: $\lim\limits_{x \downarrow c} F_X(x) = F_X(c)$
    - Границю зліва прийнято позначати як $F_X(x-)$
  - Є нормованою:

$$\lim_{x \to -\infty} F_X(x) = 0 \ , \qquad \lim_{x \to \infty} F_X(x) = 1$$

- За допомогою функцій розподілу можна обчислювати ймовірності:
  - $\mathbb{P}_X\left(X \geq x\right) = 1 - F(x-)$
  - $\mathbb{P}_X\left((a; b]\right) = F(b) - F(a)$, $a, b \in \mathbb{R}$
  - $\mathbb{P}_X\left(X = x\right) = F(x) - F(x-)$
- Якщо випадкова величина неперервна, то $\mathbb{P}_X\left(X = x\right) = 0$
  - Відтак $\mathbb{P}_X\left(X \geq x\right) = 1 - F(x)$
  - Відтак $\mathbb{P}_X\left((a; b]\right) = \mathbb{P}_X\left([a; b)\right) = \mathbb{P}_X\left((a; b)\right) = \mathbb{P}_X\left([a; b]\right) = F(b) - F(a) = \int_a^b f(x)\,dx$, $a, b \in \mathbb{R}$



Зв’язок між функціями розподілу та щільностями

- Для **неперервної** величини X маємо

$$F_X(x) = \mathbb{P}_X\left(X \leq x\right) = \int_{-\infty}^{x} f_X(x)\, dx$$

- Іншими словами, для всіх точок x, у яких функція F_X диференційовна:

$$f_X(x) = F_X'(x) \tag{2.5}$$

Функції від випадкових величин

- Нехай X має функцію розподілу F_X
- Нехай g — деяка вимірна[9] функція
- Функцію розподілу F_Y випадкової величини $Y = g(X)$ можна дістати за визначенням:

$$F_Y(y) = \mathbb{P}_Y(Y \leq y) = \mathbb{P}_X(g(X) \leq y) \tag{2.6}$$

- Якщо X дискретна,

$$\mathbb{P}_Y(Y = y) = \mathbb{P}_X(g(X) = y) = \sum_{x: g(x) = y} \mathbb{P}_X(X = x) \tag{2.7}$$

- Якщо X неперервна, то справедлива формула заміни змінних:

$$f_Y(y) = f_X(g^{-1}(y)) \cdot \left| \frac{d}{dy} g^{-1}(y) \right| \cdot \mathbb{1}\{y \in g(\text{supp}(X))\} \tag{2.8}$$

  - Тут f_X неперервна на $\text{supp}(X)$
  - g взаємно однозначна на $\text{supp}(X)$
  - $g^{-1}(y)$ неперервно диференційовна на $g(\text{supp}(X))$

[9] Для наших цілей мова про довільну функцію

Випадкові вектори

Визначення 2.12

- Вимірну функцію $\mathbf{X} : \Omega \to \mathbb{R}^k$ називають **випадковим вектором**
- Можна довести, що $\mathbf{X}$ є випадковим вектором тоді й тільки тоді, коли X_i є випадковими величинами, $i = 1, \dots, k$
- Випадковий вектор $\mathbf{X} = (X_1, \dots, X_k)^\top : (\Omega, \mathcal{A}, \mathbb{P}) \to (\mathbb{R}^k, \mathcal{B}^k)$ утворює розподіл $\mathbb{P}_{\mathbf{X}}$

$$\mathbb{P}_{\mathbf{X}}(A) \equiv \mathbb{P}_{\mathbf{X}}\left((X_1, \dots, X_k)^\top \in A\right), \qquad A \in \mathcal{B}^k \tag{2.9}$$

- Такий розподіл називають **спільним розподілом** (joint distribution)
- Функція відповідного розподілу має вид

$$F_{\mathbf{X}}(x_1, \dots, x_k) = \mathbb{P}_{\mathbf{X}}\left(X_1 \le x_1, \dots, X_k \le x_k\right) \tag{2.10}$$

- Таку функцію розподілу називають **спільною функцією розподілу** (joint distribution function)

Визначення 2.13

- **Добутком** двох вимірних просторів, $(X, \mathcal{X})$ і $(Y, \mathcal{Y})$ (product space) називають декартів добуток $X \times Y$
- На $X \times Y$ можна визначити **добуток мір** (product measure) $\pi(A \times B) = \mu(A)\nu(B)$, $A \in \mathcal{X}$, $B \in \mathcal{Y}$

Теорема 2.14 (Теорема Фубіні (Fubini Theorem)[10])

- Нехай маємо $X \times Y$ і (σ-скінченний) добуток мір $\mu \times \nu \equiv \pi$
- Якщо $f : X \times Y \to \mathbb{R}$ невід'ємна або інтегровна (майже напевно), то

$$\int_{X \times Y} f(x, y)\, d\pi = \int_X \left(\int_Y f(x, y)\, d\nu \right) d\mu = \int_Y \left(\int_X f(x, y)\, d\mu \right) d\nu \tag{2.11}$$

- Інтеграл зліва називають – **подвійним** (double), а два інші — **повторними** (iterated)

[10] Гвідо Фубіні (Guido Fubini, 1879–1943) – італійський математик

Спільні функція ймовірности та щільність

Визначення 2.15

- Нехай маємо **дискретний** випадковий вектор $\mathbf{X} : (\Omega, \mathcal{A}, \mathbb{P}) \to (\mathbb{R}^k, \mathcal{B}^k)$
- Його розподіл абсолютно неперервний відносно лічної міри $\# \times \ldots \times \# \equiv \#_k$
- Тоді **спільною функцією ймовірности** (joint probability mass function) є

$$p_{\mathbf{X}}(x_1, \ldots, x_k) = \mathbb{P}_{\mathbf{X}}\left(X_1 = x_1, \ldots, X_k = x_k\right) \cdot \mathbb{1}\left\{(x_1, \ldots, x_k)^\top \in \operatorname{supp}\left(\mathbf{X}\right)\right\} \tag{2.12}$$

- До того ж $\displaystyle\sum_{(x_1,\ldots,x_k)^\top \in \operatorname{supp}(\mathbf{X})} p_{\mathbf{X}}(x_1, \ldots, x_k) = 1$

□

Визначення 2.16

- Нехай розподіл деякого випадкового вектора $\mathbf{X} : (\Omega, \mathcal{A}, \mathbb{P}) \to (\mathbb{R}^k, \mathcal{B}^k)$ **абсолютно неперервний** відносно міри Лебега $\lambda \times \ldots \times \lambda \equiv \lambda_k$
- Тоді **спільною щільністю розподілу** (joint probability density function) є функція $f_{\mathbf{X}}$ така, що

$$\mathbb{P}_{\mathbf{X}}\left(A\right) = \int_A f_{\mathbf{X}}\, d\lambda_k = \int_A f_{\mathbf{X}}(x_1, \ldots, x_k)\, dx_1 \ldots dx_k\ , \qquad \forall A \in \mathcal{B}^k \tag{2.13}$$

□

Властивості спільної щільности

- За аналогією з одновимірним випадком маємо властивість

$$F_{\mathbf{X}}(x_1, \dots, x_k) = \int_{-\infty}^{x_k} \dots \int_{-\infty}^{x_1} f_{\mathbf{X}}(t_1, \dots, t_k)\, dt_1 \dots dt_k \tag{2.14}$$

- Відповідно, маємо

$$f_{\mathbf{X}}(x_1, \dots, x_k) = \frac{\partial^k}{\partial x_1 \dots \partial x_k} F_{\mathbf{X}}(x_1, \dots, x_k) \tag{2.15}$$

- Також можна довести, що:
  - $f_{\mathbf{X}} \geq 0$
  - $\int_{\mathbb{R}^k} f_{\mathbf{X}} d\lambda_k = 1$

Маржинальні розподіли

Визначення 2.17

- Розподіл $\mathbb{P}_j$ випадкової величини X_j — j-ої координати деякого випадкового вектора $\mathbf{X} = (X_1, \dots, X_k)$ — називають **маржинальним розподілом** (marginal distribution)
- Маржинальним розподілам відповідають **маржинальні функції розподілу** (marginal distribution function):

$$\begin{aligned} F_{X_j}(x) &= \mathbb{P}_{X_j}\left(X_j \leq x\right) \\ &= \mathbb{P}_{\mathbf{X}}\left(X_1 \in \mathbb{R}, \dots, X_{j-1} \in \mathbb{R}, X_j \leq x, X_{j+1} \in \mathbb{R}, \dots, X_k \in \mathbb{R}\right) \end{aligned} \tag{2.16}$$

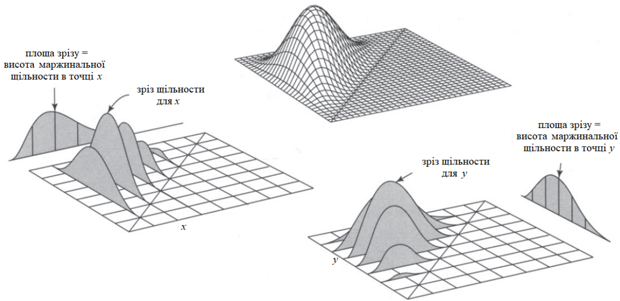
□

Маржинальні щільності розподілів

Твердження 2.18

- Нехай $\mathbb{P}_{\mathbf{X}}$ має щільність $f_{\mathbf{X}}$
- Тоді розподіл $\mathbb{P}_{X_j}$ має **маржинальну щільність розподілу** (marginal probability density function)

$$f_j(x) = \int_{\mathbb{R}^{k-1}} f_{\mathbf{X}}(x_1, \dots, x_{j-1}, x, x_{j+1}, \dots, x_k)\, dx_1 \dots dx_{j-1} dx_{j+1} \dots dx_k \quad (2.17)$$



- На практиці зустрічаються ситуації, коли випадковий вектор $\mathbf{X}$ складається з величин, деякі з яких є неперервними, а деякі — дискретними
- Наприклад, нас може цікавити аналіз ціни X квартири, яка має площу Y та має Z кімнат
- Спільний розподіл $(X, Y, Z)^\top$ **абсолютно неперервний** відносно добутку двох мір Лебега і лічної міри: $\lambda \times \lambda \times \#$
- Тому й обчислення спільних імовірностей буде приблизно таке:

$$\mathbb{P}_{X,Y,Z}\left((X,Y,Z)^\top \in A\right) = \int_{\mathbb{R}}\int_{\mathbb{R}} \sum_{z \in \text{supp}(Z)} f_{X,Y,Z}(x,y,z) \cdot \mathbb{1}\left\{(x,y,z)^\top \in A\right\} dxdy$$

Визначення 2.19

- Нехай маємо вимірний простір $(\Omega, \mathcal{A})$
- Нехай $A, B \in \mathcal{A}$ — деякі події, $\mathbb{P}(B) > 0$
- Тоді **умовна ймовірність** A за умови B (conditional probability of A given B):

$$\mathbb{P}(A \mid B) = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)} \tag{2.18}$$

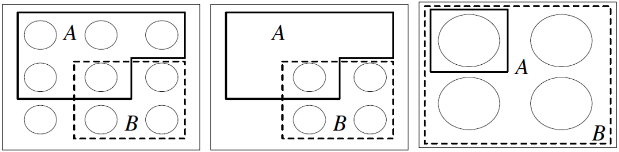
- $\mathbb{P}(A)$ — **апріорна ймовірність** (prior)
- $\mathbb{P}(A \mid B)$ — **апостеріорна ймовірність** (posterior)

$\square$

- **Не існує такої події, як** $A \mid B$! **Це просто позначення!**
- Можна довести, що умовна ймовірність — це повноцінна ймовірнісна міра

Обумовлення як перенормалізація

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



- Тут $\mathbb{P}(A) = 5/9$, проте $\mathbb{P}(A \mid B) = 1/4$

Теорема 2.20

- Нехай $\mathbb{P}\left(A_1 \cap \ldots \cap A_n\right) > 0$
- Тоді

$$\begin{aligned} \mathbb{P}\left(A_1 \cap \ldots \cap A_n\right) &= \mathbb{P}\left(A_1\right) \mathbb{P}\left(A_2 \mid A_1\right) \mathbb{P}\left(A_3 \mid A_1, A_2\right) \ldots \mathbb{P}\left(A_n \mid A_1, \ldots, A_{n-1}\right) \\ &= \mathbb{P}\left(A_2\right) \mathbb{P}\left(A_1 \mid A_2\right) \mathbb{P}\left(A_1 \mid A_2, A_3\right) \ldots \mathbb{P}\left(A_n \mid A_1, \ldots, A_{n-1}\right) \\ &= \ldots \end{aligned} \tag{2.19}$$

- І т.д. для всіх можливих комбінацій подій

Теорема Беєса та закон повної ймовірности

Теорема 2.21 (Теорема Беєса (Bayes' Theorem)[11])

- Нехай маємо вимірний простір $(\Omega, \mathcal{A})$
- Для подій $A, B \in \mathcal{A}$, $\mathbb{P}(B) > 0$, справедливо:

$$\mathbb{P}(A \mid B) = \frac{\mathbb{P}(B \mid A)\,\mathbb{P}(A)}{\mathbb{P}(B)} \tag{2.20}$$

Теорема 2.22 (Закон повної ймовірности (Law of total probability))

- Нехай маємо вимірний простір $(\Omega, \mathcal{A})$
- Розгляньмо розбиття: $\Omega = \bigcup_{i=1}^{n} A_i$, $A_i \cap A_j = \emptyset$, $i \neq j$
- Тут $\mathbb{P}(A_i) > 0$ для всіх i
- Тоді

$$\mathbb{P}(B) = \sum_{i=1}^{n} \mathbb{P}(B \mid A_i)\,\mathbb{P}(A_i) \tag{2.21}$$

[11] Томас Беєс (Thomas Bayes, 1701–1761) — англійський статистик, філософ і священник

Обумовлення декількома подіями

- Аналогічні результати можна показати, якщо обумовити додатковою подією
- Нехай $A, B, C \in \mathcal{A}$, $\mathbb{P}\left(B \cap C\right) > 0$
- Тоді

$$\mathbb{P}\left(A \mid B, C\right) = \frac{\mathbb{P}\left(B \mid A, C\right) \mathbb{P}\left(A \mid C\right)}{\mathbb{P}\left(B \mid C\right)} \tag{2.22}$$

- Нехай $A_1, \ldots, A_n$ утворюють розбиття Ω, і $\mathbb{P}\left(A_i \cap B\right) > 0$ для всіх i
- Тоді

$$\mathbb{P}\left(B \mid C\right) = \sum_{i=1}^{n} \mathbb{P}\left(B \mid A_i, C\right) \mathbb{P}\left(A_i \mid C\right) \tag{2.23}$$

Обумовлення випадковою величиною

- На практиці ми скрізь стикаємося з ситуаціями, коли треба порахувати ймовірності подій (у тому числі з випадковими величинами) **за умови іншої випадкової величини**
- Наприклад, X відповідає ціні квартири (у грн), а Y — її площі (у кв. м)
- Нас може цікавити ймовірність $\mathbb{P}_{X|Y}\left(X > 100\,000 \mid Y = 70\right)$
- Проте ми знаємо, що для неперервної величини $\mathbb{P}_{Y}\left(Y = 70\right) = 0$
- Щоб обійти це обмеження, ми формально розглядаємо обумовлення σ-алгеброю, яку породжує Y
  - Деталі можна прочитати в КЛТЙ 12.2
  - Фактично, мова про те, що

$$\mathbb{P}_{X|Y}\left(X > 100\,000 \mid Y = 70\right) = \lim_{h_n \to 0} \mathbb{P}_{X|Y}\left(X > 100\,00 \mid Y \in (70 - h_n; 70 + h_n)\right)$$

- На практиці нас у першу чергу цікавлять формули обчислення відповідних імовірностей

Умовні розподіли

Визначення 2.23

- Нехай випадкові величини X та Z мають спільний **дискретний** розподіл зі щільністю з функцією ймовірности $p_{Z,X} \equiv \mathbb{P}_{Z,X}\left((Z = z) \cap (X = x)\right)$
- Тоді **умовною функцією ймовірности** Z **за умови** X (conditional probability mass function of Z given X) є

$$\mathbb{P}_{Z|X}\left(Z = z \mid X = x\right) = \frac{\mathbb{P}_{Z,X}\left((Z = z) \cap (X = x)\right)}{\mathbb{P}_X\left(X = x\right)} \tag{2.24}$$

- Якщо $\mathbb{P}_X\left(X = x\right) = 0$ для деякого x, умовну ймовірність **не визначено**

$\square$

Визначення 2.24

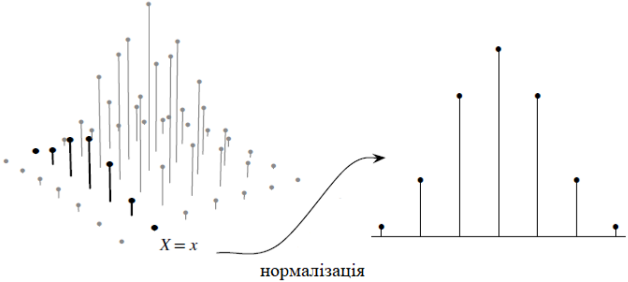
- Нехай випадкові величини X та Z мають спільний **неперервний** розподіл зі щільністю $f_{Z,X}$
- Тоді **умовною щільністю розподілу** Z **за умови** X (conditional probability density of Z given X) є

$$f_{Z|X}(z \mid x) = \begin{cases} \dfrac{f_{Z,X}(z,x)}{f_X(x)}\,, & f_X(x) \neq 0 \\ \xi(z)\,, & f_X(x) = 0 \end{cases} \tag{2.25}$$

- Тут ξ є деякою абсолютно довільною щільністю розподілу

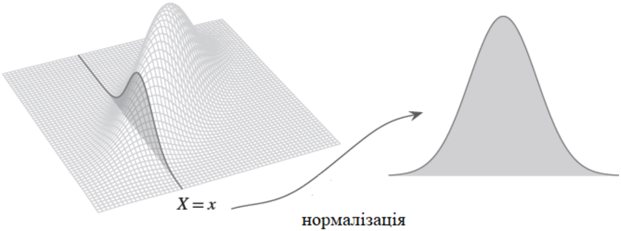
Інтуїтивна інтерпретація умовної функції ймовірности

- Ми «вирізаємо» зі спільної функції ймовірности частину, яка відповідає $X = x$
- Ми додатково **ділимо** на $\mathbb{P}_X(X = x)$ для нормалізації новоутвореної (умовної) функції ймовірности, щоб сума її значень дорівнювала 1



Інтуїтивна інтерпретація умовної щільности розподілу

- Ми «вирізаємо» зі спільної щільности розподілу частину, яка відповідає $X = x$
- Потім ми її ділимо на $f_X(x)$ для нормалізації новоутвореної (умовної) щільности розподілу, щоб її інтеграл дорівнював 1



- Для умовної щільности також, із цілком очевидних міркувань, справедлива теорема Беєса:

$$f_{Z|X}(z \mid x) = \frac{f_{X|Z}(x \mid z) f_Z(z)}{f_X(x)} \,, \quad f_X(x) > 0 \tag{2.26}$$

Незалежність подій

Визначення 2.25

- Події A і B називають **незалежними** (independent)($A \perp\!\!\!\perp B$), якщо

$$\mathbb{P}(A \cap B) = \mathbb{P}(A)\,\mathbb{P}(B) \tag{2.27}$$

- Якщо $\mathbb{P}(A) > 0$ і $\mathbb{P}(B) > 0$, то маємо $\mathbb{P}(A \mid B) = \mathbb{P}(A)$, $\mathbb{P}(B \mid A) = \mathbb{P}(B)$

□

- Можна довести, що якщо $A \perp\!\!\!\perp B$, то $A \perp\!\!\!\perp B^c$, $A^c \perp\!\!\!\perp B$, $A^c \perp\!\!\!\perp B^c$

Визначення 2.26

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

□

Визначення 2.27

Події A і B **умовно незалежні** (conditionally independent) за умови настання події C ($A \perp\!\!\!\perp B \mid C$), якщо $\mathbb{P}(A \cap B \mid C) = \mathbb{P}(A \mid C)\,\mathbb{P}(B \mid C)$ □

Визначення 2.28

- Випадкові вектори $\mathbf{X}_1 : (\Omega, \mathcal{A}, \mathbb{P}) \to \mathbb{R}^{d_1}$, ..., $\mathbf{X}_k : (\Omega, \mathcal{A}, \mathbb{P}) \to \mathbb{R}^{d_k}$ **незалежні** (independent), якщо

$$\mathbb{P}_{\mathbf{X}}\left(\mathbf{X}_1 \in H_1, \dots, \mathbf{X}_k \in H_k\right) = \mathbb{P}_{\mathbf{X}_1}\left(\mathbf{X}_1 \in H_1\right) \cdot \dots \cdot \mathbb{P}_{\mathbf{X}_k}\left(\mathbf{X}_k \in H_k\right)\ ,\ H_i \in \mathcal{B}^{d_i}\ , \tag{2.28}$$

□

- Можна довести, що $X_1, \dots, X_k$ незалежні **тоді й тільки тоді**, коли
  - $F_{\mathbf{X}}(x_1, \dots, x_k) = F_{X_1}(x_1) \cdot \dots \cdot F_{X_k}(x_k)\ , \quad x_i \in (-\infty; \infty) \cup \{\infty\}$
  - Якщо існують щільності, то $f_{\mathbf{X}}(x_1, \dots, x_k) = f_1(x_1) \cdot \dots \cdot f_k(x_k)\ , \qquad x_i \in \mathbb{R}$
  - Зокрема, для дискретних величин $\mathbb{P}_{\mathbf{X}}\left(X_1 = x_1, \dots, X_k = x_k\right) = \mathbb{P}_{X_1}\left(X_1 = x_1\right) \cdot \dots \cdot \mathbb{P}_{X_k}\left(X_k = x_k\right)$
- Також можна довести, що якщо незалежні випадкові величини, то незалежні й **функції** від них

Визначення 2.29

Сподіванням (expectation) будь-якої випадкової величини X є її інтеграл:

$$\mathbb{E}\left[X\right] = \int X \, d\mathbb{P} \tag{2.29}$$

- $\mathbb{E}\left[X\right] \in \mathbb{R}$, тобто воно є (сталим) **числом**
- Поняття сподівання подібне до поняття центру мас у фізиці

Обчислення сподівань

- Для **дискретної** величини X з функцією ймовірности p_X сподівання $g(X)$ дорівнює

$$\mathbb{E}\left[g(X)\right] = \int g(x) \cdot p_X(x)\, d\# = \sum_{x \in \mathrm{supp}(X)} g(x) \cdot p_X(x) \tag{2.30}$$

- Для **неперервної** величини X зі щільністю f_X сподівання $g(X)$ дорівнює

$$\mathbb{E}\left[g(X)\right] = \int g(x) \cdot f_X(x)\, d\lambda = \int_{-\infty}^{\infty} g(x) \cdot f_X(x)\, dx \tag{2.31}$$

- Сподівання може бути нескінченним
- Сподівання може взагалі не існувати (наприклад, у розподілу Коші)
- Якщо сподівання скінченне, кажуть, що воно **існує** і що X **інтегровна**
- Для інтегровних X та Y можна довести такі властивості:
  - Якщо $X \geq 0$ майже напевно, то $\mathbb{E}\left[X\right] \geq 0$
  - **Лінійність сподівання**: для всіх $a, b \in \mathbb{R}$ виконується $\mathbb{E}\left[aX + bY\right] = a\mathbb{E}\left[X\right] + b\mathbb{E}\left[Y\right]$
  - **Монотонність**: якщо $X \leq Y$ майже напевно, то $\mathbb{E}\left[X\right] \leq \mathbb{E}\left[Y\right]$
  - Якщо $X = Y$ майже напевно, то $\mathbb{E}\left[X\right] = \mathbb{E}\left[Y\right]$

Дисперсія випадкової величини

Визначення 2.30

- **Дисперсією** (variance) випадкової величини X є

$$\text{Var}\left(X\right) = \mathbb{E}\left[\left(X - \mathbb{E}\left[X\right]\right)^2\right] \tag{2.32}$$

- **Середньоквадратичним відхиленням** (standard deviation) випадкової величини X є

$$\text{sd}\left(X\right) = \sqrt{\text{Var}\left(X\right)} \tag{2.33}$$

□

- Часто $\text{Var}\left(X\right)$ позначають через σ_X^2, і тоді $\text{sd}\left(X\right)$ має позначення σ_X
- Дисперсія показує, наскільки **далеко (в середньому)** стоять від $\mathbb{E}\left[X\right]$ значення величини X
- Тобто **наскільки сильно** значення величини X **розкидано** відносно $\mathbb{E}\left[X\right]$
- Можна показати, що $\text{Var}\left(X\right) = \mathbb{E}\left[X^2\right] - \left(\mathbb{E}\left[X\right]\right)^2$
- Варто звернути увагу, що дисперсія не є лінійною:

$$\text{Var}\left(aX + b\right) = a^2\text{Var}\left(X\right)$$

Нерівність Єнсена

Теорема 2.31 (Нерівність Єнсена (Jensen's inequality)[12])

- Нехай функція $\varphi : \mathbb{R} \to \mathbb{R}$ **опукла**
- Нехай випадкові величини X і $\varphi(X)$ інтегровні
- Тоді

$$\mathbb{E}\left[\varphi(X)\right] \geq \varphi(\mathbb{E}\left[X\right]) \tag{2.34}$$

- Якщо функція **вгнута** (concave), тобто опукла вгору, то $\mathbb{E}\left[\varphi(X)\right] \leq \varphi(\mathbb{E}\left[X\right])$
- $\mathbb{E}\left[\varphi(X)\right] = \varphi(\mathbb{E}\left[X\right])$ тоді й тільки тоді, коли $\varphi(X) = a + bX$ майже напевно
- Конкретні приклади:
  - $\mathbb{E}\left[|X|\right] \geq |\mathbb{E}\left[X\right]|$
  - $\mathbb{E}\left[X^r\right] \geq (\mathbb{E}\left[X\right])^r$ для $r > 1$ для додатних випадкових величин X
  - $\mathbb{E}\left[\frac{1}{X}\right] \geq \frac{1}{\mathbb{E}\left[X\right]}$ для додатних випадкових величин X
  - $\mathbb{E}\left[\ln X\right] \leq \ln(\mathbb{E}\left[X\right])$ для додатних випадкових величин X
  - $\mathbb{E}\left[X^r\right] \leq (\mathbb{E}\left[X\right])^r$ для $0 < r < 1$

[12] Йоган Єнсен (Johan Ludwig William Valdemar Jensen, 1859–1925) — данський математик

Нерівність Маркова

Теорема 2.32 (Нерівність Маркова (Markov's inequality)[13])

- Нехай $g : \mathbb{R} \to [0; \infty)$ — неспадна вимірна функція
- Тоді для **будь-якої** випадкової величини X має місце

$$\mathbb{P}_X\left(X \geq c\right) \leq \frac{\mathbb{E}\left[g(X)\right]}{g(c)}, \qquad c > 0 \tag{2.35}$$

- Класичне формулювання, якщо $X \geq 0$ майже напевно, а $g(x) = x$:

$$\mathbb{P}_X\left(X \geq c\right) \leq \frac{\mathbb{E}\left[X\right]}{c}, \qquad c > 0 \tag{2.36}$$

- Можна прибрати вимогу про невід'ємність величини X, розглянувши $|X|$:

$$\mathbb{P}_X\left(|X| \geq c\right) \leq \frac{\mathbb{E}\left[|X|\right]}{c}, \qquad c > 0$$

- Або, що те ж саме:

$$\mathbb{P}_X\left(|X| \geq c \cdot \mathbb{E}\left[|X|\right]\right) \leq \frac{1}{c}, \qquad c > 0$$

[13] Андрєй Марков (1856–1922) — російський математик

Нерівність Чебишова

- Чи не **найважливішим** наслідком нерівности Маркова є **нерівність Чебишова** (Chebyshev's inequality)[14]:

$$\mathbb{P}_X\left(\left|X - \mathbb{E}\left[X\right]\right| \geq c\right) \leq \frac{\mathsf{Var}\left(X\right)}{c^2}\,, \qquad c > 0 \tag{2.37}$$

- Альтернативна форма запису цієї нерівности:

$$\mathbb{P}_X\left(\left|X - \mathbb{E}\left[X\right]\right| \geq c \cdot \sigma_X\right) \leq \frac{1}{c^2}\,, \qquad c > 0$$

- Випадкова величина **не може** суттєво **відхилятися** від свого сподівання на деяку відстань, що вимірюється в середньоквадратичних відхиляннях
- Якщо порівняти нерівності Маркова й Чебишова, то нерівність Чебишова пропонує **точнішу** оцінку
- Проте вона вимагає існування **скінченної дисперсії** (Марков вимагає тільки існування скінченного **сподівання**)

[14] Пафнутій Чебишов (1821–1894) — російський математик

Незалежність сподівань

Теорема 2.33

- Нехай X і Y — дві незалежні випадкові величини
- Нехай $f, g : \mathbb{R} \to \mathbb{R}$ — вимірні функції такі, що вони невід'ємні чи інтегровні
- Тоді

$$\mathbb{E}\left[f(X)g(Y)\right] = \mathbb{E}\left[f(X)\right]\mathbb{E}\left[g(Y)\right] \tag{2.38}$$

- Частковий випадок: якщо $X_1, \ldots, X_k$ незалежні, невід'ємні чи інтегровні, то

$$\mathbb{E}\left[\prod_{i=1}^{k} X_i\right] = \prod_{i=1}^{k} \mathbb{E}\left[X_i\right] \tag{2.39}$$

Сподівання випадкового вектора

Визначення 2.34

Сподіванням (expectation) випадкового вектора $\mathbf{X} = (X_1, \ldots, X_k)^\top : (\Omega, \mathcal{A}, \mathbb{P}) \to \mathbb{R}^k$ є вектор

$$\mathbb{E}[\mathbf{X}] = \begin{pmatrix} \mathbb{E}[X_1] \\ \vdots \\ \mathbb{E}[X_k] \end{pmatrix} \qquad (2.40)$$

□

- За властивістю лінійности (одновимірного) сподівання:

$$\mathbb{E}[\mathbf{A}\mathbf{X} + \mathbf{b}] = \mathbf{A}\mathbb{E}[\mathbf{X}] + \mathbf{b}\,, \qquad \mathbf{A} \in \mathbb{R}^n \times \mathbb{R}^k\,, \mathbf{b} \in \mathbb{R}^n \qquad (2.41)$$

Коваріація

Визначення 2.35

Коваріацією (covariance) двох випадкових величин X і Y є

$$\text{Cov}\left(X, Y\right) = \mathbb{E}\left[(X - \mathbb{E}\left[X\right])(Y - \mathbb{E}\left[Y\right])\right] \tag{2.42}$$

□

- За аналогією з позначенням дисперсії σ_X^2, коваріацію часто позначають як σ_{XY}
- Можна довести такі властивості коваріацій:
  - $\text{Cov}\left(X, Y\right) = \text{Cov}\left(Y, X\right)$
  - $\text{Cov}\left(X, X\right) = \text{Var}\left(X\right)$
  - $\text{Cov}\left(X, Y\right) = \mathbb{E}\left[XY\right] - \mathbb{E}\left[X\right]\mathbb{E}\left[Y\right]$
  - $\text{Cov}\left(X, a\right) = 0, a \in \mathbb{R}$
  - $\text{Cov}\left(aX + bY, cV + dW\right) =$
    $ac\text{Cov}\left(X, V\right) + ad\text{Cov}\left(X, W\right) + bc\text{Cov}\left(Y, V\right) + bd\text{Cov}\left(Y, W\right)$
  - $\text{Var}\left(X + Y\right) = \text{Var}\left(X\right) + \text{Var}\left(Y\right) + 2\text{Cov}\left(X, Y\right)$
  - $\text{Var}\left(\sum_{i=1}^{k} X_i\right) = \sum_{i=1}^{k} \text{Var}\left(X_i\right) + 2\sum_{i<j} \text{Cov}\left(X_i, X_j\right)$
- Можна довести, що якщо $X \perp\!\!\!\perp Y$, то $\text{Cov}\left(X, Y\right) = 0$
- Такі випадкові величини називають **некорельованими** (uncorrelated)
- При цьому дуже важливо пам'ятати, що **зворотне твердження в загальному випадку несправедливе**

Кореляція

Визначення 2.36

Коефіцієнтом кореляції (correlation) X та Y є

$$\mathrm{Corr}\left(X, Y\right) = \frac{\mathrm{Cov}\left(X, Y\right)}{\sqrt{\mathrm{Var}\left(X\right)\mathrm{Var}\left(Y\right)}} \tag{2.43}$$

- Коефіцієнт кореляції часто позначають як ρ_{XY}
- Можна показати, що $\rho_{XY} \in [-1; 1]$
- Коваріація $\mathrm{Cov}\left(X, Y\right)$ показує ступінь **лінійного** зв'язку між X та Y
  - Якщо $\mathrm{Cov}\left(X, Y\right) > 0$, то між величинами існує **додатний** лінійний зв'язок: **більші** значення однієї величини свідчать про **більші** значення іншої
  - Якщо $\mathrm{Cov}\left(X, Y\right) < 0$, то між величинами існує **від'ємний** лінійний зв'язок: **більші** значення однієї величини свідчать про **менші** значення іншої
  - Якщо $\mathrm{Cov}\left(X, Y\right) = 0$, то **лінійного** зв'язку між величинами **немає**
- Аналогічно для $\mathrm{Corr}\left(X, Y\right)$: що ближчі значення $\mathrm{Corr}\left(X, Y\right)$ за модулем до 1, то сильніша лінійна залежність

Матриця коваріацій

Визначення 2.37

- Нехай $\mathbf{X} = (X_1, \ldots, X_k)^\top : (\Omega, \mathcal{A}, \mathbb{P}) \to \mathbb{R}^k$ — деякий випадковий вектор
- Тоді його **матрицею коваріацій** (covariance matrix) є матриця

$$\begin{aligned}\mathbf{Cov}\left(\mathbf{X}\right) &= \mathbb{E}\left[(\mathbf{X} - \mathbb{E}\left[\mathbf{X}\right])(\mathbf{X} - \mathbb{E}\left[\mathbf{X}\right])^\top\right] \\ &= \begin{pmatrix} \mathrm{Var}\left(X_1\right) & \mathrm{Cov}\left(X_1, X_2\right) & \ldots & \mathrm{Cov}\left(X_1, X_k\right) \\ \mathrm{Cov}\left(X_2, X_1\right) & \mathrm{Var}\left(X_2\right) & \ldots & \mathrm{Cov}\left(X_2, X_k\right) \\ \vdots & \vdots & \ddots & \vdots \\ \mathrm{Cov}\left(X_k, X_1\right) & \mathrm{Cov}\left(X_k, X_2\right) & \ldots & \mathrm{Var}\left(X_k\right) \end{pmatrix}\end{aligned} \tag{2.44}$$

□

- Можна показати, що матриця коваріацій має такі властивості:
  - **Симетричність**: $\Sigma^\top = \Sigma$
  - $\mathbf{Cov}\left(\mathbf{AX} + \mathbf{b}\right) = \mathbf{A}\Sigma\mathbf{A}^\top$
  - **Невід'ємна визначеність**: $\mathbf{a}^\top\Sigma\mathbf{a} \geq 0\ , \qquad \mathbf{a} \in \mathbb{R}^k$

Умовне сподівання

- Нехай випадкові величини X та Z мають неперервний спільний розподіл зі щільністю $f_{Z,X}$
- Нехай функція $g(Z)$ інтегровна
- Тоді **умовне сподівання** (conditional expectation) $g(Z)$ за умови X дорівнює

$$\mathbb{E}\left[g(Z) \mid X = x\right] = \int_{-\infty}^{\infty} g(z) \cdot f_{Z|X}(z \mid x)\, dz \tag{2.45}$$

- Варто звернути увагу, що умовне сподівання є **функцією від** z, а не сталою
- Тобто умовне сподівання є **випадковою величиною**!
- Для дискретної величини маємо

$$\mathbb{E}\left[g(Z) \mid X = x\right] = \sum_{z \in \operatorname{supp}(Z)} g(z) \mathbb{P}\left(Z = z \mid X = x\right)$$

- Нехай події $B_1, B_2, \ldots$ утворюють не більш ніж зліченне розбиття Ω
- Тоді умовне сподівання

$$\mathbb{E}\left[X \mid B_i\right] = \frac{\displaystyle\int_{B_i} X \, d\mathbb{P}}{\mathbb{P}\left(B_i\right)} \tag{2.46}$$

- Якщо $\mathbb{P}\left(B_i\right) = 0$, то (стале) значення $\mathbb{E}\left[X \mid B_i\right]$ може бути довільне

Теорема 2.38 (Закон повного сподівання (Law of total expectation))

- Нехай події $B_1, B_2, \dots$ утворюють не більш ніж злічене розбиття Ω
- Тоді **безумовне** сподівання випадкової величини X дорівнює

$$\mathbb{E}[X] = \sum_i \mathbb{E}[X \mid B_i] \mathbb{P}(B_i) \tag{2.47}$$

Закон ітерованих сподівань

- Можна показати, що якщо Y і XY інтегровні, то майже напевно

$$\mathbb{E}\left[XY \mid X\right] = X\mathbb{E}\left[Y \mid X\right] \tag{2.48}$$

Теорема 2.39 (Закон ітерованих сподівань (Law of iterated expectations))

- Нехай X інтегровна
- Тоді

$$\mathbb{E}\left[X\right] = \mathbb{E}\left[\mathbb{E}\left[X \mid Y\right]\right] \tag{2.49}$$

- Можна помітити, що $\mathbb{P}_X\left(X \in A\right) = \mathbb{E}\left[\mathbb{1}\left\{X \in A\right\}\right]$
- Відтак справедливий закон повної ймовірности в неперервному випадку:

$$\mathbb{P}_X\left(X \in A\right) = \int_{-\infty}^{\infty} \mathbb{P}_{X|Y}\left(X \in A \mid Y = y\right) \cdot f_Y(y)\, dy \tag{2.50}$$

Умовна дисперсія

- За аналогією з умовними сподіванням також можна визначити умовні моменти довільного порядку
- Зокрема, **умовною дисперсію** (conditional variance) величини Z за умови величини X є

$$\operatorname{Var}(Z \mid X) = \mathbb{E}\left[(Z - \mathbb{E}[Z \mid X])^2 \mid X\right] \tag{2.51}$$

Теорема 2.40 (Закон повної дисперсії (Law of total variance))

- Нехай X та Z — дві випадкові величини
- Тоді дисперсію $\operatorname{Var}(X)$, якщо вона існує, можна розписати так:

$$\operatorname{Var}(X) = \mathbb{E}[\operatorname{Var}(X \mid Z)] + \operatorname{Var}(\mathbb{E}[X \mid Z]) \tag{2.52}$$

План лекції

Загальні міркування

- Ми згадаємо деякі найважливіші розподіли, які часто виринають на практиці
- В R усі ці розподіли реалізовані в пакеті `stats`
- Для кожного розподілу існує 4 функції, які мають однакову назву, але різні префікси:
  - Функція з префіксом `p` (від probability) реалізує функцію розподілу
  - Функція з префіксом `d` (від density) реалізує щільність розподілу / функцію ймовірности
  - Функція з префіксом `q` (від quantile) реалізує функцію квантилів
  - Функція з префіксом `r` (від random) реалізує генератор псевдовипадкових чисел згідно з розподілом

Розподіл Бернуллі

- Чи не найпростішим розподілом є **розподіл Бернуллі** (Bernoulli distritubion)[15]
- Це дискретний розподіл із функцією ймовірности

$$p_X(x) = p^x (1-p)^{1-x} \cdot \mathbb{1}\left\{x \in \{0, 1\}\right\} \tag{3.1}$$

- Тобто X може набувати тільки значень 1 або 0 з імовірностями p або $1-p$ відповідно
- Це позначають через $X \sim \text{Bern}\left(p\right)$
- Можна порахувати, що $\mathbb{E}\left[X\right] = p$, а $\text{Var}\left(X\right) = p(1-p)$

[15] Яків Бернуллі (Jacob Bernoulli, 1654–1705) — швейцарський математик

Біномний розподіл

- Нехай маємо $X_1, \dots, X_n \overset{\text{i.i.d.}}{\sim} \text{Bern}\,(p)$
  - Фактично маємо так звану схему Бернуллі
  - Тут i.i.d. означає **independent and identically distributed** (незалежні й однаково розподілені)
- Тоді $X = \sum_{i=1}^{n} X_i$ має **біномний розподіл** із параметрами n і p
- Це позначають через $X \sim \text{Binom}\,(n, p)$
- X відповідає загальному числу «успіхів» (подій, що стаються з імовірністю p) із-посеред n можливих
- Функція ймовірности дорівнює

$$p_X(k) = \binom{n}{k} p^k (1-p)^{n-k} \cdot \mathbb{1}\left\{k \in \{0, 1, \dots, n\}\right\} \tag{3.2}$$

- Очевидно, що якщо $X \sim \text{Bern}\,(p)$ і $Y \sim \text{Binom}\,(1, p)$, то $X \overset{d}{=} Y$
- Можна порахувати, що $\mathbb{E}\left[X\right] = np$, а $\text{Var}\,(X) = np(1-p)$
- Можна показати, що якщо $X_i \sim \text{Binom}\,(n_i, p)$, $i = 1, \dots, k$, і всі незалежні, то $\sum_{i=1}^{k} X_i \sim \text{Binom}\left(\sum_{i=1}^{k} n_i, p\right)$

Біномний розподіл і розподіл Бернуллі в R

- Оскільки розподіл Бернуллі є частковим випадком біномного, в пакеті `stats` існує одна сім'я функцій `binom`
- Наприклад, нехай компанія виробляє гайки пакетами по 10 штук
- Нехай гайки можуть бути з дефектом з імовірністю 0.1 незалежно одна від одної
- Тоді X =«число дефективних гайок у пакеті» має розподіл $X \sim \text{Binom}\left(10, 0.1\right)$
- Наприклад, імовірність того, що в пакеті 2 дефективні гайки дорівнює $\mathbb{P}_X\left(X = 2\right) = \binom{10}{2} 0.1^2 (1 - 0.1)^{10-2}$
- В R це можна порахувати так

```
dbinom(2, size = 10, prob = 0.1)
```

```
## [1] 0.1937102
```

- Компанія поверне кошти, якщо в пакеті буде більше 1 дефективної гайки
- За яку частку пакетів повернуть кошти?
- Відповідь дорівнює $\mathbb{P}_X\left(X > 1\right) = 1 - \mathbb{P}_X\left(X \leq 1\right) = 1 - \mathbb{P}_X\left(X = 0\right) - \mathbb{P}_X\left(X = 1\right)$
- Або

```
1 - dbinom(0, size = 10, prob = 0.1) - dbinom(1, size = 10, prob = 0.1)
```

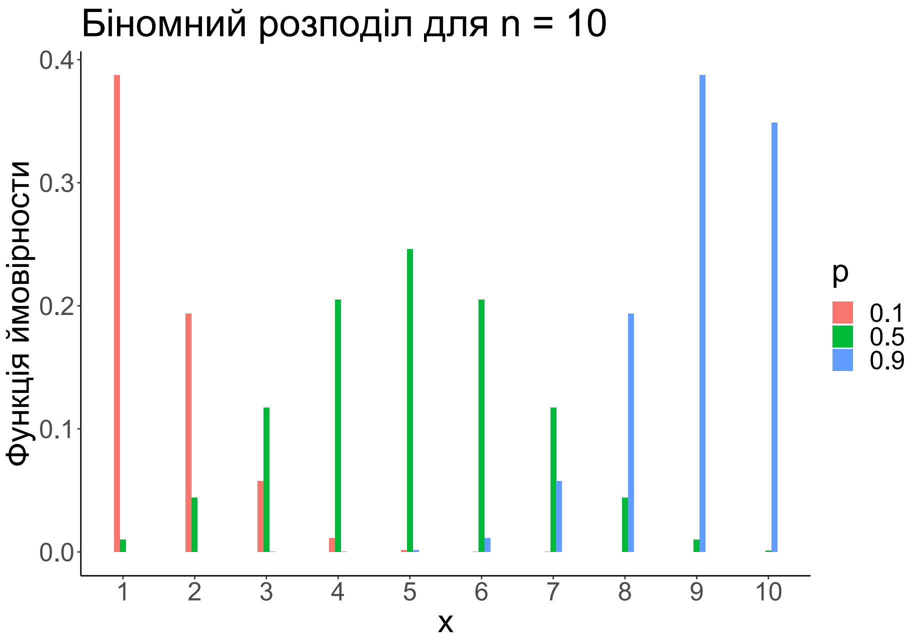
```
## [1] 0.2639011
```

```
1 - pbinom(1, size = 10, prob = 0.1)
```

```
## [1] 0.2639011
```

```
pbinom(1, size = 10, prob = 0.1, lower.tail = FALSE)
```

```
## [1] 0.2639011
```


Біномний розподіл для n = 10
Функція ймовірности
0.0
0.1
0.2
0.3
0.4
1
2
3
4
5
6
7
8
9
10
x
p
0.1
0.5
0.9

Розподіл Пуассона

- Випадкова величина X має **розподіл Пуассона** (Poisson distribution)[16] із параметром $\lambda > 0$, якщо її функція ймовірности має вид

$$p_X(k) = \frac{e^{-\lambda}\lambda^k}{k!} \cdot \mathbb{1}\left\{k \in \mathbb{N} \cup \{0\}\right\} \tag{3.3}$$

- Це позначають через $X \sim \text{Pois}\left(\lambda\right)$
- Розподіл Пуассона використовують для моделювання явищ, які стаються нечасто
  - Кількість електронних листів, отримуваних за годину
  - Кількість шматочків шоколаду в печиві
  - Число землетрусів на рік у деякому регіоні планети
  - Число типографських помилок на окремо взятій сторінці
- Параметр λ можна інтерпретувати як частоту появи деяких подій
- Можна порахувати, що $\mathbb{E}\left[X\right] = \text{Var}\left(X\right) = \lambda$
- Можна показати, що якщо $X_i \sim \text{Pois}\left(\lambda_i\right)$, $i = 1, \ldots, k$, і всі незалежні, то $\sum_{i=1}^k X_i \sim \text{Pois}\left(\sum_{i=1}^k \lambda_i\right)$

[16] Симеон Дені Пуассон (Sim'eon Denis Poisson, 1781–1840) — французький математик

Розподіл Пуассона в R

- У пакеті `stats` для розподілу Пуассона існує сім'я функцій `pois`
- Наприклад, нехай число типографських помилок на окремо взятій сторінці книжки X має розподіл Пуассона $\text{Pois}\left(0.5\right)$
- Тобто в середньому 1 помилка на 2 сторінки
- Тоді ймовірність того, що на сторінці буде 2 помилки, дорівнює $\mathbb{P}_X\left(X=2\right)=\dfrac{e^{-0.5}0.5^2}{2!}$
- В R це можна порахувати так

```
dpois(2, lambda = 0.5)
```

```
## [1] 0.07581633
```

- Чому дорівнює ймовірність **хоча б однієї** помилки на сторінці?
- Відповідь дорівнює $\mathbb{P}_X\left(X\geq 1\right)=1-\mathbb{P}_X\left(X<1\right)=1-\mathbb{P}_X\left(X=0\right)$
- Або

```
1 - dpois(0, lambda = 0.5)
```

```
## [1] 0.3934693
```

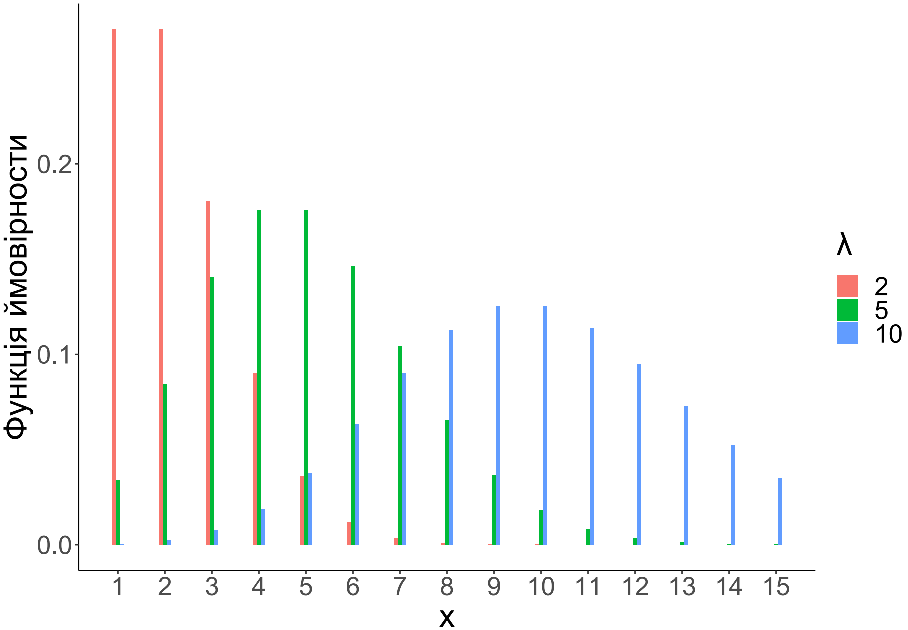
```
1 - ppois(0, lambda = 0.5) # X <= 0 <=> X < 1
```

```
## [1] 0.3934693
```

```
ppois(0, lambda = 0.5, lower.tail = FALSE)
```

```
## [1] 0.3934693
```

Функції ймовірности розподілу Пуассона



Зв'язок між розподілом Пуассона і біномним розподілом

- Можна довести таке твердження
- Нехай $X \sim \text{Binom}\,(n, p)$, $n \to \infty$, $p_n \to 0$, але $np_n \to \lambda$
- Тоді функція ймовірности X прямує до функції ймовірности $Y \sim \text{Pois}\,(\lambda)$
- Приклади застосування:
  - Число осіб у громаді, які доживуть до 100 років
  - Число неправильно набраних телефонних номерів на день
  - Число пакетів собачого корму, проданих у деякому магазині за день
  - Число клієнтів поштового відділку на день
  - Число вакансій, що відкриваються за рік у судовій системі
  - Число α-частинок, випромінюваних протягом фіксованого періоду часу
- Нехай $X =$«число відвідувачів вебсайту на день»
- Щодня відвідати сайт незалежно одне від одного вирішує 100 користувачів, з імовірністю відвідання $p = 0.02$
- Строго формально $X \sim \text{Binom}\,(100, p)$
- Проте наближено $X \sim \text{Pois}\,(100 \cdot 0.02) = \text{Pois}\,(2)$
- Справді, ймовірності, що сайт відвідають щонайменше 3 користувачі, такі:

```
pbinom(2, size = 10^2, prob = 2*10^{-2}, lower.tail = FALSE) # X >= 3 <=> X > 2
```

```
## [1] 0.3233144
```

```
ppois(2, lambda = 2, lower.tail = FALSE)
```

```
## [1] 0.3233236
```

Рівномірний розподіл

- Розгляньмо тепер неперервні випадкові величини
- Величина X має **рівномірний розподіл** (uniform distribution) на проміжку $(a; b)$, якщо її щільність дорівнює

$$f_X(x) = \frac{1}{b-a} \cdot \mathbb{1}\left\{x \in (a; b)\right\} \tag{3.4}$$

- Це позначають через $X \sim \mathsf{U}\left((a; b)\right)$
- Функція рівномірного розподілу має такий вид:

$$F_X(x) = \int_{-\infty}^{x} \mathbb{1}\left\{t \in (a; b]\right\} dt = \begin{cases} 0\,, & x < a \\ \dfrac{x-a}{b-a}\,, & a \le x < b \\ 1\,, & x \ge b \end{cases} \tag{3.5}$$

- Можна показати, що якщо $U \sim \mathsf{U}\left((0; 1]\right)$, то $X = a + (b-a)U$
- Можна порахувати, що $\mathbb{E}\left[X\right] = \dfrac{a+b}{2}$, а $\mathrm{Var}\left(X\right) = \dfrac{(b-a)^2}{12}$

Рівномірний розподіл в R

- У пакеті `stats` для розподілу Пуассона існує сім'я функцій `unif`
- Наприклад, нехай між аеропортом та містом що 15 хвилин курсує шатл
- Нехай час очікування шатла X має рівномірний розподіл $X \sim \mathtt{U}\left((0;15)\right)$
- Чому дорівнює ймовірність, що пасажиру доведеться ждати шатл понад 6 хвилин?
- Відповідь дорівнює $\mathbb{P}_X\left(X \geq 6\right) = 1 - F_X(6) = 1 - \dfrac{6-0}{15-0}$
- Або

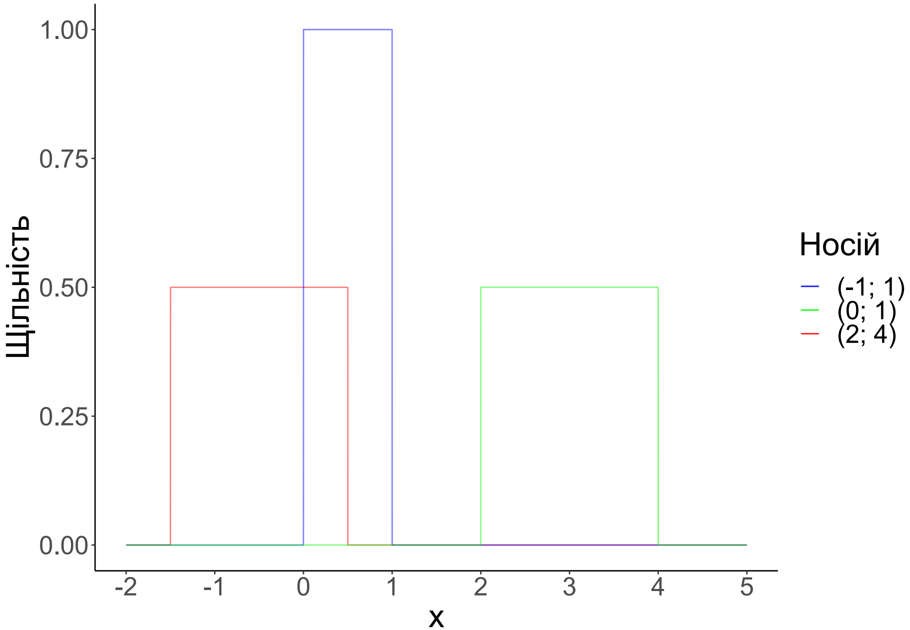
```
1 - punif(6, min = 0, max = 15)
```

```
## [1] 0.6
```

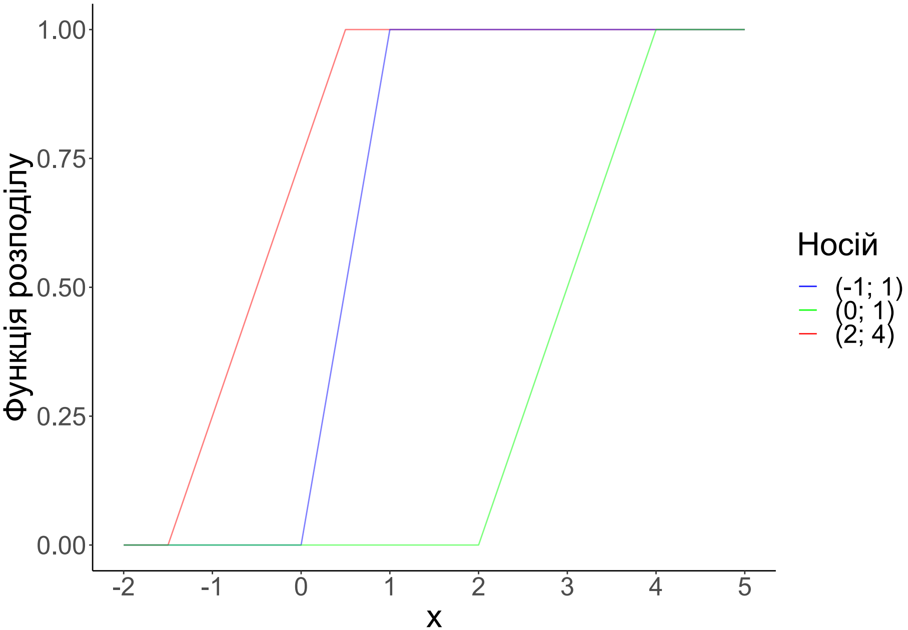
```
punif(6, min = 0, max = 15, lower.tail = FALSE)
```

```
## [1] 0.6
```

Щільності рівномірного розподілу



Функції рівномірного розподілу



Експоненційний розподіл

- Величина X має **експоненційний розподіл** (exponential distribution) із параметром $\lambda > 0$, якщо її щільність дорівнює

$$f_X(x) = \lambda e^{-\lambda x} \cdot \mathbb{1}\{x > 0\} \tag{3.6}$$

- Це позначають через $X \sim \text{Exp}(\lambda)$
- Функція експоненційного розподілу має такий вид:

$$F_X(x) = \int_{-\infty}^{x} \lambda e^{-\lambda x} \cdot \mathbb{1}\{x > 0\}\, dt = \begin{cases} 0\,, & x < 0 \\ 1 - e^{-\lambda x}\,, & x \geq 0 \end{cases} \tag{3.7}$$

- Інтерпретація X — **час очікування** настання нової події у потоці подій з інтенсивністю λ
- Можна порахувати, що $\mathbb{E}[X] = \frac{1}{\lambda}$, а $\text{Var}(X) = \frac{1}{\lambda^2}$
- Експоненційний розподіл — **єдиний неперервний**, який не має пам'яти:

$$\mathbb{P}_X(X \geq s + t \mid X \geq s) = \mathbb{P}_X(X \geq t)\,, \quad s, t \geq 0$$

Експоненційний розподіл в R

- У пакеті `stats` для розподілу Пуассона існує сім'я функцій `exp`
- Наприклад, нехай тривалість X телефонної розмови (у хвилинах) має експоненційний розподіл із параметром $\lambda = 0.1$
- Тобто в послідовності телефонних дзвінків, де наступний починається відразу після завершення попереднього, завершення розмови стається (в середньому) 0.1 разів на хвилину, або 1 раз на 10 хвилин
- Тоді ймовірність того, що щойно розпочата розмова протриває понад 10 хвилин, дорівнює

$$\mathbb{P}_X\left(X > 10\right) = 1 - F(10) = 1 - \left(1 - e^{-0.1\cdot 10}\right) = e^{-1}$$

- Або

```
1 - pexp(10, rate = 0.1)
```

```
## [1] 0.3678794
```

```
pexp(10, rate = 0.1, lower.tail = FALSE)
```

```
## [1] 0.3678794
```

- А ймовірність того, що тривалість буде від 10 до 20 хвилин, дорівнює

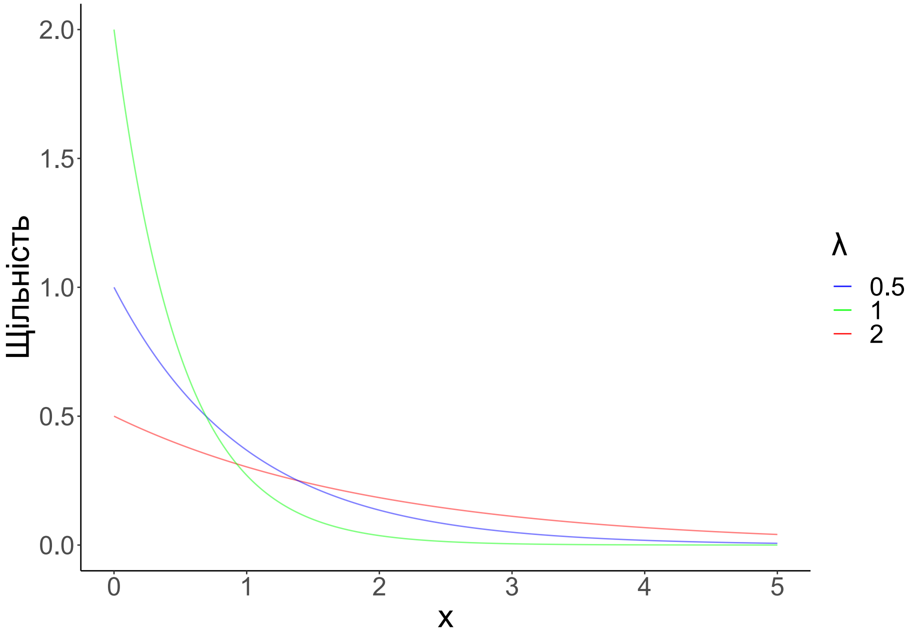
$$\mathbb{P}_X\left(10 < X < 20\right) = F(20) - F(10) = e^{-1} - e^{-2}$$

- Або

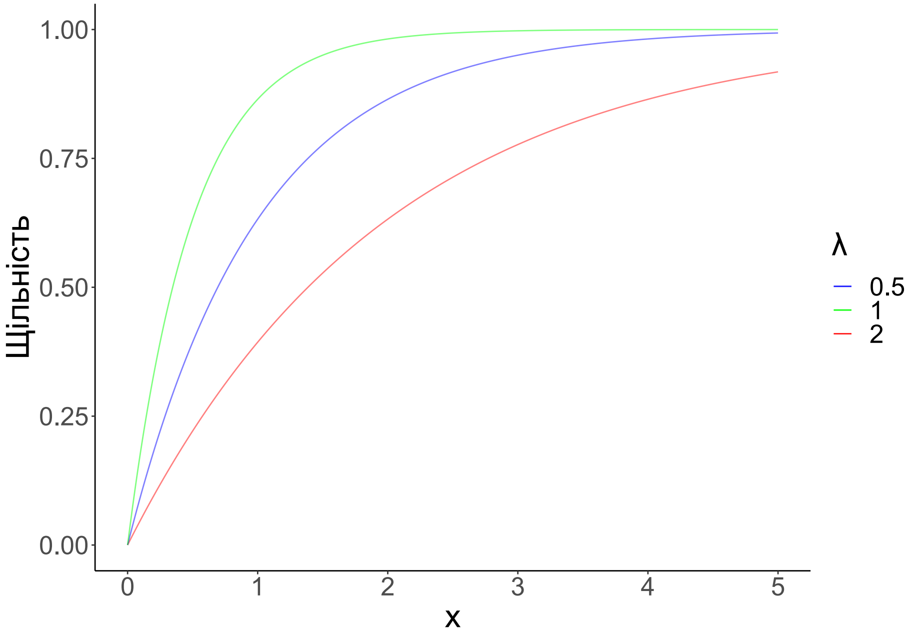
```
pexp(20, rate = 0.1) - pexp(10, rate = 0.1)
```

```
## [1] 0.2325442
```

Щільності експоненційного розподілу



Функції експоненційного розподілу



Стандартний нормальний розподіл

- Без перебільшення, **найбільш важливим розподілом у теорії ймовірностей та статистиці** є **нормальний розподіл** (normal distribution)
- Випадкова величина Z має **стандартний** (standard) нормальний розподіл, якщо її щільність розподілу має вид

$$\phi(z) = \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}} \tag{3.8}$$

- Це позначають через $Z \sim N(0, 1)$
- Корисні властивості ϕ:
  - Вона є парною функцією: $\phi(-z) = \phi(z)$
  - $\phi'(z) = -z\phi(z)$
- Функція стандартного нормального розподілу, $\Phi(z) = \int_{-\infty}^{z} \phi(t)\, dt$, **не має аналітичного виразу**
  - Її рахують чисельно
- Корисні властивості Φ:
  - $\Phi(z) = 1 - \Phi(-z)$
  - $\Phi^{-1}(p) = -\Phi^{-1}(1-p), p \in (0; 1)$
- Можна порахувати, що $\mathbb{E}[X] = 0$, а $\operatorname{Var}(X) = 1$
  - Звідси позначення $N(0, 1)$

Нормальний розподіл

- Випадкова величина X має **нормальний розподіл** (normal distribution) зі сподіванням μ та дисперсією σ^2 ($\sigma > 0$), якщо $X = \mu + \sigma Z$, де $Z \sim N(0, 1)$
- Це позначають через $X \sim N(\mu, \sigma^2)$
- Отже завжди можна перейти від X до Z, віднявши μ і поділивши на σ
  - Тоді сподівання буде 0, а дисперсія — 1
  - Це справедливо також для інших величин
  - У цьому контексті Z називають Z**-оцінкою** (Z-score)
- Існує прямий зв'язок між щільностями й функціями розподілу $N(0, 1)$ і $N(\mu, \sigma^2)$:
  - $F(x) = \Phi\left(\frac{x-\mu}{\sigma}\right)$
  - $f(x) = \frac{1}{\sigma} \cdot \phi\left(\frac{x-\mu}{\sigma}\right) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$
- Можна показати, що $aX + b \sim N(a\mu + b, a^2\sigma^2)$
- Можна порахувати, що $\mathbb{E}\left[X\right] = \mu$, а $\text{Var}\left(X\right) = \sigma^2$
  - Звідси позначення $N(\mu, \sigma^2)$
- Можна показати, що якщо $X_i \sim N(\mu_i, \sigma_i^2)$, $i = 1, \ldots, k$, і всі незалежні, то $\sum_{i=1}^k X_i \sim N\left(\sum_{i=1}^k \mu_i, \sum_{i=1}^k \sigma_i^2\right)$

Нормальний розподіл в R

- У пакеті `stats` для нормального розподілу існує сім'я функцій `norm`
- Наприклад, нехай розподіл IQ X серед деякої групи осіб є нормальний зі сподіванням $\mu = 105$ та середньоквадратичним відхиленням $\sigma = 20$
  - Тобто $X \sim N(105, 20^2)$
- Тоді IQ щонайбільше 80 у частки осіб

$$\mathbb{P}_X(X \leq 80) = \mathbb{P}_X\left(\frac{X - 105}{20} \leq \frac{80 - 105}{20}\right) = \Phi(-1.25)$$

- Або

```
pnorm(80, mean = 105, sd = 20)
## [1] 0.1056498
pnorm((80 - 105) / 20)
## [1] 0.1056498
```

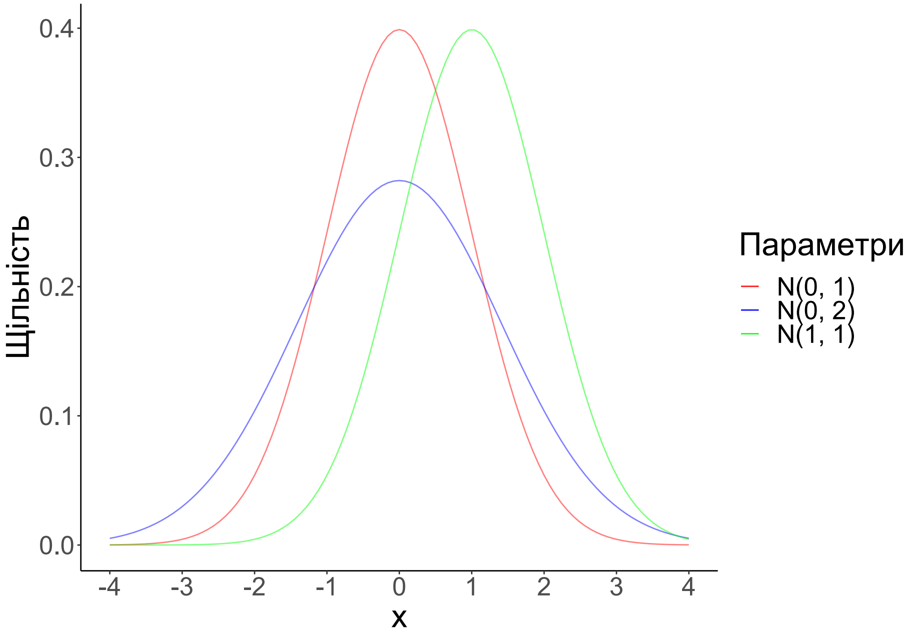
- А, наприклад, IQ лежить між 95 і 125 у частки осіб

$$\mathbb{P}_X(95 \leq X \leq 125) = \mathbb{P}_X\left(\frac{95 - 105}{20} \leq \frac{X - 105}{20} \leq \frac{125 - 105}{20}\right) = \Phi(1) - \Phi(-0.5)$$

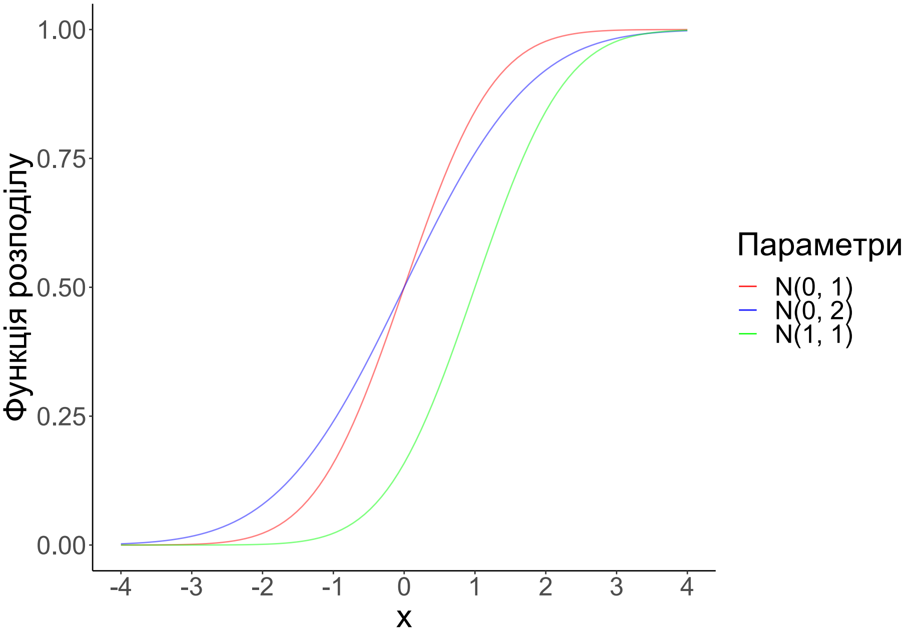
- Або

```
pnorm(125, mean = 105, sd = 20) - pnorm(95, mean = 105, sd = 20)
## [1] 0.5328072
pnorm((125 - 105) / 20) - pnorm((95 - 105) / 20)
## [1] 0.5328072
```

Щільності нормального розподілу



Функції нормального розподілу



Правило трьох сигм (1)

- Надзвичайно корисним на практиці є так зване **правильо трьох сигм** (the three sigma rule)
- Наприклад, якщо $X \sim N(\mu, \sigma^2)$,

$$\mathbb{P}_X\left(|X-\mu| \leq \sigma\right) = \mathbb{P}_X\left(\left|\frac{X-\mu}{\sigma}\right| \leq 1\right) = \mathbb{P}_Z\left(-1 \leq Z \leq 1\right)$$

- Можна обчислити

```
pnorm(1) - pnorm(-1)
```

```
## [1] 0.6826895
```

- Аналогічно для $\mathbb{P}_X\left(|X-\mu| \leq 2\sigma\right)$ і $\mathbb{P}_X\left(|X-\mu| \leq 3\sigma\right)$

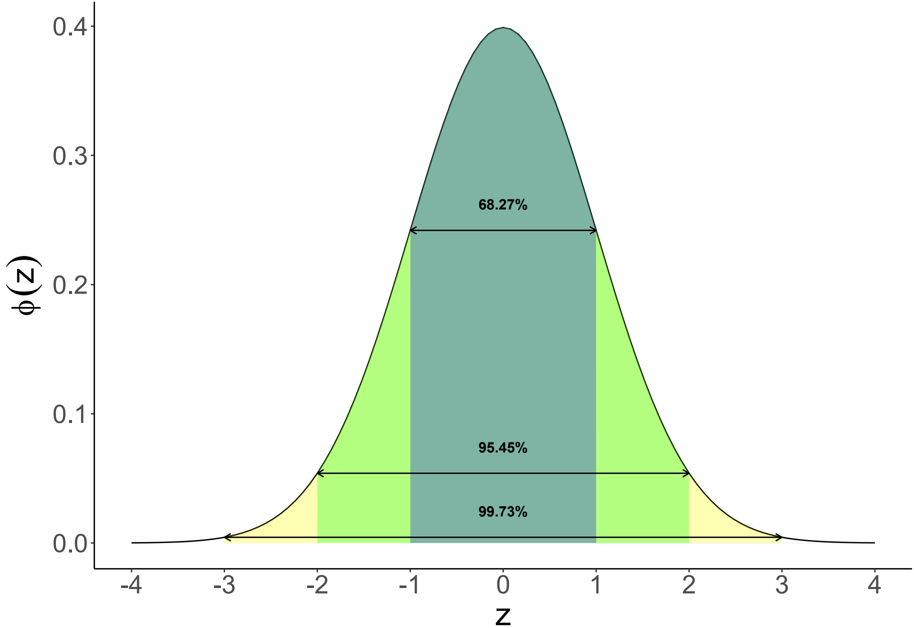
```
pnorm(2) - pnorm(-2)
```

```
## [1] 0.9544997
```

```
pnorm(3) - pnorm(-3)
```

```
## [1] 0.9973002
```

Правило трьох сигм (2)



Багатовимірний нормальний розподіл (1)

- Випадковий вектор $\mathbf{X} = (X_1, \dots, X_k)^\top$ має **багатовимірний нормальний розподіл** (multivariate normal distribution), якщо

$$\mathbf{X} = \mathbf{A} \cdot \mathbf{Z} + \vec{\mu}$$

  - $\Sigma = \mathbf{A}\mathbf{A}^\top$, вона є симетричною і додатно визначеною
  - $\mathbf{A} \in \mathbb{R}^k \times \mathbb{R}^n$ — матриця така, що Σ невироджена
  - $\mathbf{Z} = (Z_1, \dots Z_n)^\top$, $Z_i \sim N(0, 1)$, усі Z_i незалежні
  - $\vec{\mu} = (\mu_1, \dots, \mu_n)^\top \in \mathbb{R}^n$
- Це позначають через $\mathbf{X} \sim N(\vec{\mu}, \Sigma)$
- **Альтернативне визначення**: $\mathbf{X} \sim N(\vec{\mu}, \Sigma)$ якщо будь-яка лінійна комбінація $Y = \mathbf{c}^\top \mathbf{X}$ його координат, $\mathbf{c} \in \mathbb{R}^k$, має нормальний розподіл
- Зверніть увагу: маржинальними розподілами для багатовимірного нормального є нормальні
  - Але зворотне твердження **у загальному випадку** не виконується
- Спільна щільність $\mathbf{X}$ дорівнює

$$f_{\mathbf{X}}(x_1, \dots, x_k) = \frac{1}{\sqrt{(2\pi)^k |\Sigma|}} e^{-\frac{1}{2}(\mathbf{x}-\vec{\mu})^\top \Sigma^{-1} (\mathbf{x}-\vec{\mu})} \tag{3.9}$$

- Можна порахувати, що $\mathbb{E}\left[\mathbf{X}\right] = \vec{\mu}$, **Cov** $\left(\mathbf{X}\right) = \Sigma$

Багатовимірний нормальний розподіл (2)

- Якщо $\mathbf{X} = (X_1, \ldots, X_k)^\top \sim N(\vec{\mu}, \Sigma)$, то $\mathbf{X}' = (X_{i_1}, \ldots, X_{i_m})^\top \sim N(\vec{\mu}', \Sigma')$, де

$$\vec{\mu}' = \begin{pmatrix} \mu_{i_1} \\ \mu_{i_2} \\ \vdots \\ \mu_{i_m} \end{pmatrix}, \qquad \Sigma' = \begin{pmatrix} \Sigma_{i_1,i_1} & \Sigma_{i_1,i_2} & \ldots & \Sigma_{i_1,i_m} \\ \Sigma_{i_2,i_1} & \Sigma_{i_2,i_2} & \ldots & \Sigma_{i_2,i_m} \\ \vdots & \vdots & \ddots & \vdots \\ \Sigma_{i_m,i_1} & \Sigma_{i_m,i_2} & \ldots & \Sigma_{i_m,i_m} \end{pmatrix}$$

- Наприклад, нехай $(X_1, X_2, X_3)^\top \sim N\left((\mu_1, \mu_2, \mu_3)^\top, \Sigma\right)$
  - Тоді, скажімо, $(X_1, X_3)^\top \sim N\left((\mu_1, \mu_3)^\top, \begin{pmatrix} \Sigma_{11} & \Sigma_{13} \\ \Sigma_{31} & \Sigma_{33} \end{pmatrix}\right)$
- Також для багатовимірного нормального розподілу справедлива **унікальна** властивість
- Нехай $\mathbf{X} \sim N(\vec{\mu}, \Sigma)$ можна подати як конкатенацію $\mathbf{X} = \left(\mathbf{X}_1^\top, \mathbf{X}_2^\top\right)^\top$
- Тоді якщо кореляція між будь-якою координатою $\mathbf{X}_1$ і будь-якою координатою $\mathbf{X}_2$ нульова, то $\mathbf{X}_1 \perp\!\!\!\perp \mathbf{X}_2$

Розподіл χ^2

- На основі нормального можна утворити інші розподіли, винятково важливі в статистиці
- $Z \sim N(0,1)$, а $X = Z^2$
- Тоді X має **розподіл χ^2 з одним ступенем свободи**
- Функцію розподілу дуже просто визначити:

$$\begin{aligned} F_X(x) &= \mathbb{P}_Z\left(Z^2 \le x\right) = \left(\Phi(\sqrt{x}) - \Phi(-\sqrt{x})\right) \cdot \mathbb{1}\left\{x \ge 0\right\} \\ &= \left(2\Phi(\sqrt{x}) - 1\right) \cdot \mathbb{1}\left\{x \ge 0\right\} \end{aligned}$$

- Звідси $f_X(x) = \left((2\Phi(\sqrt{x}) - 1) \cdot \mathbb{1}\left\{x \ge 0\right\}\right)' = 2\phi(\sqrt{x}) \cdot \frac{1}{2\sqrt{x}} \cdot \mathbb{1}\left\{x > 0\right\}$
- У загальному випадку, сума $X = \sum_{i=1}^r Z_i^2$, $Z_i \sim N(0,1)$, $i = 1, \ldots, r$, де всі величини незалежні, має **розподіл χ^2 з r ступенями свободи**
- Це позначають через $X \sim \chi^2_r$
- Її щільність дорівнює[17]

$$f_X(x) = \frac{1}{2^{\frac{r}{2}}\Gamma\left(\frac{r}{2}\right)} x^{\frac{r}{2}-1} e^{-\frac{x}{2}} \cdot \mathbb{1}\left\{x > 0\right\} \tag{3.10}$$

- Можна порахувати, що $\mathbb{E}\left[\chi^2_r\right] = r$, Var $\left(\chi^2_r\right) = 2r$
- У пакеті `stats` для розподілу χ^2 існує сім'я функцій `chisq`

[17] Що можна легко вивести з властивостей Гамма-розподілу (КЛТЙ 10.2)

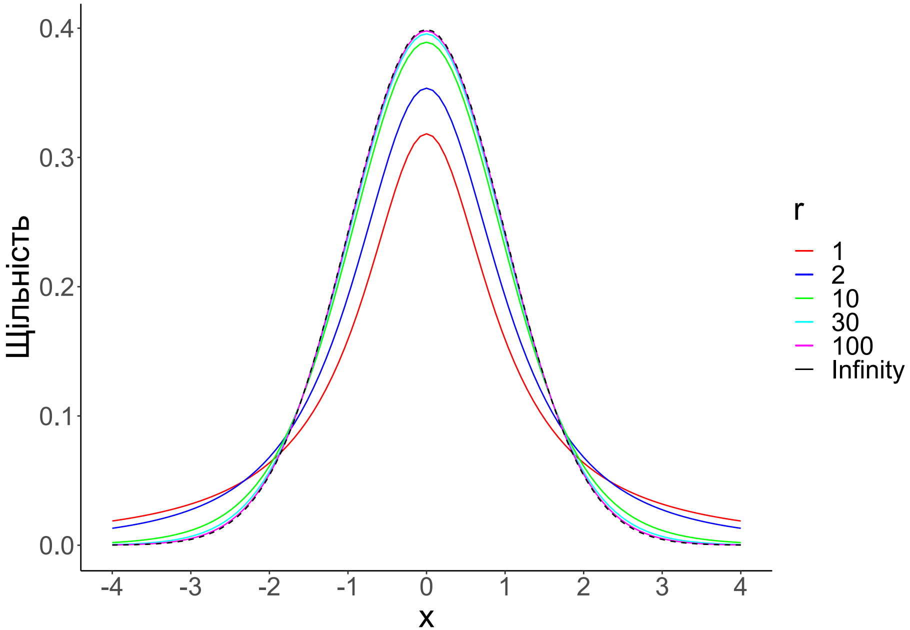
Розподіл t (1)

- На основі розподілу χ^2 можна сформулювати ще два важливі розподіли, які нам знадобляться
- Випадкова величина $T = \frac{Z}{\sqrt{Y/r}}$, де $Z \sim N(0,1)$, $Y \sim \chi^2_r$, має t**-розподіл з** r **ступенями свободи** (він же розподіл Ст'юдента, Student's t-distribution)
- Це позначають через $T \sim t_r$
- Можна показати, що якщо $r = 1$, то T має розподіл Коші
- Щільність розподілу t_r дорівнює

$$f(t) = \frac{\Gamma\left(\frac{r+1}{2}\right)}{\sqrt{r\pi}\Gamma\left(\frac{r}{2}\right)}\left(1 + \frac{t^2}{r}\right)^{-\frac{r+1}{2}} \tag{3.11}$$

- У пакеті `stats` для t-розподілу існує сім'я функцій `t`
- Розподіл t прямує до стандартного нормального розподілу, коли $r \to \infty$, і до того ж дуже швидко

Розподіл t (2)



Розподіл F

- Випадкова величина $F = \frac{X/r_1}{Y/r_2}$, де $X \sim \chi^2_{r_1}$, $Y \sim \chi^2_{r_2}$, має F**-розподіл з** r_1 **і** r_2 **ступенями свободи** (він же розподіл Фішера-Снедекора, Fisher–Snedecor distribution)
- Це позначають через $F \sim F_{r_1, r_2}$
- Щільність розподілу F_{r_1, r_2} дорівнює

$$F(x) = \frac{1}{B\left(\frac{r_1}{2}, \frac{r_2}{2}\right)} \left(\frac{r_1}{r_2}\right)^{r_1/2} x^{\frac{r_1}{2}-1} \left(1 + \frac{r_1}{r_2}x\right)^{-\frac{r_1+r_2}{2}} \tag{3.12}$$

- У пакеті `stats` для F-розподілу існує сім’я функцій `f`