Конспект лекцій з теорії ймовірностей та математичної статистики

Каніовська І.Ю.

Зміст

Частина I

Теорія ймовірностей

Розділ 1

Випадкові події

1.1 Основні поняття теорії ймовірностей

1.1.1 Теоретико-множинний підхід до основних понять ТЙ

Означення 1.1.1. *Стохастичним експериментом (випробуванням)* називається експеримент, який можна повторювати неодноразово, зберігаючи певні умови, і результат якого експерименту заздалегідь передбачити неможливо.

Означення 1.1.2. Будь-який результат СЕ називається *подією.*

Приклад. 1. СЕ — кидання кубика один раз, подія — випало 6 очок.
2. СЕ — кидання кубика двічі, подія — сума очок, що випала, дорівнює 6.

Події бувають:
1. Випадкові — можуть відбутися чи не відбутися при проведенні СЕ.
2. Неможливі — ніколи не відбуваються в даному СЕ.
3. Вірогідні — завжди відбуваються в даному СЕ.

Будемо вважати, що кожному СЕ можна поставити у відповідність деяку множину, що називається *простором елементарних подій* Ω. Під *елементарними подіями* ω будемо розуміти єдині логічно можливі події СЕ, що виключають одна одну.

Приклад. 1. СЕ — кидання кубика один раз. $\Omega = \{\omega_1, \omega_2, ..., \omega_6\}$, де події $\omega_k = \{\text{випало } k \text{ очок}\}, k = 1, ..., 6$.
2. СЕ — кидання монетки до першої появи герба. $\Omega = \{\omega_1, \omega_2, ..., \omega_k, ...\}$, де $\omega_i = \{\text{герб випав на } i\text{-тому киданні}\}, i \in \mathbb{N}$.
3. СЕ — зустріч двох осіб, що домовилися зустрітися протягом години. x — час приходу першої особи, y — час приходу другої, $0 \le x, y \le 1$. $\Omega = \{(x, y) : 0 \le x, y \le 1\} \subset \mathbb{R}^2$.

В подальшому випадкові події позначатимемо $A, B, C, ...$. *Випадкова подія — підмножина* Ω. У прикладі з киданням кубика один раз $A = \{\text{випало 6 очок}\} = \{\omega_6\}$. В загальному випадку маємо $A = \{\omega_{k1}, \omega_{k2}, ..., \omega_{kn}, ...\} \subset \Omega$. Неможлива подія — $\varnothing$, вірогідна — Ω.

Зауваження. Розглядаємо події в межах фіксованих СЕ та простору елементарних подій.

1. *Включення* $A \subset B$ означає, що якщо відбулася подія A, то обов'язково відбудеться подія B. Наприклад, $A = \{\text{витягнуто даму пік}\}, B = \{\text{витягнуто карту чорної масті}\}$. Очевидно, $A \subset B$.
   Рівність подій: $(A \subset B, B \subset A) \iff (A = B)$.
2. *Об'єднання подій* $A \cup B$ — це подія, яка відбувається тоді, коли відбувається принаймні одна з подій A чи B.
   Властивості: $A \cup A = A$, $A \cup B = B \cup A$, $(A \subset B) \Rightarrow (A \cup B = B)$, $A \cup \Omega = \Omega$, $A \cup \varnothing = A$, $(A \cup B) \cup C = A \cup (B \cup C)$.
   Операція узагальнюється на скінченну або зліченну кількість подій: $A = \bigcup\limits_{i=1}^{n(\infty)} A_i$.
3. *Перетин подій* $A \cap B$ — це подія, яка відбувається тоді, коли A і B відбуваються одночасно.

Властивості: $A \cap A = A$, $A \cap B = B \cap A$, $(A \subset B) \Rightarrow (A \cap B = A)$, $A \cap \Omega = A$, $A \cap \varnothing = \varnothing$, $(A \cap B) \cap C = A \cap (B \cap C)$, $A \cap B \subset A$, $A \cap B \subset B$, $(A \cup B) \cap C = (A \cap C) \cup (B \cap C)$, $(A \cap B) \cup C = (A \cup C) \cap (B \cup C)$.

Операція узагальнюється на скінченну або зліченну кількість подій: $A = \bigcap\limits_{i=1}^{n(\infty)} A_i$.

Означення 1.1.3. Події A та B називається *несумісними*, якщо вони не відбуваються одночасно: $A \cap B = \varnothing$. Узагальнення: події $A_1, A_2, ..., A_n, ...$ називаються *попарно несумісними*, якщо $A_i \cap A_j = \varnothing$ для $i \neq j$.

Означення 1.1.4. Події $A_1, A_2, ..., A_n, ...$ утворюють *повну групу подій*, якщо вони попарно несумісні та $\bigcup\limits_{i=1}^{n(\infty)} A_i = \Omega$.

4. *Протилежна подія* $\overline{A}$ — це подія, яка відбувається тоді, коли A не відбувається. Властивості: $A \cup \overline{A} = \Omega$, $A \cap \overline{A} = \varnothing$, $\overline{(A \cup B)} = \overline{A} \cap \overline{B}$, $\overline{(A \cap B)} = \overline{A} \cup \overline{B}$.
5. *Різниця подій* $A \backslash B$ — це подія $A \cap \overline{B}$. Для протилежної події маємо $\overline{A} = \Omega \backslash A$.

Означення 1.1.5. Непорожня система підмножин $\mathcal{F}$ простору елементарних подій Ω утворює *алгебру подій*, якщо:
1. $\Omega \in \mathcal{F}$;
2. $(A, B \in \mathcal{F}) \Rightarrow (A \cup B \in \mathcal{F})$;
3. $(A \in \mathcal{F}) \Rightarrow \left(\overline{A} \in \mathcal{F}\right)$.

Узагальнення: якщо Ω містить зліченну кількість подій, то означення σ-алгебра отримаємо заміною умови 2 на $(\forall n \in \mathbb{N} : A_n \in \mathcal{F}) \Rightarrow \left(\bigcup\limits_{n=1}^{\infty} A_n \in \mathcal{F}\right)$.

Пара $\{\Omega, \mathcal{F}\}$ називається *вимірним простором стохастичного експерименту.*

1.1.2 Класична модель ймовірності

Дослідника завжди цікавить кількісна характеристика появи тієї чи іншої події.

Нехай Ω — скінченний чи зліченний. Поставимо у відповідність кожній елементарній події ω_k число $p_k \geq 0$ так, що $\sum\limits_{k=0}^{n(\infty)} p_k = 1$. Тоді $\mathbb{P}(A) = \sum\limits_{\omega_k \in A} p_k$ — кількісна характеристика, ймовірність події A.

Приклад. *Класична модель ймовірності.* Якщо простір елементарних подій Ω СЕ скінченний та всі ω_k рівноможливі, то такий СЕ називається *класичним.* В такому випадку $p_1 = p_2 = ... = p_n = \frac{1}{n}$, де $n = \text{card}(\Omega)$.

$$\mathbb{P}(A) = \sum_{\omega_k \in A} p_k = \frac{m}{n} = \frac{\text{кількість елементарних подій в } A}{\text{загальна кількість елементарних подій}} \tag{1}$$

Ймовірності, що розраховуються за формулою (1), називаються *класичними.*

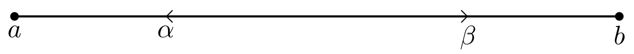
Приклад.
1. «Задачі про вибір» — коли з великої кількості чогось вибирають певну кількість. Наприклад, з урни з 10 кульками, 3 чорними та 7 білими, навмання витягають 5 кульок. Обчислимо ймовірність події $A = \{$серед них 2 чорних кульки$\}$: $\mathbb{P}(A) = \frac{C_3^2 \cdot C_7^3}{C_{10}^5} = \frac{3 \cdot 35}{252} = \frac{5}{12}$.
2. «Задачі про ліфт». 5 осіб одночасно зайшли в ліфт 11-поверхового будинку. Яка ймовірність того, що вони всі вийдуть на різних поверхах, починаючи з другого? $\mathbb{P}(A) = \frac{A_{10}^5}{\widetilde{A}_{10}^5} = \frac{10 \cdot 9 \cdot 8 \cdot 7 \cdot 6}{10^5} = 0.3024$. Тут A_n^k та $\widetilde{A}_n^k$ — кількості розміщень без повторень та з повтореннями відповідно.

Властивості класичної ймовірності:
1. $\forall A \in \mathcal{F} : 0 \leq \mathbb{P}(A) \leq 1$.
2. $\mathbb{P}(\Omega) = 1$.
3. $\mathbb{P}(\overline{A}) = 1 - \mathbb{P}(A)$.
4. $\mathbb{P}(\varnothing) = 0$.
5. $\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B) - \mathbb{P}(A \cap B)$, для несумісних A і B $\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B)$.

6. $(A \subset B) \Rightarrow (\mathbb{P}(A) \le \mathbb{P}(B))$.
7. Якщо $A_1, A_2, ..., A_n$ — повна група подій СЕ, то $\mathbb{P}\left(\bigcup\limits_{i=1}^{n} A_i\right) = 1$.

1.1.3 Геометрична модель ймовірності

Приклад. Нехай точка кидається навмання на відрізок $[a; b]$. Яка ймовірність її потрапляння в $\langle\alpha; \beta\rangle \subset [a; b]$? Розглянемо подію $A = \{$точка потрапила в $\langle\alpha; \beta\rangle\}$.

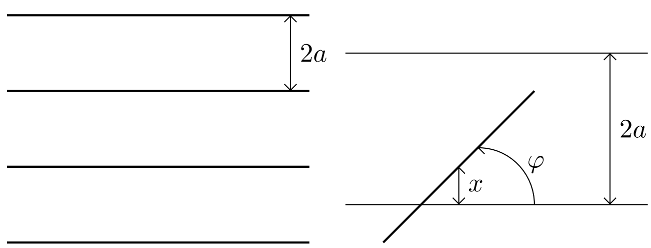


Логічно припустити, що $\mathbb{P}(A) = k \cdot l_{\langle\alpha;\beta\rangle}$ для деякого $k > 0$. Тоді $\mathbb{P}(\Omega) = 1 = k \cdot l_{[a;b]}$. Таким чином, $k = \frac{1}{l_{[a;b]}} = \frac{1}{b-a}$, тому $\mathbb{P}(A) = \frac{l_{\langle\alpha;\beta\rangle}}{l_{[a;b]}}$.

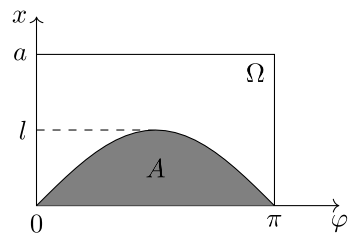
Нехай простір елементарних подій інтерпретується як замкнена область в $\mathbb{R}^n$, а випадкові події — її підмножини. В якості σ-алгебри подій $\mathcal{F}$ беремо підмножини, що мають міру μ. Тоді в якості ймовірності деякої події A беремо $\mathbb{P}(A) = \frac{\mu(A)}{\mu(\Omega)}$. Наприклад, в $\mathbb{R}^1$ беремо міру «довжина», в $\mathbb{R}^2$ — «площа», а в $\mathbb{R}^3$ — «об'єм».

Ймовірності, що знаходяться таким чином називаються *геометричними*, а сама модель називається *геометричною моделлю ймовірності.* Геометрична модель може використовуватись, коли Ω має геометричну інтерпретацію як замкнена область в $\mathbb{R}^n$, а елементарні події — рівноможливі.

Приклад (задача Бюффона)**.** На площині накреслені паралельні прямі на відстані $2a$, на них кидається голка довжиною $2l$, $l \le a$. Яка ймовірність того, що голка перетне яку-небудь пряму?



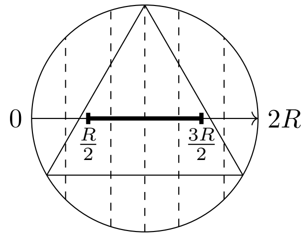
Достатньо розглядати лише дві прямі. $\Omega = \left\{(\varphi, x) \in \mathbb{R}^2 : \varphi \in [0; \pi], x \in [0; a]\right\}$
При такій побудові простору елементарних подій подія $A = \{$голка перетне пряму$\} =$ $\{(\varphi, x) \in \Omega : x \le l \sin\varphi\}$



$\mathbb{P}(A) = \frac{S(A)}{S(\Omega)}$, $S(\Omega) = \pi a$, $S(A) = \int\limits_0^\pi l \sin\varphi \, d\varphi = l\,(-\cos\varphi)|_0^\pi = 2l$. Отже, $\mathbb{P}(A) = \frac{2l}{\pi a}$.
Якщо провести n кидань голки, у m з яких голка потрапить на пряму, то за допомогою приблизної рівності $\frac{m}{n} \approx \frac{2l}{\pi a}$ можна приблизно обчислити число π. Наприклад, якщо взяти $n = 5000, m = 2532, l/a = 0.8$, то отримаємо $\pi \approx 3.1596$.

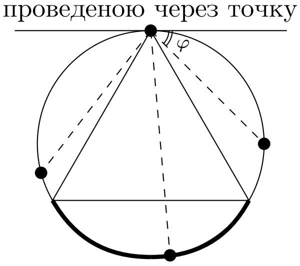
Приклад (парадокс Бертрана)**.** Нехай в колі радіусом R навмання обирається хорда. Яка ймовірність того, що її довжина буде більшою за довжину сторони правильного трикутника, вписаного в це коло?
Існують три способи розв'язання цієї задачі, причому кожен з них дає різний результат.

Спосіб 1. З міркувань симетрії обирається якийсь фіксований діаметр кола і розглядається всі перпендикулярні до нього хорди. Серед них і обирається навмання довільна хорда. Очевидно, що кожна хорда у цьому випадку може бути однозначно визначена своєю серединою, тобто кожній хорді можна взаємно однозначно поставити у відповідність координату її середини, якщо за початок відліку взяти якийсь з кінців фіксованого діаметру.



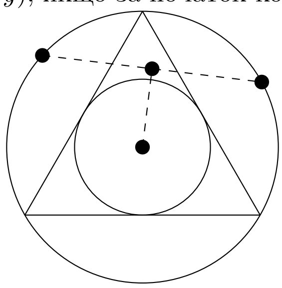
Таким чином, множина всіх значень координати середини хорди $\Omega = [0; 2R]$. Множина, що відповідає події — це відрізок $A = \left[\frac{R}{2}; \frac{3R}{2}\right]$. В якості міри візьмемо довжину. Тоді $\mathbb{P}(A) = \frac{l(A)}{l(\Omega)} = \frac{R}{2R} = \frac{1}{2}$.

Спосіб 2. В цьому способі пропонується розглядати тільки хорди з одним закріпленим кінцем. Кожній хорді поставимо у відповідність ту частину дуги кола, яка потрапляє у кут φ, що утворює хорда з дотичною, проведеною через точку закріплення кінця всіх хорд.



Тоді множина $\Omega = [0; \pi]$, а $A = \left[\frac{\pi}{3}; \frac{2\pi}{3}\right]$. Знову візьмемо за міру довжину і отримаємо $\mathbb{P}(A) = \frac{\pi/3}{\pi} = \frac{1}{3}$.

Спосіб 3. Розглядаються всі хорди кола. Кожній з них взаємно однозначно ставиться у відповідність точка її середини (x, y), якщо за початок координат прийняти центр кола.



В такому випадку $\Omega = \left\{(x, y) \in \mathbb{R}^2 : x^2 + y^2 \leq R^2\right\}$, а $A = \left\{(x, y) \in \Omega : x^2 + y^2 \leq \frac{R^2}{4}\right\}$. Тоді $\mathbb{P}(A) = \frac{\pi R^2/4}{\pi R^2} = \frac{1}{4}$.

Жозеф Бертран був противником геометричного означення ймовірності. Він казав, що воно не дає можливості однозначно обчислити ймовірність однієї і тієї ж випадкової події та використовував вищенаведений приклад для доведення своєї правоти. Дійсно, було отримано три різних відповіді при розв'язанні однієї задачі. Так в чому ж справа? Насправді, помилка полягає у різних трактуваннях поняття «навмання обрана хорда». В кожному з трьох способів це трактування було різним, що й стало причиною різних відповідей.

1.2 Система аксіом Колмогорова

1.2.1 Аксіоми теорії ймовірностей

I. Побудова вимірного простору $\{\Omega, \mathcal{F}\}$:

A1: Будь-якому стохастичному експерименту можна поставити у відповідність простір елементарних подій Ω.

A2: $(\forall n \in \mathbb{N} : A_n \in \mathcal{F}) \Rightarrow \left(\bigcup_{n=1}^{n(\infty)} A_n \in \mathcal{F}\right)$.

A3: $(A \in \mathcal{F}) \Rightarrow \left(\overline{A} \in \mathcal{F}\right)$.

Дві останні аксіоми стосуються побудови алгебри чи σ-алгебри подій.

II. Властивості ймовірності:

P1: $\forall A \in \mathcal{F} : \mathbb{P}(A) \geq 0$.

P2: $\mathbb{P}(\Omega) = 1$ — аксіома нормування.

P3: $\forall A_1, A_2, ..., A_n, ... \in \mathcal{F}, A_i \cap A_j = \varnothing$ при $i \neq j$: $\mathbb{P}\left(\bigcup\limits_{n=1}^{\infty} A_n\right) = \sum\limits_{n=1}^{\infty} \mathbb{P}(A_n)$ — аксіома зліченної адитивності. Іноді у випадках «простої» алгебри (а не σ-алгебри) достатньо аксіоми скінченної адитивності **P3′**: $\forall A_1, A_2, ..., A_n \in \mathcal{F}, A_i \cap A_j = \varnothing$ при $i \neq j$: $\mathbb{P}\left(\bigcup\limits_{k=1}^{n} A_k\right) = \sum\limits_{k=1}^{n} \mathbb{P}(A_k)$.

Означення 1.2.1. Нормована міра $\mathbb{P}$, що введена на вимірному просторі $\{\Omega, \mathcal{F}\}$, називається *ймовірністю.* Всі події беруться з σ-алгебри подій. Трійка $\{\Omega, \mathcal{F}, \mathbb{P}\}$ називається *ймовірнісним простором* стохастичного експерименту.

Ця система аксіом несуперечна, бо існують стохастичні експерименти, які задовольняють цим аксіомам, але неповна, бо в різних задачах теорії ймовірностей розглядаються різні ймовірнісні простори.

1.2.2 Властивості ймовірності, що випливають з аксіом

1. Якщо $A_1, A_2, ..., A_n, ... \in \mathcal{F}$ утворюють повну групу подій, то $\mathbb{P}\left(\bigcup\limits_{k=1}^{\infty} A_k\right) = 1$.

   Доведення. Випливає з аксіоми **P2**. ▲

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

   Доведення. $A \cup \overline{A} = \Omega \Rightarrow 1 \overset{\mathbf{P2}}{=} \mathbb{P}(\Omega) = \mathbb{P}(A \cup \overline{A}) \overset{\mathbf{P3'}}{=} \mathbb{P}(A) + \mathbb{P}(\overline{A})$. ▲

3. $\mathbb{P}(\varnothing) = 0$.

   Доведення. Наслідок з властивості 2 ($\varnothing = \overline{\Omega}$). ▲

   Зауваження. Якщо ймовірність події дорівнює нулю, то це не означає, що подія неможлива.

   Приклад. СЕ — кидання точки на відрізок $[a; b]$. А $= \{$точка потрапила в певну $x \in [a; b]\}$. A не є неможливою, проте $\mathbb{P}(A) = 0$.

4. $A \subset B \Rightarrow \mathbb{P}(A) \leq \mathbb{P}(B)$.

   Доведення. $A \subset B \Rightarrow B = A \cup (B \backslash A) \overset{\mathbf{P3'}}{\Rightarrow} \mathbb{P}(B) = \mathbb{P}(A) + \mathbb{P}(B \backslash A) \overset{\mathbf{P1}}{\geq} \mathbb{P}(A)$. ▲

5. $\forall A \in \mathcal{F} : \mathbb{P}(A) \leq 1$.

   Доведення. Наслідок з властивості 4, $A \subset \Omega$ та **P2**. ▲

6. $A \subset B : \mathbb{P}(B \backslash A) = \mathbb{P}(B) - \mathbb{P}(A)$.

   Доведення. Наслідок з доведення властивості 4. ▲

7. $\forall A, B \in \mathcal{F} : \mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B) - \mathbb{P}(A \cap B)$.

   Доведення. $A \cup B = (A \backslash (A \cap B)) \cup (A \cap B) \cup (B \backslash (A \cap B))$ — попарно несумісні події. З аксіоми **P3′**: $\mathbb{P}(A \cup B) = \mathbb{P}(A \backslash (A \cap B)) + \mathbb{P}(A \cap B) + \mathbb{P}(B \backslash (A \cap B)) \overset{6}{=} \mathbb{P}(A) - \mathbb{P}(A \cap B) + \mathbb{P}(A \cap B) + \mathbb{P}(B) - \mathbb{P}(A \cap B) = \mathbb{P}(A) + \mathbb{P}(B) - \mathbb{P}(A \cap B)$. ▲

8. Узагальнення властивості 7 — формула включення-виключення для ймовірностей: $\forall A_1, A_2, \ldots, A_n \in \mathcal{F} : \mathbb{P}\left(\bigcup\limits_{i=1}^{n} A_i\right) = \sum\limits_{i=1}^{n} \mathbb{P}(A_i) - \sum\limits_{i<j}^{n} \mathbb{P}(A_i \cap A_j) + \sum\limits_{i<j<k}^{n} \mathbb{P}(A_i \cap A_j \cap A_k) - ... + (-1)^{n-1} \mathbb{P}\left(\bigcap\limits_{i=1}^{n} A_i\right)$.

   Вправа. Довести цю властивість.

Приклад (задача про неуважну секретарку)**.** Секретарка поклала n листів в n чистих конвертів, заклеїла ці конверти і тільки після цього написала адреси. Яка ймовірність того, що хоча б один з листів дійде за призначенням?

$A_i = \{i\text{-тий лист дійшов за призначенням}\}, i = 1, ..., n$. $\mathbb{P}(A_i) = \frac{1}{n} = \frac{(n-1)!}{n!}$.
$A = \{\text{хоча б один із листів дійшов за призначенням}\}, A = \bigcup\limits_{i=1}^{n} A_i$.
$\mathbb{P}(A_i \cap A_j) = \frac{(n-2)!}{n!} = \frac{1}{n(n-1)}, i \neq j$.
$\mathbb{P}(A_i \cap A_j \cap A_k) = \frac{1}{n(n-1)(n-2)}, i \neq j \neq k$.
...
$\mathbb{P}(A_1 \cap ... \cap A_n) = \frac{1}{n!}$.
За формулою включення-виключення $\mathbb{P}(A) = n \cdot \frac{1}{n} - C_n^2 \cdot \frac{1}{n(n-1)} - \cdots + (-1)^{n-1}\frac{1}{n!} = 1 - \frac{1}{2!} + \frac{1}{3!} - \cdots + (-1)^{n-1} \cdot \frac{1}{n!} \approx 1 - \frac{1}{e} \approx 0.63$.

9. $\forall A_1, A_2, ..., A_n, ... \in \mathcal{F} : \mathbb{P}\left(\bigcup\limits_{k=1}^{\infty} A_k\right) \leq \sum\limits_{k=1}^{\infty} \mathbb{P}(A_k)$

   Доведення. Введемо події $B_1 = A_1, B_2 = \overline{A_1} \cap A_2, B_3 = \overline{A_1} \cap \overline{A_2} \cap A_3, ..., B_n = \overline{A_1} \cap \overline{A_2} \cap ... \cap \overline{A_{n-1}} \cap A_n, ...$ Ці події попарно несумісні, $\bigcup\limits_{k=1}^{\infty} B_k = \bigcup\limits_{k=1}^{\infty} A_k$, $B_k \subset A_k$ для всіх $k \in \mathbb{N}$.
   $\mathbb{P}\left(\bigcup\limits_{k=1}^{\infty} A_k\right) = \mathbb{P}\left(\bigcup\limits_{k=1}^{\infty} B_k\right) \overset{P3}{=} \sum\limits_{k=1}^{\infty} \mathbb{P}(B_k) \overset{4}{\leq} \sum\limits_{k=1}^{\infty} \mathbb{P}(A_k)$. ▲

10. $\forall A_1, A_2, ..., A_n, ... \in \mathcal{F} : \mathbb{P}\left(\bigcap\limits_{k=1}^{\infty} A_k\right) = 1 - \mathbb{P}\left(\overline{\bigcap\limits_{k=1}^{\infty} A_k}\right) = 1 - \mathbb{P}\left(\bigcup\limits_{k=1}^{\infty} \overline{A_k}\right) \geq 1 - \sum\limits_{k=1}^{\infty} \mathbb{P}(\overline{A_k})$.

1.2.3 Теореми неперервності ймовірності

Теорема 1. *Нехай є монотонно неспадна послідовність подій* $A_1 \subset A_2 \subset ... \subset A_n ... \in \mathcal{F}$. *Тоді* $\mathbb{P}\left(\bigcup\limits_{n=1}^{\infty} A_n\right) = \lim\limits_{n\to\infty} \mathbb{P}(A_n)$.

Доведення. $\bigcup\limits_{n=1}^{\infty} A_n = A_1 \cup (A_2 \backslash A_1) \cup (A_3 \backslash A_2) \cup ... \cup (A_n \backslash A_{n-1}) \cup ...$
З аксіоми **P3**: $\mathbb{P}\left(\bigcup\limits_{n=1}^{\infty} A_n\right) = \mathbb{P}(A_1) + \sum\limits_{n=2}^{\infty} \mathbb{P}(A_n \backslash A_{n-1})$.
$S_n = \mathbb{P}(A_1) + \sum\limits_{k=2}^{n} \mathbb{P}(A_k \backslash A_{k-1}) = \mathbb{P}(A_1) + \mathbb{P}(A_2) - \mathbb{P}(A_1) + \mathbb{P}(A_3) - \mathbb{P}(A_2) + ... + \mathbb{P}(A_n) - \mathbb{P}(A_{n-1}) = \mathbb{P}(A_n)$. Отже, $\mathbb{P}\left(\bigcup\limits_{n=1}^{\infty} A_n\right) = \lim\limits_{n\to\infty} S_n = \lim\limits_{n\to\infty} \mathbb{P}(A_n)$. ▲

Теорема 2. *Нехай є монотонно спадна послідовність подій* $A_1 \supset A_2 \supset ... \supset A_n ... \in \mathcal{F}$. *Тоді* $\mathbb{P}\left(\bigcap\limits_{n=1}^{\infty} A_n\right) = \lim\limits_{n\to\infty} \mathbb{P}(A_n)$.

Доведення. $\mathbb{P}\left(\bigcap\limits_{n=1}^{\infty} A_n\right) = 1 - \mathbb{P}\left(\bigcup\limits_{n=1}^{\infty} \overline{A_n}\right) \overset{\text{т. }1}{=} 1 - \lim\limits_{n\to\infty} \mathbb{P}(\overline{A_n}) = \lim\limits_{n\to\infty} (1 - \mathbb{P}(\overline{A_n})) = \lim\limits_{n\to\infty} \mathbb{P}(A_n)$. ▲

1.2.4 Аксіома неперервності

Теорема 3. *Аксіома зліченної адитивності **P3** еквівалентна аксіомі неперервності **P4**:*

$$\forall n \in \mathbb{N} : B_n \in \mathcal{F}, B_1 \supset B_2 \supset ... \supset B_n \supset ..., \bigcap_{n=1}^{\infty} B_n = \varnothing \Rightarrow \lim_{n\to\infty} \mathbb{P}(B_n) = 0$$

Доведення. За теоремою 2 маємо наслідок **P3** $\Rightarrow$ **P4**.

Доведемо **P4** $\Rightarrow$ **P3**. Для подій $A_1, A_2, ..., A_n, ... \in \mathcal{F}, A_i \cap A_j = \varnothing$ при $i \neq j$ введемо події $B_n = \bigcup\limits_{k=n+1}^{\infty} A_k$. Для них виконується $B_1 \supset B_2 \supset ... \supset B_n \supset ...$ та $\bigcap\limits_{n=1}^{\infty} B_n = \varnothing$, тому з аксіоми **P4** $\lim\limits_{n\to\infty} \mathbb{P}(B_n) = 0$.

З іншого боку, маємо таку рівність:

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

Перейшовши у правій частині до границі при $n \to \infty$, отримаємо $\mathbb{P}\left(\bigcup\limits_{k=1}^{\infty} A_k\right) = \sum\limits_{k=1}^{\infty} \mathbb{P}(A_k)$, що є твердженням аксіоми зліченної адитивності. ▲

1.3 Умовні ймовірності. Незалежність подій. Формула Баєса.

1.3.1 Умовні ймовірності

Припустимо, що спостерігається деякий експеримент, що описується класичною моделлю, а для події B $\mathbb{P}(B) > 0$. Умовна ймовірність $\mathbb{P}(A/B)$ — це ймовірність події A за умови, що відбулась подія B. Наприклад, експеримент — витягання карт з колоди, $A = \{$витягнуто даму пік$\}$, $B = \{$витягнуто карту чорної масті$\}$.

Оскільки експеримент описується класичною моделлю, то можемо позначити m_B та $m_{A\cap B}$ кількості елементарних подій, що сприяють появам подій B та $A \cap B$ відповідно. Тоді $\mathbb{P}(A/B) = \frac{m_{A\cap B}}{m_B} = \frac{m_{A\cap B}/n}{m_B/n} = \frac{\mathbb{P}(A\cap B)}{\mathbb{P}(B)}$, де $n = \mathrm{card}(\Omega)$.

Ця формула справджується й для довільних ймовірнісних просторів.

Означення 1.3.1. Якщо подія B має додатну ймовірність $\mathbb{P}(B) > 0$, то *умовна ймовірність* події A за умови, що відбулась подія B, обчислюється за формулою

$$\mathbb{P}(A/B) = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)} \tag{1}$$

Властивості умовної ймовірності:
1. $\mathbb{P}(A/B) \geq 0$.
2. $\mathbb{P}(\Omega/B) = \mathbb{P}(B/B) = 1$.
3. $\forall A_1, A_2, ..., A_n, ... \in \mathcal{F} : A_i \cap A_j = \varnothing$ при $i \neq j$:
$$\mathbb{P}\left(\left(\bigcup_{n=1}^{\infty} A_n\right)/B\right) = \frac{\mathbb{P}\left(\left(\bigcup\limits_{n=1}^{\infty} A_n\right)\cap B\right)}{\mathbb{P}(B)} = \frac{\mathbb{P}\left(\bigcup\limits_{n=1}^{\infty}(A_n\cap B)\right)}{\mathbb{P}(B)} \overset{\mathbf{P3}}{=} \frac{\sum\limits_{n=1}^{\infty}\mathbb{P}(A_n\cap B)}{\mathbb{P}(B)} = \sum_{n=1}^{\infty} \mathbb{P}(A_n/B).$$

Таким чином, для умовної ймовірності діють всі аксіоми ТЙ. Фактично, розглядається звуження ймовірнісного простору $\{B, \mathcal{F}_B, \mathbb{P}(\cdot/B)\}$, в якому B стає вірогідною подією.

Теореми множення

З формули (1) випливає, що

$$\mathbb{P}(A \cap B) = \mathbb{P}(A) \cdot \mathbb{P}(B/A) = \mathbb{P}(B) \cdot \mathbb{P}(A/B) \tag{2}$$

Для трьох подій: $\mathbb{P}(A \cap B \cap C) = \mathbb{P}(A \cap B) \cdot \mathbb{P}(C/(A \cap B)) = \mathbb{P}(A) \cdot \mathbb{P}(B/A) \cdot \mathbb{P}(C/(A \cap B))$. За методом математичної індукції неважко довести, що

$$\mathbb{P}\left(\bigcap_{k=1}^{n} A_k\right) = \mathbb{P}(A_1) \cdot \mathbb{P}(A_2/A_1) \cdot \mathbb{P}(A_3/(A_1 \cap A_2)) \cdot ... \cdot \mathbb{P}\left(A_n/\left(\bigcap_{k=1}^{n-1} A_k\right)\right) \tag{3}$$

Приклад. На 10 картках написано по одній букві так, що можна скласти слово «математика». Дитина навмання витягує без повернення по одній картці. Яка ймовірність того, що після витягання 4 карток утвориться слово «мама»?

Треба обчислити ймовірність $\mathbb{P}(M^{(1)} \cap A^{(2)} \cap M^{(3)} \cap A^{(4)})$, де верхній індекс означає крок, на якому витягнуто літеру. $\mathbb{P}(M^{(1)}) = \frac{2}{10}$, $\mathbb{P}(A^{(2)}/M^{(1)}) = \frac{3}{9}$, $\mathbb{P}(M^{(3)}/(M^{(1)} \cap A^{(2)})) = \frac{1}{8}$, $\mathbb{P}(A^{(4)}/(M^{(1)} \cap A^{(2)} \cap M^{(3)})) = \frac{2}{7}$. Тому за формулою (3) $\mathbb{P}(M^{(1)} \cap A^{(2)} \cap M^{(3)} \cap A^{(4)}) = \frac{2\cdot3\cdot1\cdot2}{10\cdot9\cdot8\cdot7}$.

1.3.2 Незалежність подій

Означення 1.3.2. Дві події A та B називаються *незалежними*, якщо

$$\mathbb{P}(A \cap B) = \mathbb{P}(A) \cdot \mathbb{P}(B) \tag{4}$$

Властивості незалежних подій:
1. Якщо події A та B незалежні, то $\mathbb{P}(A/B) = \frac{\mathbb{P}(A\cap B)}{\mathbb{P}(B)} = \frac{\mathbb{P}(A)\cdot\mathbb{P}(B)}{\mathbb{P}(B)} = \mathbb{P}(A)$.
   Аналогічно $\mathbb{P}(B/A) = \mathbb{P}(B)$.

Зауваження. Несумісні події залежні, якщо їх ймовірності не нульові.

2. Якщо події A та B незалежні, то пари A та $\overline{B}$, $\overline{A}$ та B, $\overline{A}$ та $\overline{B}$ — теж незалежні.

   Доведення. Доведемо спочатку для пари A та $\overline{B}$. $\mathbb{P}(A \cap \overline{B}) = \mathbb{P}(A \backslash (A \cap B)) = \mathbb{P}(A) - \mathbb{P}(A \cap B) = \mathbb{P}(A) - \mathbb{P}(A) \cdot \mathbb{P}(B) = \mathbb{P}(A) \cdot (1 - \mathbb{P}(B)) = \mathbb{P}(A) \cdot \mathbb{P}(\overline{B})$. Отже, A та $\overline{B}$ — незалежні. Доведення для інших пар аналогічне. ▲

3. Якщо A та B, A та C незалежні, а B та C несумісні, то A та $B \cup C$ — незалежні події.

   Доведення. $\mathbb{P}(A \cap (B \cup C)) = \mathbb{P}((A \cap B) \cup (A \cap C)) = \mathbb{P}(A \cap B) + \mathbb{P}(A \cap C) = \mathbb{P}(A) \cdot \mathbb{P}(B) + \mathbb{P}(A) \cdot \mathbb{P}(C) = \mathbb{P}(A) \cdot (\mathbb{P}(B) + \mathbb{P}(C)) = \mathbb{P}(A) \cdot \mathbb{P}(B \cup C)$. ▲

Означення 1.3.3. Події $A_1, A_2, ..., A_n \in \mathcal{F}$ називаються *незалежними у сукупності*, якщо

$$\forall i_1, i_2, ...i_r \in \{1, ..., n\} : \mathbb{P}\left(\bigcap_{k=1}^{r} A_{i_k}\right) = \prod_{k=1}^{r} \mathbb{P}\left(A_{i_k}\right) \tag{5}$$

Зокрема:
1. $\forall i, j \in 1, .., n, i \neq j : \mathbb{P}(A_i \cap A_j) = \mathbb{P}(A_i) \cdot \mathbb{P}(A_j)$ — попарна незалежність.
2. $\mathbb{P}\left(\bigcap\limits_{k=1}^{n} A_k\right) = \prod\limits_{k=1}^{n} \mathbb{P}(A_k)$.

Зауваження. З незалежності подій у сукупності випливає попарна незалежність, але наслідку в іншу сторону, взагалі кажучи, немає.

Приклад (приклад Бернштейна)**.** Дитина кидає на підлогу тетраедр, три грані якого розфарбовані відповідно зеленим, синім та червоним кольором, а четверта — усіма трьома кольорами. Введемо події З, С, Ч, які означають, що випала грань, на якій є зелений, синій або червоний колір відповідно.

$\mathbb{P}(\text{З}) = \mathbb{P}(\text{С}) = \mathbb{P}(\text{Ч}) = \frac{1}{2}$, $\mathbb{P}(\text{З} \cap \text{С}) = \mathbb{P}(\text{З} \cap \text{Ч}) = \mathbb{P}(\text{С} \cap \text{Ч}) = \frac{1}{4}$, тому ці події є попарно незалежними. Але $\mathbb{P}(\text{З} \cap \text{С} \cap \text{Ч}) = \frac{1}{4} \neq \frac{1}{8}$, тому незалежності у сукупності немає.

Теорема додавання для незалежних у сукупності подій. $A_1, A_2, ..., A_n \in \mathcal{F}$ — незалежні у сукупності. $\mathbb{P}\left(\bigcup\limits_{k=1}^{n} A_k\right) = ?$

$\mathbb{P}\left(\bigcup\limits_{k=1}^{n} A_k\right) = 1 - \mathbb{P}\left(\overline{\bigcup\limits_{k=1}^{n} A_k}\right) = 1 - \mathbb{P}\left(\bigcap\limits_{k=1}^{n} \overline{A_k}\right) = 1 - \prod\limits_{k=1}^{n} \mathbb{P}(\overline{A_k})$.

Зокрема, для двох незалежних подій A і B: $\mathbb{P}(A \bigcup B) = 1 - \mathbb{P}(\overline{A}) \cdot \mathbb{P}(\overline{B})$

1.3.3 Формула повної ймовірності

Нехай деякому СЕ поставлено у відповідність простір елементарних подій Ω.

$H_1, H_2, ..., H_n \in \mathcal{F}$ (може бути й зліченна кількість) — повна група подій, які називаємо *гіпотезами (припущеннями)*. Нехай подія A відбувається разом з якоюсь гіпотезою. Тоді має місце *формула повної ймовірності*:

$$\mathbb{P}(A) = \sum_{k=1}^{n} \mathbb{P}(H_k) \cdot \mathbb{P}(A/H_k) \tag{6}$$

Доведення. $\mathbb{P}(A) = \mathbb{P}(A \cap \Omega) = \mathbb{P}\left(A \cap \left(\bigcup\limits_{k=1}^{n} H_k\right)\right) = \mathbb{P}\left(\bigcup\limits_{k=1}^{n} (A \cap H_k)\right) = \sum\limits_{k=1}^{n} \mathbb{P}(A \cap H_k) = \sum\limits_{k=1}^{n} \mathbb{P}(H_k) \cdot \mathbb{P}(A/H_k)$. ▲

Приклад. 1. У магазин постачають 80% телефонів з Китаю, 15% з В'єтнаму та 5% з Кореї, причому бракованих відповідно 1%, 0.1% та 0.01%. Знайти ймовірність події $A = \{$куплений телефон буде бракованим$\}$.

За умовою введемо гіпотези $H_1 = \{$телефон з Китаю$\}$, $H_2 = \{$телефон з В'єтнаму$\}$ та $H_3 = \{$телефон з Кореї$\}$. $\mathbb{P}(H_1) = 0.8$, $\mathbb{P}(H_2) = 0.15$, $\mathbb{P}(H_3) = 0.05$, $\mathbb{P}(A/H_1) = 0.01$, $\mathbb{P}(A/H_2) = 0.001$, $\mathbb{P}(A/H_3) = 0.0001$. Тоді за формулою (6) $\mathbb{P}(A) = 0.8 \cdot 0.01 + 0.15 \cdot 0.001 + 0.05 \cdot 0.0001 = 0.008155$.

2. Серед N екзаменаційних білетів n «щасливих», $n < N$. У якого студента ймовірність витягнути «щасливий» білет більша — у того, хто тягне першим, чи того, хто тягне другим?
   Позначимо через A і B події, що «щасливий» білет витягнув перший та другий студент відповідно. За умовою $\mathbb{P}(A) = \frac{n}{N}$. Щоб скористатися формулою (6) для обчислення $\mathbb{P}(B)$, введемо гіпотези $H_1 = A$ та $H_2 = \overline{A}$. $\mathbb{P}(H_1) = \frac{n}{N}$, $\mathbb{P}(H_2) = \frac{N-n}{N}$.
   $\mathbb{P}(B) = \mathbb{P}(H_1) \cdot \mathbb{P}(B/H_1) + \mathbb{P}(H_2) \cdot \mathbb{P}(B/H_2) = \frac{n}{N} \cdot \frac{n-1}{N-1} + \frac{N-n}{N} \cdot \frac{n}{N-1} = \frac{n^2-n+N\cdot n-n^2}{N\cdot(N-1)} = \frac{n\cdot(N-1)}{N\cdot(N-1)} = \frac{n}{N}$.

1.3.4 Формула Баєса

Як і раніше, нехай $H_1, H_2, ..., H_n \in \mathcal{F}$ — повна група подій деякого СЕ, які називаємо гіпотезами, причому перед проведенням експерименту відомі їх *апріорні* ймовірності $\mathbb{P}(H_1), \mathbb{P}(H_2), ..., \mathbb{P}(H_n)$. В результаті проведення експерименту відбулась деяка подія A. Постає питання: чому рівні *апостеріорні* ймовірності $\mathbb{P}(H_1/A), \mathbb{P}(H_2/A), ..., \mathbb{P}(H_n/A)$? Тобто, чому дорівнюють ймовірності, що мала місце кожна з гіпотез за умови, що подія A відбулась? На це дає відповідь *формула Баєса*:

$$\mathbb{P}(H_i/A) = \frac{\mathbb{P}(H_i \cap A)}{\mathbb{P}(A)} = \frac{\mathbb{P}(H_i) \cdot \mathbb{P}(A/H_i)}{\sum\limits_{k=1}^{n} \mathbb{P}(H_k) \cdot \mathbb{P}(A/H_k)}, i = 1, ..., n \qquad (7)$$

Приклад. В групі з 10 осіб троє вчаться на «5», четверо на «4», двоє на «3» та один на «2». Викладач підготував на екзамен 20 питань. Студенти, які вчаться на «5», знають відповіді на всі, на «4» — на 16, на «3» — на 10, на «2» — на 5. Екзаменаційний білет містить 3 питання. Деякий студент відповів правильно на всі 3. Яка ймовірність того, що він вчиться на «2»?

Введемо гіпотези $H_1 = \{$студент вчиться на «5»$\}$, $H_2 = \{$студент вчиться на «4»$\}$, $H_3 = \{$студент вчиться на «3»$\}$, $H_4 = \{$студент вчиться на «2»$\}$. За умовою $\mathbb{P}(H_1) = \frac{3}{10}$, $\mathbb{P}(H_2) = \frac{4}{10}$, $\mathbb{P}(H_3) = \frac{2}{10}$, $\mathbb{P}(H_4) = \frac{1}{10}$. Позначимо $A = \{$студент відповів на всі три питання$\}$. Тоді $\mathbb{P}(A/H_1) = 1$, $\mathbb{P}(A/H_2) = \frac{16\cdot15\cdot14}{20\cdot19\cdot18}$, $\mathbb{P}(A/H_3) = \frac{10\cdot9\cdot8}{20\cdot19\cdot18}$, $\mathbb{P}(A/H_4) = \frac{5\cdot4\cdot3}{20\cdot19\cdot18}$. За формулою (7) шукана ймовірність $\mathbb{P}(H_4/A) = \frac{\mathbb{P}(H_4)\cdot\mathbb{P}(A/H_4)}{\sum\limits_{k=1}^{4} \mathbb{P}(H_k)\cdot\mathbb{P}(A/H_k)} = \frac{0.1\cdot5\cdot4\cdot3}{0.3\cdot20\cdot19\cdot18+0.4\cdot16\cdot15\cdot14+0.2\cdot10\cdot9\cdot8+0.1\cdot5\cdot4\cdot3} = \frac{60}{35460} = \frac{1}{591}$.

1.4 Поліноміальна модель ймовірності. Схема Бернуллі.

1.4.1 Поліноміальна схема

Припустимо, що події $A_1, A_2, ..., A_k$ утворюють повну групу подій деякого СЕ, причому відомі ймовірності $\mathbb{P}(A_i) = p_i$, $\sum\limits_{i=1}^{k} p_i = 1$. Експеримент проводиться n разів, в кожному з яких може відбутись одна з подій A_i. Можливими результатами n раз проведеного експерименту будуть події $A_{i_1}^{(1)} \cap A_{i_2}^{(2)} \cap ... \cap A_{i_n}^{(n)}$, де $A_{i_r}^{(r)}$, $r = 1, ..., n$, $i_r = 1, ..., k$ — один із можливих результатів r-того випробування.

Означення 1.4.1. Випробування, що проводяться, називаються *незалежними*, якщо

$$\mathbb{P}\left(A_{i_1}^{(1)} \cap A_{i_2}^{(2)} \cap ... \cap A_{i_n}^{(n)}\right) = \mathbb{P}\left(A_{i_1}^{(1)}\right) \cdot \mathbb{P}\left(A_{i_2}^{(2)}\right) \cdot ... \cdot \mathbb{P}\left(A_{i_n}^{(n)}\right)$$

Основна задача — знайти ймовірність того, що подія A_1 відбудеться m_1 разів, подія A_2 відбудеться m_2 разів і т.д., подія A_k відбудеться m_k разів, причому $\sum\limits_{j=1}^{k} m_j = n$. Такі ймовірності будемо позначати як $\mathbb{P}_n(m_1, m_2, ..., m_k)$.

$$\mathbb{P}\left(\underbrace{A_1 \cap ... \cap A_1}_{m_1 \text{разів}} \cap \underbrace{A_2 \cap ... \cap A_2}_{m_2 \text{разів}} \cap ... \cap \underbrace{A_k \cap ... \cap A_k}_{m_k \text{разів}}\right) = p_1^{m_1} p_2^{m_2} ... p_k^{m_k}$$

Це є лише один із можливих варіантів послідовності появи подій в серії випробувань. Всього таких варіантів рівно $C_n^{m_1,m_2,...,m_k} = \frac{n!}{m_1!m_2!...m_k!}$. Таким чином отримуємо:

$$\mathbb{P}_n(m_1,...,m_k) = \frac{n!}{m_1!m_2!...m_k!} p_1^{m_1} p_2^{m_2} ... p_k^{m_k} \tag{1}$$

Означення 1.4.2. Ймовірності, що обчислюємо за даною формулою, називаються *поліноміальними*, а сама схема — *поліноміальною схемою ймовірності.*

Зауваження. Звідки назва «поліноміальна»?

$$1 = (p_1 + p_2 + ... + p_k)^n = \sum_{m_1+...+m_k=n} \mathbb{P}_n(m_1,...,m_k)$$

Приклад. Студент ІПСА за рівнем підготовки з ймовірністю 0.3 вважається слабким студентом, з ймовірністю 0.5 вважається середнім студентом та з ймовірністю 0.2 — сильним студентом. Яка ймовірність того, що з 6 навмання обраних студентів кількість слабких та сильних буде однаковою?

Розглядаємо подію $A = \{$кількість слабких рівна кількості сильних$\}$. Позначимо ймовірності того, що студент має певний рівень підготовки, таким чином: $p_1 = 0.3$, $p_2 = 0.5$, $p_3 = 0.2$. $\mathbb{P}(A) = ?$

Скористаємось поліноміальною схемою: $\mathbb{P}(A) = \mathbb{P}_6(3,0,3) + \mathbb{P}_6(2,2,2) + \mathbb{P}_6(1,4,1) + \mathbb{P}_6(0,6,0) = \frac{6!}{3!3!}\cdot(0.3)^3\cdot(0.2)^3 + \frac{6!}{2!2!2!}\cdot(0.3)^2\cdot(0.5)^2\cdot(0.2)^2 + \frac{6!}{1!4!1!}\cdot 0.3\cdot(0.5)^4\cdot 0.2 + \frac{6!}{6!}\cdot(0.5)^6 = 0.213445.$

1.4.2 Схема Бернуллі

Означення 1.4.3. Поліноміальна схема, в кожному випробуванні якої є тільки дві події A_1 та A_2, називається *біноміальною схемою* або *схемою Бернуллі.*

A_1 та A_2 утворюють повну групу подій $\Rightarrow A_2 = \overline{A_1}$. Позначимо $A_1 = A$ — «успіх», $A_2 = \overline{A}$ — «невдача», $\mathbb{P}(A) = p, \mathbb{P}(\overline{A}) = 1 - p = q$.

Яка ймовірність того, що в n випробуваннях «успіх» з'явиться m разів? На це питання дає відповідь *формула Бернуллі*:

$$\mathbb{P}_n(m) = \frac{n!}{m!(n-m)!} p^m q^{n-m} = C_n^m p^m q^{n-m} \tag{2}$$

Ймовірності, що обчислюються за цією формулою, називаються *біноміальними.*

Приклад. Спортсмен 5 разів стріляє по мішені. Ймовірність влучення при кожному пострілі — 0.2. Яка ймовірність того, що він влучив не менше 3 разів?

$\mathbb{P}\{$влучив ≥ 3 разів$\} = \mathbb{P}_5(3) + \mathbb{P}_5(4) + \mathbb{P}_5(5) = C_5^3(0.2)^3(0.8)^2 + C_5^4(0.2)^4 0.8 + C_5^5(0.2)^5 = 0.05792.$

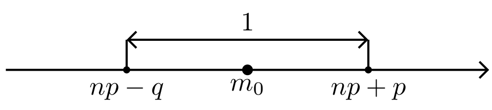
1.4.3 Найбільш імовірна кількість успіхів в схемі Бернуллі

Означення 1.4.4. Натуральне число m_0, при якому $\mathbb{P}_n(m)$ набуває найбільшого значення, називається *найбільш імовірною кількістю успіхів.*

Розглянемо систему:

$$\begin{cases} \mathbb{P}_n(m_0 - 1) \leq \mathbb{P}_n(m_0) \\ \mathbb{P}_n(m_0 + 1) \leq \mathbb{P}_n(m_0) \end{cases} \Leftrightarrow \begin{cases} \frac{n!}{(m_0-1)!(n-m_0+1)!} p^{m_0-1} q^{n-m_0+1} \leq \frac{n!}{m_0!(n-m_0)!} p^{m_0} q^{n-m_0} \\ \frac{n!}{(m_0+1)!(n-m_0-1)!} p^{m_0+1} q^{n-m_0-1} \leq \frac{n!}{m_0!(n-m_0)!} p^{m_0} q^{n-m_0} \end{cases} \Leftrightarrow$$

$$\Leftrightarrow \begin{cases} \frac{q}{n-m_0+1} \leq \frac{p}{m_0} \\ \frac{p}{m_0+1} \leq \frac{q}{n-m_0} \end{cases} \Leftrightarrow \begin{cases} m_0 \leq np + p \\ m_0 \geq np - q \end{cases}$$



Таким чином отримуємо два варіанти:

1. $np + p$ – не ціле $\Rightarrow m_0 = [np + p]$
2. $np + p$ – ціле $\Rightarrow \begin{cases} m_0^{(1)} = np + p \\ m_0^{(2)} = np - q \end{cases}$

Приклад. Знайти найбільш імовірну кількість успіхів для прикладу (1.4.2).

$np + p = 1.2 \Rightarrow m_0 = [np + p] = [1.2] = 1$. Ймовірність лише одного влучення буде найбільшою, рівною 0.904.

1.4.4 Асимптотичні наближення формули Бернуллі

Формула Бернуллі не є зручною для великих значень m та n. Існують формули, якими можна скористатись для наближення результатів формули Бернуллі.

Твердження (асимптотична формула Пуассона)**.** *Нехай n — достатньо велике натуральне число, а p — достатньо мале, так що $a = np \in (1, 20)$. Тоді для наближення формули Бернуллі можна використати асимптотичну формулу Пуассона:*

$$\mathbb{P}_n(m) \approx \frac{(np)^m}{m!} e^{-np}$$

Доведення (нестроге). $\mathbb{P}_n(m) = \frac{n(n-1)...(n-m+1)}{m!} \left(\frac{a}{n}\right)^m (1 - \frac{a}{n})^{n-m} = \frac{a^m}{m!} 1(1 - \frac{1}{n})(1 - \frac{2}{n})...(1 - \frac{m+1}{n})(1 - \frac{a}{n})^n (1 - \frac{a}{n})^{-m} \underset{n\to\infty}{\longrightarrow} \frac{a^m}{m!} e^{-a} = \frac{(np)^m}{m!} e^{-np}$. ▲

Твердження (формула Муавра-Лапласа)**.** *Нехай $npq \geq 20$. Тоді для наближення формули Бернуллі можна використати формулу Муавра-Лапласа:*

$$\mathbb{P}_n(m) \approx \frac{1}{\sqrt{2\pi npq}} e^{-\frac{(m-np)^2}{2npq}}$$

Доведення. Ґрунтується на локальній теоремі Муавра-Лапласа. ▲

Детальне доведення обох тверджень буде в темі «Граничні теореми теорії ймовірностей» (ст. 74).

Розділ 2

Випадкові величини

2.1 Поняття випадкової величини та її задання

Означення 2.1.1. *Випадковою величиною* називається дійснозначна вимірна функція, що здійснює відображення з Ω в $\mathbb{R}$, тобто $\xi = \xi(\omega) : \Omega \to \mathbb{R}$.

Зауваження. Вимірність функції $\xi(\omega)$ означає, що

$$\forall x \in \mathbb{R} : \{\omega \in \Omega \ : \ \xi(\omega) < x\} \in \mathcal{F} \tag{1}$$

В курсі функціонального аналізу доводиться, що якщо виконується (1), то $\{\omega : \xi(\omega) > x\}$, $\{\omega : \xi(\omega) \leq x\}$, $\{\omega : \xi(\omega) \geq x\}$, $\{\omega : \xi(\omega) = x\}$ та $\{\omega : \xi(\omega) \in \langle a; b\rangle\}$ також належать $\mathcal{F}$.

Надалі розглядатимуться випадкові величини двох видів — дискретні (ДВВ) та неперервні (НВВ).

2.1.1 Функція розподілу

Універсальною ймовірнісною характеристикою будь-якої випадкової величини є функція розподілу випадкової величини.

Означення 2.1.2. Дійснозначна функція дійсного аргументу $F_\xi(x) = \mathbb{P}\{\omega : \xi(\omega) < x\}$ або $\mathbb{P}(\xi < x)$ називається *функцією розподілу випадкової величини.*

Властивості функції розподілу:

1. Область визначення $D(F) = \mathbb{R}$, область значень $E(F) = \langle 0; 1\rangle$.
2. $\forall x_1 > x_2 \in \mathbb{R}$: $F_\xi(x_1) \geq F_\xi(x_2)$ — монотонно неспадна.

   Доведення. Розглянемо події $A = \{\omega : \xi(\omega) < x_1\}$ та $B = \{\omega : \xi(\omega) < x_2\}$, причому $x_1 > x_2$. Отже, $B \subset A \Rightarrow \mathbb{P}(B) \leq \mathbb{P}(A) \Rightarrow F_\xi(x_1) \geq F_\xi(x_2)$. ▲

3. $\mathbb{P}\{\omega : a \leq \xi(\omega) < b\} = F_\xi(b) - F_\xi(a)$.

   Доведення. Розглянемо події $A = \{\omega : \xi(\omega) < a\}$, $B = \{\omega : \xi(\omega) < b\}$,
   $C = \{\omega : a \leq \xi(\omega) < b\}$. $B = A \cup C, A \cap C = \varnothing \Rightarrow \mathbb{P}(B) = \mathbb{P}(A) + \mathbb{P}(C) \Rightarrow$
   $\Rightarrow \mathbb{P}(C) = \mathbb{P}\{\omega : a \leq \xi(\omega) < b\} = \mathbb{P}(B) - \mathbb{P}(A) = F_\xi(b) - F_\xi(a)$. ▲

4. $\lim\limits_{x\to-\infty} F_\xi(x) = 0$, $\lim\limits_{x\to+\infty} F_\xi(x) = 1$.

   Доведення. Введемо послідовності подій $A_n = \{\omega : \xi(\omega) < -n\}$ та $B_n = \{\omega : \xi(\omega) < n\}$. Зауважимо, що A_n є монотонно спадною, а B_n — монотонно зростаючою.
   $\lim\limits_{x\to-\infty} F_\xi(x) = \lim\limits_{x\to-\infty} \mathbb{P}(\xi < x) = \lim\limits_{n\to\infty} \mathbb{P}(\xi < -n) = \begin{bmatrix}\text{теорема} \\ \text{неперервності 2}\end{bmatrix} = \mathbb{P}\left(\bigcap\limits_{n=1}^{\infty} A_n\right) =$
   $= \mathbb{P}(\varnothing) = 0.$
   $\lim\limits_{x\to+\infty} F_\xi(x) = \lim\limits_{x\to+\infty} \mathbb{P}(\xi < x) = \lim\limits_{n\to\infty} \mathbb{P}(\xi < n) = \begin{bmatrix}\text{теорема} \\ \text{неперервності 1}\end{bmatrix} = \mathbb{P}\left(\bigcup\limits_{n=1}^{\infty} B_n\right) =$
   $= \mathbb{P}(\Omega) = 1.$ ▲

5. Функція розподілу неперервна зліва: $\lim\limits_{x\to x_0-0} F_\xi(x) = F_\xi(x_0)$.

Доведення. $A_i = \{\omega : \xi(\omega) < x_i\}$, $\lim\limits_{n\to\infty} x_n = x_0$, $x_n < x_0$.

$\lim\limits_{x\to x_0-0} F_\xi(x) = \lim\limits_{n\to\infty} \mathbb{P}(\xi < x_n) =$ [теорема неперервності 2] $= \mathbb{P}\left(\bigcup\limits_{n=1}^{\infty} A_n\right) =$
$= \mathbb{P}(\xi < x_0) = F_\xi(x_0)$. ▲

Зауваження. В зарубіжній літературі функція розподілу іноді вводиться інакше, як $F_\xi(x) = \mathbb{P}\{\omega : \xi(\omega) \le x\}$. За такого означення функція розподілу є неперервною справа.

Розглянемо ймовірність $\mathbb{P}\{x \le \xi < x + \Delta x\} = F_\xi(x + \Delta x) - F_\xi(x)$. Перейдемо до границі при $\Delta x \to 0$: $\mathbb{P}\{\xi = x\} = \begin{cases} 0, & \text{якщо } x \text{ — точка неперервності} \\ \Delta\text{-стрибок}, & \text{якщо } x \text{ — точка розриву 1-го роду} \end{cases}$.

2.1.2 Дискретні випадкові величини

Означення 2.1.3. Випадкова величина $\xi = \xi(\omega)$ називається *дискретною випадковою величиною* (ДВВ), якщо вона набуває скінченну або зліченну кількість значень.

Для задання ДВВ, крім знання значень випадкової величини, необхідно знати ймовірності, з якими ці значення приймаються.

$$p_i = \mathbb{P}\{\omega : \xi(\omega) = x_i\} = \mathbb{P}(\xi = x_i), i = 1, 2, ..., \sum_{i=1}^{\infty} p_i = 1$$

Означення 2.1.4. *Законом розподілу (ймовірностей)* ДВВ називається співвідношення, яке вказує, які значення ця випадкова величина приймає та з якими ймовірностями.

Закон розподілу ДВВ записується у вигляді *ряду розподілу*:

ξ	x_1	x_2	...	x_n	...
p	p_1	p_2	...	p_n	...

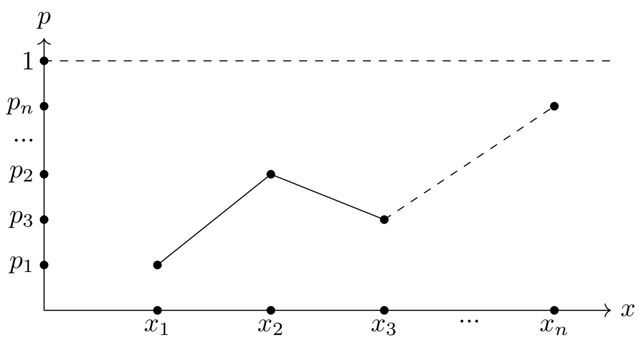
$$x_1 < x_2 < ... < x_n < ..., \sum_i p_i = 1$$

Приклад. ξ задає кількість влучень при чотирьох пострілах з імовірністю влучення $p = \frac{1}{2}$. Скласти закон розподілу цієї випадкової величини.

З формули Бернуллі $\mathbb{P}(\xi = k) = C_4^k \left(\frac{1}{2}\right)^k \left(\frac{1}{2}\right)^{4-k} = C_4^k \left(\frac{1}{2}\right)^4 = \frac{C_4^k}{2^4}, k = 0, 1, 2, 3, 4.$

ξ	0	1	2	3	4
p	$\frac{1}{2^4}$	$\frac{4}{2^4}$	$\frac{6}{2^4}$	$\frac{4}{2^4}$	$\frac{1}{2^4}$

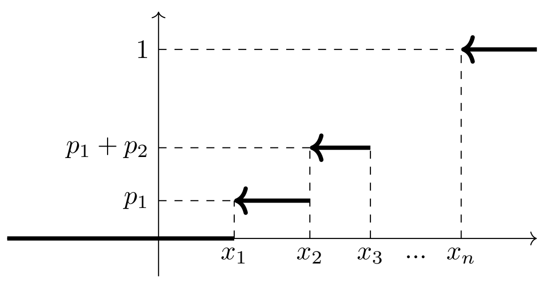
Означення 2.1.5. Графічне зображення закону розподілу ДВВ називається *полігоном розподілу ймовірностей.*



Вигляд функції розподілу для ДВВ:

ξ	x_1	x_2	...	x_n
p	p_1	p_2	...	p_n

$$x_1 < x_2 < ... < x_n, \sum_{i=1}^{n} p_i = 1$$



$$F_\xi(x) = \begin{cases} 0, & x \le x_1 \\ p_1, & x_1 < x \le x_2 \\ p_1 + p_2, & x_2 < x \le x_3 \\ \cdots & \\ \sum\limits_{i=1}^{n-1} p_i, & x_{n-1} < x \le x_n \\ \sum\limits_{i=1}^{n} p_i = 1, & x > x_n \end{cases}$$

2.1.3 Неперервні випадкові величини

Означення 2.1.6. Випадкова величина $\xi = \xi(\omega)$ називається *неперервною випадковою величиною* (НВВ), якщо її функція розподілу неперервна, диференційовна майже скрізь, можливо, за виключенням окремих ізольованих точок.

З неперервності функції розподілу для довільної точки x_0 маємо $\mathbb{P}\{\xi = x_0\} = 0$. Замість ймовірності потрапляння НВВ у окрему точку розглядається щільність розподілу ймовірностей у цій точці.

Означення 2.1.7. *Щільність розподілу ймовірностей* неперервної випадкової величини ξ дорівнює границі (якщо вона існує):

$$\lim_{\Delta x \to 0} \frac{\mathbb{P}\{x \le \xi < x + \Delta x\}}{\Delta x} \tag{2}$$

Отже, закон розподілу НВВ можна задавати щільністю $f_\xi(x) = \lim\limits_{\Delta x \to 0} \frac{\mathbb{P}\{x \le \xi < x+\Delta x\}}{\Delta x}$. Нескладно помітити *зв'язок щільності з функцією розподілу*:

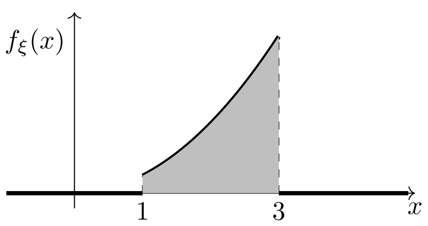
$$f_\xi(x) = \lim_{\Delta x \to 0} \frac{\mathbb{P}\{x \le \xi < x + \Delta x\}}{\Delta x} = \lim_{\Delta x \to 0} \frac{F_\xi(x + \Delta x) - F_\xi(x)}{\Delta x} = F'_\xi(x) \tag{3}$$

Графік щільності розподілу називається *кривою розподілу.*

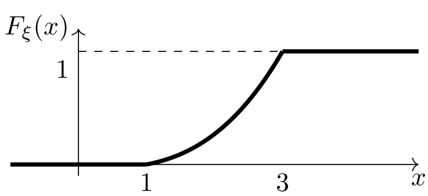
Властивості щільності розподілу:

1. Область визначення $D(f) = \mathbb{R}$, область значень $E(f) = [0; +\infty)$ — з монотонної неспадності $F_\xi(x)$.
2. $F_\xi(x) = \int\limits_{-\infty}^{x} f_\xi(t)dt$.
3. *Властивість нормування:* $\lim\limits_{x \to +\infty} F_\xi(x) = 1 \Rightarrow \int\limits_{-\infty}^{+\infty} f_\xi(x)dx = 1$. Геометрична інтерпретація цієї властивості — площа під кривою розподілу завжди рівна 1.
4. $\mathbb{P}\{\omega : a \le \xi(\omega) < b\} = F_\xi(b) - F_\xi(a) = \int\limits_a^b f_\xi(x)dx$. Ця властивість справджується і для довільних проміжків $\langle a; b\rangle$.

Приклад. 1. $f_\xi(x) = \begin{cases} 0, & x \notin [1;3] \\ A \cdot x^2, & x \in [1;3] \end{cases}$. Знайти A та $F_\xi(x)$.

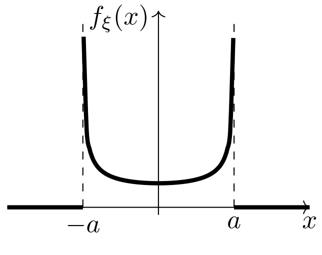


$\int\limits_{-\infty}^{+\infty} f_\xi(x)dx = A \cdot \int\limits_1^3 x^2dx = A \cdot \frac{x^3}{3}\Big|_1^3 = A \cdot \frac{26}{3}$.
З властивості нормування $A = \frac{3}{26}$.



$$F_\xi(x) = \int\limits_{-\infty}^{x} f_\xi(t)dt = \begin{cases} 0, \ x \le 1 \\ \int\limits_{-\infty}^{1} 0dt + \int\limits_1^x \frac{3}{26}t^2dt = \\ \frac{1}{26}(x^3 - 1), \ 1 < x \le 3 \\ \int\limits_{-\infty}^{1} 0dt + \int\limits_1^3 \frac{3}{26}t^2dt + \\ + \int_3^x 0dt = 1, \ x > 3 \end{cases}$$

2. «Закон арксинуса» з щільністю, що має розриви 2-го роду:



$$f_\xi(x) = \begin{cases} 0, & |x| \ge a \\ \frac{1}{\pi\sqrt{a^2 - x^2}}, & |x| < a \end{cases}$$

2.2 Числові характеристики випадкових величин

2.2.1 Математичне сподівання

Означення 2.2.1. *Математичним сподіванням* випадкової величини називається інтеграл Стілтьєса

$$\mathbb{E}\xi = \int_{-\infty}^{+\infty} x dF_\xi(x) = \begin{cases} \sum_{k=1}^{n(\infty)} x_k \mathbb{P}\{\xi = x_k\}, & \xi - \text{ДВВ} \\ \int_{-\infty}^{+\infty} x f_\xi(x) dx, & \xi - \text{НВВ} \end{cases} \tag{1}$$

Інтеграл або ряд (1) має збігатися *абсолютно*, інакше кажуть, що випадкова величина не має математичного сподівання.

Приклад. НВВ, розподілена за законом Коші з щільністю $f_\xi(x) = \frac{1}{\pi(1+x^2)}$ не має математичного сподівання, бо інтеграл $\frac{1}{\pi}\int_{-\infty}^{+\infty} \frac{x}{1+x^2} dx$ розбіжний.

Математичне сподівання — ймовірнісне середнє значення випадкової величини. Фізична інтерпретація — центр мас системи точок $\{x_1, x_2, ..., x_n, ...\}$ з масами $\{p_1, p_2, ..., p_n, ...\}$ (ДВВ) або стрижня, розподіл маси в якому задано функцією щільності (НВВ).

Властивості математичного сподівання:

1. Математичне сподівання константи — сама константа, оскільки її можна інтерпретувати як ДВВ, що приймає єдине значення з ймовірністю 1.
   $c = const, \mathbb{E}c = c$.
2. $\mathbb{E}(c \cdot \xi) = c \cdot \mathbb{E}\xi$ — властивість рядів та інтегралів.
3. $\mathbb{E}(\xi_1 + \xi_2) = \mathbb{E}\xi_1 + \mathbb{E}\xi_2$.

   Доведення для ДВВ. $\mathbb{P}\{\xi_1 = x_k\} = p_k$, $\mathbb{P}\{\xi_2 = y_j\} = p_j$, $\mathbb{P}\{\xi_1 + \xi_2 = x_k + y_j\} = p_{kj}$.
   $\mathbb{E}(\xi_1 + \xi_2) = \sum_k \sum_j (x_k + y_j) p_{kj} = \sum_k x_k \sum_j p_{kj} + \sum_j y_j \sum_k p_{kj} = \sum_k x_k p_k + \sum_j y_j p_j = \mathbb{E}\xi_1 + \mathbb{E}\xi_2$. ▲

Означення 2.2.2. Дві випадкові величини ξ_1 та ξ_2, задані на одному ймовірнісному просторі, називаються *незалежними*, якщо $\forall x, y$ події $A = \{\omega : \xi_1(\omega) < x\}$ та $B = \{\omega : \xi_2(\omega) < y\}$ є незалежними. В цьому випадку $\mathbb{P}\{\xi_1 < x, \xi_2 < y\} = \mathbb{P}(A \cap B) = \mathbb{P}(A) \cdot \mathbb{P}(B) = F_{\xi_1}(x) \cdot F_{\xi_2}(y)$.

4. Якщо ξ_1 та ξ_2 незалежні, то $\mathbb{E}\xi_1\xi_2 = \mathbb{E}\xi_1 \cdot \mathbb{E}\xi_2$.

   Доведення для ДВВ. Позначимо $p_{kj} = \mathbb{P}\{\xi_1 = x_k, \xi_2 = y_j\}$.
   $$\mathbb{E}\xi_1\xi_2 = \sum_k \sum_j x_k y_j p_{kj} = \sum_k \sum_j x_k y_j p_k p_j = \left(\sum_k x_k p_k\right) \cdot \left(\sum_j y_j p_j\right) = \mathbb{E}\xi_1 \cdot \mathbb{E}\xi_2.$$ ▲

   Властивості 3 та 4 для НВВ буде доведено в темі «Функції випадкових аргументів» (ст. 59).

Приклад. 1. Обчислити математичне сподівання двох ДВВ:

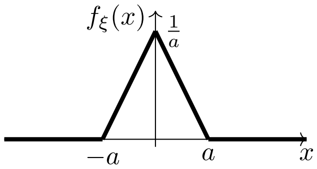
ξ_1	-1	1
p	$1/2$	$1/2$

ξ_2	-100	100
p	$1/2$	$1/2$

$\mathbb{E}\xi_1 = -1 \cdot \frac{1}{2} + 1 \cdot \frac{1}{2} = 0, \mathbb{E}\xi_2 = -100 \cdot \frac{1}{2} + 100 \cdot \frac{1}{2} = 0.$

Цей приклад показує, що попри однакове значення математичного сподівання, можливі значення цих ДВВ знаходяться на різній відстані від нього.

2. Обчислити математичне сподівання НВВ, розподіленої за законом Сімпсона.



$$f_\xi(x) = \begin{cases} \frac{1}{a}\left(1 - \frac{|x|}{a}\right), & |x| \le a \\ 0, & |x| > a \end{cases} (a > 0)$$

$\mathbb{E}\xi = \int_{-\infty}^{+\infty} x f_\xi(x) dx = \int_{-a}^{a} \frac{x}{a}\left(1 - \frac{|x|}{a}\right) dx = 0$, оскільки інтегрується непарна функція по симетричному проміжку.

2.2.2 Дисперсія

Означення 2.2.3. *Дисперсією* випадкової величини називається

$$\mathbb{D}\xi = \mathbb{E}\left(\xi - \mathbb{E}\xi\right)^2 = \int\limits_{-\infty}^{+\infty} \left(x - \mathbb{E}\xi\right)^2 dF_\xi(x) = \begin{cases} \sum\limits_{k=1}^{n(\infty)} \left(x_k - \mathbb{E}\xi\right)^2 \mathbb{P}\left\{\xi = x_k\right\}, & \xi \text{ — ДВВ} \\ \int\limits_{-\infty}^{+\infty} \left(x - \mathbb{E}\xi\right)^2 f_\xi(x) dx, & \xi \text{ — НВВ} \end{cases} \tag{2}$$

Дисперсія — характеристика розсіювання випадкової величини навколо свого математичного сподівання. Фізична інтерпретація — момент інерції маси системи точок або стрижня (як і у випадку інтерпретації математичного сподівання) відносно свого центру мас.

Приклад. 1. Обчислити дисперсії двох ДВВ:

ξ_1	-1	1
p	$1/2$	$1/2$

ξ_2	-100	100
p	$1/2$	$1/2$

$\mathbb{D}\xi_1 = (-1)^2 \cdot \frac{1}{2} + 1^2 \cdot \frac{1}{2} = 1, \mathbb{D}\xi_2 = (-100)^2 \cdot \frac{1}{2} + 100^2 \cdot \frac{1}{2} = 10000.$

2. Обчислити дисперсію НВВ, розподіленої за законом Сімпсона.

Відповідна щільність розподілу має вигляд $f_\xi(x) = \begin{cases} \frac{1}{a}\left(1 - \frac{|x|}{a}\right), & |x| \le a \\ 0, & |x| > a \end{cases}$. Математичне сподівання вже було обчислено, воно рівне 0. Тому $\mathbb{D}\xi = \int\limits_{-\infty}^{+\infty} \left(x - \mathbb{E}\xi\right)^2 f_\xi(x) dx = \int\limits_{-a}^{a} \frac{x^2}{a}\left(1 - \frac{|x|}{a}\right) dx = \frac{2}{a}\int\limits_{0}^{a} x^2 \left(1 - \frac{x}{a}\right) dx = \frac{2}{a} \cdot \left(\frac{x^3}{3} - \frac{x^4}{4a}\right)\Big|_0^a = \frac{2}{a} \cdot \left(\frac{a^3}{3} - \frac{a^3}{4}\right) = \frac{a^2}{6}$.

Властивості дисперсії:

1. $\mathbb{D}\xi \ge 0$, $\sqrt{\mathbb{D}\xi} = \sigma_\xi$ — *середньоквадратичне (стандартне) відхилення.*
2. $\mathbb{D}\xi = 0 \Leftrightarrow \xi = const$, бо $\mathbb{D}c = \mathbb{E}(c - \mathbb{E}c)^2 = 0$.
3. Для обчислення дисперсії більш зручною є наступна формула: $\mathbb{D}\xi = \mathbb{E}(\xi^2 - 2\xi\mathbb{E}\xi + (\mathbb{E}\xi)^2) = \mathbb{E}\xi^2 - 2(\mathbb{E}\xi)^2 + (\mathbb{E}\xi)^2 = \mathbb{E}\xi^2 - (\mathbb{E}\xi)^2$.
4. $\mathbb{D}(c\cdot\xi) = c^2\cdot\mathbb{D}\xi$, оскільки $\mathbb{D}(c\cdot\xi) = \mathbb{E}\left(c\cdot\xi - \mathbb{E}(c\cdot\xi)\right)^2 = \mathbb{E}\left(c\cdot\xi - c\cdot\mathbb{E}\xi\right)^2 = c^2\cdot\mathbb{E}\left(\xi - \mathbb{E}\xi\right)^2 = c^2\cdot\mathbb{D}\xi$, c — константа.
5. $\mathbb{D}(\xi + c) = \mathbb{E}(\xi + c - \mathbb{E}(\xi + c))^2 = \mathbb{E}(\xi + c - \mathbb{E}\xi - \mathbb{E}c) = \mathbb{E}(\xi - \mathbb{E}\xi)^2 = \mathbb{D}\xi$, c — константа.

Означення 2.2.4. *Центрованою* випадковою величиною, що відповідає випадковій величині ξ, називається випадкова величина $\mathring{\xi} = \xi - \mathbb{E}\xi$. Для неї $\mathbb{E}\mathring{\xi} = 0$ та $\mathbb{E}\mathring{\xi}^2 = \mathbb{D}\xi$.

6. Якщо випадкові величини ξ_1 та ξ_2 — незалежні, то $\mathbb{D}\left(\xi_1 + \xi_2\right) = \mathbb{D}\xi_1 + \mathbb{D}\xi_2$.

Доведення. $\mathbb{D}\left(\xi_1 + \xi_2\right) = \mathbb{E}((\xi_1 + \xi_2) - \mathbb{E}(\xi_1 + \xi_2))^2 = \mathbb{E}(\mathring{\xi}_1 + \mathring{\xi}_2)^2 = \mathbb{E}\mathring{\xi}_1^2 + 2\mathbb{E}\mathring{\xi}_1\mathring{\xi}_2 + \mathbb{E}\mathring{\xi}_2^2 = \mathbb{E}\mathring{\xi}_1^2 + 2\mathbb{E}\mathring{\xi}_1\mathbb{E}\mathring{\xi}_2 + \mathbb{E}\mathring{\xi}_2^2 = \mathbb{E}\mathring{\xi}_1^2 + \mathbb{E}\mathring{\xi}_2^2 = \mathbb{D}\xi_1 + \mathbb{D}\xi_2$. ▲

2.2.3 Моменти випадкової величини

Означення 2.2.5. *Початковим моментом k-того порядку* ($k \in \mathbb{N}$) ВВ ξ називається

$$\alpha_k = \mathbb{E}\xi^k = \int\limits_{-\infty}^{+\infty} x^k dF_\xi(x) = \begin{cases} \sum\limits_{m=1}^{n(\infty)} x_m^k \mathbb{P}\left\{\xi = x_m\right\}, & \xi \text{ — ДВВ} \\ \int\limits_{-\infty}^{+\infty} x^k f_\xi(x) dx, & \xi \text{ — НВВ} \end{cases} \tag{3}$$

Означення 2.2.6. *Центральним моментом k-того порядку* ($k \in \mathbb{N}$) ВВ ξ називається

$$\beta_k = \mathbb{E}\mathring{\xi}^k = \mathbb{E}\left(\xi - \mathbb{E}\xi\right)^k = \int\limits_{-\infty}^{+\infty} \left(x - \mathbb{E}\xi\right)^k dF_\xi(x) = \begin{cases} \sum\limits_{m=1}^{n(\infty)} \left(x_m - \mathbb{E}\xi\right)^k \mathbb{P}\left\{\xi = x_m\right\}, & \xi \text{ — ДВВ} \\ \int\limits_{-\infty}^{+\infty} \left(x - \mathbb{E}\xi\right)^k f_\xi(x) dx, & \xi \text{ — НВВ} \end{cases} \tag{4}$$

Зв'язок між центральними та початковими моментами

$$\beta_k = \mathbb{E}\,(\xi - \mathbb{E}\xi)^k = \mathbb{E}\left(\sum_{j=0}^{k} C_k^j \xi^j (-1)^{k-j} (\mathbb{E}\xi)^{k-j}\right) = \sum_{j=0}^{k} (-1)^{k-j} C_k^j \mathbb{E}\xi^j (\mathbb{E}\xi)^{k-j}.$$

Отже, $\beta_k = \sum\limits_{j=0}^{k} (-1)^{k-j} C_k^j \alpha_j (\alpha_1)^{k-j}$. Частковим випадком цієї формули є формула для дисперсії: $\mathbb{D}\xi = \beta_2 = \alpha_2 - \alpha_1^2 = \mathbb{E}\xi^2 - (\mathbb{E}\xi)^2$.

Вправа. Виразити через початкові моменти β_3 та β_4.

Розглядаються також абсолютні початкові моменти $\mathbb{E}|\xi|^k$, абсолютні центральні моменти $\mathbb{E}|\xi - \mathbb{E}\xi|^k$ та факторіальні моменти $\gamma_k = \mathbb{E}\,(\xi(\xi-1)...(\xi-k+1))$.

2.2.4 Мода та медіана випадкової величини

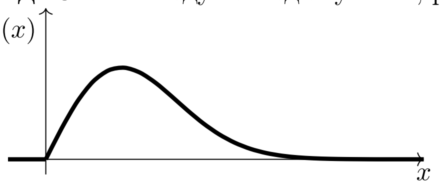
Означення 2.2.7. *Модою* $Mo\xi$ називається абсциса точки максимуму щільності розподілу у випадку НВВ та значення, ймовірність появи якого є найбільшим у випадку ДВВ.

Унімодальний закон — такий, що має лише одну моду. *Полімодальний закон* — такий, що має декілька мод. *Антимодальний закон* — такий, що не має моди. Приклад — закон арксинуса, щільність якого має лише точку мінімуму.

Означення 2.2.8. Точка x_0 називається *медіаною* $Me\xi$, якщо $\mathbb{P}\{\xi < x_0\} = \mathbb{P}\{\xi \geq x_0\} = \frac{1}{2} \Leftrightarrow F_\xi(x_0) = \frac{1}{2}$. Медіана є окремим випадком *квантиля.*

Означення 2.2.9. Точка x_0 називається *квантилем q-го порядку* якщо $F_\xi(x_0) = q$.

Приклад. Знайти моду та медіану НВВ, розподіленої за законом Релея.



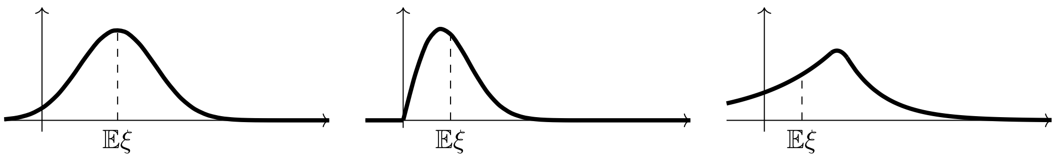
$$f_\xi(x) = \begin{cases} \frac{x}{\sigma^2} e^{-\frac{x^2}{2\sigma^2}}, & x \geq 0 \\ 0, & x < 0 \end{cases} (\sigma > 0)$$

Для визначення моди знайдемо максимум $f_\xi(x)$ при $x \geq 0$. $f'_\xi(x) = \frac{\sigma^2 - x^2}{\sigma^4} e^{-\frac{x^2}{2\sigma^2}} = 0$ при $x = \sigma$. $f''_\xi(x) = \left(-\frac{3x}{\sigma^4} + \frac{x^3}{\sigma^6}\right) e^{-\frac{x^2}{2\sigma^2}}$, $f''_\xi(\sigma) = -\frac{2}{\sigma^3} e^{-\frac{1}{2}} < 0$. Отже, в точці $x = \sigma$ дійсно максимум $f_\xi(x)$, тому $Mo\xi = \sigma$.

Медіану знайдемо з рівності $\mathbb{P}\{\xi < x_0\} = 0.5$. $\mathbb{P}\{\xi < x_0\} = \int\limits_{-\infty}^{x_0} f_\xi(x)dx = \int\limits_{0}^{x_0} \frac{x}{\sigma^2} e^{-\frac{x^2}{2\sigma^2}} dx = 1 - e^{-\frac{x_0^2}{2\sigma^2}}$. З рівняння $1 - e^{-\frac{x_0^2}{2\sigma^2}} = 0.5$ знаходимо $x_0 = \sigma\sqrt{2\ln 2}$.

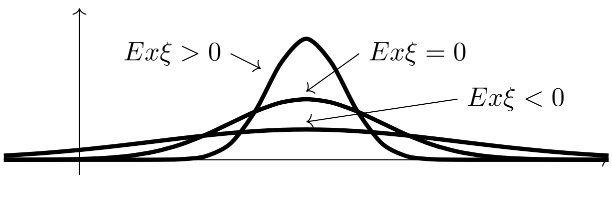
2.2.5 Асиметрія та ексцес випадкової величини

Означення 2.2.10. *Асиметрією* випадкової величини $As\xi$ називається безрозмірна числова характеристика, що дорівнює $\frac{\beta_3}{\sigma_\xi^3} = \frac{\mathbb{E}(\xi - \mathbb{E}\xi)^3}{(\mathbb{D}\xi)^{3/2}}$. Ця характеристика показує порушення чи наявність симетрії кривої розподілу відносно математичного сподівання.



$As\xi = 0$ $As\xi < 0$ $As\xi > 0$

Означення 2.2.11. *Ексцесом* випадкової величини $Ex\xi$ називається безрозмірна числова характеристика, що дорівнює $Ex\xi = \frac{\beta_4}{\sigma_\xi^4} - 3 = \frac{\mathbb{E}(\xi - \mathbb{E}\xi)^4}{(\mathbb{D}\xi)^2} - 3$. Ця характеристика показує, наскільки швидко крива розподілу прямує до точки максимуму.



2.2.6 Генератриса (твірна функція) ДВВ

Означення 2.2.12. Нехай ξ — ДВВ, що приймає цілі невід'ємні значення. *Генератрисою (твірною функцією)* цієї ДВВ називається функція комплексного аргументу

$$G_\xi(z) = \sum_{k=0}^{\infty} \mathbb{P}\{\xi = k\} z^k \tag{5}$$

Властивості генератриси:

1. Відповідний ряд рівномірно збігається принаймні в колі $|z| \le 1$.

   Доведення. $\left|\sum\limits_{k=0}^{\infty} \mathbb{P}\{\xi = k\} z^k\right| \le \sum\limits_{k=0}^{\infty} \mathbb{P}\{\xi = k\} = 1.$ ▲

2. Генератриса — аналітична функція в колі $|z| \le 1$.
3. За генератрисою можна відновити розподіл ξ: $\mathbb{P}\{\xi = k\} = \frac{G_\xi^{(k)}(0)}{k!}$.
4. За допомогою генератриси можна знайти початкові та факторіальні моменти ξ.

   Доведення. За означенням факторіальний момент $\gamma_m = \mathbb{E}(\xi(\xi-1)...(\xi-m+1))$. Для ДВВ, що розглядаються, він рівний $\sum\limits_{k=0}^{\infty} k(k-1)(k-2)...(k-m+1)\mathbb{P}\{\xi = k\}$. З іншого боку, $G'_\xi(z) = \sum\limits_{k=0}^{\infty} k\mathbb{P}\{\xi = k\} z^{k-1}$, $G''_\xi(z) = \sum\limits_{k=0}^{\infty} k(k-1)\mathbb{P}\{\xi = k\} z^{k-2}$ і так далі, $G_\xi^{(m)}(z) = \sum\limits_{k=0}^{\infty} k(k-1)(k-2)...(k-m+1)\mathbb{P}\{\xi = k\} z^{k-m}$. Отже, $\gamma_m = G_\xi^{(m)}(1)$.

   Тепер нескладно знайти, наприклад, математичне сподівання та дисперсію. Факторіальний та початковий моменти першого порядку збігаються, тому $\mathbb{E}\xi = G'_\xi(1)$. Дисперсію знайдемо за допомогою другого факторіального моменту: $\gamma_2 = \mathbb{E}(\xi(\xi-1)) = \mathbb{E}\xi^2 - \mathbb{E}\xi = G''_\xi(1)$, звідки $\mathbb{E}\xi^2 = G''_\xi(1) + G'_\xi(1)$. Отже, $\mathbb{D}\xi = \mathbb{E}\xi^2 - (\mathbb{E}\xi)^2 = G''_\xi(1) + G'_\xi(1) - \left(G'_\xi(1)\right)^2$. ▲

Вправа. Виразити через похідні генератриси центральні моменти 3-го та 4-го порядків.

2.3 Деякі закони розподілу випадкових величин

2.3.1 Біноміальний розподіл

Означення: ДВВ ξ розподілена за *біноміальним законом*, якщо набуває значень $0, 1, ..., n$ з ймовірностями

$$\mathbb{P}\{\xi = k\} = C_n^k p^k q^{n-k}, q = 1 - p \tag{1}$$

Коротке позначення: $\xi \sim \mathrm{Bin}(n, p)$. n і p — параметри закону, $n \in \mathbb{N}$, $p \in (0; 1)$.

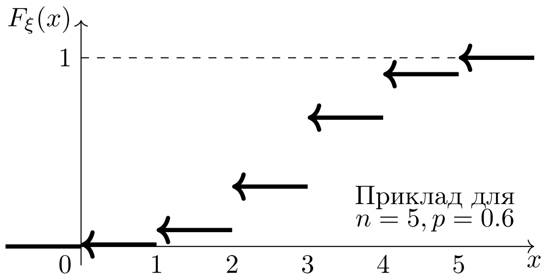
Окремим важливим випадком біноміального закону є *розподіл Бернуллі*: $\xi \sim \mathrm{Bin}(1, p)$. За законом Бернуллі розподілені випадкові величини-індикатори $I_A = \begin{cases} 0, & \text{подія A не відбулась} \\ 1, & \text{подія A відбулась} \end{cases}$ при $\mathbb{P}(A) = p$.

Ряд розподілу:

ξ	0	1	2	...	n
p	q^n	npq^{n-1}	$C_n^2 p^2 q^{n-2}$	...	p^n

Функція розподілу:

$$F_\xi(x) = \begin{cases} 0, & x \le 0 \\ q^n, & 0 < x \le 1 \\ q^n + npq^{n-1}, & 1 < x \le 2 \\ \dots & \\ 1, & x > n \end{cases}$$



Для дослідження числових характеристик скористаємося генератрисою розподілу:

$G_\xi(z) = \sum\limits_{k=0}^{n} \mathbb{P}\{\xi = k\} z^k = \sum\limits_{k=0}^{n} C_n^k p^k q^{n-k} z^k = (pz+q)^n$. $G'_\xi(z) = np(pz+q)^{n-1}$, $G''_\xi(z) = n(n-1)p^2(pz+q)^{n-2}$.

$G'_\xi(1) = np(p+q)^{n-1} = np$, $G''_\xi(1) = n(n-1)p^2 = n^2p^2 - np^2$, $G''_\xi(1) + G'_\xi(1) - \Big(G'_\xi(1)\Big)^2 = n^2p^2 - np^2 + np - n^2p^2 = np(1-p) = npq$. Отже, знайдено значення математичного сподівання та дисперсії.

Числові характеристики:

1. $\mathbb{E}\xi = np$.
2. $\mathbb{D}\xi = npq$, $\sigma_\xi = \sqrt{npq}$.
3. $Mo\xi = \begin{cases} [np+p], & \text{якщо } np+p \text{ не ціле} \\ np+p, np-q, & \text{якщо } np+p \text{ ціле} \end{cases}$ — як найбільш ймовірна кількість успіхів у схемі Бернуллі.
4. $Me\xi$ — одне зі значень $[np]-1$, $[np]$, $[np]+1$.

Застосування: якщо проводиться n незалежних випробувань з ймовірністю успіху p, то $\xi \sim \mathrm{Bin}(n,p)$ задає кількість успіхів.

2.3.2 Геометричний розподіл

Означення: ДВВ ξ розподілена за *геометричним законом*, якщо набуває значень $1, 2, 3, ..$ з ймовірностями

$$\mathbb{P}\{\xi = k\} = pq^{k-1}, q = 1-p \tag{2}$$

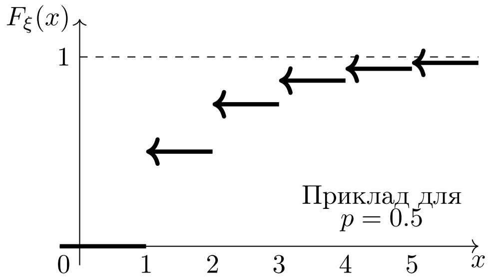
Коротке позначення: $\xi \sim \mathrm{Geom}(p)$. p — параметр закону, $p \in (0;1)$.

Ряд розподілу:

ξ	1	2	3	...	k	...
p	p	pq	pq^2	...	pq^{k-1}	...

Функція розподілу:

$$F_\xi(x) = \begin{cases} 0, & x \le 1 \\ p, & 1 < x \le 2 \\ p+pq, & 2 < x \le 3 \\ \dots \\ \sum\limits_{m=0}^{k-1} pq^m = 1-q^k, & k < x \le k+1 \\ \dots \end{cases}$$



Для дослідження числових характеристик скористаємося генератрисою розподілу:

$G_\xi(z) = \sum\limits_{k=1}^{\infty} \mathbb{P}\{\xi = k\} z^k = \sum\limits_{k=1}^{\infty} pq^{k-1}z^k = \frac{pz}{1-qz}$, якщо $|qz| < 1$. $G'_\xi(z) = \frac{p}{(1-qz)^2}$, $G''_\xi(z) = \frac{2pq}{(1-qz)^3}$. $G'_\xi(1) = \frac{1}{p}$, $G''_\xi(1) = \frac{2q}{p^2}$, $G''_\xi(1) + G'_\xi(1) - \Big(G'_\xi(1)\Big)^2 = \frac{2q}{p^2} + \frac{1}{p} - \frac{1}{p^2} = \frac{2q+p-1}{p^2} = \frac{q}{p^2}$. Отже, знайдено значення математичного сподівання та дисперсії.

Числові характеристики:

1. $\mathbb{E}\xi = \frac{1}{p}$.
2. $\mathbb{D}\xi = \frac{q}{p^2}$, $\sigma_\xi = \frac{\sqrt{q}}{p}$.
3. $Mo\xi = 1$.
4. $Me\xi = \left[\frac{-1}{\log_2(1-p)}\right]$.

Застосування: якщо проводяться незалежні випробування з ймовірністю успіху p до першого успішного, то ξ, що задає кількість проведених випробувань, має розподіл $\mathrm{Geom}(p)$.

Вправа. Записати ряд розподілу, функцію розподілу, генетратрису та обчислити числові характеристики для іншого означення геометричного розподілу, де ξ приймає значення $0, 1, 2, ...$ з ймовірностями $\mathbb{P}\{\xi = k\} = pq^k$.

Геометричний розподіл **«не має пам'яті»**: $\mathbb{P}\{\xi = n+m/\xi \ge n\} = \mathbb{P}\{\xi = m\}$. Це означає, що кількість минулих «невдач» не впливає на кількість майбутніх «невдач».

Вправа. Довести цю властивість і те, що геометричний розподіл — єдиний дискретний розподіл, що «не має пам'яті».

До геометричного закону можна звести *закон розподілу Паскаля* $\xi \sim \text{Pas}(a)$, що задається $\mathbb{P}\{\xi = k\} = \frac{a^k}{(1+a)^{k+1}}, k = 0, 1, 2, ..., a > 0$, заміною $p = \frac{1}{a+1}$. Для нього $\mathbb{E}\xi = a$, $\mathbb{D}\xi = a + a^2$.

2.3.3 Розподіл Пуассона

Означення: ДВВ ξ розподілена за *законом Пуассона*, якщо набуває значень $0, 1, 2, ..$ з ймовірностями

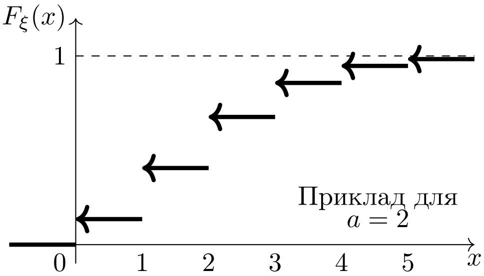
$$\mathbb{P}\{\xi = k\} = \frac{a^k}{k!} e^{-a} \tag{3}$$

Коротке позначення: $\xi \sim \text{Poiss}(a)$. a — параметр закону, $a > 0$.
Ряд розподілу:

ξ	0	1	2	...	k	...
p	e^{-a}	ae^{-a}	$\frac{a^2}{2}e^{-a}$	...	$\frac{a^k}{k!}e^{-a}$	...

Функція розподілу:

$$F_\xi(x) = \begin{cases} 0, & x \le 0 \\ e^{-a}, & 0 < x \le 1 \\ e^{-a} + ae^{-a}, & 1 < x \le 2 \\ \dots & \\ \sum\limits_{m=0}^{k-1} \frac{a^m}{m!} e^{-a} & k-1 < x \le k \\ \dots & \end{cases}$$



Для дослідження числових характеристик скористаємося генератрисою розподілу: $G_\xi(z) = \sum\limits_{k=0}^{\infty} \mathbb{P}\{\xi = k\} z^k = e^{-a} \sum\limits_{k=0}^{\infty} \frac{a^k}{k!} z^k = e^{-a(1-z)}$. $G'_\xi(z) = ae^{az-a}$, $G''_\xi(z) = a^2 e^{az-a}$. $G'_\xi(1) = a$, $G''_\xi(1) = a^2$, $G''_\xi(1) + G'_\xi(1) - \left(G'_\xi(1)\right)^2 = a^2 + a - a^2 = a$. Отже, знайдено значення математичного сподівання та дисперсії.
Числові характеристики:

1. $\mathbb{E}\xi = a$.
2. $\mathbb{D}\xi = a$, $\sigma_\xi = \sqrt{a}$.
3. $Mo\xi = [a]$.
4. $Me\xi \approx [a + 1/3 - 0.02/a]$.

2.3.4 Потік Пуассона

Потоком називають послідовність подій, які наступають в певні моменти часу одна за одною. Потік характеризується випадковою величиною $\xi(t)$ — кількістю подій, що наступили протягом проміжку часу $[0; t)$. Ймовірності $\mathbb{P}\{\xi(t) = m\}$ позначаються $p_m(t)$.

Накладемо на потік такі вимоги:

1. *Стаціонарність (однорідність)* — кількість подій, що наступають за певний проміжок часу, залежить лише від довжини проміжку і не залежить від того, де цей проміжок розташований на часовій осі.
2. *Відсутність післядії* — якщо проміжки часу не перетинаються, то кількості подій, які за ці проміжки відбулися, є незалежними подіями.
3. *Ординарність* — події наступають поодинці, тобто $\mathbb{P}\{\xi(t + \Delta t) - \xi(t) = 1\} = \lambda \cdot \Delta t + o(\Delta t)$ ($\lambda > 0$ — параметр інтенсивності) і $\mathbb{P}\{\xi(t + \Delta t) - \xi(t) \ge 2\} = o(\Delta t)$.

Означення 2.3.1. Потік, що має перелічені властивості, називається *потоком Пуассона.*

Отримаємо явний вираз для ймовірностей $p_m(t)$ у потоці Пуассона.

Позначимо $\tilde{p}_k(t + \Delta t) = \mathbb{P}\{\xi(t + \Delta t) - \xi(t) = k\}$. Тоді $p_m(t + \Delta t) = p_m(t) \cdot \tilde{p}_0(t + \Delta t) + p_{m-1}(t) \cdot \tilde{p}_1(t + \Delta t) + p_{m-2}(t) \cdot \tilde{p}_2(t + \Delta t) + ... + p_0(t) \cdot \tilde{p}_m(t + \Delta t)$ для $m \ge 1$, для $m = 0$ $p_0(t + \Delta t) = p_0(t) \cdot \tilde{p}_0(t + \Delta t)$. Застосуємо ординарність:

$p_0(t + \Delta t) = p_0(t) \cdot (1 - \lambda \cdot \Delta t + o(\Delta t))$

$p_m(t + \Delta t) = p_m(t) \cdot (1 - \lambda \cdot \Delta t + o(\Delta t)) + p_{m-1}(t) \cdot (\lambda \cdot \Delta t + o(\Delta t)) + o(\Delta t)$

Розкриємо дужки:

$p_0(t + \Delta t) - p_0(t) = -\lambda \cdot \Delta t \cdot p_0(t) + o(\Delta t)$

$p_m(t + \Delta t) - p_m(t) = -\lambda \cdot \Delta t \cdot p_m(t) + \lambda \cdot \Delta t \cdot p_{m-1}(t) + o(\Delta t)$

Поділимо на Δt та перейдемо до границі при $\Delta t \to 0$. Отримаємо систему диференціальних рівнянь:

$$\begin{cases} p_0'(t) = -\lambda p_0(t) \\ p_m'(t) = -\lambda p_m(t) + \lambda p_{m-1}(t),\ m \geq 1 \end{cases}$$

Щоб розв'язати цю систему, введемо генератрису ймовірностей $p_m(t)$:

$$G_\xi(z,t) = \sum_{m=0}^{\infty} \mathbb{P}\left\{\xi(t) = m\right\} z^m = \sum_{m=0}^{\infty} p_m(t) z^m$$

Помножимо кожне рівняння отриманої системи на z у відповідному степені та складемо:

$$\sum_{m=0}^{\infty} p_m'(t) z^m = -\lambda \sum_{m=0}^{\infty} p_m(t) z^m + \lambda \sum_{m=1}^{\infty} p_{m-1}(t) z^m = -\lambda \sum_{m=0}^{\infty} p_m(t) z^m + \lambda z \sum_{m=0}^{\infty} p_m(t) z^m$$

Отримали диференціальне рівняння для генератриси:

$$\frac{\partial G_\xi}{\partial t} = -\lambda G_\xi + \lambda z G_\xi = -\lambda(1-z) G_\xi$$

Його розв'язком буде $G_\xi(z,t) = C(z) \cdot e^{-\lambda(1-z)t}$. Оскільки $G_\xi(z,0) = 1$, то $C(z) = 1$.

Отже, генератриса рівна $G_\xi(z,t) = e^{-\lambda(1-z)t}$ — це генератриса закону Пуассона з параметром $a = \lambda t$.

Таким чином, $p_m(t) = \frac{(\lambda t)^m}{m!} e^{-\lambda t}$. Тепер зрозумілою стає назва параметру λ («інтенсивність»): $\mathbb{E}\xi = \lambda t$, $\lambda = \frac{\mathbb{E}\xi}{t}$ — середня кількість подій за одиницю часу.

2.3.5 Рівномірний розподіл

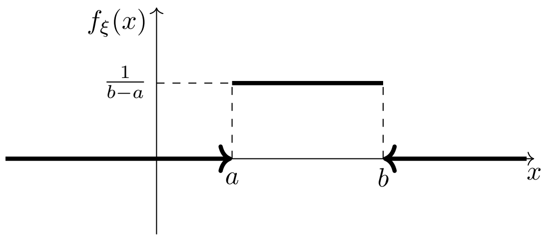
Означення: НВВ ξ розподілена за *рівномірним законом*, якщо її щільність розподілу має вигляд

$$f_\xi(x) = \begin{cases} \frac{1}{b-a}, & x \in \langle a; b \rangle \\ 0, & x \notin \langle a; b \rangle \end{cases} \tag{4}$$

Коротке позначення: $\xi \sim \mathrm{U}\langle a; b\rangle$. a і b — параметри закону, $a < b$, $a, b \in \mathbb{R}$.

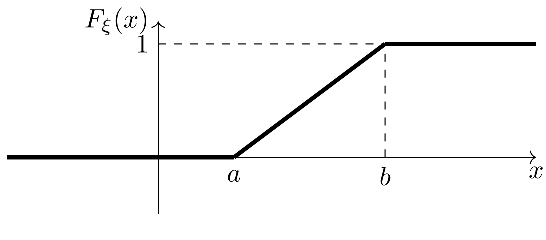
Розподіл $\mathrm{U}\langle 0; 1\rangle$ називається *стандартним рівномірним розподілом.*

Крива розподілу:



Функція розподілу:

$$F_\xi(x) = \begin{cases} 0, & x \leq a \\ \frac{x-a}{b-a}, & a < x \leq b \\ 1, & x > b \end{cases}$$



Числові характеристики:

1. $\mathbb{E}\xi = \int\limits_{-\infty}^{+\infty} x f_\xi(x) dx = \int\limits_a^b \frac{x}{b-a} dx = \frac{a+b}{2}$.
2. $\mathbb{D}\xi = \mathbb{E}\xi^2 - (\mathbb{E}\xi)^2 = \int\limits_a^b \frac{x^2}{b-a} dx - \frac{(a+b)^2}{4} = \frac{(b-a)^2}{12}$, $\sigma_\xi = \frac{b-a}{2\sqrt{3}}$.
3. $Mo\xi$ — будь-яка точка з $\langle a; b\rangle$.
4. $Me\xi = \frac{a+b}{2}$.
5. $As\xi = 0$.
6. $Ex\xi = -\frac{6}{5}$.

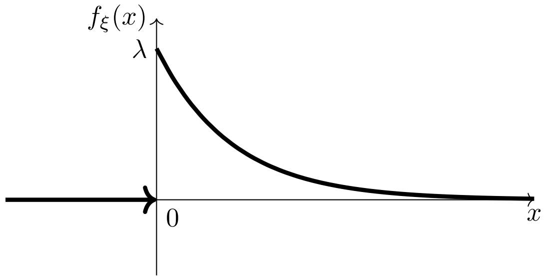
Застосування: отримання вибірок з інших законів розподілу; похибка округлення до найближчої поділки приладу при ціні поділки a має розподіл $\mathrm{U}\left\langle -\frac{a}{2}; \frac{a}{2}\right\rangle$.

2.3.6 Експоненційний (показниковий) розподіл

Означення: НВВ ξ розподілена за *експоненційним законом*, якщо її щільність розподілу має вигляд

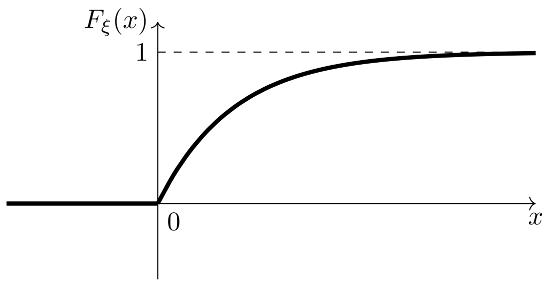
$$f_\xi(x) = \begin{cases} \lambda e^{-\lambda x}, & x \geq 0 \\ 0, & x < 0 \end{cases} \tag{5}$$

Коротке позначення: $\xi \sim \mathrm{Exp}(\lambda)$. $\lambda > 0$ — параметр закону.
Крива розподілу:



Функція розподілу:

$$F_\xi(x) = \begin{cases} \lambda \int\limits_0^x e^{-\lambda t} dt = 1 - e^{-\lambda x}, & x > 0 \\ 0, & x \leq 0 \end{cases}$$



Знайдемо всі *початкові моменти* експоненційного розподілу.
$\mathbb{E}\xi^k = \int\limits_{-\infty}^{+\infty} x^k f_\xi(x)dx = \lambda \int\limits_0^{+\infty} x^k e^{-\lambda x}dx = \left[\lambda x = t, dx = \frac{dt}{\lambda}\right] = \frac{\lambda}{\lambda^k \cdot \lambda} \int\limits_0^{+\infty} t^k e^{-t}dt = \frac{\Gamma(k+1)}{\lambda^k} = \frac{k!}{\lambda^k}$.

Числові характеристики:
1. $\mathbb{E}\xi = \frac{1}{\lambda}$.
2. $\mathbb{D}\xi = \mathbb{E}\xi^2 - (\mathbb{E}\xi)^2 = \frac{2}{\lambda^2} - \frac{1}{\lambda^2} = \frac{1}{\lambda^2}$, $\sigma_\xi = \frac{1}{\lambda}$.
3. $Mo\xi = 0$.
4. $Me\xi = \frac{\ln 2}{\lambda}$.
5. $As\xi = 2$ — не залежить від λ, бо $\beta_3 = \alpha_3 - 3\alpha_2\alpha_1 + 2\alpha_1^3 = \frac{2}{\lambda^3}$.
6. $Ex\xi = 6$.

Вправа. Дослідити характеристики інших варіантів експоненційного розподілу: $\mathrm{Exp}(\lambda, x_0)$ з щільністю $f_\xi(x) = \begin{cases} \lambda e^{-\lambda(x-x_0)}, & x \geq x_0 \\ 0, & x < x_0 \end{cases}$ (експоненційний із зсувом) та $\mathrm{Exp}(\alpha)$ зі щільністю $f_\xi(x) = \begin{cases} \frac{1}{\alpha} e^{-\frac{x}{\alpha}}, & x \geq 0 \\ 0, & x < 0 \end{cases}$, що відповідає розглянутому варіанту, але з параметром $\lambda = \frac{1}{\alpha}$.

Як і геометричний розподіл, експоненційний **«не має пам'яті»**. Доведемо: якщо проміжок часу T, розподілений за експоненційним законом, тягнувся час τ, то проміжок $T_1 = T - \tau$ розподілений так само — за експоненційним законом з тим самим параметром.

$$F_{T_1}(t) = \mathbb{P}\{T_1 < t / T \geq \tau\} = \frac{\mathbb{P}\{(T_1 < t) \cap (T \geq \tau)\}}{\mathbb{P}\{T \geq \tau\}} = \frac{\mathbb{P}\{\tau \leq T < t + \tau\}}{1 - \mathbb{P}\{T < \tau\}} =$$
$$\frac{F_T(t+\tau) - F_T(\tau)}{1 - F_T(\tau)} = \frac{1 - e^{-\lambda(t+\tau)} - (1 - e^{-\lambda\tau})}{1 - (1 - e^{-\lambda\tau})} = \frac{e^{-\lambda\tau}(1 - e^{-\lambda t})}{e^{-\lambda\tau}} = 1 - e^{-\lambda t}$$

Отже, $F_{T_1}(t) = F_T(t)$.
Застосування:
1. Час між двома сусідніми подіями в потоці Пуассона має експоненційний розподіл.

Вправа. Довести цю властивість.

2. Як правило, час безвідмовної роботи приладу має експоненційний розподіл.
Нехай в момент часу $t = 0$ увімкнули прилад. Припустимо, що умовна ймовірність виходу з ладу приладу в інтервалі часу $[t; t + \Delta t)$ за умови, що до часу t він працював, пропорційна Δt і рівна $\lambda \cdot \Delta t + o(\Delta t)$. Нехай ξ — час безвідмовної роботи, покладемо $Q(t) = \mathbb{P}\{\xi \geq t\}$. $Q(t+\Delta t) = \mathbb{P}\{\xi \geq t + \Delta t\} = \mathbb{P}\{\xi \geq t\} \cdot \mathbb{P}\{\xi \geq t + \Delta t / \xi \geq t\} = Q(t) \cdot (1 - \lambda \cdot \Delta t + o(\Delta t))$. $\frac{Q(t+\Delta t) - Q(t)}{\Delta t} = -\lambda Q(t) + \frac{o(\Delta t)}{\Delta t} \cdot Q(t)$, при $\Delta t \to 0$ отримаємо $Q'(t) = -\lambda Q(t)$. Розв'язком цього диференціального рівняння буде $Q(t) = C \cdot e^{-\lambda t}$. Оскільки вважаємо, що в момент часу $t = 0$ прилад працював, то $Q(0) = 1$ і $C = 1$.
Тоді $F_\xi(t) = \mathbb{P}\{\xi < t\} = 1 - \mathbb{P}\{\xi \geq t\} = 1 - Q(t) = \begin{cases} 1 - e^{-\lambda t}, & t \geq 0 \\ 0, & t < 0 \end{cases}$. Отже, $\xi \sim \mathrm{Exp}(\lambda)$.

Приклад. Яка ймовірність того, що людина проживе 100 років? Нехай ξ — час життя людини. Оскільки $\mathbb{P}\{\xi = 100\} = 0$, знайдемо $\mathbb{P}\{\xi \geq 100\}$. $\mathbb{P}\{\xi \geq 100\} = e^{-\lambda \cdot 100}$, $\lambda = ?$. Оскільки $\lambda = \frac{1}{\mathbb{E}\xi}$, то, прийнявши $\mathbb{E}\xi = 70$ (середня тривалість життя), отримаємо $\mathbb{P}\{\xi \geq 100\} = e^{-100/70} \approx 0.24$.

2.3.7 Гауссівський (нормальний) розподіл

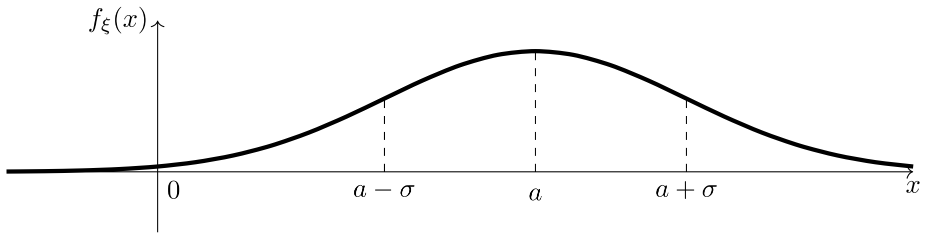
Означення: НВВ ξ розподілена за *нормальним законом*, якщо її щільність розподілу має вигляд

$$f_\xi(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-a)^2}{2\sigma^2}} \tag{6}$$

Коротке позначення: $\xi \sim \mathrm{N}(a, \sigma)$ або $\xi \sim \mathrm{N}(a, \sigma^2)$. $a \in \mathbb{R}, \sigma > 0$ — параметри закону.

Розподіл $\mathrm{N}(0, 1)$ називають стандартним гауссівським розподілом.

Крива розподілу:



Зміст параметрів a та σ^2:

1. $\mathbb{E}\xi = \frac{1}{\sqrt{2\pi}\sigma} \int\limits_{-\infty}^{+\infty} x e^{-\frac{(x-a)^2}{2\sigma^2}} dx = \left[\frac{x-a}{\sigma\sqrt{2}} = t,\ dx = \sigma\sqrt{2} dt\right] = \frac{1}{\sqrt{\pi}} \int\limits_{-\infty}^{+\infty} (a + \sigma\sqrt{2}t) e^{-t^2} dt =$
$= \frac{a}{\sqrt{\pi}} \int\limits_{-\infty}^{+\infty} e^{-t^2} dt = [\text{інтеграл Ейлера-Пуассона}] = a.$

2. Знайдемо всі центральні моменти: $\beta_k = \mathbb{E}(\xi - \mathbb{E}\xi)^k = \frac{1}{\sigma\sqrt{2\pi}} \int\limits_{-\infty}^{+\infty} (x-a)^k e^{-\frac{(x-a)^2}{2\sigma^2}} dx =$
$[\text{ідентична заміна}] = \frac{\sigma^k 2^{k/2}}{\sqrt{\pi}} \int\limits_{-\infty}^{+\infty} t^k e^{-t^2} dt.$
Якщо $k = 2l + 1$, то $\beta_k = 0$ (непарна функція під інтегралом). $\beta_3 = 0 \Rightarrow As\xi = 0$.
Візьмемо $k = 2l$:
$\beta_{2l} = \frac{\sigma^{2l} 2^l}{\sqrt{\pi}} \int\limits_{-\infty}^{+\infty} t^{2l} e^{-t^2} dt = \left[t^2 = z,\ dt = \frac{1}{2} z^{-\frac{1}{2}} dz\right] = 2\frac{\sigma^{2l} 2^l}{\sqrt{\pi}} \int\limits_{0}^{+\infty} z^l e^{-z} \cdot \frac{1}{2} z^{-\frac{1}{2}} dz =$
$= \frac{\sigma^{2l} 2^l}{\sqrt{\pi}} \int\limits_{0}^{+\infty} z^{l - \frac{1}{2}} e^{-z} dz = \frac{\sigma^{2l} 2^l}{\sqrt{\pi}} \cdot \Gamma(l + \frac{1}{2}).$

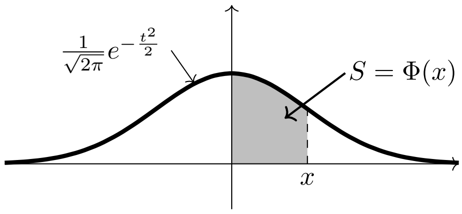
Числові характеристики:

1. $\mathbb{E}\xi = Mo\xi = Me\xi = a$.
2. $\mathbb{D}\xi = \frac{2\sigma^2}{\sqrt{\pi}} \cdot \Gamma\left(\frac{3}{2}\right) = \sigma^2$, $\sigma_\xi = \sigma$.
3. $As\xi = 0$.
4. $Ex\xi = \frac{\beta_4}{\sigma_\xi^4} - 3 = \frac{\sigma^4 \cdot 2^2 \cdot \Gamma(5/2)}{\sqrt{\pi} \cdot \sigma^4} - 3 = \frac{4 \cdot \sqrt{\pi} \cdot 3/4}{\sqrt{\pi}} - 3 = 0$. Доданок -3 у формулі ексцесу був введений саме для того, щоб нормальний розподіл мав нульовий ексцес.

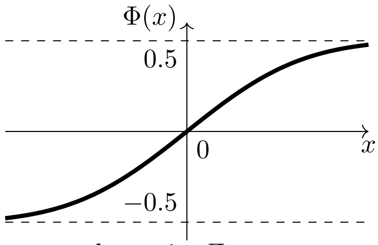
Функція розподілу:
$F_\xi(x) = \int\limits_{-\infty}^{x} f_\xi(t)dt = \frac{1}{\sqrt{2\pi}\sigma}\int\limits_{-\infty}^{x} e^{-\frac{(t-a)^2}{2\sigma^2}}dt$ — не береться в елементарних функціях.

Для зручного використання гауссівського закону в задачах введемо спеціальну функцію, значення якої занесено до таблиць.
Функція Лапласа:

$$\Phi(x) = \frac{1}{\sqrt{2\pi}}\int\limits_0^x e^{-\frac{t^2}{2}}dt$$



Властивості функції Лапласа:
1. $\Phi(-x) = -\Phi(x)$ — непарна.
2. $\Phi(0) = 0$.
3. $y = \pm 0.5$ — горизонтальні асимптоти.
4. $\forall x \geq 5 : \Phi(x) \approx 0.5$.
5. $\forall x \leq -5 : \Phi(x) \approx -0.5$.



Виразимо функцію розподілу гауссівського закону через функцію Лапласа:

$$F_\xi(x) = \frac{1}{\sqrt{2\pi}\sigma}\int\limits_{-\infty}^{x} e^{-\frac{(t-a)^2}{2\sigma^2}}dt = \left[\frac{t-a}{\sigma} = z\right] = \frac{1}{\sqrt{2\pi}}\int\limits_{-\infty}^{\frac{x-a}{\sigma}} e^{-\frac{z^2}{2}}dz = \frac{1}{2} + \Phi\left(\frac{x-a}{\sigma}\right) \tag{7}$$

Робочі формули для розв'язання задач:
1. $\mathbb{P}\{\xi \in \langle\alpha;\beta\rangle\} = F_\xi(\beta) - F_\xi(\alpha) = \Phi\left(\frac{\beta-a}{\sigma}\right) - \Phi\left(\frac{\alpha-a}{\sigma}\right)$.
2. $\mathbb{P}\{|\xi - a| < \varepsilon\} = \mathbb{P}\{\xi \in (a-\varepsilon, a+\varepsilon)\} = \Phi\left(\frac{a+\epsilon-a}{\sigma}\right) - \Phi\left(\frac{a-\epsilon-a}{\sigma}\right) = 2\Phi(\frac{\varepsilon}{\sigma})$.
3. *«Правило 3σ»*: $\mathbb{P}\{|\xi - a| < 3\sigma\} = \mathbb{P}\{\xi \in (a-3\sigma; a+3\sigma)\} = 2\Phi(3) \approx 0.9973$.
   Майже всі значення випадкової величини, розподіленої за гауссівським законом, лежать на відстані не більше трьох середньоквадратичних відхилень від її математичного сподівання.

2.3.8 Гамма-розподіл

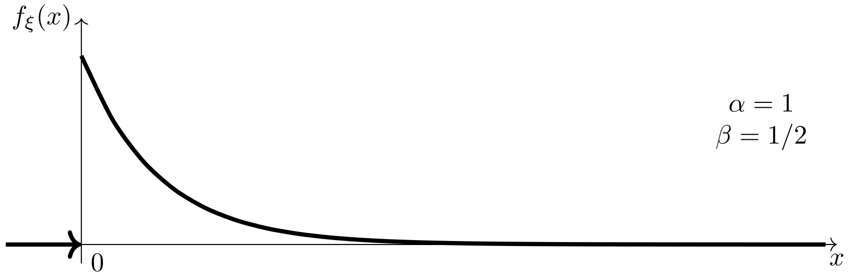
Означення: НВВ ξ розподілена за *гамма-законом*, якщо її щільність розподілу має вигляд

$$f_\xi(x) = \begin{cases} \frac{\beta^\alpha}{\Gamma(\alpha)}x^{\alpha-1}e^{-\beta x}, & x \geq 0 \\ 0, & x < 0 \end{cases} \tag{8}$$

Коротке позначення: $\xi \sim \Gamma(\alpha,\beta)$. $\alpha > 0, \beta > 0$ — параметри закону.

Зауваження. $\Gamma(1,\lambda) = \mathrm{Exp}(\lambda)$.

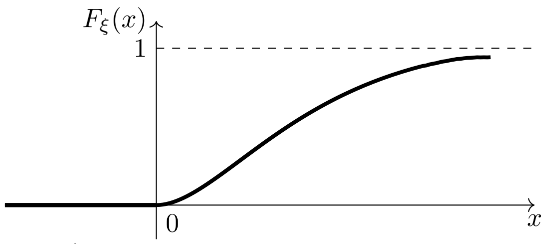
Крива розподілу:



Функція розподілу:

$$F_\xi(x) = \begin{cases} 0, & x \le 0 \\ \int\limits_0^x f_\xi(t)dt = \frac{\gamma(\alpha,\beta x)}{\Gamma(\alpha)}, & x > 0 \end{cases}$$

де $\gamma(s,x) = \int\limits_0^x t^{s-1}e^{-t}dt$



Знайдемо всі *початкові моменти* гамма-розподілу.

$\mathbb{E}\xi^k = \int\limits_{-\infty}^{+\infty} x^k f_\xi(x)dx = \frac{\beta^\alpha}{\Gamma(\alpha)} \int\limits_0^{+\infty} x^k x^{\alpha-1}e^{-\beta x}dx = [\beta x = t] = \frac{\beta^\alpha}{\Gamma(\alpha)} \cdot \frac{1}{\beta^k \cdot \beta^\alpha} \int\limits_0^{+\infty} t^{k+\alpha-1}e^{-t}dt = \frac{1}{\Gamma(\alpha)\cdot\beta^k} \int\limits_0^{+\infty} t^{k+\alpha-1}e^{-t}dt = \frac{\Gamma(\alpha+k)}{\Gamma(\alpha)\cdot\beta^k}$. Зокрема, $\mathbb{E}\xi = \frac{\Gamma(\alpha+1)}{\Gamma(\alpha)\cdot\beta} = \frac{\alpha}{\beta}$, $\mathbb{E}\xi^2 = \frac{\Gamma(\alpha+2)}{\Gamma(\alpha)\cdot\beta^2} = \frac{(\alpha+1)\alpha}{\beta^2}$.

Числові характеристики:

1. $\mathbb{E}\xi = \frac{\alpha}{\beta}$.
2. $\mathbb{D}\xi = \mathbb{E}\xi^2 - (\mathbb{E}\xi)^2 = \frac{(\alpha+1)\alpha}{\beta^2} - \frac{\alpha^2}{\beta^2} = \frac{\alpha}{\beta^2}$, $\sigma_\xi = \frac{\sqrt{\alpha}}{\beta}$.
3. $Mo\xi = \frac{\alpha-1}{\beta}$.
4. $Me\xi$ — немає виразу у замкненій формі.
5. $As\xi = \frac{2}{\sqrt{\alpha}}$.
6. $Ex\xi = \frac{6}{\alpha}$.

Застосування: гамма-розподіл застосовується для моделювання складних потоків подій, в економіці, теорії масового обслуговування, логістиці. У випадку натурального параметру $\alpha = k \in \mathbb{N}$ за законом $\Gamma(k,\lambda)$ розподілений час очікування появи k-тої події в процесі Пуассона з інтенсивністю λ. Така версія гамма-розподілу називається *розподілом Ерланга.*

Вправа. Нехай T — час очікування появи k-тої події в процесі Пуассона з інтенсивністю λ. Довести $T \sim \Gamma(k,\lambda)$.

Розділ 3

Випадкові вектори

3.1 Дискретні випадкові вектори

Означення 3.1.1. Вимірна функція $\Omega \to \mathbb{R}^n$, задана на ймовірнісному просторі $\{\Omega, \mathcal{F}, \mathbb{P}\}$, називається *випадковим вектором (системою випадкових величин)*. Позначається $\vec{\xi} = \vec{\xi}(\omega) = (\xi_1(\omega), \xi_2(\omega), ..., \xi_n(\omega))^T$.

Зауваження. Під вимірністю мається на увазі те, що

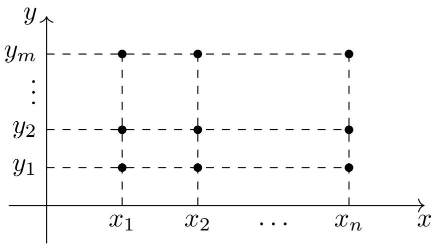
$$\forall \vec{x} \in \mathbb{R}^n : A = \{\omega \in \Omega : \xi_1(\omega) < x_1, \xi_2(\omega) < x_2, ..., \xi_n(\omega) < x_n\} \in \mathcal{F}$$

Означення 3.1.2. Випадковий вектор називається *дискретним*, якщо всі його координати — дискретні випадкові величини.

Зауваження. Випадковий вектор можна трактувати як випадкову точку в $\mathbb{R}^n$. При $n = 2$: $\vec{\xi} = (\xi_1, \xi_2)^T$ — випадкова точка на площині.

Закон розподілу двовимірного дискретного випадкового вектора задається *таблицею розподілу.*

ξ_2 \ ξ_1	x_1	x_2	...	x_n
y_1	p_{11}	p_{21}	...	p_{n1}
y_2	p_{12}	p_{22}	...	p_{n2}
...	...	...	p_{ij}	...
y_m	p_{1m}	p_{2m}	...	p_{nm}



$p_{ij} = \mathbb{P}\{\xi_1 = x_i, \xi_2 = y_j\}$, $\sum\limits_{i,j} p_{ij} = 1$.

З таблиці розподілу можна обчислити ряди розподілу ξ_1 та ξ_2:

ξ_1	x_1	x_2	...	x_n
p	$\sum\limits_{k=1}^{m} p_{1k}$	$\sum\limits_{k=1}^{m} p_{2k}$	...	$\sum\limits_{k=1}^{m} p_{nk}$

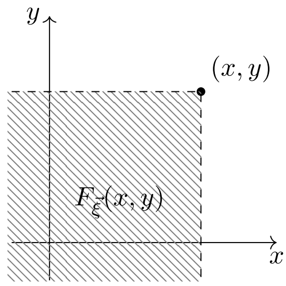
ξ_2	y_1	y_2	...	y_m
p	$\sum\limits_{k=1}^{n} p_{k1}$	$\sum\limits_{k=1}^{n} p_{k2}$	...	$\sum\limits_{k=1}^{n} p_{km}$

Зауваження. Обчислити таблицю розподілу, знаючи ряди розподілу координат, можна лише у випадку їх незалежності.

3.1.1 Сумісна функція розподілу двовимірного випадкового вектора

Означення 3.1.3. *Сумісною функцією розподілу* двовимірного випадкового вектора $\vec{\xi}$ називається $F_{\vec{\xi}}(x, y) = \mathbb{P}\{\omega : \xi_1 < x, \xi_2 < y\}$, $x, y \in \mathbb{R}$.

Геометрична інтерпретація: Значення $F_{\vec{\xi}}(x, y)$ рівне імовірності потрапляння точки в заштриховану зону.



Властивості:

1. $D(F_{\vec{\xi}}) = \mathbb{R}^2$, $E(F_{\vec{\xi}}) = \langle 0; 1\rangle$.
2. Монотонно неспадна по кожній з координат.

   Доведення. $A = \{\xi_1 < x_2,\ \xi_2 < y\}$, $B = \{x_2 \le \xi_1 < x_1,\ \xi_2 < y\}$, $C = \{\xi_1 < x_1,\ \xi_2 < y\}$, $x_1 > x_2$. $C = A \cup B \Rightarrow \mathbb{P}(C) = \mathbb{P}(A) + \mathbb{P}(B)$, $\mathbb{P}(C) = F_{\vec{\xi}}(x_1, y)$, $\mathbb{P}(A) = F_{\vec{\xi}}(x_2, y)$.
   $F_{\vec{\xi}}(x_1, y) = F_{\vec{\xi}}(x_2, y) + \mathbb{P}(B) \Rightarrow F_{\vec{\xi}}(x_1, y) \ge F_{\vec{\xi}}(x_2, y)$.
   Для другої координати аналогічно. ▲

3. $\lim\limits_{x\to-\infty} F_{\vec{\xi}}(x, y) = \lim\limits_{y\to-\infty} F_{\vec{\xi}}(x, y) = \lim\limits_{x,y\to-\infty} F_{\vec{\xi}}(x, y) = 0$.

   Доведення. Строге доведення ґрунтується на теоремах про неперервність ймовірності. $\lim\limits_{x\to-\infty} F_{\vec{\xi}}(x, y) = \lim\limits_{x\to-\infty} \mathbb{P}\{\xi_1 < x,\ \xi_2 < y\}$ = [використаємо теорему про неперервність ймовірності] $= \mathbb{P}(\varnothing \cap \{\xi_2 < y\}) = \mathbb{P}(\varnothing) = 0$. Інші випадки — аналогічно. ▲

4. *«Умови узгодженості»*: $\lim\limits_{x\to+\infty} F_{\vec{\xi}}(x, y) = F_{\xi_2}(y)$, $\lim\limits_{y\to+\infty} F_{\vec{\xi}}(x, y) = F_{\xi_1}(x)$.

   Доведення. $\lim\limits_{x\to+\infty} F_{\vec{\xi}}(x, y) = \lim\limits_{x\to+\infty} \mathbb{P}(\xi_1 < x,\ \xi_2 < y) = \mathbb{P}(\xi_1 < +\infty,\ \xi_2 < y) =$
   $= \mathbb{P}(\xi_2 < y) = F_{\xi_2}(y)$. Інша — аналогічно. ▲

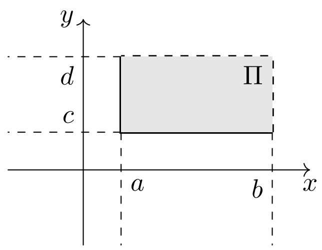
5. $\lim\limits_{x,y\to+\infty} F_{\vec{\xi}}(x, y) = 1$.

   Доведення. Випливає з умов узгодженості. ▲

6. Функція розподілу є неперервною зліва по кожному з аргументів:
   $\lim\limits_{x\to x_0-0} F_{\vec{\xi}}(x, y) = F_{\vec{\xi}}(x_0, y)$, $\lim\limits_{y\to y_0-0} F_{\vec{\xi}}(x, y) = F_{\vec{\xi}}(x, y_0)$.

   Доведення. Доведення аналогічне одновимірному випадку (твердж. 5, с. 18). ▲

7. $\Pi = [a; b) \times [c; d)$. $\mathbb{P}\left\{\vec{\xi} \in \Pi\right\} = ?$



Введемо події $A = \{\xi_1 < a, \xi_2 < c\}$, $B = \{\xi_1 \in [a; b), \xi_2 < c\}$, $C = \{\xi_1 < a, \xi_2 \in [c; d)\}$, $D = \left\{\vec{\xi} \in \Pi\right\}$, $E = \{\xi_1 < b, \xi_2 < d\}$. $E = A \cup B \cup C \cup D$, причому A, B, C, D — попарно несумісні, тому $\mathbb{P}(E) = \mathbb{P}(A) + \mathbb{P}(B) + \mathbb{P}(C) + \mathbb{P}(D)$. $\mathbb{P}(E) = F(b, d)$, $\mathbb{P}(A) = F(a, c)$, $\mathbb{P}(B) = F(b, c) - F(a, c)$, $\mathbb{P}(C) = F(a, d) - F(a, c)$.
Отже, $\mathbb{P}\left\{\vec{\xi} \in \Pi\right\} = F(b, d) + F(a, c) - F(b, c) - F(a, d)$.

3.1.2 Побудова функції розподілу дискретного випадкового вектора

Нехай двовимірний випадковий вектор задано таблицею розподілу:

$\xi_2 \backslash \xi_1$	1	2
0	0.3	0.4
1	0.2	0.1

Значення функції розподілу також зручно звести в таблицю:

$y \backslash x$	$x \le 1$	$1 < x \le 2$	$x > 2$
$y \le 0$	0	0	0
$0 < y \le 1$	0	0.3	0.7
$y > 1$	0	0.5	1

Останній рядок таблиці — значення $F_{\xi_1}(x)$, останній стовпчик — значення $F_{\xi_2}(y)$.

3.1.3 Сумісна функція розподілу n-вимірного вектора

Означення 3.1.4. *Сумісною функцією розподілу* n-вимірного випадкового вектора $\vec{\xi}$ називається $F_{\vec{\xi}}(\vec{x}) = \mathbb{P}\{\omega : \xi_1 < x_1, \xi_2 < x_2, ..., \xi_n < x_n\}$, $\vec{x} \in \mathbb{R}^n$.

Властивості:

1. Є монотонно неспадною по кожній координаті.
2. Є неперервною зліва по кожній координаті.
3. $\forall k = 1, ..., n : (x_k \to -\infty) \Rightarrow (F_{\vec{\xi}}(\vec{x}) \to 0)$.
4. $\lim\limits_{x_1,...,x_n \to +\infty} F_{\vec{\xi}}(\vec{x}) = 1$.
5. *«Умови узгодженості»*: $\forall k = 1, ..., n : F_{\xi_k}(x_k) = F_{\vec{\xi}}(+\infty, ..., +\infty, x_k, +\infty, ..., +\infty)$, $\forall k < n : F_{\xi_1\xi_2...\xi_k}(x_1, x_2, ..., x_k) = F_{\vec{\xi}}(x_1, x_2, ..., x_k, +\infty, ..., +\infty)$.

Вправа. Довести ці властивості.

3.2 Неперервні випадкові вектори

Означення 3.2.1. Вимірна функція $\vec{\xi}(\omega) : \Omega \to \mathbb{R}^n$ називається *неперервним випадковим вектором*, якщо координати ξ_i є неперервними випадковими величинами для $i = 1, ..., n$.

Функція розподілу $F_{\vec{\xi}}(\vec{x}) = \mathbb{P}\{\omega : \xi_1(\omega) < x_1, ..., \xi_n(\omega) < x_n\}$ є неперервною по всім аргументам.

3.2.1 Щільність розподілу двовимірного неперервного випадкового вектора

Означення 3.2.2. *Щільністю розподілу* двовимірного випадкового вектора називається подвійна границя

$$f_{\vec{\xi}}(x, y) = \lim_{\substack{\Delta x \to 0 \\ \Delta y \to 0}} \frac{\mathbb{P}\left\{\vec{\xi} \in \Pi\right\}}{\Delta x \Delta y} \tag{1}$$

де $\Pi = [x; x + \Delta x] \times [y; y + \Delta y]$.

Зауваження. В означенні замість прямокутника можна брати будь-яку обмежену замкнену множину D і тоді

$$f_{\vec{\xi}}(x, y) = \lim_{S(D) \to 0} \frac{\mathbb{P}\left\{\vec{\xi} \in D\right\}}{S(D)}, \text{ де } S(D) \text{ — площа множини } D$$

Зв'язок щільності розподілу з сумісною функцією розподілу:

$$\lim_{\substack{\Delta x \to 0 \\ \Delta y \to 0}} \frac{\mathbb{P}\left\{\vec{\xi} \in \Pi\right\}}{\Delta x \Delta y} = \lim_{\substack{\Delta x \to 0 \\ \Delta y \to 0}} \left(\frac{F_{\vec{\xi}}(x + \Delta x, y + \Delta y) - F_{\vec{\xi}}(x, y + \Delta y)}{\Delta x \Delta y} - \frac{F_{\vec{\xi}}(x + \Delta x, y) - F_{\vec{\xi}}(x, y)}{\Delta x \Delta y} \right) =$$

$$= [\text{формула Лагранжа про скінченні прирости}, \theta_1, \theta_2 \in (0, 1)] =$$

$$= \lim_{\substack{\Delta x \to 0 \\ \Delta y \to 0}} \left(\frac{\frac{\partial F}{\partial x}(x + \theta_1 \Delta x, y + \Delta y)\Delta x}{\Delta x \Delta y} - \frac{\frac{\partial F}{\partial x}(x + \theta_2 \Delta x, y)\Delta x}{\Delta x \Delta y} \right) =$$

$$= [\theta_3, \theta_4 \in (0, 1)] = \lim_{\substack{\Delta x \to 0 \\ \Delta y \to 0}} \frac{\frac{\partial^2 F}{\partial x \partial y}(x + \theta_3 \Delta x, y + \theta_4 \Delta y)\Delta y}{\Delta y} = \frac{\partial^2 F_{\vec{\xi}}}{\partial x \partial y}(x, y)$$

Таким чином, якщо існує неперервна друга похідна $\frac{\partial^2 F_{\vec{\xi}}}{\partial x \partial y}$, то

$$f_{\vec{\xi}}(x, y) = \frac{\partial^2 F_{\vec{\xi}}}{\partial x \partial y}(x, y) \tag{2}$$

Означення 3.2.3. Поверхня, що є графіком щільності двовимірного випадкового вектора, називається *поверхнею розподілу*.

Означення 3.2.4. Лінії, де $f_\xi(x,y) = const$, називаються *лініями рівних ймовірностей.*

Властивості щільності:

1. $f_{\vec{\xi}}(x,y) \geq 0$.
2. З формули 2: $F_{\vec{\xi}}(x,y) = \int\limits_{-\infty}^{x}\int\limits_{-\infty}^{y} f_{\vec{\xi}}(t,s)dtds$.
3. Умова нормування: $\int\limits_{-\infty}^{+\infty}\int\limits_{-\infty}^{\infty} f_{\vec{\xi}}(t,s)dtds = 1$ — об'єм під поверхнею розподілу дорівнює 1.
4. $F_{\xi_1}(x) = \int\limits_{-\infty}^{x}\int\limits_{-\infty}^{+\infty} f_{\vec{\xi}}(t,s)dtds$.

   $F_{\xi_2}(y) = \int\limits_{-\infty}^{+\infty}\int\limits_{-\infty}^{y} f_{\vec{\xi}}(t,s)dtds$.
5. Щільності розподілу окремих координат називаються *маргінальними щільностями.*

   $f_{\xi_1}(x) = F'_{\xi_1}(x) = \int\limits_{-\infty}^{+\infty} f_{\vec{\xi}}(x,y)dy$.

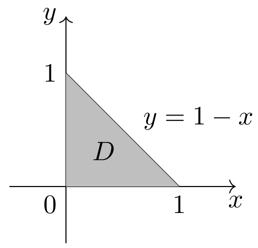
   $f_{\xi_2}(y) = F'_{\xi_2}(y) = \int\limits_{-\infty}^{+\infty} f_{\vec{\xi}}(x,y)dx$.
6. D — замкнена обмежена область в $\mathbb{R}^2$. $\mathbb{P}\left\{\vec{\xi} \in D\right\} = \iint\limits_{D} f_{\vec{\xi}}(x,y)dxdy$.
7. Якщо координати випадкового вектора ξ_1 та ξ_2 незалежні, то $F_{\vec{\xi}}(x,y) = F_{\xi_1}(x) \cdot F_{\xi_2}(y)$, і тоді $f_{\vec{\xi}}(x,y) = \frac{\partial^2 F_{\vec{\xi}}(x,y)}{\partial x \partial y} = \frac{\partial^2 (F_{\xi_1}(x)F_{\xi_2}(y))}{\partial x \partial y} = F'_{\xi_1}(x)F'_{\xi_2}(y) = f_{\xi_1}(x) \cdot f_{\xi_2}(y)$.

3.2.2 Рівномірний закон розподілу на площині

Означення 3.2.5. D — замкнена обмежена область в $\mathbb{R}^2$. Вектор $\vec{\xi} = (\xi_1, \xi_2)^T$ називається *рівномірно розподіленим в області D*, якщо

$$f_{\vec{\xi}}(x,y) = \begin{cases} \frac{1}{S(D)}, & (x,y) \in D \\ 0, & (x,y) \notin D \end{cases}$$

Приклад. $f_{\vec{\xi}}(x,y) = \begin{cases} 2, & (x,y) \in D \\ 0, & (x,y) \notin D \end{cases}$



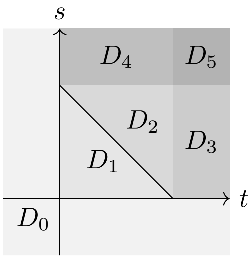
1. $f_{\xi_1}(x) = \int\limits_{-\infty}^{+\infty} f_{\vec{\xi}}(x,y)dy = \begin{cases} 0, & x \leq 0 \text{ або } x > 1 \\ \int\limits_{0}^{1-x} 2dy = 2(1-x), & 0 < x \leq 1 \end{cases}$

   $f_{\xi_2}(y) = \int\limits_{-\infty}^{+\infty} f_{\vec{\xi}}(x,y)dx = \begin{cases} 0, & y \leq 0 \text{ або } y > 1 \\ \int\limits_{0}^{1-y} 2dx = 2(1-y), & 0 < y \leq 1 \end{cases}$
2. Функції розподілу координат:

   $F_{\xi_1}(x) = \int\limits_{-\infty}^{x} f_{\xi_1}(t)dt = \begin{cases} 0, & x \leq 0 \\ 2\int\limits_{0}^{x}(1-t)dt = 2(x - \frac{x^2}{2}) = 2x - x^2, & 0 < x \leq 1 \\ 1, x > 1 \end{cases}$

   $F_{\xi_2}(y) = \int\limits_{-\infty}^{y} f_{\xi_2}(t)dt = \begin{cases} 0, & y \leq 0 \\ 2y - y^2, & 0 < y \leq 1 \\ 1, y > 1 \end{cases}$
3. Сумісна функція розподілу будується за допомогою розбиття площини $\mathbb{R}^2$ на області, в яких функція розподілу має однаковий вигляд, оскільки для рівномірного розподілу ймовірність потрапляння в певну область пропорційна її площі:

$$F_{\vec{\xi}}(x,y) = \int\limits_{-\infty}^{x}\int\limits_{-\infty}^{y} f_{\vec{\xi}}(t,s)dtds$$



$(x,y) \in D_0$
$D_0 = \{(x,y): \ x \le 0 \vee y \le 0\}$
$F_{\vec{\xi}}(x,y) = 0$

$(x,y) \in D_1$
$D_1 = \{(x,y): \ x \in [0,1], y \in [0, 1-x]\}$
$F_{\vec{\xi}}(x,y) = 2xy$

$(x,y) \in D_2$
$D_2 = \{(x,y): \ x \in [0,1], y \in [1-x, 1]\}$
$F_{\vec{\xi}}(x,y) = 2(xy - \frac{1}{2}(x-1+y)(y-1+x))$

$(x,y) \in D_3$
$D_3 = \{(x,y): \ x > 1, y \in [0,1]\}$
$F_{\vec{\xi}}(x,y) = 2\frac{1+1-y}{2}y = y(2-y)$

$(x,y) \in D_4$
$D_4 = \{(x,y): \ y > 1, x \in [0,1]\}$
$F_{\vec{\xi}}(x,y) = 2\frac{1+1-x}{2}x = x(2-x)$

$(x,y) \in D_5$
$D_5 = \{(x,y): \ x > 1 \wedge y > 1\}$
$F_{\vec{\xi}}(x,y) = 1$

Зауваження. Перевірити правильність побудови функції розподілу можна за допомогою умов узгодженості, перевірки точок стику та неперервності на лініях стику.

3.2.3 Щільність розподілу n-вимірного неперервного випадкового вектора

Означення 3.2.6. *Щільністю розподілу* n-вимірного неперервного випадкового вектора $\vec{\xi} = (\xi_1, \xi_2, ..., \xi_n)^T$ називається границя

$$f_{\vec{\xi}}(\vec{x}) = \lim_{\substack{\Delta x_i \to 0 \\ \forall i = \overline{1,n}}} \frac{\mathbb{P}\left\{\vec{\xi} \in \Pi\right\}}{\Delta x_1 ... \Delta x_n}, \tag{3}$$

де $\Pi = [x_1; x_1 + \Delta x_1] \times ... \times [x_n; x_n + \Delta x_n]$.

Якщо існує неперервна похідна $\frac{\partial^n F_{\vec{\xi}}(\vec{x})}{\partial x_1 ... \partial x_n}$, то

$$f_{\vec{\xi}}(\vec{x}) = \frac{\partial^n F_{\vec{\xi}}}{\partial x_1 ... \partial x_n}(\vec{x}) \tag{4}$$

Властивості щільності:

1. $f_{\vec{\xi}}(\vec{x}) \ge 0$.
2. $\int\limits_{-\infty}^{+\infty} ... \int\limits_{-\infty}^{+\infty} f_{\vec{\xi}}(\vec{x}) dx_1 ... dx_n = 1$.
3. $F_{\vec{\xi}}(\vec{x}) = \int\limits_{-\infty}^{x_1} ... \int\limits_{-\infty}^{x_n} f_{\vec{\xi}}(\vec{t}) dt_1 ... dt_n$.
4. $F_{\xi_1 \xi_2 ... \xi_k}(x_1, ..., x_k) = \int\limits_{-\infty}^{x_1} ... \int\limits_{-\infty}^{x_k} \int\limits_{-\infty}^{+\infty} ... \int\limits_{-\infty}^{+\infty} f_{\vec{\xi}}(\vec{t}) dt_1 ... dt_n$.

   $F_{\xi_i}(x_i) = \int\limits_{-\infty}^{+\infty} ... \int\limits_{-\infty}^{x_i} ... \int\limits_{-\infty}^{+\infty} f_{\vec{\xi}}(\vec{t}) dt_1 ... dt_n$.
5. $f_{\xi_1 ... \xi_k}(x_1, ..., x_k) = \underbrace{\int\limits_{-\infty}^{+\infty} ... \int\limits_{-\infty}^{+\infty}}_{n-k} f_{\vec{\xi}}(\vec{x}) dx_{k+1} .. dx_n$.
6. D — замкнена обмежена область в $\mathbb{R}^n$, $\mathbb{P}\left\{\vec{\xi} \in D\right\} = \int\int ... \int_D f_{\vec{\xi}}(\vec{x}) dx_1 ... dx_n$.
7. Якщо координати $\xi_1, ..., \xi_n$ незалежні у сукупності (тобто, події $\{\omega : \xi_{i_k}(\omega) < x_{i_k}\}$ незалежні для всіх x_{i_k} та $i_1, i_2, ... i_r \in \{1, ..., n\}$), то $f_{\vec{\xi}}(x_1, ..., x_n) = f_{\xi_1}(x_1) \cdot ... \cdot f_{\xi_n}(x_n)$.

Вправа. Довести властивості щільності.

3.3 Числові характеристики випадкових векторів

3.3.1 Математичне сподівання випадкових векторів

Означення 3.3.1. *Математичним сподіванням* $\mathbb{E}\vec{\xi}$ випадкового вектора $\vec{\xi}$ називається вектор $(\mathbb{E}\xi_1, ..., \mathbb{E}\xi_n)^T$.

Зауваження. Математичне сподівання випадкового вектора ще називають *центром розсіювання.*

Способи знаходження:

1. Знайти закони розподілу окремих координат, а далі — математичні сподівання цих координат.
2. Знайти математичні сподівання координат одразу. Для $\vec{\xi} = (\xi_1, \xi_2)^T$:
$$\mathbb{E}\xi_1 = \begin{cases} \sum\limits_i \sum\limits_j x_i p_{ij}, & \vec{\xi} - \text{ДВВ} \\ \int\limits_{-\infty}^{+\infty} \int\limits_{-\infty}^{+\infty} x f_{\vec{\xi}}(x, y) dx dy, & \vec{\xi} - \text{НВВ} \end{cases}, \mathbb{E}\xi_2 = \begin{cases} \sum\limits_j \sum\limits_i y_j p_{ij}, & \vec{\xi} - \text{ДВВ} \\ \int\limits_{-\infty}^{+\infty} \int\limits_{-\infty}^{+\infty} y f_{\vec{\xi}}(x, y) dx dy, & \vec{\xi} - \text{НВВ} \end{cases}.$$
В загальному випадку для ДВВ $\mathbb{E}\xi_k = \sum\limits_{i_1, i_2, ..., i_n} (x_k)_{i_k} p_{i_1 i_2 ... i_n}$, де індекс i_k пробігає кількість значень, що приймає координата x_k, для НВВ $\mathbb{E}\xi_k = \int\limits_{-\infty}^{+\infty} ... \int\limits_{-\infty}^{+\infty} x_k f_{\vec{\xi}}(\vec{x}) dx_1 ... dx_n$.

3.3.2 Мішані початкові та центральні моменти випадкових векторів

Означення 3.3.2. *Мішаним початковим моментом* порядку $k_1 + k_2 + ... + k_n$ $(k_i \in \mathbb{N})$ випадкового вектора $\vec{\xi} = (\xi_1, ..., \xi_n)^T$ називається число

$$\alpha_{k_1 + k_2 + ... + k_n} = \mathbb{E}\xi_1^{k_1} ... \xi_n^{k_n}$$

Зауваження. $\mathbb{E}\xi_1 = \alpha_{1+0+...+0}, \mathbb{E}\xi_2 = \alpha_{0+1+...+0}, ..., \mathbb{E}\xi_n = \alpha_{0+0+...+1}$.

Означення 3.3.3. *Мішаним центральним моментом* порядку $k_1 + k_2 + ... + k_n$ $(k_i \in \mathbb{N})$ випадкового вектора $\vec{\xi} = (\xi_1, ..., \xi_n)^T$ називається число

$$\beta_{k_1 + k_2 + ... + k_n} = \mathbb{E}\mathring{\xi}_1^{k_1} ... \mathring{\xi}_n^{k_n}, \mathring{\xi}_k = \xi_k - \mathbb{E}\xi_k$$

Зауваження. Всі центральні моменти 1-го порядку — нульові.

Центральні моменти порядку $0 + ... + 0 + 2 + 0 + ... + 0$ — дисперсії відповідних координат: $\forall k = \overline{1, n}$: $\beta_{0+...+0+\underset{k}{2}+0+...+0} = \mathbb{E}\mathring{\xi}_k^2 = \mathbb{D}\xi_k$

Дисперсії координат задають розсіювання вздовж відповідних координатних осей.

Означення 3.3.4. *Кореляційним моментом* або *коваріацією* випадкових величин ξ_i та ξ_j називається мішаний центральний момент порядку $0 + ... + 0 + \underset{i}{1} + 0 + ... + 0 + \underset{j}{1} + 0 + ... + 0$:

$$cov(\xi_i, \xi_j) = \mathbb{K}\xi_i\xi_j = \beta_{0+...+0+\underset{i}{1}+0+...+0+\underset{j}{1}+0+...+0} = \mathbb{E}(\xi_i - \mathbb{E}\xi_i)(\xi_j - \mathbb{E}\xi_j)$$

Зауваження. $cov(\xi_k, \xi_k) = \mathbb{E}(\xi_k - \mathbb{E}\xi_k)^2 = \mathbb{D}\xi_k$.

Означення 3.3.5. *Кореляційною матрицею* випадкового вектора $\vec{\xi}$ називається матриця, у яку зібрано усі кореляційні моменти випадкового вектора

$$\mathbb{K}\vec{\xi} = \begin{pmatrix} \mathbb{D}\xi_1 & \mathbb{K}\xi_1\xi_2 & \cdots & \mathbb{K}\xi_1\xi_n \\ \mathbb{K}\xi_1\xi_2 & \mathbb{D}\xi_2 & \cdots & \mathbb{K}\xi_2\xi_n \\ \vdots & \vdots & \ddots & \vdots \\ \mathbb{K}\xi_1\xi_n & \mathbb{K}\xi_2\xi_n & \cdots & \mathbb{D}\xi_n \end{pmatrix}$$

Зауваження. $\mathring{\vec{\xi}} = \left(\mathring{\xi}_1, ..., \mathring{\xi}_n\right)^T$. Тоді $\mathbb{K}\vec{\xi} = \mathbb{E}\,\mathring{\vec{\xi}}\mathring{\vec{\xi}}^T = \mathbb{E}\left(\vec{\xi} - \mathbb{E}\xi\right)\left(\vec{\xi} - \mathbb{E}\xi\right)^T$.

Властивості кореляційного моменту випадкових величин:

1. $\mathbb{K}\xi_i\xi_j = \mathbb{K}\xi_j\xi_i$.

Доведення. Випливає з означення кореляційного моменту. ▲

2. $\eta_1 = a_1\xi_1 + b_1$, $\eta_2 = a_2\xi_2 + b_2$. Тоді $\mathbb{K}\eta_1\eta_2 = a_1a_2\mathbb{K}\xi_1\xi_2$.

   Доведення. $\mathbb{K}\eta_1\eta_2 = \mathbb{E}(\eta_1-\mathbb{E}\eta_1)(\eta_2-\mathbb{E}\eta_2) = \mathbb{E}(a_1\xi_1+b_1-\mathbb{E}(a_1\xi_1+b_1))(a_2\xi_2+b_2-\mathbb{E}(a_2\xi_2+b_2)) = \mathbb{E}a_1(\xi_1-\mathbb{E}\xi_1)a_2(\xi_2-\mathbb{E}\xi_2) = a_1a_2\mathbb{E}(\xi_1-\mathbb{E}\xi_1)(\xi_2-\mathbb{E}\xi_2) = a_1a_2\mathbb{K}\xi_1\xi_2$. ▲

3. Зручна формула для обчислення кореляційного моменту.

$$\mathbb{K}\xi_1\xi_2 = \mathbb{E}\xi_1\xi_2 - \mathbb{E}\xi_1\mathbb{E}\xi_2 \tag{1}$$

   Доведення. $\mathbb{K}\xi_1\xi_2 = \mathbb{E}(\xi_1-\mathbb{E}\xi_1)(\xi_2-\mathbb{E}\xi_2) = \mathbb{E}(\xi_1\xi_2 - \xi_1\mathbb{E}\xi_2 - \xi_2\mathbb{E}\xi_1 + \mathbb{E}\xi_1\mathbb{E}\xi_2) = \mathbb{E}\xi_1\xi_2 - \mathbb{E}\xi_1\mathbb{E}\xi_2 - \mathbb{E}\xi_2\mathbb{E}\xi_1 + \mathbb{E}\xi_1\mathbb{E}\xi_2 = \mathbb{E}\xi_1\xi_2 - \mathbb{E}\xi_1\mathbb{E}\xi_2$ ▲

4. $\mathbb{D}(a\xi_1 \pm b\xi_2) = a^2\mathbb{D}\xi_1 \pm 2ab\mathbb{K}\xi_1\xi_2 + b^2\mathbb{D}\xi_2$.

   Доведення. $\mathbb{D}(a\xi_1 \pm b\xi_2) = \mathbb{E}(a\xi_1 \pm b\xi_2 - \mathbb{E}(a\xi_1 \pm b\xi_2))^2 = \mathbb{E}(a(\xi_1-\mathbb{E}\xi_1) \pm b(\xi_2-\mathbb{E}\xi_2))^2 = a^2\mathbb{E}\mathring{\xi}_1^2 \pm 2ab\mathbb{E}\mathring{\xi}_1\mathring{\xi}_2 + b^2\mathbb{E}\mathring{\xi}_2^2 = a^2\mathbb{D}\xi_1 \pm 2ab\mathbb{K}\xi_1\xi_2 + b^2\mathbb{D}\xi_2$. ▲

5. $|\mathbb{K}\xi_1\xi_2| \le \sigma_{\xi_1}\sigma_{\xi_2}$.

   Доведення. Два способи:
   I. $\mathbb{D}(x\xi_1 \pm \xi_2) = x^2\mathbb{D}\xi_1 \pm 2x\mathbb{K}\xi_1\xi_2 + \mathbb{D}\xi_2 \ge 0 \Rightarrow 4(\mathbb{K}\xi_1\xi_2)^2 - 4\mathbb{D}\xi_1\mathbb{D}\xi_2 \le 0 \Rightarrow$ $|\mathbb{K}\xi_1\xi_2| \le \sigma_{\xi_1}\sigma_{\xi_2}$.
   II. Введемо нормовані центровані випадкові величини $\xi_H = \frac{\xi-\mathbb{E}\xi}{\sigma_\xi}$, $\mathbb{E}\xi_H = 0$, $\mathbb{D}\xi_H = 1$. $\mathbb{E}(\xi_{1H} \pm \xi_{2H})^2 \ge 0 \Leftrightarrow \mathbb{E}(\xi_{1H}^2 \pm 2\xi_{1H}\xi_{2H} + \xi_{2H}^2) \ge 0 \Leftrightarrow 2 \pm 2\frac{\mathbb{K}\xi_1\xi_2}{\sigma_{\xi_1}\sigma_{\xi_2}} \ge 0 \Leftrightarrow$ $|\mathbb{K}\xi_1\xi_2| \le \sigma_{\xi_1}\sigma_{\xi_2}$.

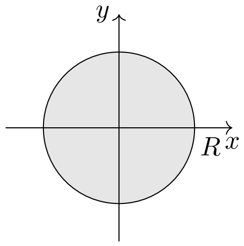
   ▲

6. Якщо ξ_1 та ξ_2 — незалежні, то $\mathbb{K}\xi_1\xi_2 = 0$.

   Доведення. Нехай ξ_1 та ξ_2 — незалежні.
   $\mathbb{K}\xi_1\xi_2 = \mathbb{E}\xi_1\xi_2 - \mathbb{E}\xi_1\mathbb{E}\xi_2 = \iint\limits_{\mathbb{R}^2} xyf_{\vec{\xi}}(x,y)dxdy - \mathbb{E}\xi_1\mathbb{E}\xi_2 = \int\limits_{-\infty}^{+\infty} xf_{\xi_1}(x)dx \cdot \int\limits_{-\infty}^{+\infty} yf_{\xi_2}(y)dy - \mathbb{E}\xi_1\mathbb{E}\xi_2 = \mathbb{E}\xi_1\mathbb{E}\xi_2 - \mathbb{E}\xi_1\mathbb{E}\xi_2 = 0$. ▲

   Обернене твердження, взагалі кажучи, не має місця. Тому випадкові величини, для яких $\mathbb{K}\xi_1\xi_2 = 0$, називаються *некорельованими*.

Приклад. Взагалі кажучи, з некорельованості не випливає незалежність.



$S = \{(x,y) \in \mathbb{R}^2 : x^2+y^2 \le R^2\}$, $\vec{\xi} \sim \mathrm{U}(S)$.

$$f_{\vec{\xi}}(x,y) = \begin{cases} \frac{1}{\pi R^2}, & (x,y) \in S \\ 0, & (x,y) \notin S \end{cases}$$

$$f_{\xi_1}(x) = \begin{cases} \int\limits_{-\sqrt{R^2-x^2}}^{\sqrt{R^2-x^2}} \frac{1}{\pi R^2}dy = \frac{2\sqrt{R^2-x^2}}{\pi R^2}, & |x| \le R \\ 0, & |x| > R \end{cases}, \quad f_{\xi_2}(y) = \begin{cases} \frac{2\sqrt{R^2-y^2}}{\pi R^2}, & |y| \le R \\ 0, & |y| > R \end{cases}$$

$\mathbb{E}\xi_1 = \mathbb{E}\xi_2 = \int\limits_{-R}^{R} t \cdot \frac{2\sqrt{R^2-t^2}}{\pi R^2}dt = 0$, бо інтегрується непарна функція по симетричному проміжку.

$\mathbb{K}\xi_1\xi_2 = \iint\limits_S xy\frac{1}{\pi r^2}dxdy - \mathbb{E}\xi_1\mathbb{E}\xi_2 = \frac{1}{\pi R^2}\int\limits_0^{2\pi} d\varphi \int\limits_0^R r^2\cos\varphi\sin\varphi r dr = \frac{1}{2\pi R^2}\int\limits_0^{2\pi}\sin 2\varphi d\varphi \int\limits_0^R r^3 dr = 0$.
Отже, $f_{\vec{\xi}}(x,y) \ne f_{\xi_1}(x)f_{\xi_2}(y)$ і координати вектора залежні, але $\mathbb{K}\xi_1\xi_2 = 0$.

3.3.3 Коваріація як скалярний добуток випадкових величин

Зафіксуємо ймовірнісний простір $\{\Omega, \mathcal{F}, \mathbb{P}\}$. Позначимо $\mathcal{L}^2(\Omega) = \{\xi : \Omega \to \mathbb{R} \mid \mathbb{E}\xi^2 < +\infty\}$. Оскільки $\mathbb{D}\xi = \mathbb{E}\xi^2 - (\mathbb{E}\xi)^2 \ge 0$, то для $\xi \in \mathcal{L}^2(\Omega)$ $|\mathbb{E}\xi| < +\infty$. Таким чином, для будь-яких $\xi, \eta \in \mathcal{L}^2(\Omega)$ існує $cov(\xi,\eta)$, бо $|cov(\xi,\eta)| \le \sqrt{\mathbb{D}\xi}\cdot\sqrt{\mathbb{D}\eta}$.

Для коваріації маємо $cov(a\xi_1 + b\xi_2, \eta) = \mathbb{E}(a\xi_1 + b\xi_2)\eta - \mathbb{E}(a\xi_1 + b\xi_2)\mathbb{E}\eta = a\mathbb{E}\xi_1\eta + b\mathbb{E}\xi_2\eta - a\mathbb{E}\xi_1\mathbb{E}\eta - b\mathbb{E}\xi_2\mathbb{E}\eta = a \cdot cov(\xi_1, \eta) + b \cdot cov(\xi_2, \eta)$. Також було доведено $cov(\xi, \eta) = cov(\eta, \xi)$ та $cov(\xi, \xi) \geq 0$. Виконуються всі умови скалярного добутку, окрім $cov(\xi, \xi) = 0 \Rightarrow \xi = 0$. Ця умова буде виконуватися, якщо розглядати інший простір випадкових величин: $\mathcal{L}_0^2(\Omega) = \left\{\xi : \Omega \to \mathbb{R} \mid \mathbb{E}\xi = 0, \mathbb{E}\xi^2 < +\infty\right\}$.

Зауваження. Якщо $\xi \in \mathcal{L}^2(\Omega)$, то $\mathring{\xi} \in \mathcal{L}_0^2(\Omega)$.

Вправа. Перевірити, що $\mathcal{L}_0^2(\Omega)$ є лінійним простором.

Висновок: $cov(\xi, \eta)$ задає *скалярний добуток* на $\mathcal{L}_0^2(\Omega)$.

Таким чином, властивість $|cov(\xi, \eta)| \leq \sqrt{\mathbb{D}\xi} \cdot \sqrt{\mathbb{D}\eta}$ — це нерівність Коші-Буняковського, а кореляційна матриця $\mathbb{K}$ — це матриця Грама системи випадкових величин $\{\xi_1, \xi_2, ..., \xi_n\}$. З курсу лінійної алгебри відомо, що якщо ці випадкові величини лінійно незалежні, то матриця $\mathbb{K}$ є додатно визначеною, і невід'ємно визначеною в загальному випадку.

3.3.4 Коефіцієнт кореляції

Означення 3.3.6. *Коефіцієнтом кореляції* називають безрозмірну числову характеристику

$$\rho_{\xi_1\xi_2} = \frac{\mathbb{K}\xi_1\xi_2}{\sigma_{\xi_1}\sigma_{\xi_2}}$$

Означення 3.3.7. Всі коефіцієнти кореляції збираються в *нормовану кореляційну матрицю*:

$$R = \begin{pmatrix} 1 & \rho_{\xi_1\xi_2} & \cdots & \rho_{\xi_1\xi_n} \\ \rho_{\xi_1\xi_2} & 1 & \cdots & \rho_{\xi_2\xi_n} \\ \vdots & \vdots & \ddots & \vdots \\ \rho_{\xi_1\xi_n} & \rho_{\xi_2\xi_n} & \cdots & 1 \end{pmatrix}$$

Властивості коефіцієнту кореляції:

1. Для некорельованих випадкових величин ξ_1 та ξ_2 $\rho_{\xi_1\xi_2} = 0$.

   Доведення. Випливає з означення коефіцієнта кореляції. ▲

2. $|\rho_{\xi_1\xi_2}| \leq 1$.

   Доведення. Випливає з властивості кореляційного моменту. ▲

3. $|\rho_{\xi_1\xi_2}| = 1 \Leftrightarrow \exists\, a, b \in \mathbb{R} : \xi_2 = a\xi_1 + b$.

   Доведення. Нехай $\xi_2 = a\xi_1 + b$. $\mathbb{K}\xi_1\xi_2 = a\mathbb{D}\xi_1$, $\sigma_{\xi_2} = \sqrt{\mathbb{D}\xi_2} = \sqrt{a^2\mathbb{D}\xi_1} = |a|\sqrt{\mathbb{D}\xi_1}$.
   Тоді $\rho_{\xi_1\xi_2} = \frac{a\mathbb{D}\xi_1}{|a|(\sqrt{\mathbb{D}\xi_1})^2} \Rightarrow |\rho_{\xi_1\xi_2}| = 1$.
   Нехай $|\rho_{\xi_1\xi_2}| = 1$. $\mathbb{E}\left(\frac{\xi_1 - \mathbb{E}\xi_1}{\sigma_{\xi_1}} \pm \frac{\xi_2 - \mathbb{E}\xi_2}{\sigma_{\xi_2}}\right)^2 = 2(1 \pm \rho_{\xi_1\xi_2}) = 0$.
   $\mathbb{E}\xi^2 = 0 \Rightarrow \mathbb{D}\xi = -(\mathbb{E}\xi)^2 \geq 0 \Rightarrow \begin{cases} \mathbb{E}\xi = 0 \\ \mathbb{D}\xi = 0 \end{cases} \Rightarrow \xi$ приймає значення 0 з ймовірністю 1.
   Отже, $\frac{\xi_1 - \mathbb{E}\xi_1}{\sigma_{\xi_1}} \pm \frac{\xi_2 - \mathbb{E}\xi_2}{\sigma_{\xi_2}} = 0$ з ймовірністю 1, звідки отримуємо лінійний зв'язок. ▲

3.4 Умовні закони розподілу ВВ

Розглядаємо випадок $n = 2$, $\vec{\xi} = (\xi_1, \xi_2)^T$.

Означення 3.4.1. *Умовним законом розподілу* ξ_1 називається закон розподілу ξ_1 за умови того, що ξ_2 набула відповідного значення (ДВВ) або потрапила в деякий проміжок (НВВ) (для ξ_2 аналогічно).

Означення 3.4.2. Універсальним умовним законом розподілу ВВ є *умовна функція розподілу*:

$$F_{\xi_1}(x/y) = \mathbb{P}(\xi_1 < x/\xi_2 < y) = \frac{\mathbb{P}\left\{\xi_1 < x, \xi_2 < y\right\}}{\mathbb{P}\left\{\xi_2 < y\right\}} = \frac{F_{\vec{\xi}}(x, y)}{F_{\xi_2}(y)}, F_{\xi_2}(y/x) = \frac{F_{\vec{\xi}}(x, y)}{F_{\xi_1}(x)}$$

Зауваження. З означення випливає, що $F_{\vec{\xi}}(x, y) = F_{\xi_1}(x)F_{\xi_2}(y/x)$ та $F_{\vec{\xi}}(x, y) = F_{\xi_2}(y)F_{\xi_1}(x/y)$.

3.4.1 Умовний закон розподілу дискретного випадкового вектора

Знову розглядаємо випадок $n = 2$ та $\vec{\xi} = (\xi_1, \xi_2)^T$.

Умовні розподіли координат задаються $\mathbb{P}\{\xi_1 = x_i/\xi_2 = y_j\} = \frac{\mathbb{P}\{\xi_1 = x_i, \xi_2 = y_j\}}{\mathbb{P}\{\xi_2 = y_j\}}$ та $\mathbb{P}\{\xi_2 = y_j/\xi_1 = x_i\} = \frac{\mathbb{P}\{\xi_1 = x_i, \xi_2 = y_j\}}{\mathbb{P}\{\xi_1 = x_i\}}$.

Приклад. Дискретний випадковий вектор має закон розподілу:

ξ_2 \ ξ_1	-1	0	1
0	0.1	0.2	0.1
1	0.2	0.3	0.1

ξ_1	-1	0	1
p	0.3	0.5	0.2

ξ_2	0	1
p	0.4	0.6

ξ_1	-1	0	1
$\mathbb{P}\{\xi_1/\xi_2 = 0\}$	1/4	2/4	1/4
$\mathbb{P}\{\xi_1/\xi_2 = 1\}$	2/6	3/6	1/6

— умовний закон розподілу (умовний ряд розподілу) ξ_1.

ξ_2	0	1
$\mathbb{P}\{\xi_2/\xi_1 = -1\}$	1/3	2/3
$\mathbb{P}\{\xi_2/\xi_1 = 0\}$	2/5	3/5
$\mathbb{P}\{\xi_2/\xi_1 = 1\}$	1/2	1/2

— умовний закон розподілу (умовний ряд розподілу) ξ_2.

3.4.2 Умовне математичне сподівання дискретного випадкового вектора

Означення 3.4.3. *Умовним математичним сподіванням* випадкової величини ξ_1 є математичне сподівання цієї випадкової величини за умови, що ξ_2 набула певного значення.

$$\mathbb{E}(\xi_1/\xi_2 = y_j) = \sum_{i=1}^{n(\infty)} x_i \mathbb{P}\{\xi_1 = x_i/\xi_2 = y_j\}, \mathbb{E}(\xi_2/\xi_1 = x_i) = \sum_{j=1}^{n(\infty)} y_j \mathbb{P}\{\xi_2 = y_j/\xi_1 = x_i\}$$

Зауваження. Якщо ξ_1 та ξ_2 незалежні, то $\mathbb{E}(\xi_1/\xi_2 = y_j) = \mathbb{E}\xi_1$, $\mathbb{E}(\xi_2/\xi_1 = x_i) = \mathbb{E}\xi_2$.

Умовні математичні сподівання координат дискретного випадкового вектора є дискретними випадковими величинами, оскільки приймають декілька значень з певними ймовірностями, тому можна скласти їх закон розподілу.

Приклад. Продовження попереднього прикладу:

$\mathbb{E}(\xi_1/\xi_2)$	$-1/6$	0
p	0.6	0.4

$\mathbb{E}(\xi_2/\xi_1)$	1/2	3/5	2/3
p	0.2	0.5	0.3

3.4.3 Умовні закони розподілу неперервних випадкових величин

У випадку $n = 2$, $\vec{\xi} = (\xi_1, \xi_2)^T$:

$$\mathbb{P}\{x \le \xi_1 < x + \Delta x/y \le \xi_2 < y + \Delta y\} = \frac{\mathbb{P}\{\vec{\xi} \in \Pi\}}{\mathbb{P}\{y \le \xi_2 < y + \Delta y\}} = \frac{f_{\vec{\xi}}(x, y)\Delta x \Delta y + o(\sqrt{\Delta x^2 + \Delta y^2})}{f_{\xi_2}(y)\Delta y + o(\Delta y)}$$

Означення 3.4.4. *Умовною щільністю розподілу* називається функція вигляду:

$$f_{\xi_1}(x/y) = \frac{f_{\vec{\xi}}(x, y)}{f_{\xi_2}(y)}, f_{\xi_2}(y/x) = \frac{f_{\vec{\xi}}(x, y)}{f_{\xi_1}(x)}$$

Зауваження. Графік умовної щільності розподілу можна інтерпретувати як лінію перетину поверхні розподілу та площини $y = y_{\text{знач.}}$ або $x = x_{\text{знач.}}$ відповідно, нормовану на одиничну площу під нею.

Властивості умовної щільності:

1. $f_{\xi_1}(x/y) \ge 0, f_{\xi_2}(y/x) \ge 0$.
2. $f_{\xi_1}(x/y) = \frac{f_{\vec{\xi}}(x,y)}{\int\limits_{-\infty}^{+\infty} f_{\vec{\xi}}(x,y)dx}, f_{\xi_2}(y/x) = \frac{f_{\vec{\xi}}(x,y)}{\int\limits_{-\infty}^{+\infty} f_{\vec{\xi}}(x,y)dy}$.
3. $\int\limits_{-\infty}^{+\infty} f_{\xi_1}(x/y)dx = \int\limits_{-\infty}^{+\infty} \frac{f_{\vec{\xi}}(x,y)}{\int\limits_{-\infty}^{+\infty} f_{\vec{\xi}}(x,y)dx} dx = 1, \int\limits_{-\infty}^{+\infty} f_{\xi_2}(y/x)dy = \int\limits_{-\infty}^{+\infty} \frac{f_{\vec{\xi}}(x,y)}{\int\limits_{-\infty}^{+\infty} f_{\vec{\xi}}(x,y)dy} dy = 1$.

Властивості умовної щільності аналогічні властивостям звичайної (безумовної) щільності.

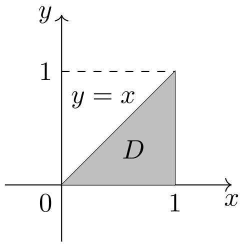
3.4.4 Умовне математичне сподівання неперервного випадкового вектора

Умовні математичні сподівання координат НВВ визначаються через умовні щільності:

$$\mathbb{E}(\xi_1/\xi_2 = y) = \int\limits_{-\infty}^{+\infty} x f_{\xi_1}(x/y)dx,\ \mathbb{E}(\xi_2/\xi_1 = x) = \int\limits_{-\infty}^{+\infty} y f_{\xi_2}(y/x)dy$$

Означення 3.4.5. Функції $\varphi(y) = \mathbb{E}(\xi_1/\xi_2 = y)$ та $\psi(x) = \mathbb{E}(\xi_2/\xi_1 = x)$ називаються *лініями регресії.*

Приклад. $f_{\vec{\xi}}(x,y) = \begin{cases} 2, & (x,y) \in D \\ 0, & (x,y) \notin D \end{cases}$



$f_{\xi_1}(x) = \begin{cases} \int\limits_0^x 2dy = 2x, & x \in [0;1] \\ 0, & x \notin [0;1] \end{cases}$, $f_{\xi_2}(y) = \begin{cases} \int\limits_y^1 2dx = 2(1-y), & y \in [0;1] \\ 0, & y \notin [0;1] \end{cases}$.

Запишемо умовні щільності:

$f_{\xi_1}(x/y) = \frac{f_{\vec{\xi}}(x,y)}{f_{\xi_2}(y)} = \begin{cases} \frac{1}{1-y}, & x \in [y;1], \\ & y \in [0;1) \\ 0, & \text{інакше} \end{cases}$, $f_{\xi_2}(y/x) = \frac{f_{\vec{\xi}}(x,y)}{f_{\xi_1}(x)} = \begin{cases} \frac{1}{x}, & y \in [0;x], \\ & x \in (0;1] \\ 0, & \text{інакше} \end{cases}$.

Умовні розподіли обох координат є рівномірними з параметрами $\langle y, 1\rangle$ та $\langle 0, x\rangle$ відповідно.

$\mathbb{E}\xi_1 = \int\limits_0^1 2x^2dx = \frac{2}{3}$, $\mathbb{E}\xi_2 = \int\limits_0^1 2y(1-y)dy = \frac{1}{3}$. Знайдемо умовні математичні сподівання.

$\mathbb{E}(\xi_1/\xi_2 = y) = \int\limits_{-\infty}^{+\infty} x f_{\xi_1}(x/y)dx = \frac{1}{1-y}\int\limits_y^1 xdx = \frac{1}{1-y} \cdot \frac{1-y^2}{2} = \frac{1+y}{2}$, $y \in [0;1)$.

$\mathbb{E}(\xi_2/\xi_1 = x) = \int\limits_{-\infty}^{+\infty} y f_{\xi_2}(y/x)dy = \frac{1}{x}\int\limits_0^x ydy = \frac{1}{x} \cdot \frac{x^2}{2} = \frac{x}{2}$, $x \in (0;1]$.

3.4.5 Формули повного математичного сподівання та дисперсії

Формула повного математичного сподівання. $\mathbb{E}(\mathbb{E}(\xi_1/\xi_2)) = \mathbb{E}\xi_1$, $\mathbb{E}(\mathbb{E}(\xi_2/\xi_1)) = \mathbb{E}\xi_2$.

Доведення. Дискретний випадок:

$\mathbb{E}(\mathbb{E}(\xi_1/\xi_2)) = \sum\limits_{j=1}^m \left(\sum\limits_{i=1}^n x_i \mathbb{P}\{\xi_1 = x_i/\xi_2 = y_j\}\right)\mathbb{P}\{\xi_2 = y_j\} = \sum\limits_{i=1}^n x_i\mathbb{P}\{\xi_1 = x_i\} = \mathbb{E}\xi_1$.

Неперервний випадок:

$\mathbb{E}(\mathbb{E}(\xi_1/\xi_2)) = \int\limits_{-\infty}^{+\infty}\left(\int\limits_{-\infty}^{+\infty} x f_{\xi_1}(x/y)dx\right) f_{\xi_2}(y)dy = \int\limits_{-\infty}^{+\infty} x f_{\xi_1}(x)dx = \mathbb{E}\xi_1$.

Для ξ_2 — аналогічно. ▲

Приклад. Продовження попередніх прикладів.

1. Дискретний випадок:
   $\mathbb{E}\xi_1 = -0.3 + 0.2 = -0.1$, $\mathbb{E}(\mathbb{E}(\xi_1/\xi_2)) = -\frac{1}{6} \cdot 0.6 = -0.1$.
   $\mathbb{E}\xi_2 = 0.6$, $\mathbb{E}(\mathbb{E}(\xi_2/\xi_1)) = 0.3 \cdot \frac{2}{3} + \frac{1}{2} \cdot \frac{3}{5} + \frac{1}{2} \cdot \frac{2}{10} = 0.6$.
2. Неперервний випадок:
   $\mathbb{E}(\mathbb{E}(\xi_1/\xi_2)) = \int\limits_0^1 \frac{1+y}{2} \cdot 2(1-y)dy = \int\limits_0^1 (1-y^2)dy = \frac{2}{3} = \mathbb{E}\xi_1$,
   $\mathbb{E}(\mathbb{E}(\xi_2/\xi_1)) = \int\limits_0^1 \frac{x}{2} \cdot 2xdx = \int\limits_0^1 x^2dx = \frac{1}{3} = \mathbb{E}\xi_2$.

Крім умовного математичного сподівання розглядають *умовні дисперсії* — міра розсіювання однієї випадкової величини за умови того, що інша набула певне значення.

Означення 3.4.6. *Умовною дисперсією* дискретної випадкової величини ξ_1 є дисперсія цієї випадкової величини за умови, що ξ_2 набула певного значення.

$$\mathbb{D}(\xi_1/\xi_2 = y_j) = \mathbb{E}((\xi_1 - \mathbb{E}(\xi_1/\xi_2 = y_j))^2/\xi_2 = y_j)$$

Формула повної дисперсії. Нехай ξ_1, ξ_2 — випадкові величини та $\mathbb{D}\xi_1 < +\infty$. Тоді $\mathbb{D}\xi_1 = \mathbb{E}(\mathbb{D}(\xi_1/\xi_2)) + \mathbb{D}(\mathbb{E}(\xi_1/\xi_2))$.

Доведення. $\mathbb{D}\xi_1 = \mathbb{E}\xi_1^2 - (\mathbb{E}\xi_1)^2 = \mathbb{E}(\mathbb{E}(\xi_1^2/\xi_2)) - (\mathbb{E}(\mathbb{E}(\xi_1/\xi_2)))^2 = \mathbb{E}(\mathbb{D}(\xi_1/\xi_2) + (\mathbb{E}(\xi_1/\xi_2))^2) - (\mathbb{E}(\mathbb{E}(\xi_1/\xi_2)))^2 = \mathbb{E}(\mathbb{D}(\xi_1/\xi_2)) + (\mathbb{E}(\mathbb{E}(\xi_1/\xi_2)^2) - (\mathbb{E}(\mathbb{E}(\xi_1/\xi_2)))^2) = \mathbb{E}(\mathbb{D}(\xi_1/\xi_2)) + \mathbb{D}(\mathbb{E}(\xi_1/\xi_2))$. ▲

3.4.6 Випадок незалежних координат випадкового вектора

Розглядаємо випадок $n = 2$, $\vec{\xi} = (\xi_1, \xi_2)^T$. Необхідною і достатньою умовою незалежності координат, як вже було розглянуто, є $F_{\vec{\xi}}(x, y) = F_{\xi_1}(x) \cdot F_{\xi_2}(y)$.

З цього випливає, що в разі незалежності координат $F_{\xi_1}(x/y) = F_{\xi_1}(x)$, $F_{\xi_2}(y/x) = F_{\xi_2}(y)$.

Дискретний випадок:

$\mathbb{P}\{\xi_1 = x_i, \xi_2 = y_j\} = \mathbb{P}\{\xi_1 = x_i\} \cdot \mathbb{P}\{\xi_2 = y_j\}$ $\forall\, i, j$. З цього випливає, що:

1. $\mathbb{P}\{\xi_1 = x_i/\xi_2 = y_j\} = \mathbb{P}\{\xi_1 = x_i\}$.
2. $\mathbb{E}(\xi_1/\xi_2 = y_j) = \mathbb{E}\xi_1$, $\mathbb{E}(\xi_2/\xi_1 = x_i) = \mathbb{E}\xi_2$.

Неперервний випадок:

$f_{\vec{\xi}}(x, y) = f_{\xi_1}(x) \cdot f_{\xi_2}(y)$. З цього випливає, що:

1. $f_{\xi_1}(x/y) = f_{\xi_1}(x)$, $f_{\xi_2}(y/x) = f_{\xi_2}(y)$.
2. $\mathbb{E}(\xi_1/\xi_2 = y) = \mathbb{E}\xi_1$, $\mathbb{E}(\xi_2/\xi_1 = x) = \mathbb{E}\xi_2$.

Для n-вимірного випадку ($n > 2$): $\forall i, j$: $F_{\xi_i\xi_j}(x_i, x_j) = F_{\xi_i}(x_i) \cdot F_{\xi_j}(x_j)$ — лише попарна незалежність. Нагадаємо, що координати $\xi_1, \xi_2, ..., \xi_n$ є *незалежними у сукупності* тоді і тільки тоді, коли $F_{\vec{\xi}}(\vec{x}) = \prod\limits_{k=1}^{n} F_{\xi_k}(x_k)$. З незалежності у сукупності випливає незалежність, а отже, некорельованість будь-якої пари координат.

Розділ 4

Характеристичні функції. Гауссівські випадкові вектори

4.1 Характеристичні функції випадкових величин

4.1.1 Поняття характеристичної функції

Означення 4.1.1. *Характеристичною функцією* випадкової величини ξ називається комплекснозначна функція $\chi_\xi(t) = \mathbb{E}e^{it\xi}, t \in \mathbb{R}$.

$$\chi_\xi(t) = \int\limits_{-\infty}^{+\infty} e^{itx} dF_\xi(x) = \begin{cases} \sum\limits_{k=1}^{n(\infty)} e^{itx_k}\mathbb{P}\left\{\xi = x_k\right\}, & \xi \text{ — ДВВ} \\ \int\limits_{-\infty}^{+\infty} e^{itx} f_\xi(x)dx, & \xi \text{ — НВВ} \end{cases} \tag{1}$$

Інтеграл $\int\limits_{-\infty}^{+\infty} e^{itx} f_\xi(x)dx$ у курсі гармонічного аналізу називається *перетворенням Фур'є* функції $f_\xi(x)$. Отже, у випадку НВВ ξ $\chi_\xi(t)$ — перетворення Фур'є щільності. Також з курсу гармонічного аналізу відомо, що за допомогою *оберненого перетворення Фур'є* можна відновити $f_\xi(x)$ за $\chi_\xi(t)$: $f_\xi(x) = \frac{1}{2\pi}\int\limits_{-\infty}^{+\infty} e^{-itx}\chi_\xi(t)dt$.

4.1.2 Властивості характеристичної функції

1. $\chi_\xi(0) = 1$, оскільки $\int\limits_{-\infty}^{+\infty} f_\xi(x)dx = 1$.

   $|\chi_\xi(t)| \le 1$, оскільки $\left|\int\limits_{-\infty}^{+\infty} e^{itx}dF_\xi(x)\right| \le \int\limits_{-\infty}^{+\infty}\left|e^{itx}\right| dF_\xi(x) = \int\limits_{-\infty}^{+\infty} dF_\xi(x) = 1$.
2. Характеристична функція є рівномірно неперервною.

   Доведення. $|\chi_\xi(t+h) - \chi_\xi(t)| = \left|\int\limits_{-\infty}^{+\infty}\left(e^{i(t+h)x} - e^{itx}\right) dF_\xi(x)\right| \le$

   $\le \int\limits_{-\infty}^{+\infty}\left|e^{i(t+h)x} - e^{itx}\right| dF_\xi(x) = \int\limits_{-\infty}^{+\infty}\left|e^{itx}\right|\left|e^{ihx} - 1\right| dF_\xi(x) = \int\limits_{-\infty}^{+\infty}\left|e^{ihx} - 1\right| dF_\xi(x)$.

   $e^{ihx} - 1 = (\cos(hx) - 1) + i\sin(hx) = -2\sin^2(\frac{hx}{2}) + 2i\sin(\frac{hx}{2})\cos(\frac{hx}{2})$,

   $\left|e^{ihx} - 1\right| = \sqrt{4\sin^4(\frac{hx}{2}) + 4\sin^2(\frac{hx}{2})\cos^2(\frac{hx}{2})} = 2\left|\sin(\frac{hx}{2})\right|$.

   $\int\limits_{-\infty}^{+\infty}\left|e^{ihx} - 1\right| dF_\xi(x) = 2\int\limits_{-\infty}^{+\infty}\left|\sin(\frac{hx}{2})\right| dF_\xi(x) =$

   $\overset{A\ge 0}{=} 2\cdot\left(\int\limits_{-\infty}^{-A}\left|\sin(\frac{hx}{2})\right| dF_\xi(x) + \int\limits_{-A}^{A}\left|\sin(\frac{hx}{2})\right| dF_\xi(x) + \int\limits_{A}^{+\infty}\left|\sin(\frac{hx}{2})\right| dF_\xi(x)\right)$.

Інтеграл $\int\limits_{-A}^{A} \left|\sin(\frac{hx}{2})\right| dF_\xi(x)$ можна зробити як завгодно малим за рахунок вибору h, а інші два — за рахунок вибору A. ▲

3. $\eta = a\xi + b$ — афінне перетворення ξ, $a, b \in \mathbb{R}$.
   $\chi_\eta(t) = \mathbb{E}e^{it\eta} = \mathbb{E}e^{it(a\xi+b)} = e^{itb} \cdot \mathbb{E}e^{ita\xi} = e^{itb} \cdot \chi_\xi(at)$.
4. Якщо $\xi_1, ..., \xi_n$ — незалежні у сукупності, то $\chi_{\sum_{k=1}^n \xi_k}(t) = \prod\limits_{k=1}^n \chi_{\xi_k}(t)$.

   Доведення. $\chi_{\sum_{k=1}^n \xi_k}(t) = \mathbb{E}\left(e^{it\sum_{k=1}^n \xi_k}\right) = \mathbb{E}\left(\prod\limits_{k=1}^n e^{it\xi_k}\right) =$
   $= \left[e^{it\xi_k} \text{ — теж незалежні у сукупності}\right] = \prod\limits_{k=1}^n \mathbb{E}e^{it\xi_k} = \prod\limits_{k=1}^n \chi_{\xi_k}(t)$. ▲
5. За допомогою характеристичної функції можна знайти початкові моменти будь-якого порядку: $\chi_\xi^{(k)}(t) = \int\limits_{-\infty}^{+\infty} (ix)^k e^{itx} dF_\xi(x)$, причому $\chi_\xi^{(k)}(0) = \int\limits_{-\infty}^{+\infty} (ix)^k dF_\xi(x) = i^k \cdot \mathbb{E}\xi^k$.
   Отже, $\mathbb{E}\xi = -i\chi'_\xi(0)$, $\mathbb{D}\xi = \mathbb{E}\xi^2 - (\mathbb{E}\xi)^2 = -\chi''_\xi(0) + \left(\chi'_\xi(0)\right)^2$.
6. Зв'язок характеристичної функції та генератриси ДВВ.
   $G_\xi(z) = \sum\limits_{k=0}^{\infty} \mathbb{P}\{\xi = k\} z^k$, $\chi_\xi(t) = \sum\limits_{k=0}^{\infty} e^{itk}\mathbb{P}\{\xi = k\}$. Отже, $\chi_\xi(t) = G_\xi(e^{it})$.
7. $\overline{\chi_\xi(t)} = \mathbb{E}e^{-it\xi} = \chi_\xi(-t) = \chi_{-\xi}(t)$.
8. Для того, щоб характеристична функція була дійснозначною, необхідно і достатньо, щоб розподіл ВВ був симетричним відносно 0.

   Доведення. Розглянемо випадок НВВ. Нехай розподіл є симетричним відносно 0, тоді $f_\xi(x)$ — парна. $\chi_\xi(t) = \int\limits_{-\infty}^{+\infty} e^{itx} f_\xi(x)dx = \int\limits_{-\infty}^{+\infty} \cos(tx)f_\xi(x)dx + i\int\limits_{-\infty}^{+\infty} \sin(tx)f_\xi(x)dx = \int\limits_{-\infty}^{+\infty} \cos(tx)f_\xi(x)dx = Re\chi_\xi(t)$, інтеграл з синусом рівний 0, бо інтегрується непарна функція по симетричному проміжку.
   Нехай $\chi_\xi(t)$ — дійснозначна, тоді $\chi_\xi(t) = \overline{\chi_\xi(t)} = \chi_\xi(-t)$. Отже, $\chi_\xi(t)$ — парна, тоді з оберненого перетворення Фур'є $f_\xi(x)$ — теж парна. ▲

4.1.3 Необхідні умови того, що функція є характеристичною

Нехай $\chi(t)$ — деяка комплекснозначна функція дійсного аргументу. Якщо вона є характеристичною функцією деякої ВВ, то для неї має виконуватися:
1. $\chi(0) = 1$, $|\chi(t)| \le 1$.
2. $\chi(t)$ — рівномірно неперервна.
3. $\overline{\chi(t)} = \chi(-t)$.

Приклад. 1. $\chi(t) = \cos(t)$ — може бути характеристичною. Оскільки $\cos(t) = \frac{1}{2}e^{it} + \frac{1}{2}e^{-it}$, то відповідна ВВ — дискретна:

ξ	-1	1
p	$1/2$	$1/2$

2. $\chi(t) = \sin(t)$ — не може бути характеристичною, бо $\chi(0) = 0 \neq 1$.
3. $\chi(t) = \cos^2(t)$ — може бути характеристичною. $\cos^2(t) = \frac{1}{2} + \frac{1}{2}\cos(2t) = \frac{1}{2} + \frac{1}{4}e^{2it} + \frac{1}{4}e^{-2it}$, тому відповідна ВВ — дискретна:

ξ	-2	0	2
p	$1/4$	$1/2$	$1/4$

4. $\chi(t) = \frac{\alpha^2}{\alpha^2+t^2}$ — може бути характеристичною. Знайдемо щільність розподілу відповідної НВВ за формулою $f_\xi(x) = \frac{1}{2\pi}\int\limits_{-\infty}^{+\infty} \frac{\alpha^2 \cdot e^{-itx}}{\alpha^2+t^2}dt$.

   При $x < 0$: $\int\limits_{-\infty}^{+\infty} \frac{e^{-itx}}{\alpha^2+t^2}dt = 2\pi i \cdot \operatorname*{Res}\limits_{z=\alpha i} \frac{e^{-izx}}{\alpha^2+z^2} = 2\pi i \cdot \frac{e^{-i\cdot\alpha i\cdot x}}{2\alpha i} = \pi \cdot \frac{e^{\alpha x}}{\alpha}$.

   При $x > 0$ аналогічно отримуємо $\pi \cdot \frac{e^{-\alpha x}}{\alpha}$. Остаточно $f_\xi(x) = \frac{\alpha}{2}e^{-\alpha|x|}$ — це щільність розподілу Лапласа.

Вправа. Нехай $\chi(t)$ — характеристична функція деякого розподілу. Чи можуть бути характеристичними функції $\overline{\chi}$, χ^2, $|\chi|^2$, $Re\chi$, $|\chi|$, $Im\chi$?

Необхідні та достатні умови того, що функція є характеристичною, дає *теорема Бохнера-Хінчина*: $\chi(t)$ має бути рівномірно неперервною, $\chi(0)=1$ та невід'ємно визначеною:

$$\forall\, t_1,t_2,...,t_n\in\mathbb{R}, c_1,c_2,...,c_n\in\mathbb{C}: \sum_{k=1}^{n}\sum_{m=1}^{n}\chi(t_k-t_m)c_k\overline{c_m}\geq 0$$

4.1.4 Характеристичні функції деяких розподілів

1. $\xi\sim\mathrm{Bin}(n,p)$ — біноміальний розподіл. $G_\xi(z)=(pz+q)^n\Rightarrow\chi_\xi(t)=\left(pe^{it}+q\right)^n$.
2. $\xi\sim\mathrm{Geom}(p)$ — геометричний розподіл. $G_\xi(z)=\frac{pz}{1-qz}\Rightarrow\chi_\xi(t)=\frac{pe^{it}}{1-qe^{it}}$.
3. $\xi\sim\mathrm{Poiss}(a)$ — розподіл Пуассона. $G_\xi(z)=e^{a(z-1)}\Rightarrow\chi_\xi(t)=e^{a(e^{it}-1)}$.
4. $\xi\sim\mathrm{U}\left<a;b\right>$ — рівномірний розподіл.
   $\chi_\xi(t)=\frac{1}{b-a}\int\limits_a^b e^{itx}dx=\frac{e^{itb}-e^{ita}}{it(b-a)}$. Зокрема, при $\xi\sim U\left<-a;a\right>$ $\chi_\xi(t)=\begin{cases}\frac{\sin(at)}{at}, & t\neq 0\\ 1, & t=0\end{cases}$.
5. $\xi\sim\mathrm{Exp}(\lambda)$ — експоненційний розподіл.
   $\chi_\xi(t)=\lambda\int\limits_0^{+\infty}e^{itx}e^{-\lambda x}dx=\lambda\int\limits_0^{+\infty}e^{-(\lambda-it)x}dx=\frac{\lambda}{\lambda-it}$.
6. $\xi\sim\mathrm{N}(a,\sigma^2)$ — нормальний розподіл. Розглянемо стандартний розподіл $\eta\sim\mathrm{N}(0,1)$, тоді $\xi=a+\sigma\cdot\eta$.
   $\chi_\eta(t)=\frac{1}{\sqrt{2\pi}}\int\limits_{-\infty}^{+\infty}e^{itx}e^{-x^2/2}dx=\frac{1}{\sqrt{2\pi}}\int\limits_{-\infty}^{+\infty}e^{-x^2/2+itx}dx=\frac{1}{\sqrt{2\pi}}\int\limits_{-\infty}^{+\infty}e^{-\frac{1}{2}(x^2-2itx+i^2t^2)+\frac{1}{2}i^2t^2}dx=$
   $\frac{1}{\sqrt{2\pi}}e^{-\frac{t^2}{2}}\int\limits_{-\infty}^{+\infty}e^{-\frac{(x-it)^2}{2}}dx=[x-it=u,dx=du]=e^{-\frac{t^2}{2}}$. Отже, $\chi_\xi(t)=e^{iat-\frac{\sigma^2t^2}{2}}$.
7. $\xi\sim\Gamma(\alpha,\beta)$ — гамма-розподіл.
   $\chi_\xi(t)=\frac{\beta^\alpha}{\Gamma(\alpha)}\int\limits_0^{+\infty}e^{itx}x^{\alpha-1}e^{-\beta x}dx=\frac{\beta^\alpha}{\Gamma(\alpha)}\int\limits_0^{+\infty}x^{\alpha-1}e^{-(\beta-it)x}dx=\left[(\beta-it)x=u,dx=\frac{du}{\beta-it}\right]=$
   $\frac{\beta^\alpha}{\Gamma(\alpha)}\cdot\frac{1}{(\beta-it)^\alpha}\int\limits_0^{+\infty}u^{\alpha-1}e^{-u}du=\frac{\beta^\alpha}{\Gamma(\alpha)}\cdot\frac{1}{(\beta-it)^\alpha}\cdot\Gamma(\alpha)=(1-\frac{it}{\beta})^{-\alpha}$.

Означення 4.1.2. *Композицією законів розподілу* називається складання закону розподілу суми незалежних випадкових величин.

Закон розподілу називається *стійким по відношенню до операції додавання*, якщо закон розподілу суми незалежних випадкових величин, розподілених за цим законом (в загальному випадку з різними параметрами), є таким самим.

Приклад. 1. Перевірити стійкість нормального розподілу. Нехай $\xi_1\sim\mathrm{N}(a_1,\sigma_1),...,\xi_n\sim\mathrm{N}(a_n,\sigma_n)$ — незалежні у сукупності.

Знайдемо характеристичну функцію суми: $\chi_{\xi_1+...+\xi_n}(t)=\prod\limits_{k=1}^{n}\chi_{\xi_k}(t)=\prod\limits_{k=1}^{n}e^{ia_kt-\frac{\sigma_k^2t^2}{2}}=$ $\exp\left\{i\cdot\left(\sum\limits_{k=1}^{n}a_k\right)\cdot t-\frac{t^2}{2}\cdot\left(\sum\limits_{k=1}^{n}\sigma_k^2\right)\right\}$ — це характеристична функція нормального розподілу. Отже, $\xi_1+...+\xi_n\sim\mathrm{N}\left(\sum\limits_{k=1}^{n}a_k,\sqrt{\sum\limits_{k=1}^{n}\sigma_k^2}\right)$.

2. Перевірити стійкість біноміального розподілу. Нехай $\xi_1\sim\mathrm{Bin}(n_1,p_1)$, $\xi_2\sim\mathrm{Bin}(n_2,p_2)$ — незалежні. $\chi_{\xi_1+\xi_2}(t)=\left(p_1e^{it}+q_1\right)^{n_1}\cdot\left(p_2e^{it}+q_2\right)^{n_2}\stackrel{?}{=}\left(\mathbb{P}e^{it}+Q\right)^N$.
   Взагалі кажучи, біноміальний розподіл не є стійким по відношенню до операції додавання. Його називають *умовно стійким* при $p_1=p_2=p$, тоді $\chi_{\xi_1+\xi_2}(t)=\left(pe^{it}+q\right)^{n_1+n_2}$ і $\xi_1+\xi_2\sim\mathrm{Bin}(n_1+n_2,p)$.

Вправа. Перевірити стійкість гамма-розподілу $\Gamma(\alpha,\beta)$.

4.2 Характеристичні функції випадкових векторів

4.2.1 Поняття характеристичної функції випадкового вектору

Означення 4.2.1. *Характеристичною функцією* випадкового вектору $\vec{\xi} = (\xi_1, \xi_2, ..., \xi_n)^T$ називається комплекснозначна функція n дійсних аргументів $\chi_{\vec{\xi}}(\vec{t}) = \mathbb{E}e^{i(\vec{\xi},\vec{t})}$, $\vec{t} \in \mathbb{R}^n$.

$$\chi_{\vec{\xi}}(\vec{t}) = \int\limits_{\mathbb{R}^n} e^{i(\vec{x},\vec{t})} dF_{\vec{\xi}}(\vec{x}) = \begin{cases} \sum\limits_{k_1,k_2,...,k_n} e^{i\sum\limits_{j=1}^{n} x_{k_j} t_j} \mathbb{P}\left\{\xi_1 = x_{k_1}, ..., \xi_n = x_{k_n}\right\}, & \vec{\xi} \text{ — ДВВ} \\ \int\limits_{\mathbb{R}^n} e^{i(\vec{x},\vec{t})} f_{\vec{\xi}}(x_1, x_2, ..., x_n) dx_1 dx_2 ... dx_n, & \vec{\xi} \text{ — НВВ} \end{cases} \tag{1}$$

Приклад. Обчислення характеристичної функції для дискретного та неперервного випадкових векторів

1.

$\xi_2 \backslash \xi_1$	1	2
0	0.1	0.3
1	0.2	0.4

   $\chi_{\vec{\xi}}(t_1, t_2) = 0.1e^{i(1\cdot t_1 + 0\cdot t_2)} + 0.3e^{i(2\cdot t_1 + 0\cdot t_2)} + 0.2e^{i(1\cdot t_1 + 1\cdot t_2)} + 0.4e^{i(2\cdot t_1 + 1\cdot t_2)}$.

2. $\vec{\xi} \sim U(D)$, $D = \left\{(x; y) \in \mathbb{R}^2 : |x| \le 1, |y| \le 1\right\}$.

   $\chi_{\vec{\xi}}(t_1, t_2) = \frac{1}{4} \iint\limits_{D} e^{i(t_1 x + t_2 y)} dxdy = \frac{1}{4} \cdot \left(\int\limits_{-1}^{1} e^{it_1 x} dx\right) \cdot \left(\int\limits_{-1}^{1} e^{it_2 y} dy\right) = \frac{e^{it_1} - e^{-it_1}}{2it_1} \cdot \frac{e^{it_2} - e^{-it_2}}{2it_2} =$
   $\begin{cases} \frac{\sin(t_1)}{t_1} \cdot \frac{\sin(t_2)}{t_2}, & t_1 \ne 0, t_2 \ne 0 \\ \frac{\sin(t_1)}{t_1}, & t_2 = 0 \\ \frac{\sin(t_2)}{t_2}, & t_1 = 0 \\ 1, & t_1 = t_2 = 0 \end{cases}$

4.2.2 Властивості характеристичних функцій векторів

1. Характеристична функція існує для будь-якого розподілу випадкового вектора, бо $\left|\chi_{\vec{\xi}}(\vec{t})\right| \le 1$, $\chi_{\vec{\xi}}(\vec{0}) = 1$.
2. Характеристична функція рівномірно неперервна по кожному з аргументів.
3. За характеристичною функцією вектора можна знайти характеристичну функцію будь-якої підсистеми цього вектора. $\vec{\xi} = (\xi_1, \xi_2, ..., \xi_n)$, $\vec{\eta} = (\xi_1, \xi_2, ..., \xi_k)$, $k < n$.
   $\chi_{\vec{\eta}}(t_1, t_2, ..., t_k) = \chi_{\vec{\xi}}(t_1, t_2, ..., t_k, 0, ..., 0)$.
4. Якщо координати $\vec{\xi}$ незалежні у сукупності, то $\chi_{\vec{\xi}}(\vec{t}) = \prod\limits_{k=1}^{n} \chi_{\xi_k}(t_k)$.

   Доведення. $\chi_{\vec{\xi}}(\vec{t}) = \mathbb{E}\left(e^{i(\xi_1 t_1 + ... + \xi_n t_n)}\right) = \mathbb{E}\left(e^{i\xi_1 t_1} ... e^{i\xi_n t_n}\right) =$ [незалежні у сукупності] $= \mathbb{E}\left(e^{i\xi_1 t_1}\right) \cdot ... \cdot \mathbb{E}\left(e^{i\xi_n t_n}\right) = \prod\limits_{k=1}^{n} \chi_{\xi_k}(t_k)$. ▲

5. $\vec{\eta} = A\vec{\xi} + \vec{b}$ — афінне перетворення $\vec{\xi}$.
   $\chi_{\vec{\eta}}(\vec{t}) = \mathbb{E}e^{i(\vec{\eta},\vec{t})} = \mathbb{E}e^{i(A\vec{\xi}+\vec{b},\vec{t})} = \mathbb{E}e^{i(\vec{b},\vec{t})} \cdot \mathbb{E}e^{i(A\vec{\xi},\vec{t})} = e^{i(\vec{b},\vec{t})} \cdot \mathbb{E}e^{i(\vec{\xi},A^*\vec{t})} = e^{i(\vec{b},\vec{t})} \cdot \chi_{\vec{\xi}}(A^*\vec{t})$.
6. Характеристична функція вектора дозволяє знайти початкові моменти будь-якого порядку:

$$\mathbb{E}\xi_1^{k_1}\xi_2^{k_2}...\xi_n^{k_n} = \frac{1}{i^{k_1+k_2+...+k_n}} \cdot \left.\frac{\partial^{k_1+k_2+...+k_n}\chi_{\vec{\xi}}(\vec{t})}{\partial t_1^{k_1}\partial t_2^{k_2}...\partial t_n^{k_n}}\right|_{\vec{t}=\vec{0}}, \quad \mathbb{E}\xi_j = \frac{1}{i} \cdot \left.\frac{\partial\chi_{\vec{\xi}}(\vec{t})}{\partial t_j}\right|_{\vec{t}=\vec{0}}$$

$$\mathbb{K}(\xi_k, \xi_j) = \mathbb{E}\xi_k\xi_j - \mathbb{E}\xi_k \cdot \mathbb{E}\xi_j = \frac{1}{i^2}\left.\frac{\partial^2\chi_{\vec{\xi}}(\vec{t})}{\partial t_k \partial t_j}\right|_{\vec{t}=\vec{0}} - \left(\frac{1}{i} \cdot \left.\frac{\partial\chi_{\vec{\xi}}(\vec{t})}{\partial t_k}\right|_{\vec{t}=\vec{0}}\right) \cdot \left(\frac{1}{i} \cdot \left.\frac{\partial\chi_{\vec{\xi}}(\vec{t})}{\partial t_j}\right|_{\vec{t}=\vec{0}}\right) =$$

$$= -\left.\frac{\partial^2\chi_{\vec{\xi}}(\vec{t})}{\partial t_k \partial t_j}\right|_{\vec{t}=\vec{0}} + \left(\left.\frac{\partial\chi_{\vec{\xi}}(\vec{t})}{\partial t_k}\right|_{\vec{t}=\vec{0}}\right) \cdot \left(\left.\frac{\partial\chi_{\vec{\xi}}(\vec{t})}{\partial t_j}\right|_{\vec{t}=\vec{0}}\right)$$

Необхідні та достатні умови того, що комплекснозначна функція n дійсних змінних є характеристичною, дає *теорема Бохнера-Хінчина*: $\chi(\vec{t})$ має бути рівномірно неперервною по кожній змінній, $\chi(\vec{0}) = 1$ та невід'ємно визначеною:

$$\forall\ \vec{t^1}, \vec{t^2}, ..., \vec{t^n} \in \mathbb{R}^n, c_1, c_2, ..., c_n \in \mathbb{C} : \sum_{k=1}^{n} \sum_{m=1}^{n} \chi(\vec{t^k} - \vec{t^m}) c_k \overline{c_m} \geq 0$$

4.3 Багатовимірний нормальний розподіл

4.3.1 Виведення характеристичної функції та щільності

Розглянемо випадковий вектор $\vec{\xi} = (\xi_1, \xi_2, ..., \xi_n)^T$, $\xi_k \sim \mathrm{N}(a_k^o, \sigma_k)$ для $k = 1, ..., n$, координати незалежні у сукупності. За властивостями характеристичної функції та щільності $\chi_{\vec{\xi}}(\vec{t}) = \prod\limits_{k=1}^{n} \chi_{\xi_k}(t_k)$, $f_{\vec{\xi}}(\vec{x}) = \prod\limits_{k=1}^{n} f_{\xi_k}(x_k)$.

$$\chi_{\vec{\xi}}(\vec{t}) = \prod_{k=1}^{n} e^{i a_k^o t_k - \frac{\sigma_k^2 t_k^2}{2}} = \exp\left\{ i(\vec{a^o}, \vec{t}) - \frac{1}{2} \sum_{k=1}^{n} \sigma_k^2 t_k^2 \right\}$$

$$f_{\vec{\xi}}(\vec{x}) = \prod_{k=1}^{n} \frac{1}{\sqrt{2\pi}\sigma_k} e^{-\frac{(x_k - a_k^o)^2}{2\sigma_k^2}} = \frac{1}{(2\pi)^{n/2} \prod_{k=1}^{n} \sigma_k} \exp\left\{ -\frac{1}{2} \sum_{k=1}^{n} \frac{(x_k - a_k^o)^2}{\sigma_k^2} \right\}$$

Введемо матрицю $D = \begin{pmatrix} \sigma_1^2 & 0 & \cdots & 0 \\ 0 & \sigma_2^2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \sigma_n^2 \end{pmatrix}$, $D^{-1} = \begin{pmatrix} 1/\sigma_1^2 & 0 & \cdots & 0 \\ 0 & 1/\sigma_2^2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 1/\sigma_n^2 \end{pmatrix}$.

Тоді отримаємо

$$\chi_{\vec{\xi}}(\vec{t}) = \exp\left\{ i(\vec{a^o}, \vec{t}) - \frac{1}{2}(D\vec{t}, \vec{t}) \right\} \tag{1}$$

$$f_{\vec{\xi}}(\vec{x}) = \frac{1}{(2\pi)^{n/2}\sqrt{\det D}} \exp\left\{ -\frac{1}{2}\left(D^{-1}(\vec{x} - \vec{a^o}), (\vec{x} - \vec{a^o}) \right) \right\} \tag{2}$$

Причому в такому випадку D — кореляційна матриця.

Тепер розглянемо вектор $\vec{\eta} = A\vec{\xi}$, де A — невироджена матриця. За властивістю $\chi_{\vec{\eta}}(\vec{t}) = \chi_{\vec{\xi}}(A^*\vec{t})$. Маємо

$$\begin{aligned} \chi_{\vec{\eta}}(\vec{t}) = \exp\left\{ i(\vec{a^o}, A^*\vec{t}) - \frac{1}{2}(DA^*\vec{t}, A^*\vec{t}) \right\} = \exp\left\{ i(A\vec{a^o}, \vec{t}) - \frac{1}{2}(ADA^*\vec{t}, \vec{t}) \right\} = \\ = \exp\left\{ i(\vec{a}, \vec{t}) - \frac{1}{2}(K\vec{t}, \vec{t}) \right\} \end{aligned} \tag{3}$$

Доведемо, що $K = ADA^*$ — симетрична й додатно визначена, тому її можна вважати кореляційною матрицею.

1. $K^* = (ADA^*)^* = (A^*)^* DA^* = ADA^* = K$.
2. $\forall\ \vec{u} \in \mathbb{R}^n\ (K\vec{u}, \vec{u}) = (ADA^*\vec{u}, \vec{u}) = (DA^*\vec{u}, A^*\vec{u}) = [A^*\vec{u} = \vec{v}] = (D\vec{v}, \vec{v}) > 0$ для $\vec{u} \neq \vec{0}$.

Нехай в n-вимірному евклідовому просторі задано вектор $\vec{a}$ та симетричну додатно визначену матрицю K. Існує ортогональне перетворення U, яке дає можливість записати $K = UDU^*$, причому на діагоналі D стоять строго додатні власні числа K. Тоді функцію вигляду $\exp\left\{ i(\vec{a}, \vec{t}) - \frac{1}{2}(K\vec{t}, \vec{t}) \right\}$ можна розглядати як характеристичну функцію випадкового вектора $\vec{\eta} = U\vec{\xi}$, де $\vec{\xi}$ — вектор, координати якого незалежні у сукупності та мають нормальний розподіл.

Означення 4.3.1. *n-вимірним гауссівським вектором* $\vec{\xi}$ називається випадковий вектор, характеристична функція якого має вигляд $\chi_{\vec{\xi}}(\vec{t}) = \exp\left\{ i(\vec{a}, \vec{t}) - \frac{1}{2}(K\vec{t}, \vec{t}) \right\}$, де K — кореляційна матриця, а $\vec{a} = (\mathbb{E}\xi_1, \mathbb{E}\xi_2, ..., \mathbb{E}\xi_n)^T$ — центр розсіювання.

Позначення: $\vec{\xi} \sim \mathrm{N}(\vec{a}, K)$. $\mathrm{N}(\vec{0}, I)$ — стандартний нормальний розподіл.

Вправа. Перевірити, що $\vec{a}$ дійсно є центром розсіювання.

Знайдемо щільність сумісну розподілу такого вектору:

$$\mathbb{P}\{\vec{\eta} \in C \subset \mathbb{R}^n\} = \mathbb{P}\left\{U\vec{\xi} \in C\right\} = \mathbb{P}\left\{\vec{\xi} \in U^{-1}C\right\} = \int\limits_{U^{-1}C} f_{\vec{\xi}}(\vec{x}) d\vec{x} =$$

$$\int\limits_{U^{-1}C} \frac{1}{(2\pi)^{n/2}\sqrt{\det D}} \exp\left\{-\frac{1}{2}\left(D^{-1}(\vec{x}-\vec{a^o}), (\vec{x}-\vec{a^o})\right)\right\} d\vec{x} = \left[\vec{x}-\vec{a^o} = U^{-1}(\vec{y}-\vec{a}) =\right.$$

$$\left.= U^*(\vec{y}-\vec{a}), d\vec{x} = d\vec{y}\right] = \frac{1}{(2\pi)^{n/2}\sqrt{\det D}} \int\limits_{C} \exp\left\{-\frac{1}{2}\left(D^{-1}(U^*\vec{y}-U^*\vec{a}), (U^*\vec{y}-U^*\vec{a})\right)\right\} d\vec{y} =$$

$$= \left[UD^{-1}U^* = K^{-1}\right] = \frac{1}{(2\pi)^{n/2}\sqrt{\det K}} \int\limits_{C} \exp\left\{-\frac{1}{2}\left(K^{-1}(\vec{y}-\vec{a}), (\vec{y}-\vec{a})\right)\right\} d\vec{y}$$

Оскільки $\mathbb{P}\{\vec{\eta} \in C\} = \int\limits_C f_{\vec{\eta}}(\vec{y}) d\vec{y}$, отримуємо **щільність розподілу** гауссівського вектора $\vec{\xi} \sim \mathrm{N}(\vec{a}, K)$:

$$f_{\vec{\xi}}(\vec{x}) = \frac{1}{(2\pi)^{n/2}\sqrt{\det K}} \exp\left\{-\frac{1}{2}\left(K^{-1}(\vec{x}-\vec{a}), (\vec{x}-\vec{a})\right)\right\} \tag{4}$$

4.3.2 Властивості гауссівських векторів

1. Якщо $\vec{\xi}$ — гауссівський вектор, то всі його координати гауссівські, а будь-яка підсистема теж є гауссівським вектором.

   Доведення. $\chi_{\vec{\xi}}(0, 0, ..., t_j, ..., 0) = \chi_{\xi_j}(t_j) = e^{ia_jt_j - \frac{1}{2}\sigma_j^2 t_j^2} \Rightarrow \xi_j \sim \mathrm{N}(a_j, \sigma_j^2)$.
   Аналогічно для будь-якої підсистеми $\vec{\eta} = (\xi_{k_1}, \xi_{k_2}, ..., \xi_{k_n})^T$. ▲

   Зауваження. Обернене твердження, взагалі кажучи, не є вірним.
   Розглянемо випадковий вектор із щільністю

   $$f_{\vec{\xi}}(x, y) = \frac{1}{2\pi}\left(\left(\sqrt{2}e^{-x^2/2} - e^{-x^2}\right)e^{-y^2} + \left(\sqrt{2}e^{-y^2/2} - e^{-y^2}\right)e^{-x^2}\right)$$

   Очевидно, це не щільність нормального розподілу. Знайдемо щільності розподілу координат ξ_1 та ξ_2.

   $$f_{\xi_1}(x) = \int\limits_{-\infty}^{+\infty} f_{\vec{\xi}}(x, y) dy = \frac{\sqrt{2}}{2\pi} e^{-x^2/2} \int\limits_{-\infty}^{+\infty} e^{-y^2} dy - \frac{1}{2\pi} e^{-x^2} \int\limits_{-\infty}^{+\infty} e^{-y^2} dy +$$

   $$+ \frac{\sqrt{2}}{2\pi} e^{-x^2} \int\limits_{-\infty}^{+\infty} e^{-y^2/2} dy - \frac{1}{2\pi} e^{-x^2} \int\limits_{-\infty}^{+\infty} e^{-y^2} dy = \frac{\sqrt{2}}{2\pi} e^{-x^2/2} \cdot \sqrt{\pi} - \frac{1}{2\pi} e^{-x^2} \cdot \sqrt{\pi} +$$

   $$+ \frac{\sqrt{2}}{2\pi} e^{-x^2} \cdot \sqrt{2\pi} - \frac{1}{2\pi} e^{-x^2} \cdot \sqrt{\pi} = \frac{1}{\sqrt{2\pi}} e^{-x^2/2} - \frac{1}{2\sqrt{\pi}} e^{-x^2} + \frac{1}{\sqrt{\pi}} e^{-x^2} - \frac{1}{2\sqrt{\pi}} e^{-x^2} =$$

   $$= \frac{1}{\sqrt{2\pi}} e^{-x^2/2} \Rightarrow \xi_1 \sim \mathrm{N}(0, 1).$$

   Аналогічно $\xi_2 \sim \mathrm{N}(0, 1)$. Отже, координати вектора мають нормальний розподіл, а сам вектор — ні.

2. Для гауссівського випадкового вектора поняття незалежності та некорельованості координат є еквівалентними.

   Доведення. Було доведено, що з незалежності координат випливає її некорельованість. Якщо координати некорельовані, то кореляційна матриця $K = \mathrm{diag}(\sigma_1^2, \sigma_2^2, ..., \sigma_n^2)$. Тоді $\left(K\vec{t}, \vec{t}\right) = \sum_{k=1}^n \sigma_k^2 t_k^2$. Тоді $\chi_{\vec{\xi}}(\vec{t}) = \exp\left\{i(\vec{a}, \vec{t}) - \frac{1}{2}\left(K\vec{t}, \vec{t}\right)\right\} = \prod\limits_{k=1}^n e^{ia_kt_k - \frac{1}{2}\sigma_k^2 t_k^2} = \prod\limits_{k=1}^n \chi_{\xi_k}(t_k)$ і координати незалежні. ▲

3. В n-вимірному евклідовому просторі $\mathbb{R}^n$ завжди можна перейти до ортонормованого базису з власних векторів матриці K, в якому K приймає діагональний вигляд. Отже, в базисі з власних векторів матриці K координати відповідного гауссівського вектора стають незалежними.
4. *Афінне перетворення гауссівських векторів.*
$\vec{\xi} \sim \mathrm{N}(\vec{a}, K)$. $A : \mathbb{R}^n \to \mathbb{R}^m$ — лінійний оператор, заданий матрицею, $\vec{b} \in \mathbb{R}^m$, $\vec{\eta} = A\vec{\xi} + \vec{b}$. За властивістю характеристичної функції $\chi_{\vec{\eta}}(\vec{t}) = e^{i(\vec{b},\vec{t})} \cdot \chi_{\vec{\xi}}(A^*\vec{t}) = \exp\left\{i(\vec{b},\vec{t})\right\} \cdot \exp\left\{i(\vec{a}, A^*\vec{t}) - \frac{1}{2}\left(KA^*\vec{t}, A^*\vec{t}\right)\right\} = \exp\left\{i(A\vec{a}+\vec{b},\vec{t}) - \frac{1}{2}\left(AKA^*\vec{t},\vec{t}\right)\right\}$.
Отже, $\vec{\eta} \sim \mathrm{N}(A\vec{a}+\vec{b}, AKA^*)$.

Приклад. 1. Задано вектор $\vec{\xi} \sim \mathrm{N}(\vec{a}, K)$. Знайти розподіл $\eta = \xi_1 + \xi_2 + ... + \xi_n$.
$\eta = \begin{pmatrix}1 & 1 & \cdots & 1\end{pmatrix} \cdot \begin{pmatrix}\xi_1 & \xi_2 & \cdots & \xi_n\end{pmatrix}^T = A\vec{\xi}$. $\mathbb{E}\eta = A\vec{a} = \mathbb{E}\xi_1 + \mathbb{E}\xi_2 + ... + \mathbb{E}\xi_n = \sum\limits_{k=1}^{n} \mathbb{E}\xi_k$.
$\mathbb{D}\eta = AKA^* = \begin{pmatrix}1 & 1 & \cdots & 1\end{pmatrix} \cdot K \cdot \begin{pmatrix}1 & 1 & \cdots & 1\end{pmatrix}^T$. Якщо $K = \mathrm{diag}(\sigma_1^2, \sigma_2^2, ..., \sigma_n^2)$, то $\mathbb{D}\eta = \sum\limits_{k=1}^{n} \mathbb{D}\xi_k$, а в загальному випадку у цій сумі ще будуть доданки вигляду $2\mathbb{K}\xi_i\xi_j$.

2. $\vec{\xi} = \begin{pmatrix}\xi_1 \\ \xi_2\end{pmatrix} \sim \mathrm{N}\left(\begin{pmatrix}0 \\ 1\end{pmatrix}, \begin{pmatrix}1 & 2 \\ 2 & 6\end{pmatrix}\right) = \mathrm{N}(\vec{a}, K)$. Знайти розподіл, характеристичну функцію, щільність розподілу $\vec{\eta} = (3\xi_1 - 2\xi_2 + 1, \xi_1 + 3\xi_2)^T$ та розподіл координати η_1.
Запишемо $\vec{\eta}$ у вигляді $\vec{\eta} = A\vec{\xi} + \vec{b}$: $\vec{\eta} = \begin{pmatrix}3 & -2 \\ 1 & 3\end{pmatrix}\begin{pmatrix}\xi_1 \\ \xi_2\end{pmatrix} + \begin{pmatrix}1 \\ 0\end{pmatrix}$. Тоді $\vec{\eta} \sim \mathrm{N}(\vec{d}, B)$, $\vec{d} = A\vec{a} + \vec{b} = \begin{pmatrix}3 & -2 \\ 1 & 3\end{pmatrix}\begin{pmatrix}0 \\ 1\end{pmatrix} + \begin{pmatrix}1 \\ 0\end{pmatrix} = \begin{pmatrix}-2 \\ 3\end{pmatrix} + \begin{pmatrix}1 \\ 0\end{pmatrix} = \begin{pmatrix}-1 \\ 3\end{pmatrix} = \begin{pmatrix}\mathbb{E}\eta_1 \\ \mathbb{E}\eta_2\end{pmatrix}$, $B = AKA^* = \begin{pmatrix}3 & -2 \\ 1 & 3\end{pmatrix}\begin{pmatrix}1 & 2 \\ 2 & 6\end{pmatrix}\begin{pmatrix}3 & 1 \\ -2 & 3\end{pmatrix} = \begin{pmatrix}9 & -19 \\ -19 & 67\end{pmatrix}$.
$\chi_{\vec{\eta}}(t_1, t_2) = \exp\left\{i(\vec{d},\vec{t}) - \frac{1}{2}\left(B\vec{t},\vec{t}\right)\right\} = \exp\left\{i(-1t_1 + 3t_2) - \frac{1}{2}(9t_1^2 - 38t_1t_2 + 67t_2^2)\right\}$.
$B^{-1} = \begin{pmatrix}67/142 & 19/142 \\ 19/142 & 9/142\end{pmatrix}$, $\det B = 142$.
$f_{\vec{\eta}}(x_1, x_2) = \frac{1}{2\pi\cdot\sqrt{142}} \cdot \exp\left\{-\frac{1}{2}\left(B^{-1}(\vec{x}-\vec{d}), (\vec{x}-\vec{d})\right)\right\} =$
$= \frac{1}{2\pi\cdot\sqrt{142}} \cdot \exp\left\{-\frac{1}{2}\left(\frac{67}{142}(x_1+1)^2 + \frac{38}{142}(x_1+1)(x_2-3) + \frac{9}{142}(x_2-3)^2\right)\right\}$.
Знайдемо розподіл η_1 за допомогою характеристичної функції: $\chi_{\eta_1}(t) = \chi_{\vec{\eta}}(t, 0) =$
$= \exp\left\{-it - \frac{1}{2}\cdot 9t^2\right\} \Rightarrow \eta_1 \sim \mathrm{N}(-1, 9)$.

4.3.3 Вироджені гауссівські вектори

Розглянемо випадок n-вимірного нормального розподілу $\vec{\xi} \sim \mathrm{N}(\vec{a}, K)$ з виродженою матрицею K, у якої $\mathrm{rang}K = m < n$. Існує ортогональне перетворення U таке, що $K = UDU^*$, $D = \mathrm{diag}(\lambda_1, \lambda_2, ..., \lambda_m, 0, ..., 0)$. Знайдемо розподіл вектора $\vec{\eta} = U^*\vec{\xi}$ за властивістю характеристичної функції:

$$\chi_{\vec{\eta}}(\vec{t}) = \chi_{\vec{\xi}}(U^{**}\vec{t}) = \chi_{\vec{\xi}}(U\vec{t}) = \exp\left\{i(\vec{a}, U\vec{t}) - \frac{1}{2}(KU\vec{t}, U\vec{t})\right\} = \exp\left\{i(U^*\vec{a}, \vec{t}) - \frac{1}{2}(U^*KU\vec{t}, \vec{t})\right\} =$$

$$= \left[U^*\vec{a} = \vec{b}, U^*KU = D\right] = \exp\left\{i(\vec{b},\vec{t}) - \frac{1}{2}(D\vec{t},\vec{t})\right\} = \exp\left\{i\sum_{k=1}^{n} b_k t_k - \frac{1}{2}\sum_{k=1}^{m} \lambda_k t_k^2\right\} =$$

$$= \exp\left\{i\sum_{k=m+1}^{n} b_k t_k\right\} \cdot \exp\left\{i\sum_{k=1}^{m} b_k t_k - \frac{1}{2}\sum_{k=1}^{m} \lambda_k t_k^2\right\}$$

Перший множник — це характеристична функція сталого вектору, а другий — характеристична функція гауссівського вектору меншої розмірності.

Позначимо $\vec{c} = (0, ..., 0, b_{m+1}, ..., b_n)^T$, $\vec{d} = (b_1, ..., b_m, 0, ..., 0)^T$, $\vec{\zeta} = (\zeta_1, ..., \zeta_m, 0, ..., 0)^T$, $\zeta_k \sim N(0, 1)$ та незалежні у сукупності, $\Lambda = \mathrm{diag}(\sqrt{\lambda_1}, \sqrt{\lambda_2}, ..., \sqrt{\lambda_m}, 1, ..., 1)$.

Тоді $\vec{\eta} = \vec{c} + \left(\Lambda\vec{\zeta} + \vec{d}\right)$. Оскільки $\vec{\eta} = U^*\vec{\xi}$, то $\vec{\xi} = U\vec{\eta} = U\Lambda\vec{\zeta} + U(\vec{c} + \vec{d}) = U\Lambda\vec{\zeta} + \vec{a}$. Позначимо $U\Lambda = L$. Остаточно: $\vec{\xi} = L\vec{\zeta} + \vec{a}$.

Що тепер можна сказати про розподіл такого вектора? Оскільки перші m координат $\vec{\zeta}$ можуть приймати довільні дійсні значення, то можливі значення вектора $L\vec{\zeta}$ — це точки з лінійної оболонки векторів $\{u_1, u_2, ..., u_m\}$ (це власні вектори, що відповідають ненульовим власним числам K). Оскільки власні вектори, що відповідають нульовому власному числу, утворюють базис $\mathrm{Ker}K$, то лінійна оболонка $\{u_1, u_2, ..., u_m\}$ — це $\mathrm{Im}K$.

Висновок: n-вимірний гауссівський вектор з виродженою кореляційною матрицею, що має ранг $m < n$, можна представити у вигляді суми свого вектора математичного сподівання та вектора, перші m координат є незалежними у сукупності та мають стандартний нормальний розподіл, а інші координати нульові, помноженого на невироджену матрицю.

Усі значення n-вимірного гауссівського вектора $\vec{\xi} \sim \mathrm{N}(\vec{a}, K)$ з виродженою кореляційною матрицею, що має ранг $m < n$, зосереджені на m-вимірному підпросторі $\mathbb{R}^n$ — $\mathrm{Im}K + \vec{a}$.

4.3.4 Нормальний розподіл на площині

Розглянемо двовимірний гауссівський вектор $\vec{\xi} = (\xi_1, \xi_2)^T \sim \mathrm{N}(\vec{a}, K)$.

Нехай $\vec{a} = \begin{pmatrix} a_1 \\ a_2 \end{pmatrix}$, $K = \begin{pmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{pmatrix}$, $\det K = \sigma_1^2\sigma_2^2 - \rho^2\sigma_1^2\sigma_2^2 = (1-\rho^2)\sigma_1^2\sigma_2^2$, $K^{-1} = \frac{1}{(1-\rho^2)\sigma_1^2\sigma_2^2}\begin{pmatrix} \sigma_2^2 & -\rho\sigma_1\sigma_2 \\ -\rho\sigma_1\sigma_2 & \sigma_1^2 \end{pmatrix}$. Окремо запишемо щільність:

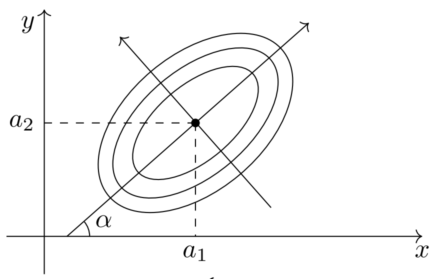
$$f_{\vec{\xi}}(x,y) = \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}} \exp\left\{-\frac{1}{2} \cdot \frac{\left(\sigma_2^2(x-a_1)^2 - 2\rho\sigma_1\sigma_2(x-a_1)(y-a_2) + \sigma_1^2(y-a_2)^2\right)}{\sigma_1^2\sigma_2^2(1-\rho^2)}\right\} =$$
$$= \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}} \exp\left\{-\frac{1}{2(1-\rho^2)} \cdot \left(\frac{(x-a_1)^2}{\sigma_1^2} - 2\rho\frac{x-a_1}{\sigma_1}\frac{y-a_2}{\sigma_2} + \frac{(y-a_2)^2}{\sigma_2^2}\right)\right\} \tag{5}$$

При $\rho = 0$ отримуємо $f_{\vec{\xi}}(x,y) = \frac{1}{\sqrt{2\pi}\sigma_1}e^{-\frac{(x-a_1)^2}{2\sigma_1^2}} \cdot \frac{1}{\sqrt{2\pi}\sigma_2}e^{-\frac{(y-a_2)^2}{2\sigma_2^2}} = f_{\xi_1}(x) \cdot f_{\xi_2}(y)$.

Поверхня, утворена графіком щільності, називається *поверхнею* або *палаткою Гаусса*.

Лінії рівня цієї поверхні задають *еліпси розсіювання*:

$$f_{\vec{\xi}}(x,y) = C \Leftrightarrow \frac{(x-a_1)^2}{\sigma_1^2} - 2\rho\frac{x-a_1}{\sigma_1}\frac{y-a_2}{\sigma_2} + \frac{(y-a_2)^2}{\sigma_2^2} = \lambda^2$$



$\mathrm{tg}2\alpha = \frac{2\rho\sigma_1\sigma_2}{\sigma_1^2 - \sigma_2^2}$

Осі еліпсів називаються *головними осями розсіювання.*

Якщо $\rho = 0$, то головні осі розсіювання паралельні осям координат.

Позначимо $\mathcal{E}_\lambda = \left\{(x;y) \in \mathbb{R}^2 : \frac{(x-a_1)^2}{\sigma_1^2} - 2\rho\frac{x-a_1}{\sigma_1}\frac{y-a_2}{\sigma_2} + \frac{(y-a_2)^2}{\sigma_2^2} \le \lambda^2\right\}$ та знайдемо ймовірність потрапляння в цю область у випадку $\rho = 0$:

$$\mathbb{P}\left\{\vec{\xi} \in \mathcal{E}_\lambda\right\} = \iint\limits_{\mathbb{E}_\lambda} f_{\vec{\xi}}(x,y)dxdy = \frac{1}{2\pi\sigma_1\sigma_2} \iint\limits_{\mathbb{E}_\lambda} \exp\left\{-\frac{1}{2}\left(\frac{(x-a_1)^2}{\sigma_1^2} + \frac{(y-a_2)^2}{\sigma_2^2}\right)\right\} dxdy =$$

$$\begin{bmatrix} x-a_1 = \lambda\sigma_1 r\cos\varphi \\ y-a_2 = \lambda\sigma_2 r\sin\varphi \\ |\mathcal{J}| = \lambda^2\sigma_1\sigma_2 r \end{bmatrix} = \frac{\lambda^2}{2\pi}\int\limits_0^{2\pi} d\varphi \int\limits_0^1 re^{-\frac{\lambda^2 r^2}{2}}dr = \int\limits_0^1 e^{-\frac{\lambda^2 r^2}{2}} d\left(\frac{\lambda^2 r^2}{2}\right) = 1 - e^{-\frac{\lambda^2}{2}}$$

Вправа. Довести, що у випадку $\rho > 0$ $\mathbb{P}\left\{\vec{\xi} \in \mathcal{E}_\lambda\right\} = 1 - e^{-\frac{\lambda^2}{2(1-\rho^2)}}$.

Приклад. $\vec{\xi} \sim \mathrm{N}\left(\begin{pmatrix}-1\\1\end{pmatrix}, \begin{pmatrix}1 & 2\\2 & 16\end{pmatrix}\right)$. Знайти рівняння еліпса розсіювання, для якого $\mathbb{P}\left\{\vec{\xi} \in \mathcal{E}_\lambda\right\} = 0.93$. З кореляційної матриці $\rho = 0.5$. Розв'яжемо відносно λ^2 рівняння $1 - e^{-\frac{\lambda^2}{2(1-0.5^2)}} = 0.93 \Leftrightarrow e^{-\frac{\lambda^2}{1.5}} = 0.07 \Rightarrow \lambda^2 \approx 4$. Тому рівняння еліпса $\frac{(x+1)^2}{1} - \frac{x+1}{1}\frac{y-1}{4} + \frac{(y-1)^2}{16} = 4$ або $\frac{(x+1)^2}{4} - \frac{(x+1)(y-1)}{16} + \frac{(y-1)^2}{64} = 1$.

Розглянемо гауссівський вектор $\vec{\xi} = (\xi_1, \xi_2)$ з незалежними координатами. Для нього $F_{\vec{\xi}}(x,y) = F_{\xi_1}(x)F_{\xi_2}(y) = \left(\frac{1}{2} + \Phi\left(\frac{x-a_1}{\sigma_1}\right)\right)\left(\frac{1}{2} + \Phi\left(\frac{y-a_2}{\sigma_2}\right)\right)$. Тоді ймовірність потрапляння в прямокутник
$Pi = \left\{(x;y) \in \mathbb{R}^2 : \alpha \le x < \beta, \gamma \le y < \delta\right\}$ дорівнює

$$\left(\Phi\left(\frac{\beta-a_1}{\sigma_1}\right) - \Phi\left(\frac{\alpha-a_1}{\sigma_1}\right)\right)\cdot\left(\Phi\left(\frac{\delta-a_2}{\sigma_2}\right) - \Phi\left(\frac{\gamma-a_2}{\sigma_2}\right)\right)$$

Вправа. Довести формулу для $\mathbb{P}\left\{\vec{\xi} \in Pi\right\}$.

Для гауссівського вектору з незалежними координатами також має місце «правило 3σ»:

$$\mathbb{P}\left\{\vec{\xi} \in (a_1 - 3\sigma_1; a_1 + 3\sigma_1) \times (a_2 - 3\sigma_2; a_2 + 3\sigma_2)\right\} =$$
$$= \mathbb{P}\left\{\xi_1 \in (a_1 - 3\sigma_1; a_1 + 3\sigma_1)\right\} \cdot \mathbb{P}\left\{\xi_2 \in (a_2 - 3\sigma_2; a_2 + 3\sigma_2)\right\} \approx 0.9973^2 \approx 0.9946$$

4.3.5 Колове розсіювання

Означення 4.3.2. Двовимірний гауссівський вектор має *колове розсіювання*, якщо $\vec{a} = \vec{0}$, $K = \begin{pmatrix}\sigma^2 & 0\\0 & \sigma^2\end{pmatrix}$. В цьому випадку $f_{\vec{\xi}}(x,y) = \frac{1}{2\pi\sigma^2}\exp\left\{-\frac{x^2+y^2}{2\sigma^2}\right\}$, а еліпси розсіювання стають колами.

Знайдемо ймовірність потрапляння такого $\vec{\xi}$ в коло $K_R = \left\{(x;y) \in \mathbb{R}^2 : x^2 + y^2 \le R^2\right\}$:

$$\mathbb{P}\left\{\vec{\xi} \in K_R\right\} = \frac{1}{2\pi\sigma^2}\iint\limits_{K_R} e^{-\frac{x^2+y^2}{2\sigma^2}}dxdy = \begin{bmatrix} x = \sigma r\cos\varphi \\ y = \sigma r\sin\varphi \\ |\mathcal{J}| = \sigma^2 r \end{bmatrix} = \frac{1}{2\pi}\int\limits_0^{2\pi} d\varphi \int\limits_0^{R/\sigma} re^{-\frac{r^2}{2}}dr =$$
$$= \int\limits_0^{R/\sigma} e^{-\frac{r^2}{2}} d\left(\frac{r^2}{2}\right) = 1 - e^{-\frac{R^2}{2\sigma^2}}$$

Таким чином, тепер відомий розподіл норми $\vec{\xi}$. $\eta = \|\vec{\xi}\| = \sqrt{\xi_1^2 + \xi_2^2}$ — випадкова величина, для якої $F_\eta(R) = \mathbb{P}\left\{\|\vec{\xi}\| < R\right\} = \begin{cases} 1 - e^{-\frac{R^2}{2\sigma^2}}, & R > 0 \\ 0, & R \le 0 \end{cases}$ — це *розподіл Релея.* Ця властивість переноситься й на довільну скінченну розмірність.

Вправа. Знайти основні числові характеристики розподілу Релея.

4.3.6 Умовний гауссівський розподіл на площині

Знайдемо одну з умовних щільностей розподілу:

$$f_{\xi_1}(x/\xi_2=y)=\frac{f_{\vec{\xi}}(x,y)}{f_{\xi_2}(y)}=\frac{\frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}}\exp\left\{-\frac{1}{2(1-\rho^2)}\cdot\left(\frac{(x-a_1)^2}{\sigma_1^2}-2\rho\frac{x-a_1}{\sigma_1}\frac{y-a_2}{\sigma_2}+\frac{(y-a_2)^2}{\sigma_2^2}\right)\right\}}{\frac{1}{\sqrt{2\pi}\sigma_2}\exp\left\{-\frac{1}{2\sigma_2^2}(y-a_2)^2\right\}}=$$

$$\frac{1}{\sqrt{2\pi}\sigma_1\sqrt{1-\rho^2}}\exp\left\{-\frac{1}{2(1-\rho^2)}\left(\frac{(x-a_1)^2}{\sigma_1^2}-2\rho\frac{(x-a_1)(y-a_2)}{\sigma_1\sigma_2}+\frac{(y-a_2)^2}{\sigma_2^2}\right)+\frac{(y-a_2)^2}{2\sigma_2^2}\right\}=$$

$$=\left[\frac{(y-a_2)^2}{\sigma_2^2}-\frac{(1-\rho^2)(y-a_2)^2}{\sigma_2^2}=\rho^2\frac{(y-a_2)^2}{\sigma_2^2}\right]=$$

$$=\frac{1}{\sqrt{2\pi}\sigma_1\sqrt{1-\rho^2}}\exp\left\{-\frac{1}{2(1-\rho^2)}\left(\frac{x-a_1}{\sigma_1}-\rho\cdot\frac{y-a_2}{\sigma_2}\right)^2\right\}=$$

$$=\frac{1}{\sqrt{2\pi}\sigma_1\sqrt{1-\rho^2}}\exp\left\{-\frac{1}{2\sigma_1^2(1-\rho^2)}\left(x-\left(a_1+\frac{\rho\sigma_1(y-a_2)}{\sigma_2}\right)\right)^2\right\}$$

Аналогічно $f_{\xi_2}(y/\xi_1=x)=\frac{1}{\sqrt{2\pi}\sigma_2\sqrt{1-\rho^2}}\exp\left\{-\frac{1}{2\sigma_2^2(1-\rho^2)}\left(y-\left(a_2+\frac{\rho\sigma_2(x-a_1)}{\sigma_1}\right)\right)^2\right\}$.
Отже, обидві умовні щільності є щільностями нормального розподілу.

$\mathbb{E}\,(\xi_1/\xi_2=y)=a_1+\frac{\rho\sigma_1(y-a_2)}{\sigma_2}$, $\mathbb{E}\,(\xi_2/\xi_1=x)=a_2+\frac{\rho\sigma_2(x-a_1)}{\sigma_1}$. Лінії регресії — прямі, $\rho\frac{\sigma_1}{\sigma_2}=\frac{K\xi_1\xi_2}{D\xi_2}$ та $\rho\frac{\sigma_2}{\sigma_1}=\frac{K\xi_1\xi_2}{D\xi_1}$ — відповідні кутові коефіцієнти прямих регресії.

$\mathbb{D}\,(\xi_1/\xi_2=y)=\sigma_1^2(1-\rho^2)$, $\mathbb{D}\,(\xi_2/\xi_1=x)=\sigma_2^2(1-\rho^2)$. Умовні дисперсії є сталими, ця властивість називається *гомоскедастичністю*.

Розділ 5

Функції випадкових аргументів

5.1 Функції одного випадкового аргументу

5.1.1 Функції від дискретного випадкового аргументу

Нехай ξ — ДВВ з рядом розподілу

ξ	x_1	x_2	...	x_n	...
p	p_1	p_2	...	p_n	...

, а φ — деяка вимірна числова функція, область визначення якої містить можливі значення ξ. Можна розглядати випадкову величину $\eta = \varphi(\xi)$.

Очевидно, що η — теж ДВВ, що приймає значення $y_k = \varphi(x_k)$, $k = 1, 2, ...$ з ймовірностями $\mathbb{P}\{\eta = y_k\} = \mathbb{P}\{\xi = x_k\} = p_k$. Однак, можливо, що $y_k = \varphi(x_k) = \varphi(x_s) = ... = \varphi(x_r)$. Тоді $\mathbb{P}\{\eta = y_k\} = p_k + p_s + ... + p_r$.

Приклад.

ξ	-2	-1	0	1	2
p	$1/9$	$2/9$	$1/9$	$2/9$	$3/9$

. Знайти закон розподілу $\eta = \xi^2$.

З ряду розподілу ξ видно, що ξ^2 приймає значення $0, 1, 4$ з ймовірностями $1/9$, $2/9 + 2/9$ та $1/9 + 3/9$. Отже, маємо ряд розподілу η:

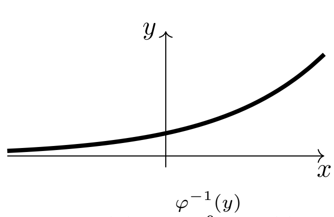
η	0	1	4
p	$1/9$	$4/9$	$4/9$

З алгоритму побудови ряду розподілу функції від дискретного випадкового аргументу $\mathbb{E}(\varphi(\xi))^k = \sum\limits_{i=1}^{n(\infty)} \varphi^k(x_i)\mathbb{P}\{\xi = x_i\}$.

5.1.2 Розподіл монотонної функції від неперервного випадкового аргументу

Нехай ξ — НВВ, $f_\xi(x)$ — її щільність розподілу, а φ — монотонна неперервна числова функція. Розглянемо два випадки.

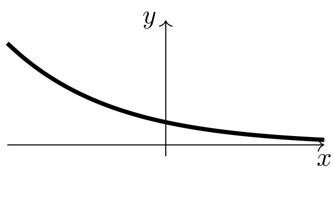
1. $\eta = \varphi(\xi)$, де φ — монотонно зростаюча.



$F_\eta(y) = \mathbb{P}\{\eta < y\} = \mathbb{P}\{\varphi(\xi) < y\}$.
Нехай $y = \varphi(x)$, тоді з монотонності φ маємо рівність подій $\{\varphi(\xi) < y\} = \{\xi < x\}$. Тому $F_\eta(y) = F_\xi(x) = \int\limits_{-\infty}^{x} f_\xi(t)dt$. φ має обернену, $x = \varphi^{-1}(y)$.

Отже, $F_\eta(y) = \int\limits_{-\infty}^{\varphi^{-1}(y)} f_\xi(t)dt$, $f_\eta(y) = F'_\eta(y) = f_\xi\left(\varphi^{-1}(y)\right) \cdot \left(\varphi^{-1}(y)\right)'$.

2. $\eta = \varphi(\xi)$, де φ — монотонно спадна.



$F_\eta(y) = \mathbb{P}\{\eta < y\} = \mathbb{P}\{\varphi(\xi) < y\}$.
Нехай $y = \varphi(x)$, тоді з монотонності φ маємо рівність подій $\{\varphi(\xi) < y\} = \{\xi > x\}$. Тому $F_\eta(y) = 1 - F_\xi(x) = 1 - \int\limits_{-\infty}^{x} f_\xi(t)dt$. φ має обернену, $x = \varphi^{-1}(y)$.

Отже, $F_\eta(y) = 1 - \int\limits_{-\infty}^{\varphi^{-1}(y)} f_\xi(t)dt$, $f_\eta(y) = F'_\eta(y) = -f_\xi\left(\varphi^{-1}(y)\right) \cdot \left(\varphi^{-1}(y)\right)'$.
Остаточно, для монотонної неперервної функції φ має місце формула:

$$f_\eta(y) = f_\xi\left(\varphi^{-1}(y)\right) \cdot \left|\left(\varphi^{-1}(y)\right)'\right|,\ \eta = \varphi(\xi)$$

При цьому обов'язково треба зазначити допустимі значення y.

Зауваження. Взагалі кажучи, на функцію φ, окрім неперервності, треба накладати й умову диференційовності φ^{-1} майже скрізь.

Приклад. $\xi \sim \mathrm{U}(-\frac{\pi}{2}; \frac{\pi}{2})$. Знайти розподіл $\eta = \sin\xi$.
$f_\xi(x) = \begin{cases} \frac{1}{\pi}, & |x| < \frac{\pi}{2} \\ 0, & |x| \geq \frac{\pi}{2} \end{cases}$, $\varphi(x) = \sin x$, $\varphi^{-1}(y) = \arcsin(y)$, $\left(\varphi^{-1}(y)\right)' = \frac{1}{\sqrt{1-y^2}}$, η може приймати значення від -1 до 1. Отже, $f_\eta(y) = \begin{cases} \frac{1}{\pi\sqrt{1-y^2}}, & |y| < 1 \\ 0, & |y| \geq 1 \end{cases}$ — це «закон арксинуса».

Приклад. Нехай ξ — довільна НВВ, $F_\xi(x)$ — її функція розподілу. Знайти закон розподілу $\eta = F_\xi(\xi)$.

$F_\eta(y) = \mathbb{P}\{\eta < y\} = \mathbb{P}\{F_\xi(\xi) < y\} = \mathbb{P}\{\xi < F_\xi^{-1}(y)\} = F_\xi(F_\xi^{-1}(y)) = \begin{cases} 0, & y \leq 0 \\ y, & 0 < y \leq 1 \\ 1, & y > 1 \end{cases}$.
Отже, $F_\xi(\xi) \sim U\left<0; 1\right>$. Ще раз зауважимо, що результат має місце для довільної НВВ ξ.

Зауваження. Якщо φ не є неперервною, то $\eta = \varphi(\xi)$ може не бути НВВ. Наведемо декілька прикладів.

1. Нехай $\xi \sim \mathrm{Exp}(1)$. Знайдемо розподіл $\eta = [\xi]$, де $[\cdot]$ — ціла частина. Ціла частина приймає значення $0, 1, 2, ...$, знайдемо відповідні ймовірності.
   $\mathbb{P}\{[\xi] = n\} = \mathbb{P}\{\xi \in [n; n+1)\} = \int\limits_{n}^{n+1} e^{-x}dx = e^{-n} - e^{-(n+1)} = e^{-n}(1 - e^{-1})$, $n = 0, 1, 2, ...$
   Отже, $\eta = [\xi] \sim \mathrm{Geom}(1 - e^{-1})$.
   За цим прикладом можна зробити висновок: якщо φ — кусково стала, а ξ — НВВ, то $\varphi(\xi)$ — ДВВ.
2. Нехай $\xi \sim \mathrm{Exp}(1)$. Знайдемо розподіл $\eta = \{\xi\} = \xi - [\xi]$. В цьому випадку φ є кусково неперервною функцією та приймає значення з інтервалу $[0; 1)$.
   Для $0 < y < 1$ $F_\eta(y) = \mathbb{P}\{\eta < y\} = \mathbb{P}\{\eta \in [0; y)\} = \mathbb{P}\{\xi \in [n; n+y), n = 0, 1, 2, ...\}$.
   Для фіксованого n $\mathbb{P}\{\xi \in [n; n+y)\} = \int\limits_{n}^{n+y} e^{-x}dx = e^{-n}(1 - e^{-y})$. Отже, $\mathbb{P}\{\eta \in [0; y)\} = \sum\limits_{n=0}^{\infty} \mathbb{P}\{\xi \in [n; n+y)\} = (1 - e^{-y}) \sum\limits_{n=0}^{\infty} e^{-n} = \frac{1-e^{-y}}{1-e^{-1}}$. Отримали функцію розподілу η: $F_\eta(y) = \begin{cases} 0, & y \leq 0 \\ \frac{1-e^{-y}}{1-e^{-1}}, & 0 < y \leq 1 \\ 1, & y > 1 \end{cases}$. Таким чином, η — НВВ.

5.1.3 Розподіл немонотонної функції від неперервного випадкового аргументу

Нехай ξ — НВВ, $f_\xi(x)$ — її щільність розподілу, а φ — немонотонна числова функція. Тоді область можливих значень ξ можна розбити на проміжки, на яких φ буде монотонною. Тоді, скориставшись результатом для монотонної функції, отримаємо формулу для щільності $\eta = \varphi(\xi)$:

$$f_\eta(y) = \sum_{k=1}^{m} f_\xi\left(\varphi_k^{-1}(y)\right) \cdot \left|\left(\varphi_k^{-1}(y)\right)'\right|$$

де m — кількість проміжків монотонності, а φ_k^{-1} — відповідні обернені функції.

Приклад. 1. $\xi \sim \mathrm{U}(-\frac{\pi}{2}, \frac{\pi}{2})$, знайти розподіл $\eta = \cos\xi$.
На проміжках $(-\frac{\pi}{2}; 0]$ та $[0; \frac{\pi}{2})$ $\cos x$ є монотонною функцією, відповідні обернені — $\varphi_1^{-1}(y) = -\arccos y$, $\varphi_2^{-1}(y) = \arccos y$.
Тоді $f_\eta(y) = \frac{1}{\pi} \cdot \left|-\frac{1}{\sqrt{1-y^2}}\right| + \frac{1}{\pi} \cdot \left|\frac{1}{\sqrt{1-y^2}}\right| = \begin{cases} \frac{2}{\pi}\frac{1}{\sqrt{1-y^2}}, & y \in [0;1) \\ 0, & y \notin [0;1) \end{cases}$.

2. $\xi \sim \mathrm{N}(a, \sigma)$, знайти розподіл $\eta = \xi^2$.
На проміжках $(-\infty; 0]$ та $[0; +\infty)$ x^2 є монотонною функцією, відповідні обернені — $\varphi_1^{-1}(y) = -\sqrt{y}$, $\varphi_2^{-1}(y) = \sqrt{y}$. Щільність розподілу ξ — $f_\xi(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-a)^2}{2\sigma^2}}$.
Тоді $f_\eta(y) = \begin{cases} \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(-\sqrt{y}-a)^2}{2\sigma^2}} \cdot \frac{1}{2\sqrt{y}} + \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(\sqrt{y}-a)^2}{2\sigma^2}} \cdot \frac{1}{2\sqrt{y}}, & y > 0 \\ 0, & y \le 0 \end{cases}$.
При $\xi \sim \mathrm{N}(0,1)$ $f_\eta(y) = \begin{cases} \frac{1}{\sqrt{2\pi y}} e^{-y/2}, y > 0 \\ 0, y \le 0 \end{cases}$, що означає $\xi^2 \sim \Gamma(\frac{1}{2}, \frac{1}{2})$.

5.1.4 Числові характеристики функції неперервного випадкового аргументу

Твердження. *$\mathbb{E}(\varphi(\xi))^k = \int\limits_{-\infty}^{+\infty} \varphi^k(x) f_\xi(x) dx$ для всіх $k \in \mathbb{N}$.*

Доведення. Достатньо довести для монотонно зростаючої φ.
$\eta = \varphi(\xi)$, $\mathbb{E}\eta^k = \int\limits_{-\infty}^{+\infty} y^k f_\eta(y) dy = \int\limits_{-\infty}^{+\infty} y^k f_\xi(\varphi^{-1}(y))(\varphi^{-1}(y))' dy = \int\limits_{-\infty}^{+\infty} y^k f_\xi(\varphi^{-1}(y)) d\varphi^{-1}(y) = \left[\varphi^{-1}(y) = x, y = \varphi(x)\right] = \int\limits_{-\infty}^{+\infty} \varphi^k(x) f_\xi(x) dx$. У випадку монотонно спадної φ під інтегралом отримаємо $-(\varphi^{-1}(y))'$, а інтеграл після заміни змінної буде від $+\infty$ до $-\infty$. У випадку немонотонної φ треба буде скористатися адитивністю інтеграла. ▲

Зауваження. Для знаходження числових характеристик функції неперервного випадкового аргументу розподіл самої функції знаходити не потрібно.

Вправа. Дослідити, за яких умов на φ ця формула має місце.

5.2 Функції кількох випадкових аргументів

5.2.1 Випадок довільної функції

Нехай $\varphi : \mathbb{R}^n \to \mathbb{R}$ — задана числова функція.

Якщо $\vec{\xi} = (\xi_1, ..., \xi_n)$ — дискретний випадковий вектор, тоді $\eta = \varphi(\vec{\xi})$ — ДВВ. Побудову закону розподілу η доцільно розглянути на прикладі.

Приклад. $\vec{\xi} = (\xi_1, \xi_2)$ задано таблицею розподілу:

ξ_2 \ ξ_1	0	1	2
-1	0.1	0.2	0.3
1	0.2	0.1	0.1

Знайти закони розподілу $\eta_1 = \xi_1\xi_2$ та $\eta_2 = \xi_1 + \xi_2$. Для цього треба визначити значення, які приймають ці величини, та обчислити відповідні ймовірності.

η_1 приймає значення -2 (коли $\xi_1 = 2$, $\xi_2 = -1$), -1 (коли $\xi_1 = 1$, $\xi_2 = -1$), 0 (коли $\xi_1 = 0$, $\xi_2 = -1$ або $\xi_2 = 1$), 1 (коли $\xi_1 = 1$, $\xi_2 = 1$), 2 (коли $\xi_1 = 2$, $\xi_2 = 1$).

η_2 приймає значення -1 (коли $\xi_1 = 0$, $\xi_2 = -1$), 0 (коли $\xi_1 = 1$, $\xi_2 = -1$), 1 (коли $\xi_1 = 0$, $\xi_2 = 1$ або $\xi_1 = 2$, $\xi_2 = -1$), 2 (коли $\xi_1 = 1$, $\xi_2 = 1$), 3 (коли $\xi_1 = 2$, $\xi_2 = 1$). Відповідні сумісні ймовірності отримуємо з таблиці розподілу $\vec{\xi}$.

η_1	-2	-1	0	1	2
p	0.3	0.2	0.3	0.1	0.1

η_2	-1	0	1	2	3
p	0.1	0.2	0.5	0.1	0.1

Якщо $\vec{\xi} = (\xi_1, ..., \xi_n)$ — неперервний випадковий вектор із щільністю $f_{\vec{\xi}}(\vec{x})$, то можна знайти функцію розподілу $\eta = \varphi(\vec{\xi})$.

$$F_\eta(y) = \mathbb{P}\{\eta < y\} = \mathbb{P}\left\{\vec{\xi} \in D_y\right\} = \underset{D_y}{\int ... \int} f_{\vec{\xi}}(\vec{x})d\vec{x}, \text{ де } D_y = \{\vec{x} \in \mathbb{R}^n : \varphi(\vec{x}) < y\}$$

Розглянемо тепер взаємно однозначне гладке перетворення $\vec{\psi} : \mathbb{R}^n \to \mathbb{R}^n$ та знайдемо щільність розподілу $\vec{\eta} = \vec{\psi}(\vec{\xi})$. Для множини $D \subset \mathbb{R}^n$ $\mathbb{P}\left\{\vec{\psi}(\vec{\xi}) \in D\right\} = \mathbb{P}\left\{\vec{\xi} \in \vec{\psi}^{-1}(D)\right\} = \int_{\vec{\psi}^{-1}(D)} f_{\vec{\xi}}(\vec{x})d\vec{x} = \left[\text{заміна } \vec{y} = \vec{\psi}(\vec{x})\right] = \int_D f_{\vec{\xi}}(\vec{\psi}^{-1}(\vec{y})) \left|\mathcal{J}^{-1}(\vec{\psi}^{-1}(\vec{y})\right| d\vec{y}$, де $\mathcal{J}(\vec{x})$ — якобіан $\vec{\psi}$. Отже,

$$f_{\vec{\eta}}(\vec{y}) = f_{\vec{\xi}}(\vec{\psi}^{-1}(\vec{y})) \left|\mathcal{J}^{-1}(\vec{\psi}^{-1}(\vec{y})\right|$$

Приклад. Нехай A — невироджена матриця розміру $n \times n$, $\vec{b} \in \mathbb{R}^n$ — деякий вектор, $\vec{\xi}$ — неперервний випадковий вектор. Знайти щільність розподілу $\vec{\eta} = A\vec{\xi} + \vec{b}$.

Тут $\vec{y} = \vec{\psi}(\vec{x}) = A\vec{x} + \vec{b}$, тому $\vec{\psi}^{-1}(\vec{y}) = A^{-1}(\vec{y} - \vec{b})$. Якобіан $\vec{\psi}$ рівний $|\det A|$. Отже, $f_{\vec{\eta}}(\vec{y}) = \frac{1}{|\det A|} f_{\vec{\xi}}\left(A^{-1}(\vec{y} - \vec{b})\right)$

5.2.2 Закон розподілу мінімуму та максимуму

Нехай $\xi_1, ..., \xi_n$ незалежні та розподілені однаково, як деяка випадкова величина ξ. Знайдемо розподіл їх мінімуму та максимуму.

1. $\mu_1 = \min\{\xi_1, \xi_2, ..., \xi_n\}$.
   $F_{\mu_1}(x) = \mathbb{P}\{\min\{\xi_1, \xi_2, ..., \xi_n\} < x\} = 1 - \mathbb{P}\{\min\{\xi_1, \xi_2, ..., \xi_n\} \geq x\} =$
   $= 1 - \mathbb{P}\{\xi_1 \geq x, \xi_2 \geq x, ..., \xi_n \geq x\} = 1 - \mathbb{P}\{\xi_1 \geq x\} \cdot \mathbb{P}\{\xi_2 \geq x\} \cdot ... \cdot \mathbb{P}\{\xi_n \geq x\} =$
   $= 1 - (\mathbb{P}\{\xi \geq x\})^n = 1 - (1 - F_\xi(x))^n$.
   Якщо $\xi_1, ..., \xi_n$ — неперервні, то $f_{\mu_1}(x) = (F_{\mu_1}(x))' = n(1 - F_\xi(x))^{n-1} f_\xi(x)$.
2. $\mu_2 = \max\{\xi_1, \xi_2, ..., \xi_n\}$.
   $F_{\mu_2}(x) = \mathbb{P}\{\max\{\xi_1, \xi_2, ..., \xi_n\} < x\} = \mathbb{P}\{\xi_1 < x, \xi_2 < x, ..., \xi_n < x\} =$
   $= \mathbb{P}\{\xi_1 < x\} \cdot \mathbb{P}\{\xi_2 < x\} \cdot ... \cdot \mathbb{P}\{\xi_n < x\} = (\mathbb{P}\{\xi < x\})^n = (F_\xi(x))^n$.
   Якщо $\xi_1, ..., \xi_n$ — неперервні, то $f_{\mu_2}(x) = (F_{\mu_2}(x))' = n(F_\xi(x))^{n-1} f_\xi(x)$.

Приклад. Нехай $\xi_1, ..., \xi_n$ незалежні та мають розподіл $\text{Exp}(\lambda)$. Знайти розподіл їх мінімуму.
$f_{\min}(x) = n(1 - (1 - e^{-\lambda x}))^{n-1} \lambda e^{-\lambda x} = n\lambda e^{-n\lambda x}$ при $x \geq 0$ та 0 при $x < 0$.
Отже, $\min\{\xi_1, ..., \xi_n\} \sim \text{Exp}(n\lambda)$.

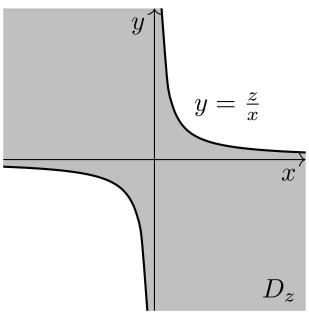
5.2.3 Закон розподілу добутку двох НВВ

Нехай $\vec{\xi} = (\xi_1, \xi_2)$ — неперервний випадковий вектор, щільність $f_{\vec{\xi}}(x, y)$ відома.
Задача: знайти розподіл $\eta = \xi_1 \xi_2$.

Розпишемо функцію розподілу за вже відомою схемою:
$F_\eta(z) = \iint\limits_{D_z} f_{\vec{\xi}}(x, y)dxdy$, $D_z = \left\{(x, y) \in \mathbb{R}^2 : xy < z\right\}$

Розглянемо два варіанти:

1.

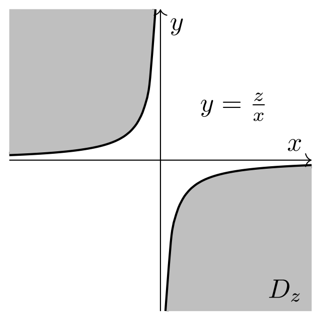


$z > 0$, тоді $D_z = \left\{x < 0, y > \frac{z}{x}\right\} \cup \left\{x > 0, y < \frac{z}{x}\right\}$.

$F_\eta(z) = \int\limits_{-\infty}^{0} dx \int\limits_{\frac{z}{x}}^{+\infty} f_{\vec{\xi}}(x, y)dy + \int\limits_{0}^{+\infty} dx \int\limits_{-\infty}^{\frac{z}{x}} f_{\vec{\xi}}(x, y)dy.$

Продиференціюємо обидві частини по z:

$f_\eta(z) = -\int\limits_{-\infty}^{0} \frac{1}{x} f_{\vec{\xi}}(x, \frac{z}{x})dx + \int\limits_{0}^{+\infty} \frac{1}{x} f_{\vec{\xi}}(x, \frac{z}{x})dx.$

2. 

$z < 0$, тоді $D_z = \left\{x < 0, y > \frac{z}{x}\right\} \cup \left\{x > 0, y < \frac{z}{x}\right\}$.

$F_\eta(z) = \int\limits_{-\infty}^{0} dx \int\limits_{\frac{z}{x}}^{+\infty} f_{\vec{\xi}}(x,y)dy + \int\limits_{0}^{+\infty} dx \int\limits_{-\infty}^{\frac{z}{x}} f_{\vec{\xi}}(x,y)dy$

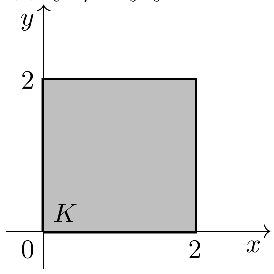
Продиференціюємо обидві частини по z:

$f_\eta(z) = -\int\limits_{-\infty}^{0} \frac{1}{x} f_{\vec{\xi}}(x, \frac{z}{x})dx + \int\limits_{0}^{+\infty} \frac{1}{x} f_{\vec{\xi}}(x, \frac{z}{x})dx$

Отже, $f_{\xi_1\xi_2}(z) = -\int\limits_{-\infty}^{0} \frac{1}{x} f_{\vec{\xi}}(x, \frac{z}{x})dx + \int\limits_{0}^{+\infty} \frac{1}{x} f_{\vec{\xi}}(x, \frac{z}{x})dx$.

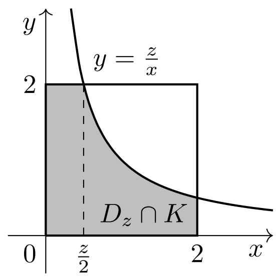
Зауваження. Якщо ξ_1 та ξ_2 незалежні, то $f_{\vec{\xi}}(x, \frac{z}{x}) = f_{\xi_1}(x) f_{\xi_2}(\frac{z}{x})$.

Приклад. Нехай $\vec{\xi} = (\xi_1, \xi_2)$ рівномірно розподілений в квадраті $[0;2] \times [0;2]$. Знайти закон розподілу $\eta = \xi_1\xi_2$.



$$f_{\vec{\xi}}(x,y) = \begin{cases} \frac{1}{4}, & (x,y) \in K \\ 0, & (x,y) \notin K \end{cases}$$

$F_\eta(z) = \mathbb{P}\{\xi_1\xi_2 < z\} = \iint\limits_{\{xy<z\}\cap K} f_{\vec{\xi}}(x,y)dxdy$. При $z < 0$ $F_\eta(z) = 0$, а при $z \geq 4$ $F_\eta(z) = 1$.



$$\begin{aligned}
F_\eta(z) &= \frac{1}{4} S_{D_z \cap K} = \frac{1}{4}\left(4 - \int\limits_{\frac{z}{2}}^{2} dx \int\limits_{\frac{z}{x}}^{2} dy\right) = \\
&= \frac{1}{4}\left(4 - \int\limits_{\frac{z}{2}}^{2} \left(2 - \frac{z}{x}\right) dx\right) \\
&= 1 - \frac{1}{4}\left(2x - z\ln x\right)\Big|_{\frac{z}{2}}^{2} = \\
&= 1 - \frac{1}{4}(4 - z\ln 2 - z + z\ln\frac{z}{2}) = \\
&= \frac{1}{4}(z + z\ln 2 - z\ln z + z\ln 2) = \\
&= \frac{1}{4}(z + 2z\ln 2 - z\ln z) \text{ при } z \in [0;4).
\end{aligned}$$

Отже, $f_\eta(z) = \begin{cases} 0, & z \notin (0,4] \\ \frac{1}{4}(2\ln 2 - \ln z), & z \in (0,4] \end{cases}$.

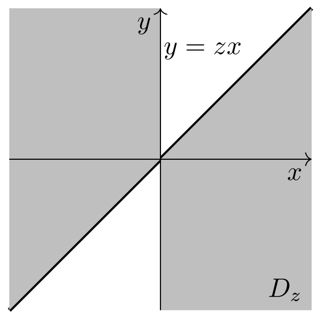
5.2.4 Закон розподілу частки двох НВВ

Нехай $\vec{\xi} = (\xi_1, \xi_2)$ — неперервний випадковий вектор, щільність $f_{\vec{\xi}}(x,y)$ відома.

Задача: знайти розподіл $\eta = \frac{\xi_2}{\xi_1}$.

Розпишемо функцію розподілу за вже відомою схемою:

$F_\eta(z) = \iint\limits_{D_z} f_{\vec{\xi}}(x,y)dxdy$, $D_z = \left\{(x,y) \in \mathbb{R}^2 : \frac{y}{x} < z\right\}$

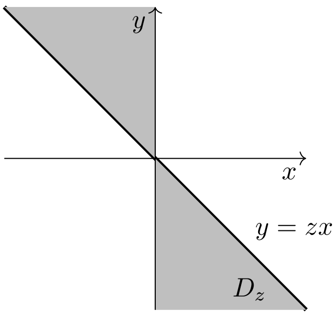
1. 

$z > 0$, тоді $D_z = \{x < 0, y > zx\} \cup \{x > 0, y < zx\}$

$F_\eta(z) = \int\limits_{-\infty}^{0} dx \int\limits_{zx}^{+\infty} f_{\vec{\xi}}(x,y)dy + \int\limits_{0}^{+\infty} dx \int\limits_{-\infty}^{zx} f_{\vec{\xi}}(x,y)dy$

Продиференціюємо обидві частини по z:

$f_\eta(z) = -\int\limits_{-\infty}^{0} x f_{\vec{\xi}}(x, zx)dx + \int\limits_{0}^{+\infty} x f_{\vec{\xi}}(x, zx)dx$

2. 

$z < 0$, тоді $D_z = \{x < 0, y > zx\} \cup \{x > 0, y < zx\}$
$F_\eta(z) = \int\limits_{-\infty}^{0} dx \int\limits_{zx}^{+\infty} f_{\vec{\xi}}(x,y)dy + \int\limits_{0}^{+\infty} dx \int\limits_{-\infty}^{zx} f_{\vec{\xi}}(x,y)dy$
Продиференціюємо обидві частини по z:
$f_\eta(z) = -\int\limits_{-\infty}^{0} x f_{\vec{\xi}}(x,zx)dx + \int\limits_{0}^{+\infty} x f_{\vec{\xi}}(x,zx)dx$

Отже, $f_{\frac{\xi_2}{\xi_1}}(z) = -\int\limits_{-\infty}^{0} x f_{\vec{\xi}}(x,zx)dx + \int\limits_{0}^{+\infty} x f_{\vec{\xi}}(x,zx)dx$.

Зауваження. Якщо ξ_1 та ξ_2 незалежні, то $f_{\vec{\xi}}(x,zx) = f_{\xi_1}(x) f_{\xi_2}(zx)$.

Вправа. Перевірити, що щільність розподілу $\eta = \frac{\xi_1}{\xi_2}$ має вигляд
$f_\eta(z) = -\int\limits_{-\infty}^{0} y f_{\vec{\xi}}(zy,y)dy + \int\limits_{0}^{+\infty} y f_{\vec{\xi}}(zy,y)dy$.

Приклад. ξ_1 та ξ_2 — незалежні випадкові величини, що мають розподіл $\mathrm{N}(0,\sigma)$.
Знайти закон розподілу $\eta = \frac{\xi_2}{\xi_1}$.
Скористаємося знайденою формулою, враховуючи незалежність:

$$f_\eta(z) = -\int\limits_{-\infty}^{0} x f_{\xi_1}(x) f_{\xi_2}(zx)dx + \int\limits_{0}^{+\infty} x f_{\xi_1}(x) f_{\xi_2}(zx)dx =$$

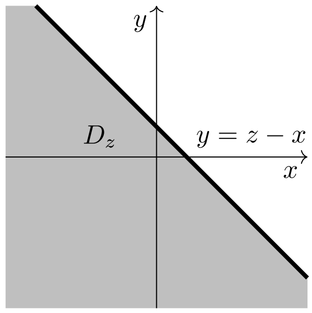
$$= -\frac{1}{2\pi\sigma^2}\int\limits_{-\infty}^{0} x e^{-\frac{x^2}{2\sigma^2}(1+z^2)}dx + \frac{1}{2\pi\sigma^2}\int\limits_{0}^{+\infty} x e^{-\frac{x^2}{2\sigma^2}(1+z^2)}dx =$$

$$= \left[\begin{array}{c} \frac{x^2}{2\sigma^2}(1+z^2) = t \\ \frac{x}{\sigma^2}(1+z^2)dx = dt \end{array}\right] = \frac{-\sigma^2}{2\pi\sigma^2(1+z^2)}\int\limits_{+\infty}^{0} e^{-t}dt + \frac{\sigma^2}{2\pi\sigma^2(1+z^2)}\int\limits_{0}^{+\infty} e^{-t}dt =$$

$$= \frac{1}{\pi(1+z^2)}\int\limits_{0}^{+\infty} e^{-t}dt = \frac{1}{\pi(1+z^2)}, z \in \mathbb{R} \text{ — це щільність розподілу Коші.}$$

5.2.5 Закон розподілу суми двох НВВ

Нехай $\vec{\xi} = (\xi_1, \xi_2)$ — неперервний випадковий вектор, щільність $f_{\vec{\xi}}(x,y)$ відома.
Задача: знайти розподіл $\eta = \xi_1 + \xi_2$.
Розпишемо функцію розподілу за вже відомою схемою:
$F_\eta(z) = \iint\limits_{D_z} f_{\vec{\xi}}(x,y)dxdy$, $D_z = \left\{(x,y) \in \mathbb{R}^2 : x + y < z\right\}$



$F_\eta(z) = \int\limits_{-\infty}^{+\infty} dx \int\limits_{-\infty}^{z-x} f_{\vec{\xi}}(x,y)dy$
Продиференціюємо обидві частини по z:
$f_\eta(z) = \int\limits_{-\infty}^{+\infty} f_{\vec{\xi}}(x, z-x)dx$

Зауваження. Якщо ξ_1 та ξ_2 незалежні, то $f_\eta(z) = \int\limits_{-\infty}^{+\infty} f_{\xi_1}(x) f_{\xi_2}(z-x)dx = (f_{\xi_1} * f_{\xi_2})(z)$. Тут $*$ позначає операцію згортки.

Вправа. Перевірити, що щільність розподілу $\eta_1 = \xi_1 - \xi_2$ має вигляд $f_{\eta_1}(z) = \int\limits_{-\infty}^{+\infty} f_{\vec{\xi}}(x, x-z)dx$, а $\eta_2 = \xi_2 - \xi_1$ — $f_{\eta_2}(z) = \int\limits_{-\infty}^{+\infty} f_{\vec{\xi}}(x, x+z)dx$.

Приклад. ξ_1 та ξ_2 — незалежні випадкові величини, $\xi_i \sim \mathrm{Exp}(\lambda_i)$, $i = 1, 2$, $\lambda_1 \neq \lambda_2$.
Знайти закон розподілу $\eta = \xi_1 + \xi_2$.
Скористаємося знайденою формулою, враховуючи незалежність:

$$f_\eta(z) = [z - x > 0 \Rightarrow x < z] = \int\limits_0^z \lambda_1 e^{-\lambda_1 x} \lambda_2 e^{-\lambda_2(z-x)} dx = \lambda_1\lambda_2 e^{-\lambda_2 z} \int\limits_0^z e^{-x(\lambda_1-\lambda_2)} dx =$$

$$= \lambda_1\lambda_2 e^{-\lambda_2 z} \left(\frac{1}{-(\lambda_1 - \lambda_2)} e^{-x(\lambda_1-\lambda_2)} \right)\Bigg|_0^z = \frac{\lambda_1\lambda_2}{\lambda_2 - \lambda_1} e^{-\lambda_2 z}(e^{-z(\lambda_1-\lambda_2)} - 1) =$$

$$= \frac{\lambda_1\lambda_2}{\lambda_2 - \lambda_1}(e^{-\lambda_1 z} - e^{-\lambda_2 z}),\ z \geq 0 \text{ та } 0 \text{ інакше.}$$

Це щільність закону Ерланга 2-го порядку. При $\lambda_1 = \lambda_2 = \lambda$ отримаємо $f_\eta(z) = \lambda^2 z e^{-\lambda z}$ при $z \geq 0$ та 0 інакше. Це вже буде щільність гамма-розподілу $\Gamma(2, \lambda)$.

5.2.6 Числові характеристики функції багатьох випадкових аргументів

Розглядаємо неперервний випадковий вектор $\vec{\xi} = (\xi_1, \xi_2, ..., \xi_n)$ та випадкову величину $\eta = \varphi(\vec{\xi})$.

Вправа. Довести, що для НВВ η, що приймає невід'ємні значення, $\mathbb{E}\eta = \int\limits_0^{+\infty} (1 - F_\eta(x))dx$.

Твердження. $\mathbb{E}\eta^k = \int_{\mathbb{R}^n} \varphi^k(\vec{x}) f_{\vec{\xi}}(\vec{x}) d\vec{x}$ *для всіх* $k \in \mathbb{N}$.

Доведення. Доведемо у випадку $n = 2$, та $\varphi(x, y) \geq 0$.

$$\mathbb{E}\eta^k = \int\limits_0^{+\infty} \left(1 - F_{\eta^k}(z)\right) dz = \int\limits_0^{+\infty} \left(\iint_{\varphi^k(x,y) \geq z} f_{\vec{\xi}}(x, y) dxdy \right) dz =$$

$$= \iint\limits_{\mathbb{R}^2} \left(\int\limits_0^{\varphi^k(x,y)} dz \right) f_{\vec{\xi}}(x, y) dxdy = \iint\limits_{\mathbb{R}^2} \varphi^k(x, y) f_{\vec{\xi}}(x, y) dxdy$$

▲

Вправа. Довести твердження в більш загальному вигляді.

Приклад. ξ_1, ξ_2 — НВВ, знайдемо $\mathbb{E}(\xi_1 + \xi_2)$ та $\mathbb{E}\xi_1\xi_2$ у випадку незалежності цих НВВ.

$$E(\xi_1 + \xi_2) = \iint\limits_{\mathbb{R}^2} (x + y) f_{\vec{\xi}}(x, y) dxdy = \iint\limits_{\mathbb{R}^2} x f_{\vec{\xi}}(x, y) dxdy + \iint\limits_{\mathbb{R}^2} y f_{\vec{\xi}}(x, y) dxdy = \mathbb{E}\xi_1 + \mathbb{E}\xi_2$$

Якщо ξ_1, ξ_2 незалежні, то $f_{\vec{\xi}}(x, y) = f_{\xi_1}(x) f_{\xi_2}(y)$.

$$\mathbb{E}\xi_1\xi_2 = \iint\limits_{\mathbb{R}^2} xy f_{\xi_1}(x) f_{\xi_2}(y) dxdy = \left(\int\limits_{\mathbb{R}} x f_{\xi_1}(x) dx \right) \left(\int\limits_{\mathbb{R}} y f_{\xi_2}(y) dy \right) = \mathbb{E}\xi_1 E\xi_2$$

5.3 Деякі нерівності

5.3.1 Нерівність Єнсена

Твердження. Нехай ξ — деяка випадкова величина з $\mathbb{E}|\xi| < \infty$, а $\varphi(x)$ — опукла функція. Тоді $\varphi(\mathbb{E}\xi) \le \mathbb{E}\varphi(\xi)$.

Доведення. Оскільки $\varphi(x)$ — опукла, то $\forall\ y \in \mathbb{R}\ \exists\ C(y) : \forall\ x \in \mathbb{R} : \varphi(x) - \varphi(y) \ge c(y)(x-y)$. Покладемо $y = \mathbb{E}\xi$, $x = \xi$ (оскільки ξ приймає дійсні значення), і, взявши з обох сторін нерівності математичне сподівання, отримаємо $\mathbb{E}\varphi(\xi) \ge \varphi(\mathbb{E}\xi)$. ▲

Зауваження. Для увігнутої $\varphi(x)$ нерівність виконується в іншу сторону: $\mathbb{E}\varphi(\xi) \le \varphi(\mathbb{E}\xi)$.

Приклад. Нехай $0 < s < t$. Розглянемо $\varphi(x) = |x|^{\frac{t}{s}}$, яка є опуклою. Скористаємося нерівністю Єнсена для $\eta = |\xi|^s$: $\varphi(\mathbb{E}\eta) \le \mathbb{E}\varphi(\eta) \Leftrightarrow (\mathbb{E}\,|\xi|^s)^{\frac{t}{s}} \le \mathbb{E}\,|\xi|^t \Leftrightarrow (\mathbb{E}\,|\xi|^s)^{\frac{1}{s}} \le (\mathbb{E}\,|\xi|^t)^{\frac{1}{t}}$.

Маємо важливий **наслідок**: якщо у випадкової величини ξ існує скінченний абсолютний момент $\mathbb{E}\,|\xi|^m$ $(m \in \mathbb{N})$, то

$$\mathbb{E}\,|\xi| \le (\mathbb{E}\,|\xi|^2)^{\frac{1}{2}} \le (\mathbb{E}\,|\xi|^3)^{\frac{1}{3}} \le ... \le (\mathbb{E}\,|\xi|^m)^{\frac{1}{m}}$$

Тобто, існування скінченного моменту $\mathbb{E}\,|\xi|^m$ гарантує існування як початкових, так і центральних моментів нижчих порядків (оскільки $|\mathbb{E}\xi|^k \le \mathbb{E}\,|\xi|^k$).

5.3.2 Нерівність Гельдера

Твердження. Нехай $1 < p < \infty$, $\frac{1}{p} + \frac{1}{q} = 1$. Якщо для випадкових величин ξ та η $\mathbb{E}\,|\xi|^p$ та $\mathbb{E}\,|\eta|^q$ скінченні, то $\mathbb{E}\,|\xi\eta|$ теж скінченне, причому $\mathbb{E}\,|\xi\eta| \le (\mathbb{E}\,|\xi|^p)^{\frac{1}{p}} \cdot (\mathbb{E}\,|\eta|^q)^{\frac{1}{q}}$.

Доведення. Якщо $\mathbb{E}\,|\xi|^p = 0$ або $\mathbb{E}\,|\eta|^q = 0$, то нерівність, очевидно, виконується. Нехай $\mathbb{E}\,|\xi|^p > 0$ та $\mathbb{E}\,|\eta|^q > 0$. Позначимо $\xi_0 = \frac{|\xi|}{(\mathbb{E}|\xi|^p)^{\frac{1}{p}}}$, $\eta_0 = \frac{|\eta|}{(\mathbb{E}|\eta|^q)^{\frac{1}{q}}}$, причому $\mathbb{E}\xi_0^p = \mathbb{E}\eta_0^q = 1$. З опуклості функції $f(x) = -\ln x$ маємо $xy = \exp\{\ln xy\} = \exp\{\frac{\ln x^p}{p} + \frac{\ln y^q}{q}\} \le \exp\{\ln\left(\frac{x^p}{p} + \frac{y^q}{q}\right)\} = \frac{x^p}{p} + \frac{y^q}{q}$, тому $\mathbb{E}\,(\xi_0\eta_0) \le \frac{1}{p}\mathbb{E}\xi_0^p + \frac{1}{q}\mathbb{E}\eta_0^q = \frac{1}{p} + \frac{1}{q} = 1$, звідки $\mathbb{E}\,|\xi\eta| \le (\mathbb{E}\,|\xi|^p)^{\frac{1}{p}} \cdot (\mathbb{E}\,|\eta|^q)^{\frac{1}{q}}$. ▲

Зауваження. При $p = q = 2$ отримуємо вже знайому нерівність Коші-Буняковського:
$|\mathbb{E}\xi\eta| \le \mathbb{E}\,|\xi\eta| \le \sqrt{\mathbb{E}\xi^2} \cdot \sqrt{\mathbb{E}\eta^2}$.

5.3.3 Нерівність Мінковського

Твердження. Нехай для випадкових величин ξ та η і $1 \le p < \infty$ маємо скінченні $\mathbb{E}\,|\xi|^p$ та $\mathbb{E}\,|\eta|^p$. Тоді $(\mathbb{E}\,|\xi+\eta|^p)^{\frac{1}{p}} \le (\mathbb{E}\,|\xi|^p)^{\frac{1}{p}} + (\mathbb{E}\,|\eta|^p)^{\frac{1}{p}}$.

Доведення. Для $p = 1$ нерівність є наслідком нерівності $|x+y| \le |x| + |y|$ для дійсних чисел. Нехай $p > 1$, тоді $|\xi+\eta|^p = |\xi+\eta| \cdot |\xi+\eta|^{p-1} \le |\xi| \cdot |\xi+\eta|^{p-1} + |\eta| \cdot |\xi+\eta|^{p-1}$. Покладемо $q = \frac{p}{p-1}$. Тоді $\frac{1}{p} + \frac{1}{q} = 1$, $\mathbb{E}\left(|\xi+\eta|^{p-1}\right)^q = \mathbb{E}|\xi+\eta|^p$ скінченне, бо для дійсних чисел виконується нерівність $|x+y|^p \le C(p) \cdot (|x|^p + |y|^p)$. Тоді з нерівності Гельдера: $\mathbb{E}\left(|\xi| \cdot |\xi+\eta|^{p-1}\right) \le (\mathbb{E}|\xi|^p)^{\frac{1}{p}} \cdot \left(\mathbb{E}|\xi+\eta|^{(p-1)q}\right)^{\frac{1}{q}} = (\mathbb{E}|\xi|^p)^{\frac{1}{p}} \cdot (\mathbb{E}|\xi+\eta|^p)^{\frac{1}{q}}$ і, аналогічно, $\mathbb{E}\left(|\eta| \cdot |\xi+\eta|^{p-1}\right) \le (\mathbb{E}|\eta|^p)^{\frac{1}{p}} \cdot (\mathbb{E}|\xi+\eta|^p)^{\frac{1}{q}}$. Отримуємо $\mathbb{E}\,|\xi+\eta|^p \le (\mathbb{E}|\xi+\eta|^p)^{\frac{1}{q}} \cdot \left((\mathbb{E}|\xi|^p)^{\frac{1}{p}} + (\mathbb{E}|\eta|^p)^{\frac{1}{p}}\right)$.
У випадку $\mathbb{E}\,|\xi+\eta|^p = 0$ виконання нерівності очевидне, а якщо $\mathbb{E}\,|\xi+\eta|^p > 0$, то, поділивши на $(\mathbb{E}|\xi+\eta|^p)^{\frac{1}{q}}$, отримаємо $(\mathbb{E}\,|\xi+\eta|^p)^{1-\frac{1}{q}} \le (\mathbb{E}|\xi|^p)^{\frac{1}{p}} + (\mathbb{E}|\eta|^p)^{\frac{1}{p}}$. Залишилося зауважити, що $1 - \frac{1}{q} = \frac{1}{p}$. ▲

Розділ 6

Основні розподіли математичної статистики

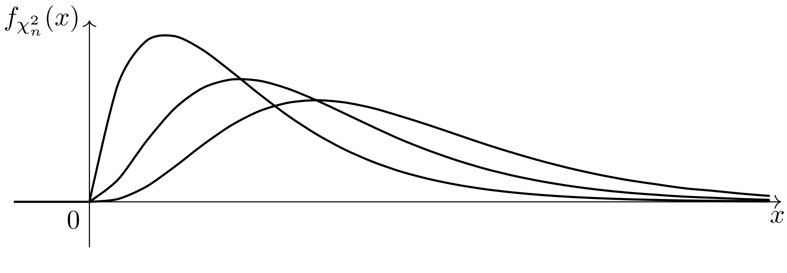
У цьому розділі буде наведено виведення законів розподілу, що застосовуються в задачах математичної статистики, та їх числових характеристик.

6.1 Розподіл χ^2 (Пірсона)

Означення: нехай $\xi_k \sim \mathrm{N}(a_k, \sigma_k)$, $k = 1, ..., n$ — незалежні у сукупності. Тоді $\xi = \sum\limits_{k=1}^{n} \left(\frac{\xi_k - a_k}{\sigma_k}\right)^2$ має розподіл χ^2 *(хі-квадрат, Пірсона) з n ступенями вільності.*
$\mathring{\xi}_k = \frac{\xi_k - a_k}{\sigma_k} \sim \mathrm{N}(0, 1)$, тому можна ще записати $\xi = \sum\limits_{k=1}^{n} (\mathring{\xi}_k)^2$.
Коротке позначення: $\xi \sim \chi^2_n$, $n \in \mathbb{N}$ — кількість ступенів вільності.
Щільність розподілу: відомо, що якщо $\eta \sim \mathrm{N}(0, 1)$, то $\eta^2 \sim \Gamma\left(\frac{1}{2}, \frac{1}{2}\right)$. Гамма-розподіл стійкий при $\beta_1 = \beta_2 = ... = \beta_n$, $\mathring{\xi}_k$ незалежні у сукупності, тому $\xi = \sum\limits_{k=1}^{n} (\mathring{\xi}_k)^2 \sim \Gamma\left(\frac{n}{2}, \frac{1}{2}\right) = \chi^2_n$.

$$f_{\chi^2_n}(x) = \begin{cases} \frac{1}{2^{\frac{n}{2}}\Gamma\left(\frac{n}{2}\right)} x^{\frac{n}{2}-1} e^{-\frac{x}{2}}, & x \geq 0 \\ 0, & x < 0 \end{cases}$$

Крива розподілу: графіки для різних значень n.



Числові характеристики:
1. $\mathbb{E}\chi^2_n = \frac{n/2}{1/2} = n$.
2. $\mathbb{D}\chi^2_n = \frac{n/2}{1/4} = 2n$.

6.2 Розподіл χ

Означення: нехай випадкова величина ξ має розподіл χ^2_n. Тоді $\eta = \sqrt{\xi}$ має розподіл χ *(хі) з n ступенями вільності.*
Коротке позначення: $\eta \sim \chi_n$, $n \in \mathbb{N}$ — кількість ступенів вільності.

Щільність розподілу: скористаємося формулою для визначення щільності розподілу функції від випадкової величини. $\eta = \sqrt{\xi}$, тому позначимо $\varphi(x) = \sqrt{x}$, $\varphi^{-1}(y) = y^2$, $\left(\varphi^{-1}(y)\right)' = 2y$.

$$f_{\chi_n}(y) = f_{\chi^2_n}\left(\varphi^{-1}(y)\right) \cdot \left|\left(\varphi^{-1}(y)\right)'\right| = f_{\chi^2_n}(y^2) \cdot 2y = \begin{cases} \frac{1}{2^{\frac{n}{2}-1}\Gamma\left(\frac{n}{2}\right)} y^{n-1} e^{-\frac{y^2}{2}}, & y \geq 0 \\ 0, & y < 0 \end{cases}$$

Числові характеристики:

1. $\mathbb{E}\chi_n = \sqrt{2} \cdot \frac{\Gamma\left(\frac{n+1}{2}\right)}{\Gamma\left(\frac{n}{2}\right)}$.
2. $\mathbb{D}\chi_n = n - \left(\mathbb{E}\chi_n\right)^2$.

Зауваження. Нескладно помітити, що χ_2 — це розподіл Релея.

Знайдемо ще розподіл $\zeta = \frac{\chi_n}{\sqrt{n}}$. $\varphi(y) = \frac{y}{\sqrt{n}}$, $\varphi^{-1}(z) = z\sqrt{n}$, $\left(\varphi^{-1}(z)\right)' = \sqrt{n}$.

$$f_{\frac{\chi_n}{\sqrt{n}}}(z) = f_{\chi_n}\left(\varphi^{-1}(z)\right) \cdot \left|\left(\varphi^{-1}(y)\right)'\right| = f_{\chi_n}(z\sqrt{n}) \cdot \sqrt{n} = \begin{cases} \frac{n^{\frac{n}{2}}}{2^{\frac{n}{2}-1}\Gamma\left(\frac{n}{2}\right)} z^{n-1} e^{-\frac{nz^2}{2}}, & z \geq 0 \\ 0, & z < 0 \end{cases}$$

Вправа. Записати числові характеристики випадкової величини, що має розподіл $\frac{\chi_n}{\sqrt{n}}$.

6.3 Розподіл Стьюдента (t-розподіл)

Означення: якщо $\xi \sim \mathrm{N}(0,1)$ та $\eta \sim \frac{\chi_n}{\sqrt{n}}$ незалежні, то $\zeta = \frac{\xi}{\eta} = \frac{\xi}{\chi_n/\sqrt{n}}$ має *розподіл Стьюдента з n ступенями вільності.*

Коротке позначення: $\zeta \sim \mathrm{St}_n$, $n \in \mathbb{N}$ — кількість ступенів вільності.

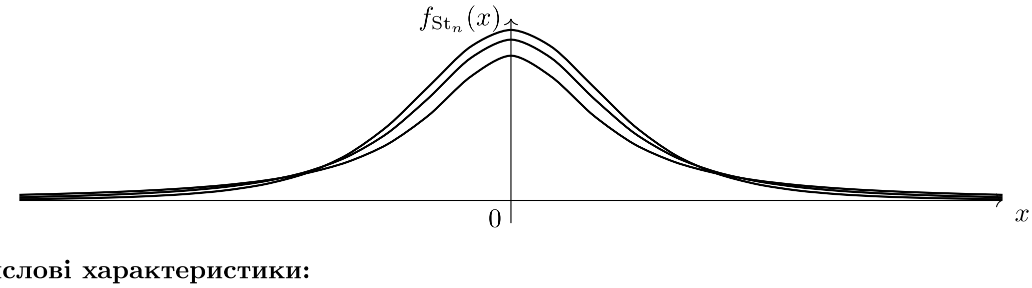
Щільність розподілу: скористаємося формулою для визначення щільності розподілу частки двох незалежних НВВ.

$$f_{\mathrm{St}_n}(z) = \int\limits_0^{+\infty} x f_\xi(zx) f_{\frac{\chi_n}{\sqrt{n}}}(x) dx = \frac{1}{\sqrt{2\pi}} \cdot \frac{n^{\frac{n}{2}}}{2^{\frac{n}{2}-1}\Gamma\left(\frac{n}{2}\right)} \int\limits_0^{+\infty} x e^{-\frac{z^2x^2}{2}} x^{n-1} e^{-\frac{nx^2}{2}} dx =$$

$$= \frac{n^{\frac{n}{2}}}{\sqrt{\pi} 2^{\frac{n-1}{2}} \Gamma\left(\frac{n}{2}\right)} \int\limits_0^{+\infty} x^n e^{-\frac{x^2}{2}(z^2+n)} dx = \left[\frac{x^2}{2}(z^2+n) = t, x = \frac{\sqrt{2}\sqrt{t}}{\sqrt{z^2+n}}, dx = \frac{1}{\sqrt{2}\sqrt{z^2+n}} \frac{dt}{\sqrt{t}}\right] =$$

$$= \frac{n^{\frac{n}{2}}}{\sqrt{\pi} 2^{\frac{n-1}{2}} \Gamma\left(\frac{n}{2}\right)} \cdot \frac{2^{\frac{n}{2}}}{(z^2+n)^{\frac{n}{2}}} \frac{1}{2^{\frac{1}{2}}(z^2+n)^{\frac{1}{2}}} \int\limits_0^{+\infty} t^{\frac{n-1}{2}} e^{-t} dt = \frac{n^{\frac{n}{2}} \Gamma\left(\frac{n+1}{2}\right)}{\sqrt{\pi}\Gamma\left(\frac{n}{2}\right)} \cdot \frac{1}{(z^2+n)^{\frac{n+1}{2}}}, \ z \in \mathbb{R}$$

Крива розподілу: графіки для різних значень n, називаються *кривими Стьюдента.* Вони схожі на криву гауссівського розподілу, але повільніше прямують до 0 на нескінченності.



Числові характеристики:

1. $\mathbb{E}\mathrm{St}_n = 0$.
2. $\mathbb{D}\mathrm{St}_n = \frac{n}{n-2}$, $n > 2$.

Зауваження. Нескладно помітити, що St_1 — це розподіл Коші.

6.4 Розподіл Фішера-Снедекора (F-розподіл)

Означення: випадкова величина $\eta = \frac{\chi^2_{n_1}/n_1}{\chi^2_{n_2}/n_2}$, чисельник та знаменник якої незалежні, має *розподіл Фішера-Снедекора з n_1, n_2 ступенями вільності.*
Коротке позначення: $\eta \sim \mathrm{F}(n_1, n_2)$, $n_1, n_2 \in \mathbb{N}$ — кількість ступенів вільності.
Щільність розподілу: скористаємося формулою для визначення щільності розподілу частки двох незалежних НВВ. Нагадаємо, що $f_{\chi^2_n}(x) = \begin{cases} \frac{1}{2^{\frac{n}{2}}\Gamma\left(\frac{n}{2}\right)} x^{\frac{n}{2}-1} e^{-\frac{x}{2}}, & x \geq 0 \\ 0, & x < 0 \end{cases}$.

Тоді $f_{\chi^2_n/n}(y) = f_{\chi^2_n}(ny) \cdot n = \begin{cases} \frac{n^{\frac{n}{2}}}{2^{\frac{n}{2}}\Gamma\left(\frac{n}{2}\right)} y^{\frac{n}{2}-1} e^{-\frac{ny}{2}}, & y \geq 0 \\ 0, & y < 0 \end{cases}$.

$$f_{\mathrm{F}(n_1,n_2)}(z) = \int_0^{+\infty} x f_{\chi^2_{n_1}/n_1}(zx) f_{\chi^2_{n_2}/n_2}(x) dx =$$

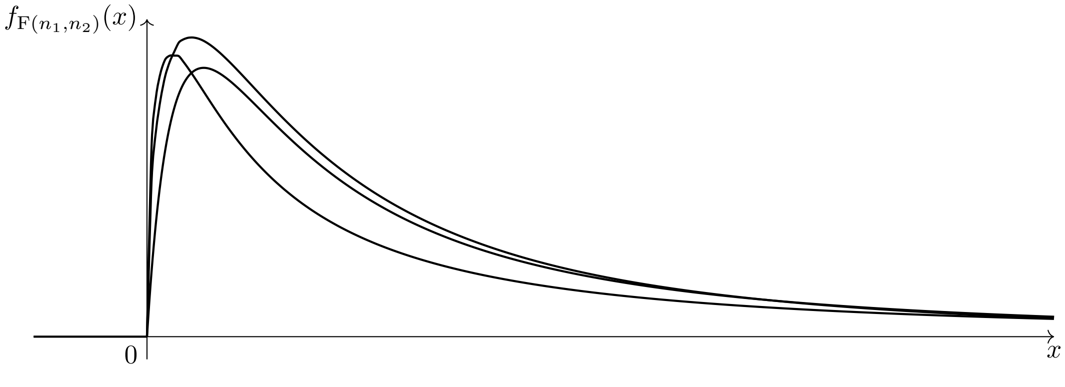
$$= \frac{n_1^{\frac{n_1}{2}} n_2^{\frac{n_2}{2}}}{2^{\frac{n_1+n_2}{2}}\Gamma\left(\frac{n_1}{2}\right)\Gamma\left(\frac{n_2}{2}\right)} \int_0^{+\infty} x z^{\frac{n_1}{2}-1} x^{\frac{n_1}{2}-1} e^{-\frac{n_1 z x}{2}} x^{\frac{n_2}{2}-1} e^{-\frac{n_2 x}{2}} dx =$$

$$= \frac{n_1^{\frac{n_1}{2}} n_2^{\frac{n_2}{2}}}{2^{\frac{n_1+n_2}{2}}\Gamma\left(\frac{n_1}{2}\right)\Gamma\left(\frac{n_2}{2}\right)} \cdot z^{\frac{n_1}{2}-1} \int_0^{+\infty} x^{\frac{n_1+n_2}{2}-1} e^{-\frac{x}{2}(n_1 z + n_2)} dx = \left[\frac{x}{2}(n_1 z + n_2) = t, x = \frac{2t}{n_1 z + n_2}\right] =$$

$$= \frac{n_1^{\frac{n_1}{2}} n_2^{\frac{n_2}{2}}}{2^{\frac{n_1+n_2}{2}}\Gamma\left(\frac{n_1}{2}\right)\Gamma\left(\frac{n_2}{2}\right)} \cdot z^{\frac{n_1}{2}-1} \cdot 2^{\frac{n_1+n_2}{2}} \cdot \frac{1}{(n_1 z + n_2)^{\frac{n_1+n_2}{2}}} \int_0^{+\infty} t^{\frac{n_1+n_2}{2}-1} e^{-t} dt =$$

$$= n_1^{\frac{n_1}{2}} n_2^{\frac{n_2}{2}} \frac{\Gamma\left(\frac{n_1+n_2}{2}\right)}{\Gamma\left(\frac{n_1}{2}\right)\Gamma\left(\frac{n_2}{2}\right)} \cdot \frac{z^{\frac{n_1}{2}-1}}{(n_1 z + n_2)^{\frac{n_1+n_2}{2}}}, \ z \geq 0 \text{ та } 0 \text{ інакше.}$$

Крива розподілу: графіки для різних значень n_1, n_2, називаються *кривими Фішера.*



Числові характеристики:

1. $\mathbb{E}\mathrm{F}(n_1, n_2) = \frac{n_2}{n_2-2}$, $n_2 > 2$.
2. $\mathbb{D}\mathrm{F}(n_1, n_2) = \frac{2n_2^2(n_1+n_2-2)}{n_1(n_2-2)^2(n_2-4)}$, $n_2 > 4$.

Зауваження. Якщо $\eta \sim \mathrm{F}(n_1, n_2)$, то $\frac{1}{\eta} \sim \mathrm{F}(n_2, n_1)$.

Розділ 7

Граничні теореми теорії ймовірностей

7.1 Послідовності випадкових величин

7.1.1 Нерівності Маркова та Чебишова

Теорема (нерівність Маркова). *Нехай модуль випадкової величини ξ має скінченне математичне сподівання: $\mathbb{E}|\xi| < +\infty$. Тоді*

$$\forall\, \varepsilon > 0 : \mathbb{P}\left\{|\xi| \geq \varepsilon\right\} \leq \frac{\mathbb{E}|\xi|}{\varepsilon} \tag{1}$$

Доведення. Запишемо випадкову величину $|\xi|$ через події-індикатори: $|\xi| = |\xi| \cdot I\left\{|\xi| \geq \varepsilon\right\} + |\xi| \cdot I\left\{|\xi| < \varepsilon\right\} \geq \varepsilon \cdot I\left\{|\xi| \geq \varepsilon\right\}$. Звідси $\mathbb{E}|\xi| \geq \mathbb{E}\left(\varepsilon \cdot I\left\{|\xi| \geq \varepsilon\right\}\right) = \varepsilon \cdot \mathbb{P}\left\{|\xi| \geq \varepsilon\right\}$. ▲

Зауваження. Еквівалентною нерівністю є $\mathbb{P}\left\{|\xi| < \varepsilon\right\} \geq 1 - \frac{\mathbb{E}|\xi|}{\varepsilon}$.

Приклад. Багатьма спостереженнями з'ясовано, що середня кількість сонячних днів у Києві складає 220. Оцінити ймовірність того, що сонячних днів за рік буде не менше 300.
Позначимо ξ кількість сонячних днів. За умовою ξ невід'ємна та $\mathbb{E}\xi = 220$, тому за нерівністю Маркова $\mathbb{P}\left\{\xi \geq 300\right\} \leq \frac{220}{300} = \frac{11}{15}$.

Теорема (нерівність Чебишова). *Нехай випадкова величина ξ має скінченні математичне сподівання та дисперсію. Тоді*

$$\forall\, \varepsilon > 0 : \mathbb{P}\left\{|\xi - \mathbb{E}\xi| \geq \varepsilon\right\} \leq \frac{\mathbb{D}\xi}{\varepsilon^2} \tag{2}$$

Доведення. $\mathbb{P}\left\{|\xi - \mathbb{E}\xi| \geq \varepsilon\right\} = \mathbb{P}\left\{(\xi - \mathbb{E}\xi)^2 \geq \varepsilon^2\right\}$. Застосуємо нерівність Маркова:
$\mathbb{P}\left\{(\xi - \mathbb{E}\xi)^2 \geq \varepsilon^2\right\} \leq \frac{\mathbb{E}(\xi-\mathbb{E}\xi)^2}{\varepsilon^2} = \frac{\mathbb{D}\xi}{\varepsilon^2}$. ▲

Зауваження. Еквівалентною нерівністю є $\mathbb{P}\left\{|\xi - \mathbb{E}\xi| < \varepsilon\right\} \geq 1 - \frac{\mathbb{D}\xi}{\varepsilon^2}$.

Приклад. Отримаємо «правило 3σ» для довільної випадкової величини зі скінченними математичним сподіванням та дисперсією. $\mathbb{P}\left\{|\xi - \mathbb{E}\xi| < 3\sigma\right\} \geq 1 - \frac{\mathbb{D}\xi}{9\sigma^2} = 1 - \frac{\sigma^2}{9\sigma^2} = \frac{8}{9}$.

Розглянемо застосування *нерівності Чебишова в схемі Бернуллі.* Нехай $\xi \sim \mathrm{Bin}(n, p)$, $\mathbb{E}\xi = np$, $\mathbb{D}\xi = npq$. Відношення $\frac{\xi}{n}$ називається відносною частотою появи успіху або частістю. $\mathbb{E}\left(\frac{\xi}{n}\right) = p$, $\mathbb{D}\left(\frac{\xi}{n}\right) = \frac{pq}{n}$. З нерівності Чебишова $\mathbb{P}\left\{\left|\frac{\xi}{n} - p\right| \geq \varepsilon\right\} \leq \frac{pq}{n\epsilon^2}$.

7.1.2 Послідовності випадкових величин

Розглядаємо фіксований ймовірнісний простір $\{\Omega, \mathcal{F}, \mathbb{P}\}$ та послідовність випадкових величин $\{\xi_n(\omega)\}_{n=1}^{\infty}$. Якщо для кожного $n \in \mathbb{N}$ випадкові величини $\xi_1, \xi_2, ..., \xi_n$ *незалежні у сукупності*, то послідовність $\{\xi_n(\omega)\}_{n=1}^{\infty}$ називається *послідовністю незалежних випадкових величин.*

Послідовності випадкових величин можна задавати різними способами:

1. Нехай ξ — деяка випадкова величина, можна задати $\xi_n = f_n(\xi)$, де f_n — деяка числова функція. Наприклад: $\xi_n = \xi^n$, $\xi_n = \cos(n\xi)$.
2. n може входити як параметр закону розподілу ξ_n. Наприклад, $\xi_n \sim \mathrm{Exp}(n)$, $\xi_n \sim \mathrm{N}(0, \frac{1}{n})$.
3. Для послідовностей ДВВ n може входити як в значення, що приймає ξ_n, так і у відповідні ймовірності. Наприклад:

ξ_n	$-\sqrt{n}$	0	$\sqrt{n}$
p	$1/n$	$1-2/n$	$1/n$

В курсі функціонального аналізу вводяться різні види збіжності послідовності вимірних функцій та зв'язок між цими видами збіжності.

7.1.3 Види збіжності послідовності випадкових величин

Нагадаємо класичне означення границі числової послідовності. Число a називають границею послідовності $\{a_n\}_{n=1}^{\infty}$, якщо

$$\forall\, \varepsilon > 0\ \exists\, N(\varepsilon) \in \mathbb{N} : \forall\, n \geq N(\varepsilon)\ \ |a_n - a| < \varepsilon$$

Оскільки послідовність випадкових величин $\{\xi_n\}_{n=1}^{\infty}$ є послідовністю функцій з Ω в $\mathbb{R}$, то це означення не є застосовним, бо $|\xi_n - \xi| < \varepsilon$ є випадковою подією, що виконується, взагалі кажучи, не для всіх елементарних подій $\omega \in \Omega$. Тому ми маємо ввести інше означення границі (та збіжності) послідовності випадкових величин. Виявляється, що таких означень можна запропонувати декілька, причому вони не є еквівалентними одне одному.

1. **Збіжність майже напевно (сильна збіжність, збіжність з ймовірністю 1).** $\{\xi_n\}_{n=1}^{\infty}$ збігається до ξ *майже напевно*, якщо $\mathbb{P}\left\{\omega : \lim\limits_{n\to\infty} \xi_n(\omega) = \xi(\omega)\right\} = 1$. Це еквівалентно умові $\mathbb{P}\left\{\omega : \lim\limits_{n\to\infty} \xi_n(\omega) \neq \xi(\omega)\right\} = 0$.

   Позначення: $\xi_n \xrightarrow{\text{P1}} \xi, n \to \infty$ або $\xi_n \xrightarrow{\text{м.н.}} \xi, n \to \infty$.

   Вправа. Довести ще одне еквівалентне означення цієї збіжності: $\xi_n \xrightarrow{\text{P1}} \xi, n \to \infty$, якщо $\forall\, \varepsilon > 0 : \mathbb{P}\left(\bigcap\limits_{n=1}^{\infty} \bigcup\limits_{k=n}^{\infty} \{\omega : |\xi_k(\omega) - \xi(\omega)| > \varepsilon\}\right) = 0$.

   Це найбільш природний з інтуїтивної точки зору вид збіжності. Зауважимо, однак, що він має доволі дивні властивості. Наприклад, можна навести приклад послідовності ξ_n, що не збігається до деякої ξ, але будь-яка її підпослідовність ξ_{n_k} містить свою підпослідовність $\xi_{n_{k_l}}$, яка все ж таки збігається до ξ.

   Твердження (лема Бореля-Кантеллі)**.** *Якщо для послідовності подій $\{A_n\}_{n=1}^{\infty}$ ряд $\sum\limits_{n=1}^{\infty} \mathbb{P}(A_n)$ збігається, то $\mathbb{P}\left(\bigcap\limits_{n=1}^{\infty} \bigcup\limits_{k=n}^{\infty} A_k\right) = 0$. Це означає, що ймовірність того, що відбудеться нескінченна кількість цих подій, є нульовою.*

   Доведення. Послідовність подій $B_n = \bigcup\limits_{k=n}^{\infty} A_k$ монотонно спадна, тому за теоремою неперервності 2 $\mathbb{P}\left(\bigcap\limits_{n=1}^{\infty} B_n\right) = \lim\limits_{n\to\infty} \mathbb{P}(B_n)$. $\mathbb{P}(B_n) = \mathbb{P}\left(\bigcup\limits_{k=n}^{\infty} A_k\right) \leq \sum\limits_{k=n}^{\infty} \mathbb{P}(A_k) \to 0$ при $n \to \infty$ зі збіжності ряду. Тому $\lim\limits_{n\to\infty} \mathbb{P}(B_n) = \mathbb{P}\left(\bigcap\limits_{n=1}^{\infty} \bigcup\limits_{k=n}^{\infty} A_k\right) = 0$. ▲

   Застосуванням цієї леми до послідовності подій $A_n = \{\omega : |\xi_n(\omega) - \xi(\omega)| > \varepsilon\}$ отримаємо зручну для використання *ознаку збіжності з ймовірністю 1*: якщо для будь-якого $\varepsilon > 0$ ряд $\sum\limits_{n=1}^{\infty} \mathbb{P}\{|\xi_n - \xi| > \varepsilon\}$ збігається, то $\xi_n \xrightarrow{\text{P1}} \xi, n \to \infty$.

   Приклад. Довести, що послідовність $\xi_n \sim \mathrm{U}(-\frac{1}{n}, \frac{1}{n})$ при $n \to \infty$ збігається до 0 майже напевно.

   Візьмемо довільне $\varepsilon > 0$ та знайдемо $\mathbb{P}\{|\xi_n| > \varepsilon\}$. Для $\varepsilon \geq 1$ ця ймовірність, очевидно, рівна 0. В іншому випадку, для кожного $\varepsilon \in (0; 1)$ можна знайти такий номер N, для якого ε буде більше за $\frac{1}{n}$ при $n \geq N$. Тому для будь-якого $\varepsilon > 0$ ймовірності $\mathbb{P}\{|\xi_n| > \varepsilon\}$ рівні 0, починаючи з якогось n. Отже, для будь-якого $\varepsilon > 0$ ряд $\sum\limits_{n=1}^{\infty} \mathbb{P}\{|\xi_n| > \varepsilon\}$ збігається і $\xi_n \xrightarrow{\text{P1}} 0, n \to \infty$.

2. **Збіжність за ймовірністю.** $\{\xi_n\}_{n=1}^{\infty}$ збігається до ξ *за ймовірністю*, якщо $\forall\ \varepsilon > 0$: $\lim\limits_{n\to\infty} \mathbb{P}\{|\xi_n - \xi| \geq \varepsilon\} = 0$. Це еквівалентно умові $\forall\ \varepsilon > 0$: $\lim\limits_{n\to\infty} \mathbb{P}\{|\xi_n - \xi| \leq \varepsilon\} = 1$. У функціональному аналізі така збіжність називається «збіжністю за мірою».

   Позначення: $\xi_n \xrightarrow{\mathrm{P}} \xi, n \to \infty$.

   Приклад. Нехай ξ_n — послідовність ДВВ:

ξ_n	0	n^7
p	$1-1/n$	$1/n$

   . Перевірити збіжність $\xi_n \xrightarrow{\mathrm{P}} 0, n \to \infty$.

   Для $\varepsilon > 0$ $\mathbb{P}\{|\xi_n - 0| \geq \varepsilon\} = \mathbb{P}\{\xi_n \geq \varepsilon\}$. $\forall\ \varepsilon > 0\ \exists\ N : \forall n \geq N : n^7 > \varepsilon$, тому з якогось номера $\mathbb{P}\{\xi_n \geq \varepsilon\} = \mathbb{P}\left\{\xi_n = n^7\right\} = \frac{1}{n} \to 0, n \to \infty$, тому $\xi_n \xrightarrow{\mathrm{P}} 0, n \to \infty$.

3. **Збіжність в середньому.** $\{\xi_n\}_{n=1}^{\infty}$ збігається до ξ *в середньому*, якщо $\lim\limits_{n\to\infty} \mathbb{E}|\xi_n - \xi| = 0$.

   Позначення: $\xi_n \xrightarrow{\mathrm{C}} \xi, n \to \infty$.

4. **Збіжність в середньому квадратичному.** $\{\xi_n\}_{n=1}^{\infty}$ збігається до ξ *в середньому квадратичному*, якщо $\lim\limits_{n\to\infty} \mathbb{E}(\xi_n - \xi)^2 = 0$.

   Позначення: $\xi_n \xrightarrow{\mathrm{CK}} \xi, n \to \infty$.

5. **Збіжність за розподілом (слабка збіжність, збіжність в основному).** $\{\xi_n\}_{n=1}^{\infty}$ збігається до ξ *за розподілом*, якщо функціональна послідовність $\{F_{\xi_n}(x)\}_{n=1}^{\infty}$ збігається до $F_\xi(x)$ в точках її неперервності.

   Позначення: $\xi_n \xrightarrow{\mathrm{F}} \xi, n \to \infty$.

   Ця збіжність за характером відрізняється від інших тим, що не враховує залежність або незалежність ξ_n. Є також *еквівалентне означення збіжності* за розподілом: якщо для будь-якої обмеженої неперервної функції $\varphi : \mathbb{R} \to \mathbb{R}$ $\lim\limits_{n\to\infty} \mathbb{E}\varphi(\xi_n) = \mathbb{E}\varphi(\xi)$. Доведення еквівалентності цих двох означень виходить за рамки курсу: його ідея полягає в тому, що $F_\xi(x) = \mathbb{P}\{\xi < x\} = \mathbb{E}1_{(-\infty,x)}(\xi)$, де $1_{(-\infty,x)}$ — індикатор множини $(-\infty, x)$, і такі функції-індикатори можна наблизити з будь-якою заданою точністю неперервними функціями та навпаки.

   Приклад. Нехай $\xi_n \sim \mathrm{Exp}(\frac{1}{n})$. Перевірити $\xi_n \xrightarrow{\mathrm{F}} 0, n \to \infty$. Тут під 0 розуміється ДВВ, що приймає значення 0 з ймовірністю 1.

   $F_{\xi_n}(x) = \begin{cases} 1 - e^{-\frac{x}{n}}, & x > 0 \\ 0, & x \leq 0 \end{cases}$. Видно, що при $n \to \infty$ $F_{\xi_n}(x) \to F_0(x) = \begin{cases} 1, & x > 0 \\ 0, & x \leq 0 \end{cases}$ — функція розподілу 0.

Зауваження. Збіжності в середньому та середньому квадратичному є частковими випадками *збіжності порядку k*, для якої $\lim\limits_{n\to\infty} \mathbb{E}|\xi_n - \xi|^k = 0$. Оскільки для $0 < s < t$ має місце $\left(\mathbb{E}\,|\xi|^s\right)^{\frac{1}{s}} \leq \left(\mathbb{E}\,|\xi|^t\right)^{\frac{1}{t}}$ то збіжність порядку k гарантує збіжність порядків менше k.

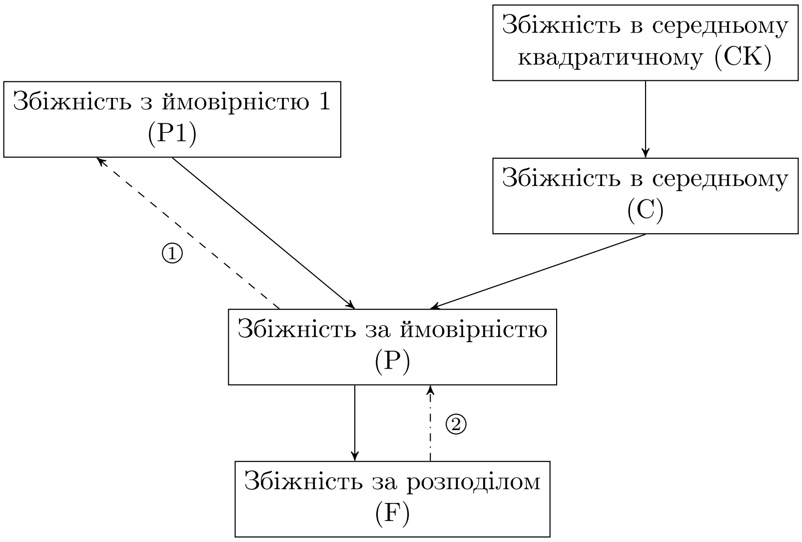
Твердження. *Нехай послідовність $\{\xi_n\}_{n=1}^{\infty}$ збігається за ймовірністю до двох випадкових величин ξ та η. Тоді $\mathbb{P}\{\xi = \eta\} = 1$.*

Доведення. Для будь-якого $\varepsilon > 0 : \{|\xi - \eta| > \varepsilon\} \subset \left\{|\xi - \xi_n| > \frac{\varepsilon}{2}\right\} \cup \left\{|\eta - \xi_n| > \frac{\varepsilon}{2}\right\}$, тому $\mathbb{P}\{|\xi - \eta| > \varepsilon\} \leq \mathbb{P}\left\{|\xi - \xi_n| > \frac{\varepsilon}{2}\right\} + \mathbb{P}\left\{|\eta - \xi_n| > \frac{\varepsilon}{2}\right\} \to 0, n \to \infty$. Отже, $\mathbb{P}\{|\xi - \eta| > \varepsilon\} = 0$, і за довільністю $\varepsilon > 0$ маємо $\mathbb{P}\{\xi \neq \eta\} = 0$, або ж $\mathbb{P}\{\xi = \eta\} = 1$. ▲

Вправа. Довести, що це твердження виконується для збіжностей з ймовірністю 1, в середньому та середньому квадратичному, але не виконується для збіжності за розподілом.

Можна довести, що усім збіжностям, крім збіжності за розподілом, притаманні відомі арифметичні властивості: збіжність суми $\xi_n + \eta_n$ та добутку $\xi_n\eta_n$ до $\xi + \eta$ та $\xi\eta$ відповідно за умови збіжностей ξ_n до ξ та η_n до η. Також, $\varphi(\xi_n)$ збігається до $\varphi(\xi)$ за умови неперервності φ.

В курсі функціонального аналізу встановлюється зв'язок між розглянутими видами збіжності:



Штрихова лінія ① означає, що у кожній послідовності $\{\xi_n\}_{n=1}^{\infty}$, що збігається за ймовірністю, міститься підпослідовність $\{\xi_{n_k}\}_{k=1}^{\infty}$, що збігається з імовірністю 1. Штрих-пунктирна лінія ② показує, що відповідний перехід справедливий, коли гранична випадкова величина ξ є константою. Доведемо деякі з цих тверджень.

Твердження. *Зі збіжності в середньому квадратичному випливає збіжність в середньому.*

Доведення. $\mathbb{E}|\xi_n - \xi| \le \sqrt{\mathbb{E}(\xi_n - \xi)^2}$, як було сказано вище, тому з $\mathbb{E}(\xi_n - \xi)^2 \to 0$ випливає $\mathbb{E}|\xi_n - \xi| \to 0, n \to \infty$. ▲

Твердження. *Зі збіжності в середньому випливає збіжність за ймовірністю.*

Доведення. З нерівності Маркова (1) $\forall\, \varepsilon > 0 : \mathbb{P}\left\{|\xi_n - \xi| \ge \varepsilon\right\} \le \frac{\mathbb{E}|\xi_n - \xi|}{\varepsilon}$, тому з $\mathbb{E}|\xi_n - \xi| \to 0$ маємо $\mathbb{P}\left\{|\xi_n - \xi| \ge \varepsilon\right\} \to 0, n \to \infty$. ▲

Твердження. *Зі збіжності за розподілом* до константи *випливає збіжність за ймовірністю.*

Доведення. Нехай $\xi_n \xrightarrow{\text{F}} c$, $c \in \mathbb{R}$. Візьмемо $\varepsilon > 0$. $\mathbb{P}\left\{|\xi_n - c| \le \varepsilon\right\} = \mathbb{P}\left\{c - \varepsilon \le \xi_n \le c + \varepsilon\right\} \ge \mathbb{P}\left\{c - \varepsilon \le \xi_n < c + \varepsilon\right\} = F_{\xi_n}(c + \varepsilon) - F_{\xi_n}(c - \varepsilon)$, тому $\lim\limits_{n\to\infty} \mathbb{P}\left\{|\xi_n - c| \le \varepsilon\right\} \ge \lim\limits_{n\to\infty} F_{\xi_n}(c + \varepsilon) - \lim\limits_{n\to\infty} F_{\xi_n}(c - \varepsilon) = F_c(c + \varepsilon) - F_c(c - \varepsilon) = 1$, звідки $\xi_n \xrightarrow{\text{P}} c$. ▲

Приклад. Нехай ξ_n — послідовність ДВВ:

ξ_n	$-n$	1	$2n$
p	$1/2n$	$1 - 1/n$	$1/2n$

Зі збільшенням n ξ_n все з більшою ймовірністю набувають значення 1. Водночас, інші два можливі значення ($-n$ та $2n$) розбігаються до нескінченності. Ці два процеси спричиняють протилежні ефекти — перший наближає ξ_n до 1, а другий — віддаляє. Тому з точки зору одних видів збіжності, більш чутливих до «аномальних викидів», послідовність ξ_n не буде мати границі, а з точки зору інших, менш чутливих, границя дорівнюватиме 1. Дійсно:

$$\mathbb{E}\,|\xi_n - 1| = \frac{1}{2n}|-n-1| + \left(1 - \frac{1}{n}\right)|1-1| + \frac{1}{2n}|2n-1| \to \frac{3}{2}$$

$$\mathbb{E}\,(\xi_n - 1)^2 = \frac{1}{2n}(-n-1)^2 + \left(1 - \frac{1}{n}\right)(1-1)^2 + \frac{1}{2n}(2n-1)^2 \to \infty$$

Отже, збіжності до 1 ні в середньому, ні в середньому квадратичному немає. Зауважимо, що якби «викиди» прямували до нескінченності повільніше або відповідні ймовірності прямували до нуля швидше, то ці збіжності могли б бути: наприклад, якщо замість $-n$ та $2n$ ξ_n набували значення $\sqrt{n}$ та $2\sqrt{n}$ з ймовірностями $\frac{1}{2n^2}$, то послідовність збігалася б до 1 і в середньому, і в середньому квадратичному. З іншого боку, для будь-якого $\varepsilon > 0$ значення $-n$ та $2n$ рано чи пізно вийдуть за межі відрізку $[1 - \varepsilon; 1 + \varepsilon]$, і тому $\xi_n \xrightarrow{\text{P}} 1$. Тепер стає зрозуміло, чому перевіряли збіжність в середньому та середньому квадратичному лише до 1: якби якась з цих границь існувала, то вона була б і границею за ймовірністю.

Зауваження. Для збіжності $\xi_n \xrightarrow{\text{СК}} C$ (C — стала) необхідно і достатньо, щоб $\lim\limits_{n\to\infty} \mathbb{E}\xi_n = C$ та $\lim\limits_{n\to\infty} \mathbb{D}\xi_n = 0$. $\lim\limits_{n\to\infty} \mathbb{E}\xi_n = C \Leftrightarrow \lim\limits_{n\to\infty} (\mathbb{E}\xi_n - \mathbb{E}C) = 0$, $\mathbb{E}(\xi_n - C)^2 = \mathbb{D}\xi_n + (\mathbb{E}(\xi_n - C))^2$, звідки отримуємо твердження при $n \to \infty$, оскільки $\mathbb{D}\xi_n \geq 0$ та $(\mathbb{E}(\xi_n - C))^2 \geq 0$. Зауважимо, що ці дві умови є достатніми для збіжності $\xi_n \xrightarrow{\text{P}} C$.

Приклад. Нехай $\xi_n \sim \text{N}\left(5 + \frac{1}{n}, \sigma = \frac{1}{n}\right)$. Перевірити $\xi_n \xrightarrow{\text{P}} 5$.
Оскільки $\mathbb{E}\xi_n = 5 + \frac{1}{n} \to 5$, $\mathbb{D}\xi_n = \frac{1}{n^2} \to 0$ при $n \to \infty$, то $\xi_n \xrightarrow{\text{СК}} 5 \Rightarrow \xi_n \xrightarrow{\text{P}} 5$.

Наведемо без доведення важливу теорему, що стосується збіжності за розподілом.

Теорема (теорема неперевності Леві). *Збіжність за розподілом послідовності випадкових величин $\{\xi_n\}_{n=1}^{\infty}$ еквівалентна поточковій збіжності їх характеристичних функцій:*

$$\xi_n \xrightarrow{\text{F}} \xi \iff \chi_{\xi_n}(t) \to \chi_{\xi}(t)\ \forall\, t \in \mathbb{R} \tag{3}$$

7.2 Закон великих чисел

Закон великих чисел (ЗВЧ) — загальна назва низки теорем та фактів, які встановлюють умови, за яких середнє арифметичне випадкових величин зі зростанням кількості доданків втрачає свою «випадковість» і може бути передбачено з наперед заданою точністю.

7.2.1 Теорема Чебишова (ЗВЧ у формі Чебишова).

Теорема. *Нехай $\{\xi_n\}_{n=1}^{\infty}$ — послідовність незалежних випадкових величин, таких, що існують скінченні $\mathbb{E}\xi_k = a_k$ та $\mathbb{D}\xi_k = \sigma_k^2$ для кожного $k \in \mathbb{N}$, причому дисперсії рівномірно обмежені: тобто $\exists\, C < +\infty : \mathbb{D}\xi_k \leq C\ \forall\, k \in \mathbb{N}$. Тоді:*

$$\forall\, \varepsilon > 0 \lim_{n\to\infty} \mathbb{P}\left\{\left|\frac{1}{n}\sum_{k=1}^{n} \xi_k - \frac{1}{n}\sum_{k=1}^{n} \mathbb{E}\xi_k\right| \geq \varepsilon\right\} = 0 \text{ (у випадку} < \varepsilon \text{ — рівна 1)} \tag{1}$$

Це означає

$$\frac{1}{n}\sum_{k=1}^{n} \xi_k - \frac{1}{n}\sum_{k=1}^{n} \mathbb{E}\xi_k \xrightarrow{\text{P}} 0,\ n \to \infty \tag{2}$$

Доведення. Позначимо $\eta_n = \frac{1}{n}\sum\limits_{k=1}^{n} \xi_k$, тоді $\mathbb{E}\eta_n = \mathbb{E}\left(\frac{1}{n}\sum\limits_{k=1}^{n} \xi_k\right) = \frac{1}{n}\sum\limits_{k=1}^{n} \mathbb{E}\xi_k$,
$\mathbb{D}\eta_n = \mathbb{D}\left(\frac{1}{n}\sum\limits_{k=1}^{n} \xi_k\right) = [\xi_k \text{ — незалежні}] = \frac{1}{n^2}\sum\limits_{k=1}^{n} \mathbb{D}\xi_k \leq \frac{C}{n}$. Тепер скористаємося нерівністю Чебишова (2): $\forall\ \varepsilon > 0 : \mathbb{P}\{|\eta_n - \mathbb{E}\eta_n| \geq \varepsilon\} \leq \frac{\mathbb{D}\eta_n}{\varepsilon^2} \leq \frac{C}{n\varepsilon^2} \to 0, n \to \infty$ — що і треба було довести. ▲

Наслідок. Припустимо, що всі ВВ ξ_n розподілені однаково і виконуються всі умови теореми. Тоді $\frac{1}{n}\sum\limits_{k=1}^{n} \mathbb{E}\xi_k = a$ і $\frac{1}{n}\sum\limits_{k=1}^{n} \xi_k \xrightarrow{\text{P}} a, n \to \infty$. Розглянемо природну інтерпретацію цього факту. Нехай *невипадкову* величину a може бути виміряно за допомогою деякого пристрою. Внаслідок наявності похибок вимірювання цей пристрій вимірює не точне значення a, а лише деяку *випадкову* величину ξ, в якомусь сенсі близьку до a. Будемо, однак, вважати, що в середньому пристрій повертає правильний результат: $\mathbb{E}\xi = a$ (іноді в прикладних науках це називають відсутністю систематичних похибок). Як в цій ситуації отримати якомога більш точне значення a? Отриманий наслідок теореми Чебишова обґрунтовує інтуїтивну відповідь на це питання: провести декілька вимірювань і обчислити їх середнє арифметичне.

Зауваження. Для однаково розподілених ξ_n припущення щодо дисперсії, виявляється, є зайвим. Для спрощення розглянемо випадок $\mathbb{E}\xi_n = 0$. Нехай $\chi(t)$ — характеристична функція розподілу всіх ξ_n, тоді для $\eta_n = \frac{1}{n}\sum\limits_{k=1}^{n} \xi_k$ маємо характеристичну функцію $\chi_{\eta_n}(t) = \chi^n\left(\frac{t}{n}\right)$. Тепер $|\chi_{\eta_n}(t) - 1| = \left|\chi^n\left(\frac{t}{n}\right) - 1\right| \leq n\left|\chi\left(\frac{t}{n}\right) - 1\right|$. Нерівність отримано з простого факту: для

$z, w \in \mathbb{C}$ з $|z| \le 1$ та $|w| \le 1$ $|z^n - w^n| = |z - w| \cdot \left|z^{n-1} + z^{n-2}w + ... + zw^{n-2} + w^{n-1}\right| \le n \cdot |z - w|$. Оскільки 1 — це характеристична функція нульової випадкової величини, а $\lim\limits_{n\to\infty} n\left(\chi(\frac{t}{n}) - 1\right) = t \cdot \chi'(0) = [\mathbb{E}\xi_n = 0] = 0$, то за теоремою Леві $\xi_n \xrightarrow{\text{F}} 0$, а тому й $\xi_n \xrightarrow{\text{P}} 0$. Залишилося зауважити, що у випадку $\mathbb{E}\xi_n = a \ne 0$ можемо розглядати послідовність $\xi_n - a$, і $\xi_n - a \xrightarrow{\text{P}} 0 \Leftrightarrow \xi_n \xrightarrow{\text{P}} a$.

Наведемо ще один варіант формулювання закону великих чисел.

Теорема (теорема Маркова)**.** *Нехай $\{\xi_n\}_{n=1}^{\infty}$ — послідовність випадкових величин (можливо, залежних), таких, що існують скінченні $\mathbb{E}\xi_k = a_k$ та $\mathbb{D}\xi_k = \sigma_k^2$ для кожного $k \in \mathbb{N}$, причому $\mathbb{D}\left(\sum\limits_{k=1}^{n} \xi_k\right) = o(n^2)$ при $n \to \infty$. Тоді $\frac{1}{n}\sum\limits_{k=1}^{n} \xi_k - \frac{1}{n}\sum\limits_{k=1}^{n} \mathbb{E}\xi_k \xrightarrow{\text{P}} 0,\ n \to \infty$.*

Вправа. Довести теорему Маркова.

Приклад. 1. Скільки вимірювань величини a треба провести, щоб з ймовірністю 0.998 стверджувати, що похибка середнього арифметичного результатів вимірювань не перевищує $\frac{1}{100}$, якщо середньоквадратичне відхилення кожного вимірювання $\sigma = 0.03$? Позначимо $\eta_n = \frac{1}{n}\sum\limits_{k=1}^{n} \xi_k$, де ξ_k — результат k-того випробування. $\mathbb{D}\eta_n = \frac{1}{n^2}\sum\limits_{k=1}^{n} \mathbb{D}\xi_k = \frac{1}{n^2} \cdot n\sigma^2 = \frac{\sigma^2}{n}$. $\mathbb{P}\{|\eta_n - a| \le \underbrace{0.01}_{\varepsilon}\} \ge 1 - \frac{1}{0.01^2} \cdot \mathbb{D}\eta_n = 1 - \frac{0.03^2}{n \cdot 0.01^2} = 0.998 \Rightarrow \frac{9}{n} = 0.002$, тому $n = 4500$.

2. Нехай ξ_n — послідовність випадкових величин, що мають розподіл $\mathrm{U}[-1;1]$. Знайти границю за ймовірністю послідовності $\eta_n = \frac{1}{n}\sum\limits_{k=1}^{n} e^{\xi_k}$.

   За ЗВЧ $\eta_n \xrightarrow{\text{P}} \mathbb{E}e^{\xi}$, де $\xi \sim \mathrm{U}[-1;1]$. $\mathbb{E}e^{\xi} = \frac{1}{2}\int\limits_{-1}^{1} e^x dx = \sinh 1$, тому $\eta_n \xrightarrow{\text{P}} \sinh 1$.

Вправа. Знайти границю за ймовірністю послідовності $\eta_n = \sqrt[n]{\xi_1 \cdot ... \cdot \xi_n}$, де ξ_n незалежні та мають розподіл $\mathrm{U}[0;1]$.

7.2.2 Посилений закон великих чисел

Посилений закон великих чисел — це загальна назва ЗВЧ, що стосуються збіжності з ймовірністю 1. Наведемо без доведення приклад такого закону.

Теорема (теорема Колмогорова)**.** *Нехай $\{\xi_n\}_{n=1}^{\infty}$ — послідовність незалежних однаково розподілених випадкових величин, що мають скінченне математичне сподівання a. Тоді без припущень щодо дисперсії*

$$\frac{1}{n}\sum_{k=1}^{n} \xi_k \xrightarrow{\text{P1}} a,\ n \to \infty \tag{3}$$

7.2.3 Закон великих чисел у схемі Бернуллі

Розглянемо частковий випадок ЗВЧ Чебишова для схеми Бернуллі.

Теорема (теорема Бернуллі)**.** *Нехай ξ_n задає кількість успіхів в схемі Бернуллі з n випробуваннями зі сталою ймовірністю успіху p, $\xi_n \sim \mathrm{Bin}(n,p)$. Тоді $\frac{\xi_n}{n} \xrightarrow{\text{P}} p, n \to \infty$, тобто*

$$\forall\, \varepsilon > 0 : \lim_{n\to\infty} \mathbb{P}\left\{\left|\frac{\xi_n}{n} - p\right| \ge \varepsilon\right\} = 0 \tag{4}$$

Доведення. Зведемо до теореми Чебишова. $\xi_n = \sum\limits_{k=1}^{n} \eta_k$, де η_k — індикатор появи успіху в k-тому випробуванні, $\eta_k \sim \mathrm{Bin}(1,p)$. η_k — незалежні та однаково розподілені, $\mathbb{E}\eta_k = p$ тому за наслідком ЗВЧ Чебишова для однаково розподілених ВВ $\frac{\xi_n}{n} = \frac{1}{n}\sum\limits_{k=1}^{n} \eta_k \xrightarrow{\text{P}} p, n \to \infty$. ▲

Зауваження. Для практичного використання цієї теореми важливою є оцінка збіжності $\frac{\xi_n}{n}$ до p:

$$\mathbb{P}\left\{\left|\frac{\xi_n}{n}-p\right|\geq\varepsilon\right\}\leq\frac{\mathbb{D}\eta_k}{n\varepsilon^2}=\frac{p(1-p)}{n\varepsilon^2}\leq\frac{1}{4n\varepsilon^2}$$

Остання нерівність пояснюється тим, що максимальне значення функції $f(t)=t(1-t)=-\left(t-\frac{1}{2}\right)^2+\frac{1}{4}$ на відрізку $[0;1]$ дорівнює $\frac{1}{4}$.

Приклад. Оцінити ймовірність того, що при 10^4 підкиданнях симетричної монети частість випадіння герба відхилиться від $\frac{1}{2}$ на 0.01 і більше.

Нехай ξ_n задає кількість гербів, що випали за n підкидань. Тоді $\mathbb{P}\left\{\left|\frac{\xi_{10000}}{10000}-\frac{1}{2}\right|\geq 0.01\right\}\leq\frac{1}{4\cdot 10^4\cdot 0.01^2}=\frac{1}{4}$.

Наведемо узагальнення теореми Бернуллі.

Теорема. *Нехай ξ_n задає кількість успіхів в схемі Бернуллі з n випробуваннями та ймовірністю успіху p_n, $\xi_n\sim\mathrm{Bin}(n,p_n)$. Тоді $\frac{\xi_n}{n}-\frac{1}{n}\sum\limits_{k=1}^{n}p_k\xrightarrow{\mathrm{P}}0,\ n\to\infty$.*

Вправа. Довести це узагальнення.

7.2.4 Методи Монте-Карло

Методом (або методами) Монте-Карло називають широкий клас підходів, що дозволяють наближено розв'язувати детерміновані («невипадкові») задачі ймовірнісними методами. Історично одним з перших застосувань такого підходу було наближене обчислення числа π за допомогою голки — задача Бюффона. Проілюструємо метод Монте-Карло задачі наближеного інтегрування.

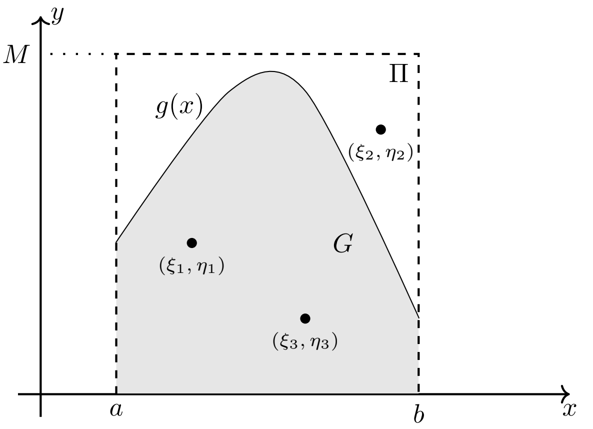
Нехай $g\,[a,b]\to[0,+\infty)$ — деяка неперервна невід'ємна функція, для якої потрібно наближено обчислити значення $\int\limits_a^b g(x)dx$. Запропонуємо два способи такого обчислення.

Спочатку розглянемо послідовність випадкових величин ξ_n, що мають спільний розподіл $\mathrm{U}[a;b]$. $\mathbb{E}g(\xi_1)=\frac{1}{b-a}\int\limits_a^b g(x)dx$, тому $\int\limits_a^b g(x)dx=(b-a)\cdot\mathbb{E}g(\xi_1)$. З теореми Колмогорова (3) $\frac{1}{n}\sum\limits_{k=1}^{n}g(\xi_k)\xrightarrow{\mathrm{P1}}\mathbb{E}g(\xi_1),n\to\infty$, тому при достатньо великих n виконується наближена рівність:

$$\int\limits_a^b g(x)dx\approx\frac{b-a}{n}\sum_{k=1}^{n}g(\xi_k)$$

Другий спосіб розв'язання цієї задачі заснований на інтерпретації інтеграла як площі під графіком функції. Оскільки g є неперервною функцією на відрізку, вона також є обмеженою: існує таке $M>0$, що $g(x)\leq M$ для будь-якого $x\in[a,b]$. Тому, кидаючи випадкові точки в прямокутник $\Pi=[a;b]\times[0;M]$, ми будемо потрапляти у підграфік G функції g якраз з ймовірністю

$$p=\frac{\text{площа } G}{\text{площа } \Pi}=\frac{\int\limits_a^b g(x)dx}{M(b-a)}$$



Для формалізації введемо дві незалежні послідовності незалежних випадкових величин ξ_n та η_n, де $\xi_n \sim \mathrm{U}[a;b]$, $\eta_n \sim \mathrm{U}[0;M]$ для всіх $n \in \mathbb{N}$. Розглянемо схему Бернуллі, де «успіхом» в n-тому випробуванні будемо вважати потрапляння точки (ξ_n, η_n) в область G, причому його ймовірність рівна заданому вище p. Знову за теоремою Колмогорова відношення кількості успіхів до загальної кількості проведених випробувань з ймовірністю 1 прямує до p. Тому при достатньо великих n виконується наближена рівність

$$\int\limits_a^b g(x)dx \approx M \cdot (b-a) \cdot \nu_n$$

де ν_n — відношення кількості успіхів до загальної кількості проведених випробувань. Зрозуміло, що такий підхід узагальнюється на кратні інтеграли.

7.3 Центральна гранична теорема

Як було показано вище, для послідовності незалежних ВВ $\{\xi_n\}_{n=1}^{\infty}$ зі скінченними математичними сподіваннями та обмеженими в сукупності дисперсіями $\frac{1}{n}\sum\limits_{k=1}^{n}(\xi_k - \mathbb{E}\xi_k) \xrightarrow{\mathrm{P}} 0$. Виявляється, що якщо знаменник цих сум буде прямувати до 0 повільніше, то границя (хоча й в слабшому сенсі) вже не буде нульовою.

7.3.1 Теорема Ляпунова

Теорема (теорема Ляпунова)**.** *Нехай $\{\xi_n\}_{n=1}^{\infty}$ — послідовність незалежних випадкових величин таких, що існують скінченні $\mathbb{E}\xi_k = a_k$, $\mathbb{D}\xi_k = \sigma_k^2$ та $m_k = \mathbb{E}\left|\xi_k - a_k\right|^3$ і виконується умова Ляпунова:*

$$\lim_{n\to\infty} \frac{\sum\limits_{k=1}^{n} m_k}{\left(\sum\limits_{k=1}^{n} \sigma_k^2\right)^{3/2}} = 0$$

Тоді рівномірно відносно x:

$$\lim_{n\to\infty} \mathbb{P}\left\{\frac{\sum\limits_{k=1}^{n}(\xi_k - a_k)}{\sqrt{\sum\limits_{k=1}^{n}\sigma_k^2}} < x\right\} = F_{\mathrm{N}(0,1)}(x) = \Phi(x) + \frac{1}{2} = \frac{1}{\sqrt{2\pi}}\int\limits_0^x e^{-\frac{t^2}{2}}dt + \frac{1}{2} \qquad (1)$$

або:

$$\frac{\sum\limits_{k=1}^{n}(\xi_k - a_k)}{\sqrt{\sum\limits_{k=1}^{n}\sigma_k^2}} \xrightarrow{\mathrm{F}} \xi \sim \mathrm{N}(0,1),\ n \to \infty$$

Доведення. Позначимо $\eta_n = \frac{\sum\limits_{k=1}^{n}(\xi_k - a_k)}{\sqrt{\sum\limits_{k=1}^{n}\sigma_k^2}}$, $\xi_k - a_k = \overset{\circ}{\xi}_k$ $(\mathbb{E}\overset{\circ}{\xi}_k = 0)$, $B_n = \sqrt{\sum\limits_{k=1}^{n}\sigma_k^2}$. Тоді $\eta_n = \frac{1}{B_n}\sum\limits_{k=1}^{n}\overset{\circ}{\xi}_k$. $\mathbb{E}\eta_n = 0$, $\mathbb{D}\eta_n = \frac{1}{B_n^2}\sum\limits_{k=1}^{n}\mathbb{D}\overset{\circ}{\xi}_k = \left[\mathbb{D}\xi_k = \mathbb{D}\overset{\circ}{\xi}_k\right] = 1$.

Доведення ґрунтується на теоремі Леві (3): збіжність за розподілом послідовності випадкових величин $\{\xi_n\}_{n=1}^{\infty}$ еквівалентна поточковій збіжності їх характеристичних функцій.

Нехай $\chi_k(t)$ — характеристична функція $\overset{\circ}{\xi}_k$, а $F_k(x)$ — функція розподілу $\overset{\circ}{\xi}_k$. За властивостями характеристичних функцій $\chi_{\eta_n}(t) = \prod\limits_{k=1}^{n}\chi_k(\frac{t}{B_n})$. Доведемо, що $\chi_{\eta_n}(t) \to e^{-t^2/2} = \chi_{\mathrm{N}(0,1)}(t)$ при $n \to \infty$.

$$\chi_k\left(\frac{t}{B_n}\right) = \int\limits_{-\infty}^{+\infty} e^{i\frac{t}{B_n}x}dF_k(x) = \int\limits_{-\infty}^{+\infty}\left(1 + i\frac{t}{B_n}x - \frac{t^2}{2B_n^2}x^2 - i\frac{t^3}{6B_n^3}x^3 + ...\right)dF_k(x) =$$

$$= [\theta_k \in (0;1)] = 1 + \frac{it}{B_n}\mathbb{E}\overset{\circ}{\xi}_k - \frac{t^2}{2B_n^2}\mathbb{E}\overset{\circ}{\xi}_k{}^2 + \theta_k\frac{|t|^3}{6B_n^3}\mathbb{E}|\overset{\circ}{\xi}_k|^3 = 1 - \frac{t^2}{2B_n^2}\sigma_k^2 + \theta_k\frac{|t|^3}{6B_n^3}m_k$$

$$\ln\chi_{\eta_n}(t) = \sum_{k=1}^{n}\ln\chi_k\left(\frac{t}{B_n}\right) = \sum_{k=1}^{n}\ln\left(1 - \frac{t^2}{2B_n^2}\sigma_k^2 + \theta_k\frac{|t|^3}{6B_n^3}m_k\right)$$

З умови Ляпунова $\lim\limits_{n\to\infty}\frac{\sum\limits_{k=1}^{n}m_k}{B_n^3} = 0$, тому з відомої властивості $\ln(1+\alpha_n) \sim \alpha_n$, де α_n — нескінченно мала при $n \to \infty$, маємо $\ln\chi_{\eta_n}(t) \sim \sum\limits_{k=1}^{n}\left(-\frac{t^2\sigma_k^2}{2B_n^2} + \theta_k\frac{|t|^3}{6B_n^3}m_k\right) = -\frac{t^2}{2}\frac{\sum\limits_{k=1}^{n}\sigma_k^2}{B_n^2} + \sum\limits_{k=1}^{n}\theta_k\frac{|t|^3}{6B_n^3}m_k = -\frac{t^2}{2} + \sum\limits_{k=1}^{n}\theta_k\frac{|t|^3}{6B_n^3}m_k \to -\frac{t^2}{2}$ при $n \to \infty$.

Таким чином, $\chi_{\eta_n}(t) \to e^{-\frac{t^2}{2}}$ при $n \to \infty$ і з теореми Леві маємо, що $\eta_n \xrightarrow{\mathrm{F}} \xi \sim \mathrm{N}(0,1)$. ▲

Зауваження. Можна довести, що для послідовності неперервних випадкових величин також має місце збіжність щільностей розподілу η_n до $\frac{1}{\sqrt{2\pi}}e^{-\frac{x^2}{2}}$ — щільності розподілу $\mathrm{N}(0,1)$.

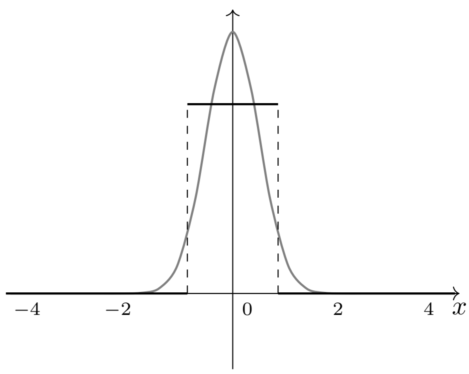
Наслідок. Нехай всі ξ_n *однаково розподілені*, $\mathbb{E}\xi_k = a$, $\mathbb{D}\xi_k = \sigma^2$, $m_k = \mathbb{E}\left|\xi_k - a\right|^3 = m$. Тоді умова Ляпунова виконується автоматично:

$$\lim_{n\to\infty}\frac{\sum\limits_{k=1}^{n}m_k}{\left(\sum\limits_{k=1}^{n}\sigma_k^2\right)^{3/2}} = \lim_{n\to\infty}\frac{n\cdot m}{(n\cdot\sigma^2)^{3/2}} = 0$$

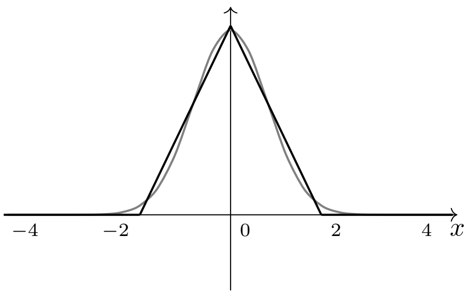
В цьому випадку

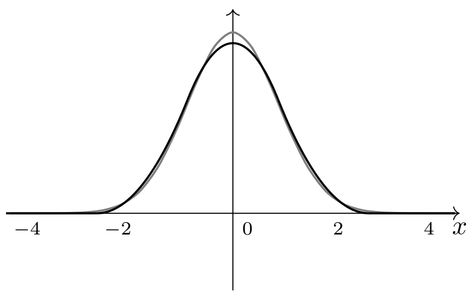
$$\frac{\sum\limits_{k=1}^{n}(\xi_k - a_k)}{\sqrt{\sum\limits_{k=1}^{n}\sigma_k^2}} = \frac{1}{\sqrt{n}}\sum_{k=1}^{n}\left(\frac{\xi_k - a}{\sigma}\right) \xrightarrow{\mathrm{F}} \xi \sim \mathrm{N}(0,1),\ n \to \infty$$

Це твердження означає, що яким би не був розподіл ξ_n, $\zeta_n = \xi_1 + ... + \xi_n$ матиме «приблизно нормальний» розподіл $\mathrm{N}(na, n\sigma^2)$. Проілюструємо це на прикладі рівномірного розподілу. На рисунку нижче зображено графіки щільностей розподілів $\mathrm{U}[-1;1]$ та $\mathrm{N}\left(0, \frac{1}{3}\right)$. Зауважимо, що $\frac{1}{3}$ — це дисперсія $\mathrm{U}[-1;1]$.



Розглянемо тепер щільності розподілу сум двох (зліва) та трьох (справа) незалежних ВВ, кожна з яких має розподіл $U[-1;1]$, і щільності нормальних розподілів $N\left(0,\frac{2}{3}\right)$ та $N\left(0,1\right)$ відповідно.





Видно, що графік щільності суми трьох ВВ дуже схожий на графік щільності нормального розподілу $N\left(0,1\right)$.

Вправа. Записати в явному вигляді щільності розподілу $\zeta_2 = \xi_1 + \xi_2$ та $\zeta_3 = \xi_1 + \xi_2 + \xi_3$, де $\xi_k \sim U[-1;1], k = 1,2,3$.

Варто пояснити практичне значення умови Ляпунова $\lim\limits_{n\to\infty} \frac{\sum_{k=1}^{n} \mathbb{E}|\xi_k - \mathbb{E}\xi_k|^3}{\left(\sum_{k=1}^{n} \mathbb{D}\xi_k\right)^{3/2}} = 0$ в доведенні теореми. Вона означає, що дисперсії величин у послідовності мають бути приблизно однакового порядку. Можна навести приклад з життя: якщо в багатоквартирному будинку не буде квартири, що використовує значно більше електроенергії, ніж інші, то сумарний розподіл кількості використаної електроенергії буде приблизно нормальним.

7.3.2 Умова Ліндеберга

Замість умови Ляпунова в доведенні однойменної теореми можна вимагати виконання умови Ліндеберга:

$$\forall\, \varepsilon > 0 : \lim_{n\to\infty} \frac{1}{B_n^2} \sum_{k=1}^{n} \mathbb{E}\left((\xi_k - a_k)^2 \cdot 1_{\{|\xi_k - a_k| > \varepsilon \cdot B_n\}} \right) = 0$$

Тут $a_k = \mathbb{E}\xi_k$, $B_n = \sqrt{\sum\limits_{k=1}^{n} \mathbb{D}\xi_k}$, 1_A — індикатор події A. Ймовірнісний зміст цієї умови такий: нехай $A_k = \{|\xi_k - a_k| > \varepsilon \cdot B_n\}$, тоді

$$\mathbb{P}\left(\bigcup_{k=1}^{n} A_k\right) \le \sum_{k=1}^{n} \mathbb{P}(A_k) = \sum_{k=1}^{n} \int_{|x-a_k|>\varepsilon\cdot B_n} dF_k(x) \le \sum_{k=1}^{n} \int_{|x-a_k|>\varepsilon\cdot B_n} \frac{(x-a_k)^2}{\varepsilon^2 B_n^2} dF_k(x) =$$
$$= \frac{1}{\varepsilon^2} \cdot \frac{1}{B_n^2} \sum_{k=1}^{n} \mathbb{E}\left((\xi_k - a_k)^2 \cdot 1_{\{|\xi_k - a_k| > \varepsilon \cdot B_n\}} \right) \to 0,\ n \to \infty$$

Це означає, що кожний доданок $\frac{\xi_k - a_k}{B_n}$ має рівномірно малий внесок в суму $\frac{1}{B_n} \sum\limits_{k=1}^{n} (\xi_k - a_k)$. Повне доведення ЦГТ з умовою Ліндеберга виходить за рамки курсу.

7.4 Застосування ЦГТ до схеми Бернуллі

7.4.1 Інтегральна теорема Муавра-Лапласа

Теорема. *Нехай ξ_n задає кількість успіхів в схемі Бернуллі з ймовірністю успіху p, тобто $\xi_n \sim \mathrm{Bin}(n,p)$, $q = 1-p$. Тоді рівномірно відносно x:*

$$\lim_{n\to\infty} \mathbb{P}\left\{\frac{\xi_n - np}{\sqrt{npq}} < x\right\} = F_{\mathrm{N}(0,1)}(x) = \Phi(x) + \frac{1}{2} = \frac{1}{\sqrt{2\pi}}\int\limits_0^x e^{-\frac{t^2}{2}}dt + \frac{1}{2} \tag{1}$$

Доведення. $\xi_n = \sum\limits_{k=1}^n \eta_k$, де η_k — індикатор появи успіху в k-тому випробуванні, $\eta_k \sim \mathrm{Bin}(1,p)$. η_k — незалежні та однаково розподілені, $\mathbb{E}\eta_k = p$, $\mathbb{D}\eta_k = pq$, тому до них можна застосувати ЦГТ:

$$\sum_{k=1}^n (\eta_k - \mathbb{E}\eta_k) = \xi_n - np,\ \sum_{k=1}^n \mathbb{D}\eta_k = npq$$

$$\frac{\sum\limits_{k=1}^n (\eta_k - \mathbb{E}\eta_k)}{\sqrt{\sum\limits_{k=1}^n \mathbb{D}\eta_k}} = \frac{\xi_n - np}{\sqrt{npq}} \xrightarrow{\mathrm{F}} \xi \sim \mathrm{N}(0,1),\ n \to \infty$$

▲

Робочі формули. При великих n:

1. $\mathbb{P}\left\{\frac{\xi_n - np}{\sqrt{npq}} < x\right\} \approx \Phi(x) + \frac{1}{2}$.
2. $\forall\, a, b \in \mathbb{R},\ a < b : \mathbb{P}\left\{a < \frac{\xi_n - np}{\sqrt{npq}} < b\right\} \approx \Phi(b) - \Phi(a)$.
3. $\forall\, m_1, m_2 \in \mathbb{R}, m_1 < m_2 : \mathbb{P}\{m_1 < \xi_n < m_2\} = \mathbb{P}\left\{\frac{m_1 - np}{\sqrt{npq}} < \frac{\xi_n - np}{\sqrt{npq}} < \frac{m_2 - np}{\sqrt{npq}}\right\} \approx$ $\approx \Phi\left(\frac{m_2 - np}{\sqrt{npq}}\right) - \Phi\left(\frac{m_1 - np}{\sqrt{npq}}\right)$.
4. $\forall\ \varepsilon > 0 : \mathbb{P}\left\{\left|\frac{\xi_n}{n} - p\right| < \varepsilon\right\} = \mathbb{P}\left\{-\varepsilon < \frac{\xi_n - np}{n} < \varepsilon\right\} = \mathbb{P}\left\{-\varepsilon\sqrt{\frac{n}{pq}} < \frac{\xi_n - np}{\sqrt{npq}} < \varepsilon\sqrt{\frac{n}{pq}}\right\} \approx$ $\approx \Phi\left(\varepsilon\sqrt{\frac{n}{pq}}\right) - \Phi\left(-\varepsilon\sqrt{\frac{n}{pq}}\right) = 2\Phi\left(\varepsilon\sqrt{\frac{n}{pq}}\right)$.

Зауваження. В усіх наведених наближених рівностях будь-яку строгу нерівність можна замінити на нестрогу. На практиці у формулі 3 іноді застосовують *«поправку на неперервність»*, використовуючи $m_2 + \frac{1}{2}$ та $m_1 - \frac{1}{2}$ замість m_2 та m_1 відповідно. Емпірично з'ясовано, що це дозволяє підвищити точність наближення. Також важливо зауважити, що на практиці ці формули застосовують при $n \geq 50$ та $npq \geq 10$ (як і для наступних формул, різні джерела можуть давати інші межі, наприклад $npq \geq 20$).

Приклад. Стрілець робить 100 пострілів по мішені з ймовірністю влучення 0.7, ξ — кількість влучень. Знайти наближено $\mathbb{P}\{67 \leq \xi \leq 72\}$ та порівняти з точним значенням.
Маємо $np = 70$, $npq = 100 \cdot 0.7 \cdot 0.3 = 21 \geq 10$, $\mathbb{P}\{67 \leq \xi \leq 72\} \approx \Phi\left(\frac{2}{\sqrt{21}}\right) - \Phi\left(-\frac{3}{\sqrt{21}}\right) \approx 0.4124$.
З поправкою на неперервність отримаємо $\mathbb{P}\{67 \leq \xi \leq 72\} \approx \Phi\left(\frac{2.5}{\sqrt{21}}\right) - \Phi\left(-\frac{3.5}{\sqrt{21}}\right) \approx 0.4848$.
Підрахунок за точною формулою $\sum\limits_{k=67}^{72} C_{100}^k 0.7^k 0.3^{100-k}$ дає значення 0.4829.

7.4.2 Локальна теорема Муавра-Лапласа

Теорема. *Нехай проводиться n незалежних випробувань з ймовірністю успіху p, $q = 1-p$, причому n досить велике. Тоді ймовірність отримання m успіхів*

$$\mathbb{P}_n(m) \approx \frac{1}{\sqrt{2\pi npq}} \exp\left\{-\frac{(m-np)^2}{2npq}\right\} \tag{2}$$

Доведення. Нехай ξ задає кількість успіхів у заданій схемі Бернуллі, тоді за інтегральною теоремою Муавра-Лапласа $\mathbb{P}_n(m) = \mathbb{P}\left\{m - \frac{1}{2} < \xi < m + \frac{1}{2}\right\} \approx \Phi\left(\frac{m+\frac{1}{2}-np}{\sqrt{npq}}\right) - \Phi\left(\frac{m-\frac{1}{2}-np}{\sqrt{npq}}\right)$. За означенням похідної $\Phi(x+\varepsilon) - \Phi(x-\varepsilon) \approx 2\varepsilon\Phi'(x)$ при малих ε, причому $\Phi'(x) = \frac{1}{\sqrt{2\pi}}\exp\left\{-\frac{x^2}{2}\right\}$. Звідси $\Phi\left(\frac{m+\frac{1}{2}-np}{\sqrt{npq}}\right) - \Phi\left(\frac{m-\frac{1}{2}-np}{\sqrt{npq}}\right) \approx \frac{1}{\sqrt{2\pi npq}}\exp\left\{-\frac{(m-np)^2}{2npq}\right\}$. ▲

Зауваження. Так само, як інтегральна теорема, локальна теорема Муавра-Лапласа на практиці застосовується при $n \geq 50$ та $npq \geq 10$ (або $npq \geq 20$).

Приклад. Стрілець робить 100 пострілів по мішені з ймовірністю влучення 0.7, ξ — кількість влучень. Знайти наближено $\mathbb{P}\{\xi = 69\}$ та порівняти з точним значенням.
Маємо $npq = 21$, $\mathbb{P}\{\xi = 69\} \approx \frac{1}{\sqrt{42\pi}}\exp\left\{-\frac{1}{42}\right\} \approx 0.085$. Підрахунок за точною формулою $C_{100}^{69}0.7^{69}0.3^{31}$ дає значення 0.084.

7.4.3 Гранична теорема Пуассона

Розглянемо граничу теорему для схеми Бернуллі, яка застосовується для наближених обчислень, коли умова $npq \geq 10$ для застосування локальної теореми Муавра-Лапласа не виконується.

Теорема. *Задано нескінченну серію схем Бернуллі: перша складається з одного випробування B_1 з імовірністю успіху p_1, друга — з двох випробувань B_2^1 та B_2^2 з імовірністю успіху p_2, третя — з трьох випробувань B_3^1, B_3^2 та B_3^3 з імовірністю успіху p_3 і так далі. Випадкова величина $\xi_n \sim \mathrm{Bin}(n, p_n)$ — кількість успіхів в n-тій схемі Бернуллі. Існує таке $a > 0$, що $\lim\limits_{n\to\infty} np_n = a$. Тоді для будь-якого цілого $m \geq 0$:*

$$\mathbb{P}\{\xi_n = m\} \to \frac{a^m}{m!}e^{-a},\ n \to \infty \tag{3}$$

Доведення. Нехай $\xi \sim \mathrm{Poiss}(a)$. Покажемо $\xi_n \xrightarrow{\mathrm{F}} \xi, n \to \infty$, скориставшись твердженням теореми Леві (3). Характеристичні функції ξ_n та ξ рівні, відповідно, $\chi_{\xi_n}(t) = \left(p_n e^{it} + 1 - p_n\right)^n$ та $\chi_\xi(t) = e^{a\left(e^{it}-1\right)}$. З умови $\lim\limits_{n\to\infty} np_n = a$ та нерівності $|z^n - w^n| \leq n \cdot |z - w|$ для $z, w \in \mathbb{C}$, $|z| \leq 1$, $|w| \leq 1$:

$$\left|\left(p_n e^{it} + 1 - p_n\right)^n - \left(\frac{a}{n}e^{it} + 1 - \frac{a}{n}\right)^n\right| \leq n \cdot \left|p_n - \frac{a}{n}\right| \left|e^{it} - 1\right| \to 0,\ n \to \infty$$

Тому маємо

$$\lim_{n\to\infty} \chi_{\xi_n}(t) = \lim_{n\to\infty}\left(\frac{a}{n}e^{it} + 1 - \frac{a}{n}\right)^n = \lim_{n\to\infty}\left(1 + \frac{a}{n}\left(e^{it} - 1\right)\right)^n = e^{a\left(e^{it}-1\right)} = \chi_\xi(t)$$

Отже, $\xi_n \xrightarrow{\mathrm{F}} \xi, n \to \infty$. Функція розподілу ξ неперервна в будь-якій нецілій точці, тому для цілих $m \geq 0$:

$$\mathbb{P}\left\{m - \frac{1}{2} \leq \xi_n < m + \frac{1}{2}\right\} = F_{\xi_n}\left(m + \frac{1}{2}\right) - F_{\xi_n}\left(m - \frac{1}{2}\right) \to$$
$$\to F_\xi\left(m + \frac{1}{2}\right) - F_\xi\left(m - \frac{1}{2}\right) = \mathbb{P}\left\{m - \frac{1}{2} \leq \xi < m + \frac{1}{2}\right\} = \mathbb{P}\{\xi = m\} = \frac{a^m}{m!}e^{-a},\ n \to \infty$$

що і треба було довести. ▲

Вправа. Довести цю теорему без використання характеристичних функцій, показавши

$$C_n^m p_n^m (1 - p_n)^{n-m} \to \frac{a^m}{m!}e^{-a}, n \to \infty$$

На практиці цією теоремою користуються, якщо $n \geq 50$ та $np \leq 10$ (або n — достатньо велике, p — достатньо мале, причому $1 < np < 20$): ймовірність отримати m успіхів у схемі Бернуллі з n випробуваннями та ймовірністю успіху p $\mathbb{P}_n(m) \approx \frac{(np)^m}{m!}e^{-np}$, $m = 0, ..., n$.

Приклад. Стрілець робить 100 пострілів по мішені. Ймовірність влучення при одному пострілі становить 0.98. Знайти наближено ймовірність того, що вдалих пострілів буде не більше 97, та порівняти з точним значенням.
Щоб звести задачу до використання теореми Пуассона, шукатимемо ймовірність, що промахів буде щонайменше 3 (ймовірність промаху — 0.02). ξ — кількість промахів, $\mathbb{P}\{\xi \geq 3\} = 1 - \mathbb{P}\{\xi \leq 2\} = 1 - p_{100}(0) - p_{100}(1) - p_{100}(2) \approx 1 - \left(\frac{2^0}{0!} + \frac{2^1}{1!} + \frac{2^2}{2!}\right) e^{-2} = 1 - 5e^{-2} \approx 0.32332$.
Підрахунок за точною формулою $1 - \sum\limits_{k=0}^{2} C_{100}^k 0.02^k 0.98^{100-k}$ дасть значення 0.32331. Нескладно перевірити, що застосування локальної теореми Муавра-Лапласа дало б менш точну відповідь 0.36049.

7.5 Збіжність та граничні теореми для послідовностей випадкових векторів

В цьому розділі наведемо без доведення деякі означення та факти, що стосуються збіжності послідовностей, закону великих чисел та центральної граничної теореми для випадкових векторів. Розглядаємо послідовності випадкових векторів $\left\{\vec{\xi}_n(\omega)\right\}_{n=1}^{\infty}$ на фіксованому ймовірносному просторі.

7.5.1 Збіжність випадкових векторів

На послідовності випадкових векторів цілком природно переносяться означення збіжності за ймовірністю, за розподілом та з ймовірністю 1.

1. Збіжність з ймовірністю 1: $\left(\vec{\xi}_n \xrightarrow{\text{P1}} \vec{\xi}, n \to \infty\right) \Leftrightarrow \left(\mathbb{P}\left\{\omega : \lim\limits_{n\to\infty} \vec{\xi}_n(\omega) = \vec{\xi}(\omega)\right\} = 1\right)$.
2. Збіжність за ймовірністю: $\left(\vec{\xi}_n \xrightarrow{\text{P}} \vec{\xi}, n \to \infty\right) \Leftrightarrow \left(\forall\, \varepsilon > 0 : \lim\limits_{n\to\infty} \mathbb{P}\left\{\left\|\vec{\xi}_n - \vec{\xi}\right\| \geq \varepsilon\right\} = 0\right)$.
3. Збіжність за розподілом:
$\left(\vec{\xi}_n \xrightarrow{\text{F}} \vec{\xi}, n \to \infty\right) \Leftrightarrow \left(\lim\limits_{n\to\infty} F_{\vec{\xi}_n}(\vec{x}) = F_{\vec{\xi}}(\vec{x}) \text{ для точок неперервності } F_{\vec{\xi}}\right)$.

Зв'язки між цими видами збіжності такі ж, як у випадку випадкових величин.

7.5.2 Граничні теореми

Для простоти розглядаємо лише випадок *незалежних однаково розподілених* $\vec{\xi}_n$.

Теорема (посилений закон великих чисел). *Нехай* $\left\{\vec{\xi}_n\right\}_{n=1}^{\infty}$ *— послідовність незалежних однаково розподілених випадкових векторів, що мають скінченне математичне сподівання* $\vec{a}$ *(тобто* $\|\vec{a}\| < \infty$*). Тоді*

$$\frac{1}{n}\sum_{k=1}^{n} \vec{\xi}_k \xrightarrow{\text{P1}} \vec{a},\ n \to \infty$$

Теорема (центральна гранична теорема). *Нехай* $\left\{\vec{\xi}_n\right\}_{n=1}^{\infty}$ *— послідовність незалежних однаково розподілених випадкових векторів, що мають скінченні математичне сподівання* $\vec{a}$ *та кореляційну матрицю* K*. Тоді*

$$\sqrt{n}\left(\frac{1}{n}\sum_{k=1}^{n} \vec{\xi}_k - \vec{a}\right) = \frac{1}{\sqrt{n}}\sum_{k=1}^{n}\left(\vec{\xi}_k - \vec{a}\right) \xrightarrow{\text{F}} \vec{\eta} \sim \mathrm{N}\left(\vec{0}, K\right),\ n \to \infty$$

Частина II

Математична статистика

Розділ 8

Вибірка, точкове та інтервальне оцінювання

8.1 Вибірка та її первинний аналіз

8.1.1 Поняття вибірки

Математична статистика вивчає випадкові величини за певними дослідними даними, які отримано в ході експерименту. Коротко кажучи, методами математичної статистики можна оцінювати (але не визначати точно) числові характеристики та розподіл деякої випадкової величини, якщо відомо деякий набір значень, яких ця величина набула в ході досліду. При цьому, зазвичай, апріорних відомостей про цю випадкову величину майже немає.

Означення 8.1.1. *Генеральною сукупністю (ГС)* називають як випадкову величину ξ, що досліджується, так і множину всіх її можливих значень.

Означення 8.1.2. *Випадковою вибіркою обсягу n* називається випадковий вектор $(\xi_1, ..., \xi_n)$, координати якого є однаково розподіленими, як ГС, і незалежними в сукупності. Підмножина $G \subseteq \mathbb{R}^n$, що складається з усіх можливих значень визначеної вище випадкової вибірки, називається *вибірковим простором*, а вектори $\vec{x} \in G$ — *реалізаціями вибірки. Конкретною реалізацією вибірки* інколи називають саме ту реалізацію, з якою працюють після проведення експерименту.

Приклад. Нехай $\xi \sim \mathrm{Bin}(N, p)$, $N = 3$. В цьому випадку $G = \{0, 1, 2, 3\}^{\times n}$ — множина n-вимірних векторів, всі координати яких набувають лише значень 0, 1, 2 та 3.

Зауваження. В прикладній статистиці зазвичай не користуються таким різноманітним набором означень. Там під вибіркою розуміють як і будь-які отримані внаслідок експерименту спостереження, так і процес їх отримання. Варто навести два означення, які, хоч і не будуть застосовуватися далі в курсі, але зустрічаються в прикладній статистиці. *Репрезентативна вибірка* — така, яка має всі властивості генеральної сукупності. Інакше кажучи, всі можливі значення (або проміжки значень) ГС мають однакову ймовірність потрапити до конкретної реалізації вибірки. *Стратифікована вибірка* — така, що гарантує збереження пропорцій, наявних у ГС. Наприклад, якщо мова йде про результати якогось опитування, то репрезентативна вибірка має містити результати всіх категорій населення, а стратифікована ще й має містити їх в тих пропорціях, які ці категорії складають в усьому населенні країни. На практиці поняття репрезентативної та стратифікованої вибірки залежать від ГС, структуру та природу якої дослідник попередньо вивчає, та від мети самого дослідження: наприклад, населення всієї країни може грати роль ГС у багатьох статистичних дослідженнях, але його поділ на категорії може відрізнятися.

8.1.2 Розподіл випадкової вибірки

Нехай $\vec{\xi} = (\xi_1, ..., \xi_n)$ — випадкова вибірка, а $F_\xi(x)$ — функція розподілу ГС. Тоді для $\vec{x} \in \mathbb{R}^n$ визначено функцію розподілу $F_{\vec{\xi}}(\vec{x}) = \mathbb{P}\{\xi_1 < x_1, ..., \xi_n < x_n\} = \prod\limits_{k=1}^{n} F_\xi(x_k)$. Для досліджень вибірок, однак, зручно користуватися іншою функцією.

Означення 8.1.3. *Функцією правдоподібності* випадкової вибірки обсягу n з ГС ξ називається

$$\mathcal{L}(\vec{x}) = \prod_{k=1}^{n} \mathbb{P}\left\{\xi = x_k\right\}, \text{ якщо } \xi \text{ — ДВВ} \tag{1}$$

$$\mathcal{L}(\vec{x}) = \prod_{k=1}^{n} f_{\xi}(x_k) \text{ якщо } \xi \text{ — НВВ} \tag{2}$$

Часто аргументами функції правдоподібності вважають параметри закону розподілу ГС. В такому випадку при фіксованому значенні $\vec{x}$ ця функція фактично показує, як в залежності від параметрів розподілу змінюється ймовірність отримати саме таку реалізацію вибірки — $(x_1, ..., x_n)$. Отримаємо функції правдоподібності для основних законів розподілу. В подальшому будуть більш корисними не самі функції правдоподібності, а їх логарифми.

1. $\xi \sim \text{Bin}(N, p)$ — біноміальний розподіл, $\forall\, k \in \mathbb{N}\ x_k \in \{0, 1, ..., N\}$:

$$\mathcal{L}_{\text{Bin}}(\vec{x}, N, p) = \prod_{k=1}^{n} C_N^{x_k} p^{x_k} (1-p)^{N-x_k} = \prod_{k=1}^{n} C_N^{x_k} \cdot p^{\sum_{k=1}^{n} x_k} \cdot (1-p)^{N\cdot n - \sum_{k=1}^{n} x_k}$$

$$\ln \mathcal{L}_{\text{Bin}}(\vec{x}, N, p) = \sum_{k=1}^{n} \ln C_N^{x_k} + \ln p \cdot \sum_{k=1}^{n} x_k + \ln(1-p) \cdot \left(N \cdot n - \sum_{k=1}^{n} x_k\right)$$

2. $\xi \sim \text{Geom}(p)$ — геометричний розподіл, $\forall\, k \in \mathbb{N}\ x_k \in \{1, 2, 3, ...\}$.

$$\mathcal{L}_{\text{Geom}}(\vec{x}, p) = \prod_{k=1}^{n} p(1-p)^{x_k - 1} = p^n \cdot (1-p)^{\sum_{k=1}^{n} x_k - n}$$

$$\ln \mathcal{L}_{\text{Geom}}(\vec{x}, p) = n \ln p + \ln(1-p) \cdot \left(\sum_{k=1}^{n} x_k - n\right)$$

3. $\xi \sim \text{Pas}(a)$ — розподіл Паскаля, $\forall\, k \in \mathbb{N}\ x_k \in \{0, 1, 2, 3, ...\}$.

$$\mathcal{L}_{\text{Pas}}(\vec{x}, a) = \prod_{k=1}^{n} \frac{a^{x_k}}{(1+a)^{x_k+1}} = \frac{a^{\sum_{k=1}^{n} x_k}}{(1+a)^{n + \sum_{k=1}^{n} x_k}}$$

$$\ln \mathcal{L}_{\text{Pas}}(\vec{x}, a) = \ln a \cdot \sum_{k=1}^{n} x_k - \ln(1+a) \cdot \left(n + \sum_{k=1}^{n} x_k\right)$$

4. $\xi \sim \text{Poiss}(a)$ — розподіл Пуассона, $\forall\, k \in \mathbb{N}\ x_k \in \{0, 1, 2, 3, ...\}$.

$$\mathcal{L}_{\text{Poiss}}(\vec{x}, a) = \prod_{k=1}^{n} \frac{a^{x_k}}{(x_k)!} e^{-a} = e^{-na} \cdot \frac{a^{\sum\limits_{k=1}^{n} x_k}}{\prod\limits_{k=1}^{n} x_k!}$$

$$\ln \mathcal{L}_{\text{Poiss}}(\vec{x}, a) = -na + \ln a \cdot \sum_{k=1}^{n} x_k - \sum_{k=1}^{n} \ln x_k!$$

5. $\xi \sim \text{U}\langle a; b\rangle$ — рівномірний розподіл, $\forall\, k \in \mathbb{N}\ x_k \in \langle a; b\rangle$:

$$\mathcal{L}_{\text{U}}(\vec{x}, a, b) = \prod_{k=1}^{n} \frac{1}{b-a} = \frac{1}{(b-a)^n}$$

$$\ln \mathcal{L}_{\text{U}}(\vec{x}, a, b) = -n \ln(b-a)$$

6. $\xi \sim \text{Exp}(\lambda, b)$ — експоненційний розподіл зі зсувом, $\forall\, k \in \mathbb{N}\ x_k \geq b$:

$$\mathcal{L}_{\text{Exp}}(\vec{x}, \lambda, b) = \prod_{k=1}^{n} \lambda e^{-\lambda(x_k - b)} = \lambda^n e^{-\lambda\left(\sum\limits_{k=1}^{n} x_k - nb\right)}$$

$$\ln \mathcal{L}_{\text{Exp}}(\vec{x}, \lambda, b) = n \ln \lambda - \lambda \left(\sum_{k=1}^{n} x_k - nb\right)$$

7. $\xi \sim \mathrm{N}(a, \sigma^2)$ — нормальний розподіл, $\forall\, k \in \mathbb{N}\ x_k \in \mathbb{R}$:

$$\mathcal{L}_{\mathrm{N}}(\vec{x}, a, \sigma) = \prod_{k=1}^{n} \frac{1}{\sqrt{2\pi}\sigma} \exp\left\{-\frac{(x_k - a)^2}{2\sigma^2}\right\} = \frac{1}{(2\pi)^{\frac{n}{2}}\sigma^n} \cdot \exp\left\{-\frac{1}{2\sigma^2}\sum_{k=1}^{n}(x_k - a)^2\right\}$$

$$\ln \mathcal{L}_{\mathrm{N}}(\vec{x}, a, \sigma) = -\frac{n}{2}\ln 2\pi - n\ln\sigma - \frac{1}{2\sigma^2}\sum_{k=1}^{n}(x_k - a)^2$$

8.1.3 Дискретний варіаційний ряд

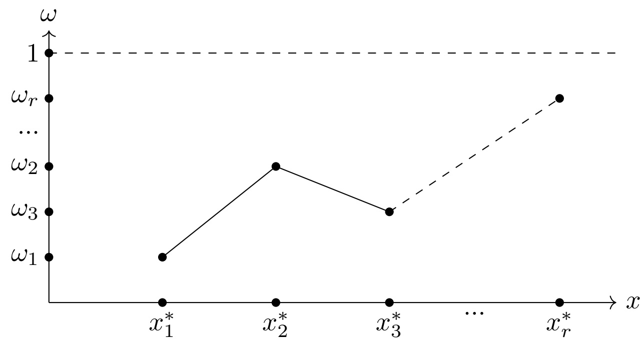
Нехай $(x_1, ..., x_n)$ — конкретна реалізація вибірки, що містить небагато (порівняно з обсягом вибірки) унікальних значень, які називаються *варіантами* та позначаються x_i^*.

Означення 8.1.4. *Дискретним варіаційним рядом (ДВР)* називається таблиця вигляду

варіанти	x_1^*	x_2^*	...	x_r^*
частоти n_i	n_1	n_2	...	n_r
частості ω_i	$\frac{n_1}{n}$	$\frac{n_2}{n}$	...	$\frac{n_r}{n}$

Тут $x_1^* < x_2^* < ... < x_r^*$, частоти — кількість входжень відповідної варіанти до реалізації вибірки, $\sum\limits_{i=1}^{r} n_i = n$, $\sum\limits_{i=1}^{r} \omega_i = 1$. Частості ще називають відносними частотами. Іноді до ДВР вносять накопичені частості $\omega_i^{\text{нак}}$: $\omega_1^{\text{нак}} = \frac{n_1}{n} = \omega_1$, $\omega_2^{\text{нак}} = \frac{n_1+n_2}{n}$ і т.д., $w_r^{\text{нак}} = 1$.

Геометричною інтерпретацію ДВР є *полігон відносних частот* з вершинами (x_i^*, ω_i):



Якщо в конкретній реалізації вибірки унікальних значень досить мало, то можна висунути припущення, що ГС має дискретний розподіл. В такому випадку вигляд полігону частостей може допомогти висунути припущення про закон розподілу. Перевірку таких припущень буде розглянуто пізніше.

8.1.4 Інтервальний варіаційний ряд

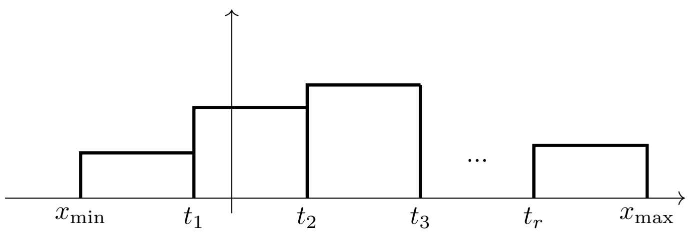
Нехай $(x_1, ..., x_n)$ — конкретна реалізація вибірки, майже всі значення якої є унікальними. В такому випадку дискретний варіаційний ряд будувати недоцільно.

Означення 8.1.5. *Інтервальний варіаційним рядом (ІВР)* називається таблиця вигляду

інтервал Δ_i	$[x_{\min}; t_1)$	$[t_1; t_2)$	...	$[t_r, x_{\max}]$
частоти n_i	n_1	n_2	...	n_{r+1}
частості ω_i	$\frac{n_1}{n}$	$\frac{n_2}{n}$	...	$\frac{n_{r+1}}{n}$

Тут $x_{\min}$ та $x_{\max}$ — мінімальне та максимальне значення цієї реалізації вибірки ($R = x_{\max} - x_{\min}$ називається *розмахом вибірки*), а t_i — деякі точки всередині відрізка $[x_{\min}; x_{\max}]$. Немає універсальної рекомендації для вибору значень та кількості t_i. На практиці часто користуються *правилом Стерджеса*: відрізок $[x_{\min}; x_{\max}]$ ділиться на $k = 1 + [3.322 \lg n]$ інтервалів рівної довжини. Іноді можна зустріти формулу $k = 1 + [\log_2 n]$, яка еквівалентна попередній, оскільки $\log_2 n = \log_2 10 \cdot \lg n \approx 3.322 \lg n$.

Геометричною інтерпретацію ІВР є *гістограма*, що складається з прямокутників, що побудовані на Δ_i та мають висоти $h_i = \frac{\omega_i}{d_i}$, де d_i — довжина Δ_i:



Якщо в конкретній реалізації вибірки багато унікальних значень, то можна висунути припущення, що ГС має неперервний розподіл. В такому випадку вигляд гістограми може допомогти висунути припущення про закон розподілу. Нескладно помітити, що сумарна площа прямокутників у гістограмі $\sum\limits_{i=1}^{r+1} d_i \cdot \frac{\omega_i}{d_i} = 1$, тому в деякому сенсі гістограма є аналогом графіка щільності розподілу. Зауважимо, що в разі невдало вибраної кількості інтервалів (забагато чи замало) гістограма буде погано відображати характер отриманої реалізації вибірки.

8.1.5 Емпірична функція розподілу

Користуючись варіаційними рядами, для конкретної реалізації вибірки можна побудувати *емпіричну функцію розподілу* $F_n^*(x)$, яка визначається наступним чином:

$$F_n^*(x) = \begin{cases} 0, & x \le x_1^* \\ \omega_1^{\text{нак}} = \frac{n_1}{n}, & x_1^* < x \le x_2^* \\ \omega_2^{\text{нак}} = \frac{n_1+n_2}{n}, & x_2^* < x \le x_3^* \\ ... & \\ 1, & x > x_r^* \end{cases} \qquad F_n^*(x) = \begin{cases} 0, & x \le x_{\min} \\ \omega_1^{\text{нак}} = \frac{n_1}{n}, & x_{\min} < x \le t_1 \\ \omega_2^{\text{нак}} = \frac{n_1+n_2}{n}, & t_1 < x \le t_2 \\ ... & \\ 1, & x > t_r \end{cases}$$

для ДВР — для ІВР

Існує й інше означення, більш коректне з теоретичної точки зору та універсальне.

Означення 8.1.6. *Емпіричною функцією розподілу (ЕФР)*, побудованою за випадковою вибіркою $\vec{\xi} = (\xi_1, ..., \xi_n)$, називається випадкова функція $F_n^* : \mathbb{R} \times \Omega \to [0; 1]$, що задана як

$$F_n^*(x) = \frac{1}{n} \sum_{k=1}^{n} 1\{\xi_k < x\} \tag{3}$$

Тут $1\{\xi_k < x\}$ — індикатор події $\{\xi_k < x\}$, що дорівнює 1 з ймовірністю $p = \mathbb{P}\{\xi_k < x\} = F_\xi(x)$ та 0 з ймовірністю $1 - p$.

Зауваження. Поняття випадкової функції тут, по суті, не є новим — це випадкова величина, що залежить від дійсного параметру (в цьому випадку x).

Таким чином, означення ЕФР, наведені спочатку, є значеннями цієї випадкової функції при фіксованих значеннях $\vec{\xi}$. Наступна теорема ілюструє важливість означення ЕФР саме як випадкової функції.

Теорема (теорема Глівенко-Кантеллі)**.**

$$\forall\, x \in \mathbb{R} : F_n^*(x) \xrightarrow{\text{P1}} F_\xi(x),\ n \to \infty$$

Доведення. Для всіх $k \in \mathbb{N}$ та $x \in \mathbb{R} : \mathbb{E}1\{\xi_k < x\} = F_\xi(x)$, причому при будь-якому $x \in \mathbb{R}$ ці індикатори є незалежними випадковими величинами. Тому згідно посиленого закону великих чисел отримуємо бажану збіжність з ймовірністю 1 до $F_\xi(x)$. ▲

Вправа. Довести теорему Глівенко-Кантеллі безпосередньо, скориставшись достатньою ознакою збіжності з ймовірністю 1: якщо для будь-якого $\varepsilon > 0$ ряд $\sum\limits_{n=1}^{\infty} \mathbb{P}\{|\xi_n - \xi| > \varepsilon\}$ збігається, то $\xi_n \xrightarrow{\text{P1}} \xi, n \to \infty$. Підказка: треба оцінити $\mathbb{P}\{|F_n^*(x) - F_\xi(x)| > \varepsilon\}$ членами збіжного ряду.

8.1.6 Вибіркові характеристики ГС

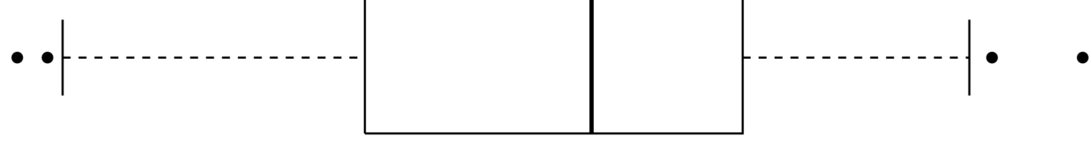
Означення 8.1.7. *Статистикою* називається випадкова величина, що є функцією від випадкової вибірки $h = h(\vec{\xi})$.

Розглянемо деякі статистики, які в деякому сенсі є аналогами числових характеристик для ГС. Пізніше буде розглянуто застосування цих статистик для висування та перевірки припущень про розподіл ГС.

1. *Емпіричні початкові моменти k-того порядку* $\mathbb{E}^*\xi^k = \frac{1}{n}\sum\limits_{i=1}^{n}\xi_i^k$. Значення на конкретній реалізації вибірки позначається як $(\mathbb{E}^*\xi^k)_{\text{зн}} = \frac{1}{n}\sum\limits_{i=1}^{n}x_i^k$. Окремим випадком емпіричних початкових моментів є *вибіркове середнє* $\overline{\xi} = \frac{1}{n}\sum\limits_{i=1}^{n}\xi_i$, значення якого позначається як $\overline{x} = \frac{1}{n}\sum\limits_{i=1}^{n}x_i$. Після складання варіаційного ряду можна записати значення вибіркового середнього як
$$\overline{x} = \begin{cases} \frac{1}{n}\sum\limits_{i=1}^{r}x_i^* n_i & \text{для ДВР} \\ \frac{1}{n}\sum\limits_{i=1}^{r+1}x_i^* n_i & \text{для ІВР} \end{cases}$$
У випадку ІВР за x_i^* береться довільний представник i-того інтервалу.
2. *Емпіричні центральні моменти k-того порядку* $\mathbb{E}^*\mathring{\xi}^k$ можуть вводитися двома способами: як $\frac{1}{n}\sum\limits_{i=1}^{n}\left(\xi_i - \mathbb{E}\xi\right)^k$, якщо $\mathbb{E}\xi$ відоме, або як $\frac{1}{n}\sum\limits_{i=1}^{n}\left(\xi_i - \overline{\xi}\right)^k$, якщо $\mathbb{E}\xi$ невідоме. Окремим випадком є *вибіркова дисперсія* $\mathbb{D}^*\xi$ при $k = 2$. Її значення позначається як $(\mathbb{D}^*\xi)_{\text{зн}} = \frac{1}{n}\sum\limits_{i=1}^{n}\left(x_i - \mathbb{E}\xi\right)^2$ або $\frac{1}{n}\sum\limits_{i=1}^{n}\left(x_i - \overline{x}\right)^2$.
3. *Вибіркова мода $Mo^*\xi$* у випадку побудови ДВР визначається як варіанта з найбільшою частістю, а у випадку ІВР — за формулою $(Mo^*\xi)_{\text{зн}} = y_i + h_i \cdot \frac{n_i - n_{i-1}}{(n_i - n_{i-1}) + (n_i - n_{i+1})}$, де y_i — початок інтервалу, якому відповідає найбільша частість (модальний інтервал), а h_i — довжина цього інтервалу, а n_i — частоти.
4. *Вибіркова медіана $Me^*\xi$* у випадку побудови ДВР визначається як середня за розташуванням варіанта (або середнє арифметичне таких варіант, якщо їх парна кількість), а у випадку ІВР — за формулою $(Me^*\xi)_{\text{зн}} = y_i + \frac{h_i}{n_i} \cdot \left(\frac{n}{2} - \sum\limits_{k=1}^{i-1} n_k\right)$, де y_i — початок інтервалу, на якому накопичена частість перевищила 0.5 (медіанний інтервал), h_i — його довжина, n_i — частоти.
5. *Вибіркова асиметрія та вибірковий ексцес* визначаються як $As^*\xi = \frac{\frac{1}{n}\sum\limits_{i=1}^{n}\left(\xi_i - \overline{\xi}\right)^3}{(\mathbb{D}^*\xi)^{3/2}}$ та $Ex^*\xi = \frac{\frac{1}{n}\sum\limits_{i=1}^{n}\left(\xi_i - \overline{\xi}\right)^4}{(\mathbb{D}^*\xi)^2} - 3$ з заміною $\overline{\xi}$ на $\mathbb{E}\xi$, якщо воно відоме.

8.1.7 Діаграми розмаху

Розглянемо спосіб візуалізації реалізацій вибірки, який часто застосовується на практиці через свою інформативність.



Для спрощення опису в межах цього пункту називатимемо конкретну реалізацію вибірки просто вибіркою. Горизонтальна вісь такої діаграми відповідає діапазону значень вибірки. Межами прямокутника (через який цю діаграму ще називають *коробковою*) є *1 та 3 квартилі вибірки*, що позначаються Q_1 та Q_3 — це медіани першої та другої половини відсортованої вибірки (які у випадку непарного обсягу розділяються медіаною вибірки). Різниця $IQR = Q_3 - Q_1$ (довжина прямокутника) називається *міжквартильним розмахом.* Відрізок всередині прямокутника позначає положення *медіани*. Вертикальні відрізки за межами прямокутника (які ще називають *вусами*) позначають «майже мінімум і максимум»: зазвичай вони знаходяться в положеннях $Q_1 - 1.5 \cdot IQR$ та $Q_3 + 1.5 \cdot IQR$. Усі значення, що менше «мінімуму» та більше «максимуму», називаються *викидами* та позначаються окремими точками. На практиці для подальшої обробки викиди прибирають з вибірки, оскільки такі значення можуть з'являтися через помилки вимірювань.

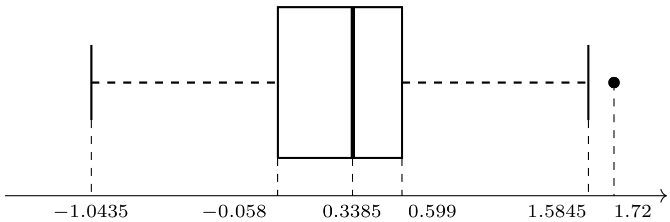
Приклад. Побудувати діаграму розмаху для вибірки ($n = 20$):

$$\begin{array}{cccccccccc} -0.005 & -0.111 & 0.384 & 0.413 & -0.152 & -0.453 & -0.273 & 0.366 & 0.331 & 1.147 \\ 0.425 & 0.166 & 0.512 & 0.789 & 0.686 & -0.205 & 0.175 & 0.092 & 1.72 & 1.43 \end{array}$$

Після сортування за зростанням отримаємо

$$\begin{array}{cccccccccc} -0.453 & -0.273 & -0.205 & -0.152 & -0.111 & -0.005 & 0.092 & 0.166 & 0.175 & 0.331 \\ 0.366 & 0.384 & 0.413 & 0.425 & 0.512 & 0.686 & 0.789 & 1.147 & 1.43 & 1.72 \end{array}$$

Медіана дорівнює $\frac{0.331+0.366}{2} = 0.3385$, а квартилі — $Q_1 = \frac{-0.111-0.005}{2} = -0.058$, $Q_3 = \frac{0.512+0.686}{2} = 0.599$. Міжквартильний розмах $IQR = Q_3 - Q_1 = 0.657$, далі обчислимо $Q_1 - 1.5 \cdot IQR = -1.0435$ та $Q_3 + 1.5 \cdot IQR = 1.5845$. Отже, значення 1.72 є викидом. Нарешті, побудуємо діаграму розмаху:



8.2 Точкові оцінки

Нехай ξ — ГС, а $F_\xi(x, \theta_1, \theta_2, ..., \theta_k)$ — її функція розподілу, де $\theta_1, \theta_2, ..., \theta_k$ — набір параметрів. Тип самої функції розподілу у випадку точкового оцінювання вважається відомим, невідомими є параметри: наприклад, λ в експоненційному законі чи a та σ в нормальному.

Означення 8.2.1. *Точковою оцінкою* θ^* невідомого параметру θ називається деяка статистика $\theta^*(\xi_1, ..., \xi_n)$, значення якої на конкретній реалізації вибірки приймається за наближене значення θ. Іноді вводиться позначення θ^*_n, щоб вказати залежність оцінки від обсягу вибірки.

Зрозуміло, що точкова оцінка θ^*, на відміну від параметру θ, є випадковою величиною, яка залежить від закону розподілу ξ та обсягу вибірки. Безумовно, можна ввести багато функцій від результатів спостережень, які можна брати в якості θ^*. Наприклад, якщо параметр θ є математичним сподіванням ξ, то за оцінку математичного сподівання за результатами спостережень можна взяти середнє арифметичне, моду, медіану, півсуму найбільшого та найменшого значень вибірки тощо. Отже, яку статистику краще обрати? Назвати «найкращою» оцінкою ту, яка найбільш близька до істинного значення оцінюваного параметру, неможливо, оскільки точкова оцінка — випадкова величина. Таким чином, робити висновки про якість оцінки варто не по її конкретним значенням, а по її розподілу. В зв'язку з цим розглянемо вимоги, що висувають до точкових оцінок.

8.2.1 Незміщеність оцінки

Означення 8.2.2. Точкова оцінка θ^* параметру θ називається *незміщеною*, якщо

$$\mathbb{E}\theta^*_n = \theta \text{ для всіх } n \in \mathbb{N} \tag{1}$$

і *асимптотично незміщеною*, якщо $\lim\limits_{n\to\infty} \mathbb{E}\theta^*_n = \theta$.

Приклад. Розглянемо деякі важливі незміщені оцінки.

1. *Вибіркове середнє* — незміщена оцінка математичного сподівання: $\overline{\xi} = \frac{1}{n}\sum\limits_{k=1}^{n}\xi_k$, $\mathbb{E}\overline{\xi} = \frac{1}{n}\sum\limits_{k=1}^{n}\mathbb{E}\xi_k = \frac{1}{n}\cdot n\cdot\mathbb{E}\xi = \mathbb{E}\xi$, оскільки всі ξ_k однаково розподілені.
2. *Вибіркова дисперсія* — незміщена оцінка дисперсії у випадку відомого математичного сподівання $\mathbb{E}\xi$: $\mathbb{D}^*\xi = \frac{1}{n}\sum\limits_{k=1}^{n}(\xi_k - \mathbb{E}\xi)^2$, $\mathbb{E}(\mathbb{D}^*\xi) = \frac{1}{n}\sum\limits_{k=1}^{n}\mathbb{E}(\xi_k - \mathbb{E}\xi)^2 = \mathbb{D}\xi$ знову через однаковий розподіл ξ_k.
3. Дослідимо вибіркову дисперсію, але у випадку невідомого математичного сподівання.

$$\mathbb{D}^*\xi = \frac{1}{n}\sum_{k=1}^{n}\left(\xi_k - \overline{\xi}\right)^2 = \frac{1}{n}\sum_{k=1}^{n}\left((\xi_k - \mathbb{E}\xi) - (\overline{\xi} - \mathbb{E}\xi)\right)^2 =$$
$$= \frac{1}{n}\sum_{k=1}^{n}(\xi_k - \mathbb{E}\xi)^2 - 2\left(\overline{\xi} - \mathbb{E}\xi\right)\cdot\underbrace{\frac{1}{n}\sum_{k=1}^{n}(\xi_k - \mathbb{E}\xi)}_{\overline{\xi}-\mathbb{E}\xi} + \left(\overline{\xi} - \mathbb{E}\xi\right)^2 =$$
$$= \frac{1}{n}\sum_{k=1}^{n}(\xi_k - \mathbb{E}\xi)^2 - \left(\overline{\xi} - \mathbb{E}\xi\right)^2$$

Перший доданок — це формула для вибіркової дисперсії при відомому математичному сподіванні, а $\mathbb{E}\xi = \mathbb{E}\overline{\xi}$, тому

$$\mathbb{E}(\mathbb{D}^*\xi) = \mathbb{D}\xi - \mathbb{E}\left(\overline{\xi} - \mathbb{E}\xi\right)^2 = \mathbb{D}\xi - \mathbb{D}\overline{\xi}$$
$$\mathbb{D}\overline{\xi} = \mathbb{D}\left(\frac{1}{n}\sum_{k=1}^{n}\xi_k\right) = \frac{1}{n^2}\sum_{k=1}^{n}\mathbb{D}\xi_k = \frac{1}{n}\mathbb{D}\xi$$

Дві останні рівності одержані через незалежність та однаковий розподіл ξ_k. Отже, маємо $\mathbb{E}(\mathbb{D}^*\xi) = \left(1 - \frac{1}{n}\right)\mathbb{D}\xi$, тому ця оцінка є лише асимптотично незміщеною. Проте, оцінка $\mathbb{D}^{**}\xi = \frac{n}{n-1}\mathbb{D}^*\xi$ буде незміщеною. Таким чином, якщо $\mathbb{E}\xi$ невідоме, то *виправлена вибіркова дисперсія* $\mathbb{D}^{**}\xi = \frac{1}{n-1}\sum\limits_{k=1}^{n}\left(\xi_k - \overline{\xi}\right)^2$ є незміщеною оцінкою дисперсії.
4. Нехай $\xi \sim \mathrm{U}\langle a; b\rangle$, перевіримо незміщеність $a_n^* = \min\limits_{1\le k\le n}\xi_k$. Знайдемо $\mathbb{E}a_n^*$. Як відомо, $f_{\min}(x) = n\left(1 - F_\xi(x)\right)^{n-1} f_\xi(x) = n\left(1 - \frac{x-a}{b-a}\right)^{n-1}\frac{1}{b-a} = n\cdot\frac{(b-x)^{n-1}}{(b-a)^n}$, якщо $x \in \langle a; b\rangle$, та 0 інакше.

$$\mathbb{E}a_n^* = \frac{n}{(b-a)^n}\int_a^b x(b-x)^{n-1}dx = [b - x = t] = \frac{n}{(b-a)^n}\int_0^{b-a}(b-t)t^{n-1}dt =$$
$$= \frac{n}{(b-a)^n}\left(\frac{bt^n}{n} - \frac{t^{n+1}}{n+1}\right)\Bigg|_0^{b-a} = \frac{n}{(b-a)^n}\left(\frac{b(b-a)^n}{n} - \frac{(b-a)^{n+1}}{n+1}\right) =$$
$$= b - \frac{n(b-a)}{n+1} = \frac{bn + b - nb + na}{n+1} = a\cdot\frac{n}{n+1} + \frac{b}{n+1} \ne a$$

Але $\lim\limits_{n\to\infty}\mathbb{E}a_n^* = a$, тому a_n^* є асимптотично незміщеною оцінкою.

Вправа. Перевірити, що $b_n^* = \max\limits_{1\le k\le n}\xi_k$ є асимптотично незміщеною оцінкою для параметра b у випадку $\xi \sim \mathrm{U}\langle a; b\rangle$. Користуючись вже дослідженими оцінками a_n^* та b_n^*, знайти незміщені оцінки для параметрів a і b (підказка: це будуть деякі лінійні комбінації a_n^* та b_n^*).

8.2.2 Консистентність оцінки

Означення 8.2.3. Точкова оцінка θ^* параметру θ називається *консистентною*, якщо

$$\forall\,\varepsilon > 0 : \lim_{n\to\infty}\mathbb{P}\left\{|\theta_n^* - \theta| \ge \varepsilon\right\} = 0 \Longleftrightarrow \theta_n^* \xrightarrow{\mathrm{P}} \theta, n \to \infty \tag{2}$$

Консистентність оцінки можна перевіряти за означенням. Але для незміщених та асимптотично незміщених оцінок є *достатня умова* консистентності.

Твердження. *Якщо θ_n^* — незміщена чи асимптотично незміщена оцінка та $\mathbb{D}\theta_n^* \to 0$, $n \to \infty$, то ця оцінка є консистентною.*

Доведення. Якщо θ_n^* незміщена, то $\mathbb{P}\left\{|\theta_n^* - \theta| \geq \varepsilon\right\} = \mathbb{P}\left\{|\theta_n^* - \mathbb{E}\theta_n^*| \geq \varepsilon\right\} \leq \frac{\mathbb{D}\theta_n^*}{\varepsilon^2} \to 0, n \to \infty$. Якщо θ_n^* асимптотично незміщена, то $\lim\limits_{n\to\infty} \mathbb{E}\theta_n^* = \theta$, звідки за критерієм збіжності в середньому квадратичному до константи маємо $\theta_n^* \xrightarrow{\text{СК}} \theta \Rightarrow \theta_n^* \xrightarrow{\text{P}} \theta, n \to \infty$. ▲

Зауваження. Також є поняття *сильно консистентної оцінки*, де збіжність за ймовірністю замінюється збіжністю майже напевно. Такі оцінки надалі розглядати не будемо через складність дослідження такої збіжності в загальному випадку.

Приклад. Перевіримо консистентність оцінок, незміщеність чи асимптотичну незміщеність яких вже дослідили.

1. *Вибіркове середнє* $\overline{\xi} = \frac{1}{n}\sum\limits_{k=1}^{n} \xi_k$ є консистентною оцінкою за законом великих чисел (і навіть сильно консистентною).
2. *Вибіркова дисперсія* $\mathbb{D}^*\xi = \frac{1}{n}\sum\limits_{k=1}^{n} \left(\xi_k - \mathbb{E}\xi\right)^2$ є незміщеною оцінкою дисперсії в разі відомого $\mathbb{E}\xi$. $\mathbb{D}\left(\mathbb{D}^*\xi\right) = \frac{1}{n^2}\sum\limits_{k=1}^{n} \mathbb{D}\left(\xi_k - \mathbb{E}\xi\right)^2 = \frac{1}{n}\mathbb{D}\left(\xi - \mathbb{E}\xi\right)^2 \to 0, n \to \infty$, тому ця оцінка консистентна.
3. *Виправлена вибіркова дисперсія* $\mathbb{D}^{**}\xi = \frac{n}{n-1}\mathbb{D}^*\xi$ є незміщеною оцінкою дисперсії в разі невідомого $\mathbb{E}\xi$. $\mathbb{D}\left(\mathbb{D}^{**}\xi\right) = \left(\frac{n}{n-1}\right)^2 \cdot \mathbb{D}\left(\mathbb{D}^*\xi\right) \to 0, n \to \infty$, тому ця оцінка консистентна.
4. Нехай $\xi \sim \mathrm{U}\left\langle a; b\right\rangle$, $a_n^* = \min\limits_{1\leq k\leq n} \xi_k$ — асимптотично незміщена оцінка a. Перевірятимемо консистентність за означенням:

$$\varepsilon > 0,\ \mathbb{P}\left\{|a_n^* - a| < \varepsilon\right\} = \mathbb{P}\left\{a - \varepsilon < a_n^* < a + \varepsilon\right\} = \int_a^{a+\varepsilon} f_{\min}(x)dx =$$
$$= \frac{n}{(b-a)^n}\int_a^{a+\varepsilon} (b-x)^{n-1}dx = -\frac{n}{(b-a)^n} \cdot \left.\frac{(b-x)^n}{n}\right|_a^{a+\varepsilon} =$$
$$= \frac{1}{(b-a)^n} \cdot \left((b-a)^n - (b-a-\varepsilon)^n\right) = 1 - \left(1 - \frac{\varepsilon}{b-a}\right)^n \to 1, n \to \infty$$

   Отже, оцінка є консистентною.

Вправа. Перевірити консистентність оцінки $b_n^* = \max\limits_{1\leq k\leq n} \xi_k$ для $\xi \sim \mathrm{U}\left\langle a; b\right\rangle$.

Зауваження. Консистентність та незміщеність вибіркового середнього, виправленої та звичайної вибіркових дисперсій було встановлено для будь-якого розподілу ξ.

8.2.3 Ефективність оцінки

Основне питання задачі оцінювання параметрів розподілу — наскільки великою є похибка. Введені означення незміщеної та консистентної оцінки показують, відповідно, чи правильні в середньому значення цієї оцінки, та чи покращується точність оцінювання зі збільшенням обсягу вибірки. Зрозуміло, що для оцінювання параметру θ можна запропонувати декілька незміщених оцінок, значення яких за визначенням зосереджені навколо справжнього значення θ. Природно вимагати від «найкращої» такої оцінки найменшої можливої дисперсії.

Означення 8.2.4. Нехай Θ_n — множина усіх незміщених оцінок параметру θ за вибірками фіксованого обсягу n. Оцінка $\theta_{\text{еф}}^*$ називається *ефективною*, якщо

$$\mathbb{D}\theta_{\text{еф}}^* = \inf_{\theta^*\in\Theta_n} \mathbb{D}\theta^* \tag{3}$$

Означення ефективності не надто сприяє дослідженню оцінки. По-перше, не завжди легко обчислити дисперсію оцінки. По-друге, навіть якщо її вдасться обчислити, то немає гарантії, що ця дисперсія буде найменшою серед дисперсій усіх оцінок з Θ_n. Перевіряти ефективність оцінок допомагає нерівність Рао-Крамера.

Теорема (нерівність Рао-Крамера). *Якщо θ^* — незміщена оцінка, то $\mathbb{D}\theta^* \geq \frac{1}{\mathcal{I}(\theta)}$, де $\mathcal{I}(\theta) = \mathbb{E}\left(\frac{\partial \ln \mathcal{L}(\vec{\xi},\theta)}{\partial\theta}\right)^2$ — «кількість інформації за Фішером». Для ефективної оцінки досягається рівність. Тут $\mathcal{L}(\vec{\xi},\theta)$ — функція правдоподібності, що залежить від випадкової вибірки $\vec{\xi}$ та невідомого параметру θ.*

Доведення. Доведемо цю нерівність у випадку неперервної ГС. За припущенням про незміщеність

$$\mathbb{E}\theta^* = \int_{\mathbb{R}^n} \theta^*(\vec{x}) f_{\vec{\xi}}(\vec{x},\theta)d\vec{x} = \theta$$

З іншого боку,

$$\theta = \theta \cdot 1 = \theta \cdot \int_{\mathbb{R}^n} f_{\vec{\xi}}(\vec{x},\theta)d\vec{x}$$

Отже, маємо рівність, яку продиференціюємо по θ:

$$\int_{\mathbb{R}^n} (\theta^*(\vec{x}) - \theta) f_{\vec{\xi}}(\vec{x},\theta)d\vec{x} = 0$$

$$\frac{\partial}{\partial\theta}\left(\int_{\mathbb{R}^n} (\theta^*(\vec{x}) - \theta) f_{\vec{\xi}}(\vec{x},\theta)d\vec{x}\right) = -\underbrace{\int_{\mathbb{R}^n} f_{\vec{\xi}}(\vec{x},\theta)d\vec{x}}_{1} + \int_{\mathbb{R}^n} (\theta^*(\vec{x}) - \theta)\frac{\partial f_{\vec{\xi}}(\vec{x},\theta)}{\partial\theta}d\vec{x} = 0$$

$$\int_{\mathbb{R}^n} (\theta^*(\vec{x}) - \theta)\frac{\partial f_{\vec{\xi}}(\vec{x},\theta)}{\partial\theta}d\vec{x} = \int_{\mathbb{R}^n} (\theta^*(\vec{x}) - \theta)\frac{\partial f_{\vec{\xi}}(\vec{x},\theta)}{\partial\theta}\frac{f_{\vec{\xi}}(\vec{x},\theta)}{f_{\vec{\xi}}(\vec{x},\theta)}d\vec{x} = 1$$

Оскільки $\frac{\partial f_{\vec{\xi}}(\vec{x},\theta)}{\partial\theta}\frac{1}{f_{\vec{\xi}}(\vec{x},\theta)} = \frac{\partial \ln f_{\vec{\xi}}(\vec{x},\theta)}{\partial\theta}$, а у випадку неперервної ГС $\mathcal{L}(\vec{\xi},\theta) = \prod\limits_{k=1}^{n} f_\xi(\xi_k,\theta) = f_{\vec{\xi}}(\vec{\xi},\theta)$, останню рівність можна записати у вигляді $1 = \mathbb{E}\eta_1\eta_2$, де $\eta_1 = \theta^*(\vec{\xi}) - \theta$, $\eta_2 = \frac{\partial \ln \mathcal{L}(\vec{\xi},\theta)}{\partial\theta}$. Оскільки $(\mathbb{E}\eta_1\eta_2)^2 \leq \mathbb{E}\eta_1^2 \cdot \mathbb{E}\eta_2^2$, маємо $1 \leq \mathbb{D}\theta^* \cdot \mathcal{I}(\theta)$, що і треба було довести. ▲

Наслідок. В нерівності $(\mathbb{E}\eta_1\eta_2)^2 \leq \mathbb{E}\eta_1^2 \cdot \mathbb{E}\eta_2^2$ рівність досягається тоді і тільки тоді, коли η_1 та η_2 лінійно залежні. Оскільки для ефективної оцінки досягається рівність, то маємо такий *критерій ефективності незміщеної оцінки*: незміщена оцінка θ^* є оптимальною тоді й тільки тоді, коли

$$\frac{\partial \ln \mathcal{L}(\vec{\xi},\theta)}{\partial\theta} = C(n,\theta) \cdot \left(\theta^*(\vec{\xi}) - \theta\right) \tag{4}$$

Тут $C(n,\theta)$ — деяке значення, що залежить лише від n та θ.

Зауваження. В умові теореми необхідно накладати деякі умови, виконання яких припускалися при доведенні: додатність та диференційовність функції правдоподібності на області визначення, існування (скінченність) $\mathcal{I}(\theta)$ та можливість диференціювання відповідного інтегралу по параметру.

Приклад. Нехай ГС $\xi \sim \text{Poiss}(a)$. $\mathbb{E}\xi = a$, тому $\overline{\xi} = \frac{1}{n}\sum\limits_{k=1}^{n}\xi_k$ — незміщена оцінка a. Перевіримо її ефективність.

$$\ln \mathcal{L}(\vec{x},a) = -na + \ln a \cdot \sum_{k=1}^{n} x_k - \sum_{k=1}^{n} \ln x_k!$$

$$\frac{\partial \ln \mathcal{L}(\vec{x},a)}{\partial a} = -n + \frac{1}{a}\cdot\sum_{k=1}^{n} x_k = -n + \frac{n}{a}\cdot\overline{x} = \frac{n}{a}\cdot(\overline{x} - a)$$

Перейдемо до випадкової вибірки: $\frac{\partial \ln \mathcal{L}(\vec{\xi},a)}{\partial a} = C(n,a)\cdot(a^* - a)$. Отже, ця оцінка є ефективною.

Розглянемо ще одну важливу теорему, що стосується ефективних оцінок.

Теорема (єдиність ефективної незміщеної точкової оцінки). *Якщо $\theta^* = \theta^*(\vec{\xi})$ — ефективна незміщена оцінка, то вона едина.*

Доведення. Нехай θ_1^* — інша ефективна незміщена оцінка, $\mathbb{E}\theta^* = \mathbb{E}\theta_1^* = \theta$, $\mathbb{D}\theta^* = \mathbb{D}\theta_1^* = \sigma^2$. Розглянемо оцінку $\theta_2^* = \frac{\theta^*+\theta_1^*}{2}$, яка теж є незміщеною. $\mathbb{D}\theta_2^* = \frac{1}{4}\left(\mathbb{D}\theta^* + \mathbb{D}\theta_1^* + 2cov(\theta^*, \theta_1^*)\right) = \frac{1}{2}\left(\sigma^2 + cov(\theta^*, \theta_1^*)\right)$. $|cov(\theta^*, \theta_1^*)| \le \sqrt{\mathbb{D}\theta^* \cdot \mathbb{D}\theta_1^*} = \sigma^2$. Отже, $\mathbb{D}\theta_2^* \le \sigma^2$. Але за умовою дисперсія ефективної оцінки θ рівна σ^2, тому $\mathbb{D}\theta_2^* = \sigma^2$. Це означає, що в $cov(\theta^*, \theta_1^*) \le \sigma^2$ досягається рівність, тому θ^* та θ_1^* є лінійно залежними, тобто $\theta_1^* = k \cdot \theta^* + b$. Тепер $\theta = \mathbb{E}\theta_1^* = k \cdot \mathbb{E}\theta^* + b = k \cdot \theta + b$, звідки $b = (1-k)\theta$. $\theta_1^* = k \cdot \theta^* + (1-k)\theta \Leftrightarrow \theta_1^* - \theta = k \cdot (\theta^* - \theta)$. Скориставшись $cov(\theta^*, \theta_1^*) = \sigma^2$, отримаємо $\sigma^2 = \mathbb{E}\left(\theta^* - \theta\right)\left(\theta_1^* - \theta\right) = k \cdot \mathbb{E}\left(\theta^* - \theta\right)^2 = k \cdot \sigma^2$, звідки $k = 1$, тому $b = 0$ і $\theta_1^* = \theta^*$. ▲

Також вводитися поняття *асимптотично ефективної оцінки*. Це означає, що $\mathbb{D}\theta^* \cdot \mathcal{I}(\theta) \to 1$ при $n \to \infty$. Оскільки за нерівністю Рао-Крамера для незміщених оцінок виконується $\mathbb{D}\theta^* \cdot \mathcal{I}(\theta) \ge 1$, то, якщо треба порівняти декілька незміщених оцінок, які не є ефективними, кращою обирається та, у якої добуток $\mathbb{D}\theta^* \cdot \mathcal{I}(\theta)$ менше.

8.2.4 Асимптотична нормальність оцінки

Означення 8.2.5. Оцінка $\theta_n^* = \theta^*(\xi_1, ..., \xi_n)$ називається *асимптотично нормальною*, якщо

$$\frac{\theta_n^* - \theta}{\sqrt{\mathbb{D}\theta_n^*}} \xrightarrow{\text{F}} \eta \sim \mathrm{N}(0,1),\ n \to \infty \tag{5}$$

Це поняття в подальшому дозволить наближувати розподіл точкових оцінок нормальним розподілом. Наприклад, для вибіркового середнього $\overline{\xi} = \frac{1}{n}\sum\limits_{k=1}^{n} \xi_k$ згідно ЦГТ маємо

$$\frac{\overline{\xi} - \mathbb{E}\xi}{\sqrt{\mathbb{D}\overline{\xi}}} = \frac{\frac{1}{n}\sum\limits_{k=1}^{n}(\xi_k - \mathbb{E}\xi)}{\sqrt{\mathbb{D}\xi/n}} = \frac{1}{\sqrt{n}}\sum_{k=1}^{n}\left(\frac{\xi_k - \mathbb{E}\xi}{\sqrt{\mathbb{D}\xi}}\right) \xrightarrow{\text{F}} \eta \sim \mathrm{N}(0,1),\ n \to \infty$$

Таким чином, при великих n можна вважати розподіл $\overline{\xi}$ приблизно рівним $\mathrm{N}\left(\mathbb{E}\xi, \frac{\mathbb{D}\xi}{n}\right)$.

8.3 Методи отримання точкових оцінок

Розглянемо методи, які допомагають знаходити точкові оцінки параметрів розподілу ГС.

8.3.1 Метод моментів

Метод моментів базується на тому, що часто деякі моменти ξ функціонально залежать від параметрів розподілу. В такому випадку, в разі «хорошої» залежності можна виразити параметри розподілу як функції від моментів і отримати точкові оцінки параметрів, замінивши в отриманих виразах моменти на їх вибіркові аналоги.

Приклад. Знайдемо декілька оцінок методом моментів.

1. $\xi \sim \mathrm{U}\langle a; b\rangle$. Відомо, що $\mathbb{E}\xi = \frac{a+b}{2}$, але щоб виразити через моменти обидва параметри, необхідне ще одне рівняння: $\mathbb{D}\xi = \frac{(b-a)^2}{12}$. Отримали систему:

$$\begin{cases} a^* + b^* = 2\overline{\xi} \\ b^* - a^* = 2\sqrt{3\mathbb{D}^{**}\xi} \end{cases} \Rightarrow \begin{cases} b^* = \overline{\xi} + \sqrt{3\mathbb{D}^{**}\xi} \\ a^* = \overline{\xi} - \sqrt{3\mathbb{D}^{**}\xi} \end{cases}$$

2. Нехай тепер $\xi \sim \mathrm{U}\langle -a; a\rangle$. Оскільки $\mathbb{E}\xi = 0$, скористаємося моментом $\mathbb{E}\xi^2 = \frac{a^2}{3}$. В такому випадку отримаємо оцінку $a^* = \sqrt{3\mathbb{E}^*\xi^2}$.
3. $\xi \sim \mathrm{Bin}(N, p)$. Відомо, що $\mathbb{E}\xi = Np$, $\mathbb{D}\xi = Np(1-p)$. В цьому випадку маємо систему

$$\begin{cases} N^* p^* = \overline{\xi} \\ N^* p^* (1 - p^*) = \mathbb{D}^{**}\xi \end{cases} \Rightarrow \begin{cases} N^* = \frac{(\overline{\xi})^2}{\overline{\xi} - \mathbb{D}^{**}\xi} \\ p^* = \frac{\overline{\xi}}{N^*} = \frac{\overline{\xi} - \mathbb{D}^{**}\xi}{\overline{\xi}} \end{cases}$$

Незважаючи на формально правильний результат, застосовувати ці оцінки недоречно: немає жодної гарантії, що N^* прийматиме лише натуральні значення. Але якщо параметр N відомий, то можна користуватися оцінкою $p^* = \frac{\overline{\xi}}{N}$.

4. $\xi \sim \mathrm{Exp}(\lambda)$. Оскільки $\mathbb{E}\xi = \frac{1}{\lambda}$ і $\mathbb{D}\xi = \frac{1}{\lambda^2}$, за допомогою методу моментів можна отримати дві оцінки λ^*: $\lambda_1^* = 1/\overline{\xi}$ та $\lambda_2^* = 1/\sqrt{\mathbb{D}^{**}\xi}$. В такому випадку порівнюють властивості отриманих оцінок та обирають кращу.

Вправа. Перевірити, що оцінка λ_1^* є асимптотично незміщеною. Чи можна зробити з неї незміщену оцінку, і, якщо так, що можна сказати про її ефективність?

Як видно з цих прикладів, розподіл оцінок, отриманих за допомогою методу моментів, може бути досить складним. Це, в свою чергу, ускладнює дослідження таких оцінок. Єдина характеристика, з якою майже не виникає складнощів — консистентність. Оскільки оцінки з методу моментів є функціями від вибіркових моментів, які є консистентними, то за властивостями збіжності за ймовірністю оцінки також є консистентними.

Зауваження. В подальшому будемо позначати оцінки, отримані за допомогою методу моментів, через $\theta^*_{\text{ММ}}$, якщо це матиме значення.

8.3.2 Метод максимальної правдоподібності

Метод максимальної правдоподібності полягає в знаходженні таких оцінок невідомих параметрів розподілу ГС, що максимізують значення функції правдоподібності. Пояснення цього методу таке: якщо провести декілька спостережень однієї ГС, то найчастіше будемо спостерігати ті значення, при яких щільність розподілу близька до свого максимуму (у випадку неперервної ГС) або ж яким відповідає найбільша ймовірність появи (у випадку дискретної ГС). Тому й шукаються такі статистики для оцінки невідомих параметрів, за яких функція правдоподібності набуватиме максимального значення. Тобто, в якості оцінки θ^* береться значення $\theta^*(\xi_1, ..., \xi_n) = \underset{\theta}{\operatorname{argmax}}\, \mathcal{L}(\vec{\xi}, \theta)$, або ж, у випадку двох або більше параметрів, $\underset{\theta_1,...,\theta_k}{\operatorname{argmax}}\, \mathcal{L}(\vec{\xi}, \theta_1, ..., \theta_k)$. Оскільки логарифм — монотонна функція, то $\ln \mathcal{L}(\vec{\xi}, \theta_1, ..., \theta_k)$ досягає максимуму по $\theta_1, ..., \theta_k$ одночасно з $\mathcal{L}(\vec{\xi}, \theta_1, ..., \theta_k)$. Таким чином, в методі максимальної правдоподібності шукають $\underset{\theta_1,...,\theta_k}{\operatorname{argmax}} \ln \mathcal{L}(\vec{\xi}, \theta_1, ..., \theta_k)$.

Надалі для спрощення сприйняття випадкову вибірку в аргументі функції правдоподібності замінимо на реалізацію вибірки. В такому випадку спочатку отримаємо точки максимуму, що залежать від $x_1, x_2, ..., x_n$ (і це фактично буде значенням оцінки), а саму оцінку, як функцію від випадкової вибірки, отримаємо зворотною заміною x_k на ξ_k.

Приклад. Знайдемо декілька оцінок методом максимальної правдоподібності.

1. $\xi \sim \mathrm{N}(a, \sigma^2)$. Відомо, що $\ln \mathcal{L}(\vec{x}, a, \sigma^2) = -\frac{n}{2} \ln 2\pi - \frac{n}{2} \ln \sigma^2 - \frac{1}{2\sigma^2} \sum\limits_{k=1}^{n} (x_k - a)^2$. Позначимо для зручності $s = \sigma^2$ та шукатимемо максимум функції двох змінних:

$$\begin{cases} \frac{\partial \ln \mathcal{L}}{\partial a} = \frac{1}{s} \sum\limits_{k=1}^{n} (x_k - a) = \frac{n}{s}(\overline{x} - a) = 0 \\ \frac{\partial \ln \mathcal{L}}{\partial s} = -\frac{n}{2s} + \frac{1}{2s^2} \sum\limits_{k=1}^{n} (x_k - a)^2 = 0 \end{cases} \Rightarrow \begin{cases} a_{\text{кр.}} = \overline{x} \\ s_{\text{кр.}} = \frac{1}{n} \sum\limits_{k=1}^{n} (x_k - a)^2 = (\mathbb{D}^*\xi)_{\text{зн}} \end{cases}$$

$$\begin{pmatrix} \frac{\partial^2 \ln \mathcal{L}}{\partial a^2} & \frac{\partial^2 \ln \mathcal{L}}{\partial a \partial s} \\ \frac{\partial^2 \ln \mathcal{L}}{\partial a \partial s} & \frac{\partial^2 \ln \mathcal{L}}{\partial s^2} \end{pmatrix} = \begin{pmatrix} -\frac{n}{s} & -\frac{n}{s^2}(\overline{x} - a) \\ -\frac{n}{s^2}(\overline{x} - a) & \frac{n}{2s^2} - \frac{1}{s^3} \sum\limits_{k=1}^{n} (x_k - a)^2 \end{pmatrix}$$

При $a = a_{\text{кр.}}$ та $s = s_{\text{кр.}}$ матриця других похідних дорівнює

$$\begin{pmatrix} -\frac{n}{s_{\text{кр.}}} & 0 \\ 0 & \frac{n}{(s_{\text{кр.}})^3} \left(\frac{s_{\text{кр.}}}{2} - \frac{1}{n} \sum\limits_{k=1}^{n} (x_k - a^*)^2 \right) \end{pmatrix} = \begin{pmatrix} -\frac{n}{s_{\text{кр.}}} & 0 \\ 0 & -\frac{n s_{\text{кр.}}}{2(s_{\text{кр.}})^3} \end{pmatrix} = \begin{pmatrix} -\frac{n}{s_{\text{кр.}}} & 0 \\ 0 & -\frac{n}{2(s_{\text{кр.}})^2} \end{pmatrix}$$

Ця матриця є від'ємно визначеною, тому $\ln \mathcal{L}(\vec{x}, a, \sigma^2)$ дійсно досягає максимуму при $a = \overline{x}$ та $\sigma^2 = (\mathbb{D}^*\xi)_{\text{зн}}$. Отже, отримали оцінки $a^* = \overline{\xi}$ та $(\sigma^2)^* = \mathbb{D}^*\xi$.

2. $\xi \sim \mathrm{Exp}(\lambda, b)$, вважатимемо λ відомим. $\ln \mathcal{L}(\vec{x}, b) = n \ln \lambda - \lambda \left(\sum\limits_{k=1}^{n} x_k - nb \right)$, але сама функція правдоподібності приймає нульові значення при $x_k < b$. Якщо просто продиференціювати функцію правдоподібності, отримаємо $\frac{\partial \ln \mathcal{L}(\vec{x}, b)}{\partial b} = \lambda n$ — цей вираз завжди

більше 0 та не залежить від вибірки. Отже, $\ln \mathcal{L}(\vec{x}, b)$ монотонно зростає відносно b. За змістом самого параметру $x_k \geq b$ для всіх $k = 1, ..., n$. Таким чином, максимальне можливе значення b — це $\min\limits_{1\leq k\leq n} x_k$. Отже, $b^* = \min\limits_{1\leq k\leq n} \xi_k$.

3. $\xi \sim \mathrm{U}[\theta; \theta+1]$, $\theta \in \mathbb{R}$. Спочатку запишемо щільність: $f_\xi(x) = \begin{cases} 1, & x \in [\theta; \theta+1] \\ 0, & x \notin [\theta; \theta+1] \end{cases}$. Отже, функція правдоподібності має вигляд $\mathcal{L}(\vec{x}, \theta) = \begin{cases} 1, & x_k \in [\theta; \theta+1], k = 1, ..., n \\ 0, & \text{інакше} \end{cases} =$ $\begin{cases} 1, & \theta \leq \min\limits_{1\leq k\leq n} x_k \leq \max\limits_{1\leq k\leq n} x_k \leq \theta + 1 \\ 0, & \text{інакше} \end{cases} = \begin{cases} 1, & \max\limits_{1\leq k\leq n} x_k - 1 \leq \theta \leq \min\limits_{1\leq k\leq n} x_k \\ 0, & \text{інакше} \end{cases}$. Вона досягає максимуму по θ в кожній точці з відрізка $\left[\max\limits_{1\leq k\leq n} x_k - 1; \min\limits_{1\leq k\leq n} x_k\right]$. Таким чином, отримуємо незліченну множину оцінок $\theta^*_\alpha = (1-\alpha)\cdot\left(\max\limits_{1\leq k\leq n} \xi_k - 1\right) + \alpha \cdot \min\limits_{1\leq k\leq n} \xi_k$, де $\alpha \in [0; 1]$.

Вправа. 1. Знайти значення α, за якого θ^*_α з останнього прикладу буде незміщеною.

2. Показати, що оцінки $a^* = \min\limits_{1\leq k\leq n} \xi_k$ та $b^* = \max\limits_{1\leq k\leq n} \xi_k$ для $\xi \sim \mathrm{U}\langle a; b\rangle$ можна знайти за допомогою методу максимальної правдоподібності.

З розглянутих прикладів видно, що на відміну від методу моментів, метод максимальної правдоподібності може дати оцінки не тільки у вигляді функцій від емпіричних моментів, а кількість оцінок може бути навіть незліченною. Іноді застосовування методу максимальної правдоподібності є досить складним: наприклад, якщо розглядати $\xi \sim \mathrm{Bin}(N, p)$ з обома невідомими параметрами, то треба буде шукати максимум функції правдоподібності за параметром N, що приймає лише скінченну кількість значень та входить в саму функцію у складі біноміальних коефіцієнтів.

Зауваження. В подальшому будемо позначати оцінки, отримані за допомогою методу максимальної правдоподібності, через $\theta^*_{\text{ММП}}$, якщо це матиме значення.

8.4 Інтервальне оцінювання

Нехай ξ — ГС, θ — якийсь параметр її розподілу. Задача *інтервального оцінювання* θ — це пошук за заданим *рівнем надійності* γ статистик $\theta^*_1(\vec{\xi})$, $\theta^*_2(\vec{\xi})$ таких, що:

$$\mathbb{P}\left\{\theta \in (\theta^*_1, \theta^*_2)\right\} = \gamma \Leftrightarrow \mathbb{P}\left\{\theta^*_1 < \theta < \theta^*_2\right\} = \gamma$$

Іноді в цих співвідношеннях пишуть $\geq \gamma$, оскільки у випадку дискретного розподілу цих статистик рівності можуть не мати сенсу.

Означення 8.4.1. (θ^*_1, θ^*_2) називається *довірчим інтервалом з рівнем надійності* γ. Більш точно — маємо справу з послідовністю таких інтервалів, що залежать від обсягу вибірки.

Зауваження. Оскільки межі інтервалу є випадковими, то запис $\theta \in (\theta^*_1, \theta^*_2)$ правильно читати не як «θ потрапляє в інтервал (θ^*_1, θ^*_2)», а як «інтервал (θ^*_1, θ^*_2) накриває θ».

Зручно шукати довірчі інтервали конкретного вигляду, наприклад, *симетричні* відносно θ з умови $\mathbb{P}\left\{|\theta - \theta^*| < \varepsilon\right\} = \gamma$, де ε — *точність* довірчого інтервалу, або *однобічні* з умов $\mathbb{P}\{\theta > \theta^*\} = \gamma$ чи $\mathbb{P}\{\theta < \theta^*\} = \gamma$. У випадку симетричного довірчого інтервалу зазвичай одразу обирається сама статистика θ^* (наприклад, якась «хороша» точкова оцінка). В такому разі пошук ε — це пошук ширини довірчого інтервалу, що забезпечує заданий рівень надійності.

Якщо розглядається конкретна реалізація вибірки, то можна обчислити межі знайденого інтервалу як значення відповідних статистик.

8.4.1 Побудова довірчих інтервалів для гауссівської ГС

Нехай ГС $\xi \sim \mathrm{N}(a, \sigma^2)$. Розглянемо чотири випадки побудови довірчих інтервалів для параметрів a та σ^2 (нагадаємо, що це математичне сподівання та дисперсія).

Довірчий інтервал для математичного сподівання при відомій дисперсії.

Шукатимемо симетричний довірчий інтервал $\mathbb{P}\left\{|a-a^*|<\varepsilon\right\}=\gamma$, де в якості оцінки a^* буде вибіркове середнє $\overline{\xi}$. З властивостей незалежних гауссівських ВВ маємо $\sum\limits_{k=1}^{n}\xi_k\sim \mathrm{N}(na,n\sigma^2)$, тому $\overline{\xi}\sim \mathrm{N}\left(a,\frac{\sigma^2}{n}\right)$. Отже, $\mathbb{P}\{|a-a^*|<\varepsilon\}=\mathbb{P}\{a-\varepsilon<\overline{\xi}<a+\varepsilon\}=2\Phi\left(\frac{\varepsilon\sqrt{n}}{\sigma}\right)=\gamma$, де $\Phi(x)$ — функція Лапласа. З таблиці її значень знайдемо відповідне значення ε і отримаємо шуканий довірчий інтервал $\left(\overline{\xi}-\varepsilon,\overline{\xi}+\varepsilon\right)$.

Довірчий інтервал для математичного сподівання при невідомій дисперсії.

Шукатимемо симетричний довірчий інтервал $\mathbb{P}\left\{|a-a^*|<\varepsilon\right\}=\gamma$. В умовах попереднього прикладу $\sqrt{n}\cdot\frac{\overline{\xi}-a}{\sigma}\sim\mathrm{N}(0,1)$, але σ тепер невідоме. Розглянемо статистику $\mathbb{D}^*\xi=\frac{1}{n}\sum\limits_{k=1}^{n}(\xi_k-\overline{\xi})^2=\frac{1}{n}\sum\limits_{k=1}^{n}(\xi_k-a)^2-(\overline{\xi}-a)^2$. $\frac{n\mathbb{D}^*\xi}{\sigma^2}=\sum\limits_{k=1}^{n}\left(\frac{\xi_k-a}{\sigma}\right)^2-\left(\sqrt{n}\cdot\frac{\overline{\xi}-a}{\sigma}\right)^2$. В теоремі нижче буде доведено незалежність доданків, що дасть право сказати, що $\frac{n\mathbb{D}^*\xi}{\sigma^2}\sim\chi^2_{n-1}$. Як наслідок,

$$\frac{\sqrt{n}\cdot\frac{\overline{\xi}-a}{\sigma}}{\sqrt{\frac{1}{n-1}\cdot\frac{n\mathbb{D}^*\xi}{\sigma^2}}}=\frac{\sqrt{n}\cdot(\overline{\xi}-a)}{\sqrt{\mathbb{D}^{**}\xi}}\sim\mathrm{St}_{n-1}$$

Тепер з рівності $\mathbb{P}\left\{\frac{\sqrt{n}\cdot|\overline{\xi}-a|}{\sqrt{\mathbb{D}^{**}\xi}}<t_\gamma\right\}=\gamma$ знайдемо значення t_γ. Шуканим значенням ε буде $\varepsilon=\frac{t_\gamma}{\sqrt{n}}\cdot\sqrt{(\mathbb{D}^{**}\xi)_{\text{зн}}}$.

Теорема (теорема Фішера)**.** *Статистики стандартної гауссівської ГС $\overline{\xi}$ та $\mathbb{D}^*\xi$ — незалежні випадкові величини.*

Доведення. $\vec{\xi}$ — випадкова вибірка, $\vec{\xi}\sim\mathrm{N}(\vec{0},I)$. Нехай C — деяка ортогональна матриця, тоді $\vec{\eta}=C\vec{\xi}$ теж має розподіл $\mathrm{N}(\vec{0},I)$, причому $\|\vec{\eta}\|=\|\vec{\xi}\|$. Розглянемо матрицю

$$C=\begin{pmatrix} c_{1,1} & c_{1,2} & \dots & c_{1,n}\\ c_{2,1} & c_{2,2} & \dots & c_{2,n}\\ \vdots & \vdots & \ddots & \vdots\\ c_{n-1,1} & c_{n-1,2} & \dots & c_{n-1,n}\\ 1/\sqrt{n} & 1/\sqrt{n} & \dots & 1/\sqrt{n}\end{pmatrix}$$

перші $n-1$ рядків якої — це елементи ортонормованого базису л.о. $\left\{\begin{pmatrix}\frac{1}{\sqrt{n}} & \cdots & \frac{1}{\sqrt{n}}\end{pmatrix}^T\right\}^\perp$. $\vec{\eta}=C\vec{\xi}$, $\eta_n=\frac{1}{\sqrt{n}}\left(\xi_1+\xi_2+...+\xi_n\right)=\sqrt{n}\cdot\overline{\xi}$. Вище у прикладі побудови довірчого інтервалу було показано, що $\mathbb{D}^*\xi=\frac{1}{n}\sum\limits_{k=1}^{n}(\xi_k-a)^2-(\overline{\xi}-a)^2$. В умовах теореми ця рівність спрощується до $\mathbb{D}^*\xi=\frac{1}{n}\sum\limits_{k=1}^{n}\xi_k^2-\overline{\xi}^2=\frac{1}{n}\|\vec{\xi}\|^2-\frac{1}{n}\eta_n^2=\frac{1}{n}\|\vec{\eta}\|^2-\frac{1}{n}\eta_n^2=\frac{1}{n}\sum\limits_{k=1}^{n}\eta_k^2-\frac{1}{n}\eta_n^2=\frac{1}{n}\sum\limits_{k=1}^{n-1}\eta_k^2$. Таким чином, $\mathbb{D}^*\xi$ залежить від перших $n-1$ координат $\vec{\eta}$, а отже — не залежить від $\overline{\xi}=\frac{1}{\sqrt{n}}\eta_n$. ▲

Приклад. Побудувати 95% довірчий інтервал для математичного сподівання гауссівської ГС, якщо $\overline{x}=2$, $n=25$, а дисперсія відома і рівна 6, а потім — якщо невідома і $(\mathbb{D}^{**}\xi)_{\text{зн}}=5.78$.

1. $\mathbb{D}\xi=6$. За умовою $\gamma=0.95$, тому ε шукаємо з $2\Phi\left(\frac{\varepsilon\sqrt{n}}{\sigma}\right)=\gamma$. $\Phi\left(\varepsilon\cdot\frac{5}{\sqrt{6}}\right)=0.475$, звідки $\varepsilon=\frac{\sqrt{6}}{5}\cdot1.96\approx0.96$. Отже, шуканий довірчий інтервал — $(1.04,2.96)$.
2. $(\mathbb{D}^{**}\xi)_{\text{зн}}=5.78$. Спочатку знайдемо $t_\gamma=2.064$, звідки $\varepsilon=\frac{2.064}{\sqrt{25}}\cdot\sqrt{5.78}\approx2.386$. Отже, шуканий довірчий інтервал — $(-0.386,4.386)$.

Довірчий інтервал для дисперсії при відомому математичному сподіванні.

В якості точкової оцінки дисперсії візьмемо $\mathbb{D}^*\xi=\frac{1}{n}\sum\limits_{k=1}^{n}(\xi_k-a)^2$, тоді $\frac{n\mathbb{D}^*\xi}{\sigma^2}=\sum\limits_{k=1}^{n}\left(\frac{\xi_k-a}{\sigma}\right)^2\sim\chi^2_n$. Шукатимемо довірчий інтервал з умови $\mathbb{P}\left\{t_1<\frac{n\mathbb{D}^*\xi}{\sigma^2}<t_2\right\}=\gamma$, де t_1 та t_2 задовольняють $\mathbb{P}\left\{\frac{n\mathbb{D}^*\xi}{\sigma^2}>t_1\right\}=\frac{1+\gamma}{2}$ та $\mathbb{P}\left\{\frac{n\mathbb{D}^*\xi}{\sigma^2}\geq t_2\right\}=\frac{1-\gamma}{2}$. Шуканим довірчим інтервалом буде $\left(\frac{n}{t_2}(\mathbb{D}^*\xi)_{\text{зн}},\frac{n}{t_1}(\mathbb{D}^*\xi)_{\text{зн}}\right)$.

Довірчий інтервал для дисперсії при невідомому математичному сподіванні.

Як було показано раніше, для $\mathbb{D}^*\xi = \frac{1}{n}\sum\limits_{k=1}^{n}(\xi_k - \overline{\xi})^2$ статистика $\frac{n\mathbb{D}^*\xi}{\sigma^2}$ має розподіл χ^2_{n-1}, тому $\frac{(n-1)\mathbb{D}^{**}\xi}{\sigma^2} \sim \chi^2_{n-1}$. Як і в попередньому випадку, шукаємо t_1 та t_2 з умов $\mathbb{P}\left\{\frac{(n-1)\mathbb{D}^{**}\xi}{\sigma^2} > t_1\right\} = \frac{1+\gamma}{2}$ та $\mathbb{P}\left\{\frac{(n-1)\mathbb{D}^{**}\xi}{\sigma^2} \geq t_2\right\} = \frac{1-\gamma}{2}$. Шуканий довірчий інтервал — $\left(\frac{n-1}{t_2}(\mathbb{D}^{**}\xi)_{\text{зн}}, \frac{n-1}{t_1}(\mathbb{D}^{**}\xi)_{\text{зн}}\right)$.

8.4.2 Наближені довірчі інтервали

Якщо статистика θ^*, що оцінює невідомий параметр, є асимптотично нормальною, то розподіл $\frac{\theta^*-\theta}{\sqrt{\mathbb{D}\theta^*}}$ можна вважати приблизно рівним $\mathrm{N}(0,1)$ (звісно, при достатньо великих n). В такому випадку довірчий інтервал можна будувати з рівності $\mathbb{P}\left\{\frac{|\theta^*-\theta|}{\sqrt{\mathbb{D}\theta^*}} < t_\gamma\right\} = \gamma$, де t_γ знаходимо з таблиці значень функції Лапласа. Щоб тепер отримати довірчий інтервал, треба розв'язати відносно θ нерівність $\frac{|\theta^*-\theta|}{\sqrt{\mathbb{D}\theta^*}} < t_\gamma$.

Приклад. Побудуємо наближений довірчий інтервал для a у випадку $\xi \sim \mathrm{Poiss}(a)$. Вважатимемо, що значення t_γ вже знайдено з рівності $\mathbb{P}\left\{\frac{|a^*-a|}{\sqrt{\mathbb{D}a^*}} < t_\gamma\right\} = \gamma$, де $a^* = \overline{\xi}$. Розв'яжемо нерівність:

$$\frac{|a^*-a|}{\sqrt{\mathbb{D}a^*}} < t_\gamma \Leftrightarrow \frac{(a^*-a)^2}{\mathbb{D}a^*} < t_\gamma^2,\ \mathbb{D}a^* = \mathbb{D}\overline{\xi} = \frac{1}{n}\mathbb{D}\xi = \frac{1}{n}a$$

$$(a^*-a)^2 < a\cdot\frac{t_\gamma^2}{n} \Leftrightarrow (a^*)^2 - 2a^*a + a^2 < a\cdot\frac{t_\gamma^2}{n} \Leftrightarrow a^2 - 2a\left(a^* + \frac{t_\gamma^2}{2n}\right) + (a^*)^2 < 0$$

Позначимо $a^* + \frac{t_\gamma^2}{2n} = b$:

$$a^2 - 2ab + (a^*)^2 < 0,\ D/4 = b^2 - (a^*)^2 \Rightarrow a \in \left(b - \sqrt{D/4}, b + \sqrt{D/4}\right)$$

$$b \pm \sqrt{D/4} = a^* + \frac{t_\gamma^2}{2n} \pm \sqrt{\left(a^* + \frac{t_\gamma^2}{2n}\right)^2 - (a^*)^2} = a^* + \frac{t_\gamma^2}{2n} \pm \sqrt{a^*\cdot\frac{t_\gamma^2}{n} + \frac{t_\gamma^4}{4n^2}}$$

Отже, шуканий довірчий інтервал:

$$\left(\overline{\xi} + \frac{t_\gamma^2}{2n} - \sqrt{\overline{\xi}\cdot\frac{t_\gamma^2}{n} + \frac{t_\gamma^4}{4n^2}}, \overline{\xi} + \frac{t_\gamma^2}{2n} + \sqrt{\overline{\xi}\cdot\frac{t_\gamma^2}{n} + \frac{t_\gamma^4}{4n^2}}\right)$$

8.4.3 Довірчий інтервал для ймовірності появи події

За точкову оцінку ймовірності p появи події A в схемі Бернуллі беруть частість $p^* = \frac{m}{n}$, де n — загальна кількість незалежних випробувань, m — кількість появ події A в цих випробуваннях. Задамо рівень надійності γ і знайдемо такі величини p_1 та p_2, щоб виконувалось співвідношення $\mathbb{P}\{p_1 < p < p_2\} = \gamma$. Інтервал (p_1, p_2) буде шуканим довірчим інтервалом. Розглянемо два випадки.

Кількість випробувань досить велика.

В цьому випадку розподіл величини m в силу граничної теореми Муавра-Лапласа можна наближено замінити $\mathrm{N}(np, \sqrt{npq})$, тому розподіл $p^* = \frac{m}{n}$ приблизно рівний $\mathrm{N}\left(p, \sqrt{\frac{pq}{n}}\right)$. Таким чином, статистика $\frac{(p^*-p)\sqrt{n}}{\sqrt{pq}}$ має наближено розподіл $\mathrm{N}(0,1)$. Користуючись таблицею значень функції Лапласа, для заданої довірчої ймовірності γ знайдемо таке t_γ, при якому $\mathbb{P}\left\{\frac{|p^*-p|\sqrt{n}}{\sqrt{pq}} < t_\gamma\right\} = \gamma$. Розв'яжемо відносно p нерівність під знаком ймовірності:

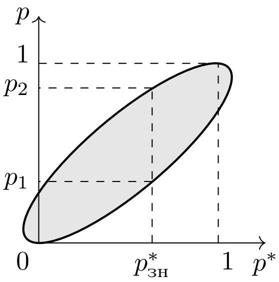
$$\frac{|p^*-p|\sqrt{n}}{\sqrt{pq}} < t_\gamma \Leftrightarrow \frac{(p^*-p)^2 n}{p(1-p)} < t_\gamma^2 \Leftrightarrow (p^*)^2 - 2p^*p + p^2 < \left(p - p^2\right)\cdot\frac{t_\gamma^2}{n} \Leftrightarrow$$

$$\Leftrightarrow \left(1 + \frac{t_\gamma^2}{n}\right)p^2 - \left(2p^* + \frac{t_\gamma^2}{n}\right)p + (p^*)^2 < 0 \Leftrightarrow$$

$$\Leftrightarrow p \in (p_1, p_2),\ p_{1,2} = \frac{\left(p^* + \frac{t_\gamma^2}{2n}\right) \pm t_\gamma\cdot\sqrt{\frac{p^*(1-p^*)}{n} + \frac{t_\gamma^2}{4n^2}}}{1 + \frac{t_\gamma^2}{n}}$$

Цей результат має геометричну інтерпретацію. Розглянемо систему координат, по осі абсцис якої відкладаємо частість p^*, а по осі ординат — ймовірність p. Повернемося до нерівності, яку вже розв'язали:

$$\left(1+\frac{t_\gamma^2}{n}\right)p^2 - 2p^*p + (p^*)^2 - \frac{t_\gamma^2}{n}p < 0$$



Ця нерівність задає внутрішню частину деякого еліпса. Таким чином, довірчим інтервалом для p при відомому значенні частості $p^*_{\text{зн}}$ буде множина точок всередині цього еліпса з абсцисою, що дорівнює $p^*_{\text{зн}}$. Важливо зауважити, що навіть якщо $p^*_{\text{зн}} = 1$, то немає підстав казати, що справжнє значення p дорівнює 1. В цьому випадку отримаємо довірчий інтервал $\left(\frac{1}{1+t_\gamma^2/n}, 1\right)$. Аналогічно, при $p^*_{\text{зн}} = 0$ отримаємо довірчий інтервал $\left(0, \frac{t_\gamma^2/n}{1+t_\gamma^2/n}\right)$.

Якщо обсяг вибірки n доволі великий, то величиною $\frac{t_\gamma^2}{n}$ можна знехтувати. Тоді межі довірчого інтервалу набувають наближених значень

$$p_{1,2} \approx p^* \pm t_\gamma\sqrt{\frac{p^*(1-p^*)}{n}}$$

Приклад. Деяка подія в серії з $n = 100$ незалежних випробувань відбулась $m = 78$ разів. Побудувати довірчий інтервал для ймовірності p появи цієї події з надійністю $\gamma = 0.9$.

За умовою $p^*_{\text{зн}} = 0.78$, а відповідне значення t_γ з таблиці функції Лапласа дорівнює 1.65. Отже,

$$p_{1,2} = \frac{0.78 + \frac{1.65^2}{200} \pm 1.65\sqrt{\frac{0.78\cdot 0.22}{100} + \frac{1.65^2}{4\cdot 100^2}}}{1+\frac{1.65^2}{100}} \Rightarrow p_1 \approx 0.7047,\ p_2 \approx 0.8404$$

Якщо обчислити межі довірчого інтервалу за наближеними формулами для великих n, отримаємо

$$p_{1,2} \approx 0.78 \pm 1.65\sqrt{\frac{0.78\cdot 0.22}{100}} \Rightarrow p_1 \approx 0.7116,\ p_2 \approx 0.8483$$

Кількість випробувань мала. В цьому випадку граничними теоремами скористатися не вийде. Згадаємо формулу для ймовірності появи події A в схемі Бернуллі з n іспитами k разів: $\mathbb{P}\{\xi = k\} = C_n^k p^k (1-p)^{n-k}$, де $\xi \sim \text{Bin}(n,p)$. Задамо рівень надійності γ та знайдемо такі p_1 та p_2, що $\mathbb{P}\{p_1 < p < p_2\} = \gamma$. Приймемо без доведення, що p_1 — розв'язок рівняння $\sum\limits_{k=0}^{m-1} C_n^k p_1^k (1-p_1)^{n-k} = \frac{1+\gamma}{2}$, а p_2 — розв'язок рівняння $\sum\limits_{k=0}^{m} C_n^k p_2^k (1-p_2)^{n-k} = \frac{1-\gamma}{2}$. В цих формулах m є конкретним числом (а не випадковою величиною), кількістю випробувань, в яких сталася подія A. Існують спеціальні таблиці для знаходження значень p_1 та p_2, що задовольняють цим рівнянням.

8.4.4 Довірчі інтервали для параметрів рівномірної ГС

Як відомо, асимптотично незміщеними та консистентними оцінками параметрів a та b рівномірної ГС є, відповідно, $a^* = \min\limits_{1\le k\le n} \xi_k$ та $b^* = \max\limits_{1\le k\le n} \xi_k$. З огляду на характер цих параметрів, шукатимемо довірчі інтервали з рівностей $\mathbb{P}\{a \le a^* \le a+\varepsilon_1\} \ge \gamma$ та $\mathbb{P}\{b-\varepsilon_2 \le b^* \le b\} \ge \gamma$, де γ — заданий рівень надійності. Користуючись щільностями розподілу a^* та b^*, обчислимо ці ймовірності:

$$\mathbb{P}\{a \le a^* \le a+\varepsilon_1\} = \int_a^{a+\varepsilon_1} \frac{n}{(b-a)^n}(b-x)^{n-1}dx = 1 - \left(1-\frac{\varepsilon_1}{b-a}\right)^n$$

$$\mathbb{P}\{b-\varepsilon_2 \le b^* \le b\} = \int_{b-\varepsilon_2}^{b} \frac{n}{(b-a)^n}(x-a)^{n-1}dx = 1 - \left(1-\frac{\varepsilon_2}{b-a}\right)^n$$

Для різниці $b-a$ з умов $a^* \le a+\varepsilon_1$ та $b^* \ge b-\varepsilon_2$ маємо оцінку $b-a \le \max\limits_{1\le k\le n} \xi_k+\varepsilon_2-\min\limits_{1\le k\le n} \xi_k+\varepsilon_1$. Далі для зручності позначатимемо $\max\limits_{1\le k\le n} \xi_k = M$, $\min\limits_{1\le k\le n} \xi_k = m$. Отримуємо систему:

$$\begin{cases} 1-\left(1-\frac{\varepsilon_1}{b-a}\right)^n \ge 1-\left(1-\frac{\varepsilon_1}{M-m+\varepsilon_2+\varepsilon_1}\right)^n = \gamma \\ 1-\left(1-\frac{\varepsilon_2}{b-a}\right)^n \ge 1-\left(1-\frac{\varepsilon_2}{M-m+\varepsilon_2+\varepsilon_1}\right)^n = \gamma \end{cases}$$

Розв'яжемо її відносно ε_1 та ε_2.

$$\begin{cases} \left(1-\frac{\varepsilon_1}{M-m+\varepsilon_2+\varepsilon_1}\right)^n = 1-\gamma \\ \left(1-\frac{\varepsilon_2}{M-m+\varepsilon_2+\varepsilon_1}\right)^n = 1-\gamma \end{cases} \Leftrightarrow \begin{cases} 1-\frac{\varepsilon_1}{M-m+\varepsilon_2+\varepsilon_1} = \sqrt[n]{1-\gamma} \\ 1-\frac{\varepsilon_2}{M-m+\varepsilon_2+\varepsilon_1} = \sqrt[n]{1-\gamma} \end{cases} \Leftrightarrow$$

$$\Leftrightarrow \begin{cases} \varepsilon_1 = (M-m+\varepsilon_2+\varepsilon_1)\left(1-\sqrt[n]{1-\gamma}\right) \\ \varepsilon_2 = (M-m+\varepsilon_2+\varepsilon_1)\left(1-\sqrt[n]{1-\gamma}\right) \end{cases}$$

Отже, $\varepsilon_1 = \varepsilon_2$, тому далі розв'язуємо лише одне рівняння:

$$\varepsilon_1 = (M-m+2\varepsilon_1)\left(1-\sqrt[n]{1-\gamma}\right) \Leftrightarrow \left(2\sqrt[n]{1-\gamma}-2+1\right)\varepsilon_1 = (M-m)\left(1-\sqrt[n]{1-\gamma}\right)$$

Таким чином, $\varepsilon_1 = \varepsilon_2 = \frac{(M-m)\left(1-\sqrt[n]{1-\gamma}\right)}{2\sqrt[n]{1-\gamma}-1}$. Оскільки $\varepsilon_1, \varepsilon_2 > 0$, то на γ треба накладати вимогу $\gamma < 1-2^{-n}$. Шукані довірчі інтервали матимуть вигляд $(a^*-\varepsilon_1, a^*)$ та $(b^*, b^*+\varepsilon_2)$.

Розділ 9

Перевірка статистичних гіпотез

9.1 Поняття статистичних гіпотез та їх перевірки

9.1.1 Статистичні гіпотези. Помилки першого та другого роду

Інформація, що отримана при обробці вибірки з деякої генеральної сукупності, може бути використана для отримання висновків про всю генеральну сукупність. Подібні висновки називають *статистичними*. Завдяки їх ймовірнісному характеру завжди можна знайти ймовірність того, що прийняте рішення буде помилковим, тобто оцінити ризик того чи іншого прийнятого рішення.

Означення 9.1.1. *Статистичною непараметричною гіпотезою* H називається припущення про вигляд розподілу генеральної сукупності, яке перевіряється за вибіркою. Часто розподіл генеральної сукупності відомий і за вибіркою треба перевірити припущення щодо значень параметрів цього розподілу. Такі статистичні гіпотези називають *параметричними*.

Гіпотези бувають прості та складні. Гіпотеза називається *простою*, якщо вона однозначно визначає розподіл генеральної сукупності, в іншому випадку гіпотеза називається *складною*. Наприклад, простою гіпотезою є припущення, що ГС має нормальний розподіл з параметрами 0 та 1. Якщо висувається гіпотеза про те, що ГС має нормальний розподіл з параметрами 1 та σ, де $\sigma \in (1, 2)$, то ця гіпотеза є складною. Теж саме стосується гіпотез про значення невідомого параметру розподілу ГС: гіпотеза виду $H : \theta = 1$ є простою, а $H : \theta \in (1, 2)$ — складною.

Означення 9.1.2. Гіпотеза, що перевіряється, називається *нульовою (основною) гіпотезою* й позначається H_0. Решта гіпотез називається *альтернативними (конкуруючими)* відносно нульової гіпотези й позначаються H_1, H_2 тощо.

Наприклад, якщо перевіряється проста гіпотеза про рівність параметра θ деякому значенню θ_0, тобто, $H_0 : \theta = \theta_0$, то в якості альтернативної гіпотези можна розглядати одну з таких: $H_1 : \theta \neq \theta_0$, $H_2 : \theta > \theta_0$, $H_3 : \theta < \theta_0$, $H_4 : \theta = \theta_1$ (де $\theta_1 \neq \theta_0$).

Зазначимо, що математична статистика не дає ніяких рекомендацій щодо вибору нульової та альтернативної гіпотези. Цей вибір повністю визначається дослідником і залежить від поставленої задачі.

Означення 9.1.3. Правило, за яким приймається чи відхиляється гіпотеза на основі вибірки, називається *статистичним критерієм* для перевірки гіпотези H. Якщо перевіряється гіпотеза про належність розподілу генеральної сукупності до якогось класу розподілів (нормального, рівномірного, Пуассона тощо), то зазначене правило називається *критерієм згоди*.

Оскільки висновок щодо прийняття або відхилення гіпотези приймається за реалізацією вибірки, то вибране рішення може бути помилковим. Розрізняють типи таких помилок.

Означення 9.1.4. Помилка, яка полягає в тому, що правильна гіпотеза H_0, згідно з вибраним критерієм, відхиляється, називається *помилкою першого роду* — це так званий «хибно позитивний висновок». *Помилка другого роду* відбувається тоді, коли справджується деяка альтернативна гіпотеза, але приймається основна гіпотеза H_0 — це так званий «хибно негативний висновок».

Ці помилки істотно різні за своєю суттю. Проілюструємо відмінність між названими помилками на прикладі перевірки медичного препарату на дієвість. В ролі ГС тут ξ — кількість випадків, коли препарат не подіяв. Висунемо дві гіпотези: H_0 — «препарат недієвий» та H_1 — «препарат дієвий». Помилка першого роду тут полягає в тому, що препарат дійсно не є дієвим, але ця гіпотеза відхиляється (і це є тим самим хибно позитивним висновком, що препарат дієвий), а помилка другого роду — в тому, що насправді дієвий препарат таким не вважають (це хибно негативний висновок про те, що препарат недієвий). Наслідком помилки першого роду є випуск у виробництво недієвого (а може, й шкідливого) препарату, а помилки другого роду — «непомічання» його дієвості та проведення наступних спроб вдосконалити препарат. Отже, помилка першого роду —- це помилка, якої важливіше уникнути.

Слід зазначити, що статистичний критерій, за яким перевіряється певна гіпотеза, не відповідає на питання, правильна гіпотеза чи ні. За критерієм ми лише вирішуємо, чи суперечать вибіркові дані висунутій гіпотезі, чи ні. Висновок «дані суперечать гіпотезі» вважаються більш вагомим, ніж «дані не суперечать гіпотезі».

9.1.2 Методика перевірки статистичних гіпотез

Нехай $\vec{\xi} = (\xi_1, \xi_2, ..., \xi_n)$ — випадкова вибірка, ξ — ГС, а $G \subseteq \mathbb{R}^n$ — вибірковий простір. В залежності від задачі формулюємо основну гіпотезу H_0 та альтернативну H_1. Сформулюємо загальний принцип побудови критеріїв перевірки гіпотез за певною реалізацією вибірки $\vec{x} = (x_1, x_2, ..., x_n)$. Критерій задають за допомогою *критичної множини* W, що є підмножиною вибіркового простору. Множину $\overline{W} = G \setminus W$ назвемо *областю прийняття гіпотези.* Рішення приймають так:

1. Якщо вибірка $\vec{x} = (x_1, x_2, ..., x_n)$ належить критичній множині W, то вважають, що *дані суперечать основній гіпотезі*, тобто H_0 відхиляють.
2. Якщо вибірка $\vec{x} = (x_1, x_2, ..., x_n)$ належить області прийняття гіпотези $G \setminus W$, то відхиляють альтернативну гіпотезу й приймають основну H_0. В цьому випадку вважають, що *дані не суперечать основній гіпотезі.*

Знайдемо ймовірності помилок першого та другого роду. Нехай гіпотези $H_i, i = 0, 1$ полягають в тому, що щільність розподілу ξ дорівнює $f_i(x)$, якщо ξ — неперервна, або $\mathbb{P}\{\xi = x_k\} = p_i(x_k)$, якщо ξ — дискретна.

Ймовірність помилки першого роду відносно нульової гіпотези H_0 дорівнює ймовірності того, що $\vec{\xi}$ потрапить в критичну множину W, тобто, $\mathbb{P}\left\{\vec{\xi} \in W/H_0\right\} = \alpha$. Величина α називається *рівнем значущості критерію.* Якщо $\mathcal{L}_{H_0}(\vec{x})$ — функція правдоподібності, побудована згідно висунутій гіпотезі H_0, то ймовірність помилки першого роду дорівнює $\int_W \mathcal{L}_{H_0}(\vec{x})d\vec{x}$. Рівень значущості, як правило, задається.

Ймовірність помилки другого роду відносно нульової гіпотези H_0 дорівнює ймовірності того, що $\vec{\xi}$ потрапить в $\overline{W}$, якщо насправді вірна гіпотеза H_1, тобто, $\mathbb{P}\left\{\vec{\xi} \in \overline{W}/H_1\right\} = \beta$. Якщо $\mathcal{L}_{H_1}(\vec{x})$ — функція правдоподібності, побудована згідно альтернативній гіпотезі H_1, то ймовірність помилки другого роду дорівнює $\int_{\overline{W}} \mathcal{L}_{H_1}(\vec{x})d\vec{x}$.

Одночасно взяти якомога малими α та β неможливо, тому критерій будують таким, щоб величина $1-\beta$, яка називається *потужністю критерію*, була найбільшою при заданому рівні значущості α. Потужність критерію — це ймовірність відхилити основну гіпотезу, якщо вона є хибною.

9.2 Критерій Пірсона (χ^2) та критерій Колмогорова

Перейдемо до найважливішої задачі математичної статистики: перевірки гіпотези про закон розподілу генеральної сукупності за певною реалізацію вибірки. Припущення про вид закону генеральної сукупності варто висувати тільки після первинної обробки статистичних даних. Параметри розподілу, як правило, невідомі, тому їх замінюємо на найкращі точкові оцінки. Очевидно, що між теоретичними та емпіричними розподілами існують розходження. Важливо зрозуміти, чи пояснюються ці розходження тільки випадковими обставинами (наприклад, обмеженість кількості спостережень), чи суттєвими (наприклад, гіпотетичний закон підібрано невдало).

9.2.1 Критерій Пірсона та алгоритм його використання

Висунемо гіпотезу $H_0 : \xi$ *має функцію розподілу* $F(x, \theta_1, \theta_2, ..., \theta_m)$. Згідно з цією гіпотезою ξ може приймати значення з множини X, яку розіб'ємо на r підмножин X_i, що попарно не перетинаються: $X = \bigcup\limits_{i=1}^{r} X_i$ (рекомендації щодо вибору цих множин та їх кількості розглянемо пізніше). В припущенні, що гіпотеза H_0 справджується, можемо обчислити ймовірності $p_i = \mathbb{P}\{\xi \in X_i / H_0\}$, причому $\sum\limits_{i=1}^{r} p_i = 1$. Якщо H_0 справджується, то частості $\frac{n_i}{n}$, де $n_i = \sum\limits_{k=1}^{n} 1\{\xi_k \in X_i\}$ (кількість значень, що потрапили в X_i), мають прямувати за ймовірністю до p_i. Критерій Пірсона з'ясовує, чи можна вважати розходження між теоретичними ймовірностями p_i та практичними значеннями $\frac{n_i}{n}$ випадковими, а не систематичними.

Розглядається статистика

$$\eta = \sum_{i=1}^{r} \frac{n}{p_i} \left(\frac{n_i}{n} - p_i \right)^2 = \sum_{i=1}^{r} \frac{(n_i - np_i)^2}{np_i} \tag{1}$$

За *теоремою Пірсона*, доведення якої буде наведено далі, якщо складна гіпотеза H_0 про закон розподілу генеральної сукупності справджується, то статистика η прямує за розподілом до розподілу χ^2_{r-s-1}, де s — кількість невідомих параметрів гіпотетичного закону розподілу, які оцінюємо, а r — кількість множин, за якими рахувалися теоретичні ймовірності p_i.

Алгоритм використання критерію Пірсона.

1. Після висування гіпотези про закон розподілу ξ розбиваємо множину X можливих значень ξ на r підмножин, що попарно не перетинаються.
2. Згідно висунутій гіпотезі обчислюємо $p_i = \mathbb{P}\{\xi \in X_i / H_0\}$, перевіряємо $\sum\limits_{i=1}^{r} p_i = 1$. Якщо $r \geq 20$, то потрібно, щоб виконувалося $np_i \geq 5, i = 1, ..., r$, а якщо $r < 20$ — $np_i \geq 10, i = 1, ...r$. Це пов'язано з тим, що заміна розподілу η на χ^2_{r-s-1} є наближеною. Якщо ці умови не виконуються, то сусідні підмножини об'єднують, зменшуючи r та збільшуючи відповідні p_i.
3. Обчислюємо $\eta_{\text{зн}}$ (з, можливо, новим значенням r) та порівнюємо з *критичним значенням* $t_{\text{кр}}$, яке для заданого рівня значущості α шукається як значення $t_{\alpha, r-s-1}$ з таблиці на ст. 123. Формально, $t_{\text{кр}}$ — це квантиль рівня $1 - \alpha$ для розподілу χ^2_{r-s-1}.
4. Критична область в критерії Пірсона є *правосторонньою*: якщо $\eta_{\text{зн}} < t_{\text{кр}}$, то дані не суперечать висунутій гіпотезі H_0, а якщо $\eta_{\text{зн}} \geq t_{\text{кр}}$, то суперечать на рівні значущості α. Це означає, що відхилення частостей від теоретичних ймовірностей p_i не можна вважати випадковими.

Наведемо декілька прикладів застосування критерію Пірсона.

Приклад. 1. Спостереження ГС наведені в інтервальному варіаційному ряді. Перевірити гіпотезу про нормальний розподіл ГС на рівні значущості $\alpha = 0.05$.

інтервал	$[-4; 0)$	$[0; 2)$	$[2; 4)$	$[4; 6]$
n_i	20	40	30	10

За реалізацією вибірки знайдемо значення найкращих оцінок параметрів гауссівського розподілу: вибіркового середнього $\overline{x} = 1.4$ та виправленої вибіркової дисперсії $D^{**}_{\text{зн}} = 4.48$. Висунемо гіпотезу $H_0 : \xi \sim \mathrm{N}(1.4, 4.48)$. Оскільки ξ за цим припущенням може приймати довільні дійсні значення, то розіб'ємо дійсну вісь на інтервали $(-\infty; -4)$, $[-4; 0)$, $[0; 2)$, $[2; 4)$, $[4; 6)$ та $(6; +\infty)$, для яких обчислимо теоретичні ймовірності p_i:

X_i	$(-\infty; -4)$	$[-4; 0)$	$[0; 2)$	$[2; 4)$	$[4; 6]$	$(6; +\infty)$
p_i	0.005	0.249	0.3575	0.2787	0.0948	0.015
np_i	0.5	24.9	35.75	27.87	9.48	1.5

Умова $np_i \geq 10$ не виконується для $(-\infty; -4)$ та $(6; +\infty)$, тому об'єднаємо їх з сусідніми інтервалами:

X_i	$(-\infty; 0)$	$[0; 2)$	$[2; 4)$	$[4; +\infty)$
p_i	0.254	0.3575	0.2787	0.1098
np_i	25.4	35.75	27.87	10.98
n_i	20	40	30	10
$n_i - np_i$	-5.4	4.25	2.13	-0.98

Тепер можемо обчислити $\eta_{\text{зн}} \approx 1.904$. За таблицею розподілу χ^2 знаходимо $t_{\text{кр}}$ для $4-2-1=1$ степені свободи, воно рівне 3.84. Отже, $\eta_{\text{зн}} < t_{\text{кр}}$, тому на рівні значущості 0.05 дані не суперечать гіпотезі про розподіл ГС $\mathrm{N}(1.4, 4.48)$.

2. Серед 2020 дводітних сімей 527 мають двох хлопчиків, 476 — двох дівчаток, а у решти 1017 сімей діти різної статі. Чи можна на рівні значущості 0.05 вважати кількість хлопчиків у сім'ї, яка має двох дітей, біноміально розподіленою випадковою величиною?
Розглянемо випадкову величину ξ, яка для кожної дводітної сім'ї набуває значення $i = 0, 1, 2$, якщо в сім'ї i хлопчиків, та висунемо гіпотезу $H_0 : \xi \sim \mathrm{Bin}(2, p)$. Параметр p розподілу невідомий, але його точковою оцінкою є $p^* = \frac{1}{2}\overline{\xi}$, $p^*_{\text{зн}} = \frac{0\cdot476+1\cdot1017+2\cdot527}{2\cdot2020} \approx 0.513$. Вибірковий простір складається з трьох елементів, $X = \{0, 1, 2\}$, тому розділимо його на три множини:

X_i	$\{0\}$	$\{1\}$	$\{2\}$
p_i	0.2371	0.4997	0.2632
np_i	478.942	1009.394	531.664
n_i	476	1017	527
$n_i - np_i$	-2.942	7.606	-4.664

Обчислимо $\eta_{\text{зн}} = \frac{(-2.942)^2}{478.942} + \frac{7.606^2}{1009.394} + \frac{(-4.664)^2}{531.664} \approx 0.116$. $r = 3$, $s = 1$, за таблицею знаходимо $t_{\text{кр}} = 3.84$. Отже, $\eta_{\text{зн}} < t_{\text{кр}}$, тому на рівні значущості 0.05 дані не суперечать гіпотезі про розподіл $\xi \sim \mathrm{Bin}(2, 0.513)$.

Окремим, більш простішим, випадком застосування критерію χ^2 є його *застосування до схеми Бернуллі*, тобто при перевірці гіпотези про ймовірність появи деякої події в послідовності незалежних випробувань. Нехай у послідовності n незалежних випробувань подія A відбулася m разів. Потрібно на рівні значущості α перевірити гіпотезу, що $\mathbb{P}(A) = p$. Дані можна розглядати як вибірку з n значень випадкової величини ξ, що є індикатором події A у випробуванні. В цьому випадку статистика η запишеться як $\eta = \frac{(m-np)^2}{np} + \frac{(n-m-nq)^2}{nq}$, де $q = 1 - p$, m — кількість появ події A у n випробуваннях. Можна дещо перетворити цю статистику:

$$\eta = \frac{(m-np)^2}{np} + \frac{(n(1-q)-m)^2}{nq} = \frac{(m-np)^2}{n}\left(\frac{1}{p} + \frac{1}{q}\right) = \frac{(m-np)^2}{npq} = \left(\frac{m-np}{\sqrt{npq}}\right)^2$$

Згідно ЦГТ $\frac{m-np}{\sqrt{npq}}$ прямує за розподілом до $\mathrm{N}(0,1)$, тому η прямує в такому ж сенсі до χ^2_1. Далі перевірка основної гіпотези проводиться аналогічно загальному випадку.

Приклад. При $n = 4040$ підкиданнях монети «герб» випав $m = 2048$ разів. Чи узгоджується на рівні значущості 0.1 з цими даними гіпотеза, що монета симетрична?

$\eta_{\text{зн}} = \frac{(2048-4040/2)^2}{4040/2} + \frac{(1992-4040/2)^2}{4040/2} \approx 0.776$. За таблицею знайдемо $t_{\text{кр}} = 2.71$, тому на рівні значущості 0.1 дані не суперечать гіпотезі про ймовірність випадання «герба» $\frac{1}{2}$.

9.2.2 Доведення теореми Пірсона

Нагадаємо «підготовчу роботу» для застосування критерію Пірсона. Висуваємо гіпотезу $H_0 : \xi$ *має функцію розподілу* $F(x, \theta_1, \theta_2, ..., \theta_m)$. Згідно з цією гіпотезою ξ може приймати значення з множини X, яку розіб'ємо на r підмножин X_i, що попарно не перетинаються: $X = \bigcup\limits_{i=1}^{r} X_i$, та обчислимо ймовірності $p_i = \mathbb{P}\left\{\xi \in X_i / H_0\right\}, i = 1, ..., r$.

$$\eta = \sum_{i=1}^{r} \frac{n}{p_i}\left(\frac{n_i}{n} - p_i\right)^2 = \sum_{i=1}^{r} \frac{(n_i - np_i)^2}{np_i}$$

Теорема (теорема Пірсона)**.** *В позначеннях, введених вище:*

1. *Якщо* ***проста*** *гіпотеза H_0 щодо закону розподілу ГС справджується, то статистика η прямує за розподілом до розподілу χ^2_{r-1} (розподіл χ^2 з $r-1$ ступенями вільності) при $n \to \infty$.*
2. *Якщо* ***складна*** *гіпотеза H_0 щодо закону розподілу ГС справджується, то статистика η прямує за розподілом до розподілу χ^2_{r-s-1} (розподіл χ^2 з $r-s-1$ ступенями вільності) при $n \to \infty$. Тут s — кількість невідомих параметрів розподілу, які оцінюються.*

Перед доведенням цієї теореми нагадаємо *ЦГТ для випадкових векторів*: нехай $\left\{\vec{\xi}_n\right\}_{n=1}^{\infty}$ — послідовність незалежних однаково розподілених випадкових векторів, що мають скінченні математичне сподівання $\vec{a}$ та кореляційну матрицю K. Тоді

$$\sqrt{n}\left(\frac{1}{n}\sum_{k=1}^{n}\vec{\xi}_k - \vec{a}\right) = \frac{1}{\sqrt{n}}\sum_{k=1}^{n}\left(\vec{\xi}_k - \vec{a}\right) \xrightarrow{\mathrm{F}} \vec{\eta} \sim \mathrm{N}\left(\vec{0}, K\right),\ n \to \infty \tag{2}$$

Наслідком ЦГТ є $\|\vec{\eta}_n\|^2 \xrightarrow{\mathrm{F}} \|\vec{\eta}\|^2, n \to \infty$, де $\vec{\eta}_n = \frac{1}{\sqrt{n}}\sum\limits_{k=1}^{n}\left(\vec{\xi}_k - \vec{a}\right)$, $\vec{\eta} \sim \mathrm{N}\left(\vec{0}, K\right)$.

Також зауважимо, що у випадку $r = 2$ теорему вже фактично було доведено, коли розглядалося застосування критерію Пірсона до схеми Бернуллі.

Доведення. З огляду на технічну складність, проведемо доведення лише для випадку простої гіпотези. План доведення: покажемо, що статистика η є квадратом норми деякого вектора, що за ЦГТ збігається за розподілом до $r-1$-вимірного стандартного гауссівського вектора, квадрат норми якого, в свою чергу, має розподіл χ^2_{r-1}, звідки отримаємо твердження теореми за наслідком ЦГТ.

Нехай H_0 справджується, гіпотетичну область значень ξ розбиваємо на $r > 2$ підмножин, що попарно не перетинаються: $X = \bigcup\limits_{i=1}^{r} X_i$, $p_i = \mathbb{P}\left\{\xi \in X_i / H_0\right\}, i = 1, ..., r$. $\vec{\xi} = (\xi_1, \xi_2, ..., \xi_n)$ — випадкова вибірка, введемо індикатори $I_{X_i}(\xi_j) = \begin{cases} 1, & \xi_j \in X_i \\ 0, & \xi_j \notin X_i \end{cases}$. Оскільки розподіл усіх ξ_j такий самий, як у ξ, то $\mathbb{E}\left(I_{X_i}(\xi_j)\right) = p_i$, $\mathbb{D}\left(I_{X_i}(\xi_j)\right) = p_i(1-p_i) = p_i q_i$.

Введемо r-вимірні вектори:

$$\overrightarrow{\mu^{(1)}} = \begin{pmatrix} \frac{I_{X_1}(\xi_1)-p_1}{\sqrt{p_1}} \\ \frac{I_{X_2}(\xi_1)-p_2}{\sqrt{p_2}} \\ \cdots \\ \frac{I_{X_r}(\xi_1)-p_r}{\sqrt{p_r}} \end{pmatrix},\ \overrightarrow{\mu^{(2)}} = \begin{pmatrix} \frac{I_{X_1}(\xi_2)-p_1}{\sqrt{p_1}} \\ \frac{I_{X_2}(\xi_2)-p_2}{\sqrt{p_2}} \\ \cdots \\ \frac{I_{X_r}(\xi_2)-p_r}{\sqrt{p_r}} \end{pmatrix},\ ...,\ \overrightarrow{\mu^{(n)}} = \begin{pmatrix} \frac{I_{X_1}(\xi_n)-p_1}{\sqrt{p_1}} \\ \frac{I_{X_2}(\xi_n)-p_2}{\sqrt{p_2}} \\ \cdots \\ \frac{I_{X_r}(\xi_n)-p_r}{\sqrt{p_r}} \end{pmatrix}$$

Внаслідок однакового розподілу та незалежності усіх ξ_j всі $\overrightarrow{\mu^{(j)}}$ теж мають однаковий розподіл та незалежні. Оскільки $\mathbb{E}\left(I_{X_i}(\xi_j)\right) = p_i$, то $\mathbb{E}\overrightarrow{\mu^{(1)}} = \mathbb{E}\overrightarrow{\mu^{(2)}} = ... = \mathbb{E}\overrightarrow{\mu^{(n)}} = \vec{0}$. Знайдемо кореляційну матрицю $\mathbb{K}$ вектора $\overrightarrow{\mu^{(1)}}$:

$$cov\left(\frac{I_{X_i}(\xi_1)-p_i}{\sqrt{p_i}}, \frac{I_{X_j}(\xi_1)-p_j}{\sqrt{p_j}}\right) = \frac{1}{\sqrt{p_i p_j}}\mathbb{E}\left(\left(I_{X_i}(\xi_1)-p_i\right)\left(I_{X_j}(\xi_1)-p_j\right)\right) =$$
$$= \frac{1}{\sqrt{p_i p_j}}\left(\mathbb{E}\left(I_{X_i}(\xi_1)I_{X_j}(\xi_1)\right) - p_i p_j\right) = \begin{cases} \frac{1}{p_i}\mathbb{D}\left(I_{X_i}(\xi_1)\right) = 1 - p_i, & i = j \\ -\sqrt{p_i p_j}, & i \neq j \end{cases}$$

Випадок $i \neq j$ тут пояснюється тим, що множини X_i та X_j не перетинаються.

$$\mathbb{K} = \begin{pmatrix} 1-p_1 & -\sqrt{p_1p_2} & \cdots & -\sqrt{p_1p_r} \\ -\sqrt{p_2p_1} & 1-p_2 & \cdots & -\sqrt{p_2p_r} \\ \vdots & \vdots & \ddots & \vdots \\ -\sqrt{p_rp_1} & -\sqrt{p_rp_2} & \cdots & 1-p_r \end{pmatrix} = \mathbb{I}_r - \begin{pmatrix} -\sqrt{p_1} \\ -\sqrt{p_2} \\ \vdots \\ -\sqrt{p_r} \end{pmatrix}\begin{pmatrix} -\sqrt{p_1} & -\sqrt{p_2} & \cdots & -\sqrt{p_r} \end{pmatrix}$$

Тут $\mathbb{I}_r$ — одична матриця розмірності $r \times r$. Виявляється, що координати $\overrightarrow{\mu^{(1)}}$ лінійно залежні:

$$\sum_{i=1}^{r}\sqrt{p_i}\mu_i^{(1)} = \sum_{i=1}^{r}\left(I_{X_i}(\xi_1) - p_i\right) = \sum_{i=1}^{r} I_{X_i}(\xi_1) - 1 = 0 \text{ з ймовірністю } 1$$

Це означає, що матриця $\mathbb{K}$ вироджена і має ранг $n-1$: одночасно більше ніж одну координату виразити через інші неможливо, оскільки $X_1, X_2, ..., X_r$ попарно не перетинаються.

Отже, можна знайти таку ортогональну матрицю U, що вектор $\widehat{\mu^{(1)}} = U\overrightarrow{\mu^{(1)}}$ матиме нульову останню координату. З міркувань про лінійну залежність координат вихідного вектора бачимо, що такою матрицею може бути та, останній рядок якої дорівнює $\begin{pmatrix}\sqrt{p_1} & \sqrt{p_2} & \cdots & \sqrt{p_r}\end{pmatrix}$

(його норма, очевидно, рівна 1), а інші складаються з ортонормованого базису ортогонального доповнення до лінійної оболонки цього рядка. Покажемо, що якими б не були перші $r-1$ рядків матриці U, кореляційна матриця $\widehat{\mathbb{K}} = U\mathbb{K}U^T$ вектора $\widehat{\mu^{(1)}}$ матиме вигляд $\begin{pmatrix} \mathbb{I}_{r-1} & 0 \\ 0 & 0 \end{pmatrix}$.

$U = (u_{ij})_{i,j=1,\dots n}$, її ортогональність означає, що для всіх $m \neq r$, $l \neq m$

$$\sum_{j=1}^{r} u_{mj}u_{rj} = \sum_{j=1}^{r} u_{mj}\sqrt{p_j} = 0,\ \sum_{j=1}^{r} u_{mj}^2 = 1,\ \sum_{j=1}^{r} u_{mj}u_{lj} = 0$$

Позначимо i, j-тий елементи матриць $\mathbb{K}$ та $\widehat{\mathbb{K}}$ через σ_{ij} та $\widehat{\sigma}_{ij}$, тоді

$$\widehat{\sigma}_{ml} = \sum_{i=1}^{r}\left(\sum_{j=1}^{r} u_{mj}\sigma_{ji}\right)u_{li} = \sum_{i=1}^{r}\left(\left(\sum_{i\neq j} -u_{mj}\sqrt{p_ip_j}\right) + u_{mi}(1-p_i)\right)u_{li} =$$

$$= \sum_{i=1}^{r}\left(\sqrt{p_i}\left(-\sum_{j\neq i} u_{mj}\sqrt{p_j} - u_{mi}\sqrt{p_i}\right) + u_{mi}\right)u_{li} =$$

$$= \sum_{i=1}^{r} u_{li}\cdot\begin{cases} u_{mi}, & m\neq r \\ 0, & m = r\end{cases} = \begin{cases} 1, & m\neq r \text{ та } m = l \\ 0, & m = r \text{ або } m \neq l\end{cases} \Longrightarrow \widehat{\mathbb{K}} = \begin{pmatrix} \mathbb{I}_{r-1} & 0 \\ 0 & 0 \end{pmatrix}$$

Розглянемо r-вимірний випадковий вектор $\overrightarrow{\eta^{(n)}} = \frac{1}{\sqrt{n}}\sum\limits_{i=1}^{n}\overrightarrow{\mu^{(i)}}$. Його перша координата дорівнює

$$\eta_1^{(n)} = \frac{1}{\sqrt{n}}\left(\frac{I_{X_1}(\xi_1) - p_1}{\sqrt{p_1}} + \frac{I_{X_1}(\xi_2) - p_1}{\sqrt{p_1}} + \ldots + \frac{I_{X_1}(\xi_n) - p_1}{\sqrt{p_1}}\right) = \frac{n_1 - np_1}{\sqrt{np_1}}$$

Аналогічно

$$\eta_2^{(n)} = \frac{n_2 - np_2}{\sqrt{np_2}}, \eta_3^{(n)} = \frac{n_3 - np_3}{\sqrt{np_3}}, \ldots, \eta_r^{(n)} = \frac{n_r - np_r}{\sqrt{np_r}}$$

Застосуємо до цього вектора матрицю U:

$$\widehat{\eta^{(n)}} = U\overrightarrow{\eta^{(n)}} = \frac{1}{\sqrt{n}}\sum_{i=1}^{n} U\overrightarrow{\mu^{(i)}} = \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\widehat{\mu^{(i)}}$$

Оскільки усі $\overrightarrow{\mu^{(i)}}$ незалежні та однаково розподілені, то усі $\widehat{\mu^{(i)}}$ — теж, причому математичні сподівання $\mathbb{E}\widehat{\mu^{(i)}} = U\mathbb{E}\overrightarrow{\mu^{(i)}} = \vec{0}$, а кореляційні матриці дорівнюють $\widehat{K} = \begin{pmatrix} \mathbb{I}_{r-1} & 0 \\ 0 & 0 \end{pmatrix}$.

Отже, за ЦГТ для випадкових векторів (2) $\widehat{\eta^{(n)}} \xrightarrow{\mathrm{F}} \vec{\eta} \sim \mathrm{N}\left(\vec{0}, \widehat{\mathbb{K}}\right)$, $n \to \infty$, а за наслідком з неї $\left\|\widehat{\eta^{(n)}}\right\|^2 \xrightarrow{\mathrm{F}} \|\vec{\eta}\|^2 \sim \chi^2_{r-1}$. Залишилося зауважити, що ортогональне перетворення зберігає норму, тому $\left\|\widehat{\eta^{(n)}}\right\|^2 = \left\|\overrightarrow{\eta^{(n)}}\right\|^2 = \sum\limits_{i=1}^{r}\frac{(n_i - np_i)^2}{np_i}$ збігається за розподілом до χ^2_{r-1}, що і треба було довести. ▲

Зауваження. Можна навести нестроге пояснення того, що у випадку складної гіпотези граничним розподілом статистики η є χ^2_{r-s-1}, а не χ^2_{r-1}. Кількість ступенів свободи у випадку простої гіпотези дорівнює $r-1$ через те, що одну з координат кожного вектора $\overrightarrow{\mu^{(j)}}$ можна було лінійно виразити через інші, але дві чи більше координат одночасно так виразити вже неможливо — це, власне, пояснює термін «степені свободи» в цьому випадку: значення однієї фіксованої координати цих векторів «автоматично» визначається, коли відомі значення інших $r-1$. У випадку складної гіпотези, коли s невідомих параметрів розподілу оцінюються, додається ще s подібних обмежень, що зменшує кількість ступенів свободи у граничного розподілу. Дещо подібне відбувалося при побудові довірчих інтервалів для дисперсії гауссівської ГС (ст. 90): там використання вибіркового середнього замість точного значення математичного сподівання теж зменшувало на 1 кількість ступенів свободи у розподілі вибіркової дисперсії.

9.2.3 Критерій незалежності χ^2

Наведемо без доведення ще один напрям застосування розподілу χ^2 для перевірки статистичних гіпотез. Нехай є n спостережень двох випадкових величин ξ та η, що позначимо відповідно $(x_1, x_2, ..., x_n)$ та $(y_1, y_2, ..., y_n)$. Перевірятимемо гіпотезу H_0 : *ξ та η незалежні.* Як побачимо далі, розподіли ξ та η знати непотрібно, тому будемо вважати, що вони приймають скінченну кількість значень $\{X_1, X_2, ..., X_l\}$ та $\{Y_1, Y_2, ..., Y_k\}$ — інакше спостереження можна згрупувати в деякі інтервали.

Якщо гіпотеза H_0 справджується, то $\mathbb{P}\{\xi = X_i, \eta = Y_j\} = \mathbb{P}\{\xi = X_i\} \cdot \mathbb{P}\{\eta = Y_j\}$. Позначимо ν_{ij} кількість таких спостережень обох величин, що одночасно спостереження ξ дорівнює X_i і спостереження η дорівнює Y_j, ν_i^ξ — кількість спостережень ξ, рівних X_i, ν_j^η – кількість спостережень η, рівних Y_j. Оскільки за ЗВЧ $\frac{\nu_i^\xi}{n} \xrightarrow{\mathrm{P}} \mathbb{P}\{\xi = X_i\}$, $\frac{\nu_j^\eta}{n} \xrightarrow{\mathrm{P}} \mathbb{P}\{\eta = Y_j\}$ та $\frac{\nu_{ij}}{n} \xrightarrow{\mathrm{P}} \mathbb{P}\{\xi = X_i, \eta = Y_j\}$ при $n \to \infty$, то вже можна здогадатися, що критерій незалежності базуватиметься на різниці між $\frac{\nu_{ij}}{n}$ та $\frac{\nu_i^\xi}{n} \cdot \frac{\nu_j^\eta}{n}$. Має місце теорема:

Теорема. *Якщо наведена гіпотеза H_0 справджується, то статистика*

$$\zeta = \sum_{i=1}^{l} \sum_{j=1}^{k} \frac{\left(\nu_{ij} - \nu_i^\xi \cdot \nu_j^\eta / n\right)^2}{\nu_i^\xi \cdot \nu_j^\eta / n} \tag{3}$$

прямує за розподілом до розподілу $\chi^2_{(l-1)(k-1)}$ при $n \to \infty$.

Усі спостереження вносять до так званої *таблиці спряженості*, застосування якої розглянемо на прикладі. Прийняття чи відхилення основної гіпотези відбувається на основі порівняння значення статистики критерію з критичним значенням, як і у випадку звичайного критерію χ^2.

Приклад. Проведено 300 спостережень одночасно над випадковими величинами ξ та η, які набувають значень $1, 2$ та $1, 2, 3$ відповідно.

$\xi \backslash \eta$	1	2	3	ν_i^ξ
1	32	68	50	150
2	40	70	40	150
ν_j^η	72	138	90	300

Перевірити за критерієм χ^2 гіпотезу про незалежність ξ та η на рівні значущості 0.01.

Спочатку знайдемо величини $m_{ij} = \nu_i^\xi \cdot \nu_j^\eta / n$:

$$m_{11} = m_{21} = \frac{150 \cdot 72}{300} = 36,\ m_{12} = m_{22} = \frac{150 \cdot 138}{300} = 69,\ m_{13} = m_{23} = \frac{150 \cdot 90}{300} = 45$$

Обчислимо квадрати відхилень $(\nu_{ij} - m_{ij})^2$:

$$m_{11} = (32 - 36)^2 = 16,\ m_{12} = (68 - 69)^2 = 1,\ m_{13} = (50 - 45)^2 = 25$$
$$m_{21} = (40 - 36)^2 = 16,\ m_{22} = (70 - 69)^2 = 1,\ m_{23} = (40 - 45)^2 = 25$$

Значення статистики критерію:

$$\zeta_{\text{зн}} = 2 \cdot \left(\frac{16}{36} + \frac{1}{69} + \frac{25}{45}\right) \approx 2.029$$

Кількість ступенів вільності дорівнює 2, тому за таблицею знаходимо $t_{\text{кр}} = 9.21$. Отже, $\zeta_{\text{зн}} < t_{\text{кр}}$ і на рівні значущості 0.01 дані не суперечать гіпотезі про незалежність ξ та η.

9.2.4 Критерій згоди Колмогорова

Розглянемо критерій згоди, що використовується для перевірки *простих* гіпотез щодо *неперервного* розподілу ГС: наприклад, $H_0 : \xi \sim \mathrm{N}(0, 1)$ чи $H_0 : \xi \sim \mathrm{U}\langle -1, 1\rangle$. Вводиться статистика $D_n = \sup\limits_{x \in \mathbb{R}} |F_n^*(x) - F_\xi(x)|$, де $F_n^*(x)$ — емпірична функція розподілу. На тому, що $F_n^*(x)$ з ймовірністю 1 прямує до $F_\xi(x)$ при $n \to \infty$, базується наступна теорема:

Теорема (теорема Колмогорова)**.** *Нехай* $\vec{\xi} = (\xi_1, \xi_2, ..., \xi_n)$ *— вибірка з неперервної ГС* ξ *з функцією розподілу* $F_\xi(x)$*,* $F_n^*(x)$ *— побудована за вибіркою емпірична функція розподілу. Тоді*

$$\lim_{n\to\infty} \mathbb{P}\left\{\sqrt{n}\cdot\sup_{x\in\mathbb{R}}|F_n^*(x) - F_\xi(x)| < \lambda\right\} = K(\lambda),\ \lambda > 0,\ \text{де}\ K(\lambda) = \sum_{j=-\infty}^{+\infty}(-1)^j e^{-2j^2\lambda^2}$$

Доводити цю теорему не будемо. Зауважимо, що особливістю статистики D_n є те, що її закон розподілу однаковий для всіх неперервних ГС і залежить лише від обсягу вибірки: якщо зробити у ній заміну $x = F_\xi^{-1}(y)$, отримаємо $D_n = \sup\limits_{y\in[0;1]}\left|F_n^*(F_\xi^{-1}(y)) - y\right|$. Оскільки випадкові величини $\eta_k = F_\xi(\xi_k)$ утворюють вибірку з розподілу $\mathrm{U}\langle 0, 1\rangle$, то

$$F_n^*(F_\xi^{-1}(y)) = \frac{1}{n}\sum_{k=1}^{n} 1\left\{\xi_k < F_\xi^{-1}(y)\right\} = \frac{1}{n}\sum_{k=1}^{n} 1\{\eta_k < y\},\ 1\{\eta_k < y\} \sim \mathrm{Bin}(1, y)\ \text{при}\ y \in [0; 1]$$

η_k незалежні, тому $\sum\limits_{k=1}^{n} 1\{\eta_k < y\} \sim \mathrm{Bin}(n, y)$. Подальші перетворення цієї випадкової величини мають невипадковий характер, тому, дійсно, розподіл D_n не залежить від розподілу ξ. Таким чином, якщо H_0 справджується, то значення D_n буде досить малим. Значення $K(\lambda)$ знаходять з відповідної таблиці.

Алгоритм використання критерію Колмогорова.

1. Висуваємо просту гіпотезу H_0 про неперервний закон розподілу ГС ξ.
2. За реалізацією вибірки будуємо емпіричну функцію розподілу $F_n^*(x)$.
3. Знаходимо значення статистики D_n та $\lambda = \sqrt{n}\cdot D_n$.
4. За заданим рівнем значущості α знаходимо таке λ_α, що $\mathbb{P}\{\sqrt{n}\cdot D_n \le \lambda_\alpha / H_0\} = 1 - K(\lambda_\alpha) = \alpha$.
5. Якщо $\lambda_{\text{зн}}$ буде більшим за критичне значення λ_α, то на рівні значущості α дослідні дані суперечать гіпотезі H_0. Інакше вважають, що гіпотезу H_0 можна прийняти.

Наведемо таблицю значень λ_α для деяких α:

α	0.1	0.05	0.025	0.01
λ_α	1.224	1.358	1.480	1.628

Приклад. Нехай реалізацію вибірки подано у вигляді інтервального варіаційного ряду:

інтервал	$[0,1)$	$[1,2)$	$[2,3)$	$[3,4)$	$[4,5)$	$[5,6)$	$[6,7)$	$[7,8)$	$[8,9)$	$[9,10]$
n_i	35	16	15	17	17	19	11	16	30	24

На рівні значущості $\alpha = 0.05$ перевірити гіпотезу $H_0 : \xi \sim \mathrm{U}\langle 0, 10\rangle$.

За інтервальним варіаційним рядом можемо обчислити значення емпіричної та теоретичної функції розподілу в правих кінцях інтервалів, а також різниці між ними:

x	1	2	3	4	5	6	7	8	9	10
$F_n^*(x)$	0.175	0.255	0.33	0.415	0.5	0.595	0.65	0.73	0.88	1
$F_\xi(x)$	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1
$F_n^*(x) - F_\xi(x)$	0.075	0.055	0.03	0.015	0	−0.004	−0.05	−0.07	−0.02	0

$(D_n)_{\text{зн}} = 0.075$, тому $\lambda_{\text{зн}} = \sqrt{200}\cdot 0.075 \approx 1.06$. За таблицею $\lambda_{0.05} = 1.358$. Отже, на рівні значущості 0.05 дані не суперечать гіпотезі про рівномірний розподіл ξ на інтервалі $\langle 0, 10\rangle$.

9.3 Перевірка параметричних гіпотез

9.3.1 Прості гіпотези. Критерій Неймана-Пірсона

Часто розподіл генеральної сукупності відомий і за вибіркою треба перевірити припущення щодо значень параметрів цього розподілу. Такі гіпотези називають *параметричними.* Методика перевірки простої параметричної гіпотези H_0 проти простої альтернативної гіпотези H_1 ґрунтується на тому, що критичну область W слід вибирати таким чином, щоб ймовірність попадання в неї статистики критерію була мінімальною і дорівнювала рівню значущості α, якщо дослідні дані не суперечать нульовій гіпотезі H_0 та максимальною у протилежному випадку. Тобто критична область повинна бути такою, щоб при заданому рівні значущості α потужність критерію $1-\beta$ була максимальною. Назвемо *найкращою критичною областю (НКО)* таку множину, що забезпечує максимальну потужність критерію. Побудова такої області ґрунтується на *лемі Неймана-Пірсона.*

Теорема (лема Неймана-Пірсона)**.** *Серед усіх критеріїв заданого рівня значущості α, які перевіряють просту параметричну гіпотезу $H_0 : \theta = \theta_0$ проти простої альтернативної гіпотези $H_1 : \theta = \theta_1$* ***критерій відношення правдоподібності*** *є найбільш потужним, тобто, найкраща критична область має вигляд*

$$W_{C_\alpha} = \left\{ \vec{x} \in \mathbb{R}^n : \frac{\mathcal{L}_{H_1}(\vec{x})}{\mathcal{L}_{H_0}(\vec{x})} \geq C_\alpha \right\} \tag{1}$$

де у чисельнику стоїть функція правдоподібності в припущенні, що справджується гіпотеза H_1, а у знаменнику — якщо справджується H_0. Стала C_α вибирається з умови

$$\mathbb{P}\left\{ \frac{\mathcal{L}_{H_1}(\vec{\xi})}{\mathcal{L}_{H_0}(\vec{\xi})} \geq C_\alpha \Big/ H_0 \right\} = \mathbb{P}\left\{ \ln \frac{\mathcal{L}_{H_1}(\vec{\xi})}{\mathcal{L}_{H_0}(\vec{\xi})} \geq \ln C_\alpha \Big/ H_0 \right\} = \alpha \tag{2}$$

Зауваження. Функція $\frac{\mathcal{L}_{H_1}(\vec{x})}{\mathcal{L}_{H_0}(\vec{x})}$ називається *відношенням правдоподібності*, а $\ln \frac{\mathcal{L}_{H_1}(\vec{x})}{\mathcal{L}_{H_0}(\vec{x})}$ — *логарифмічним відношенням правдоподібності.*

Доведення. Як було зазначено на початку розділу, ймовірність помилки першого роду при заданій критичній множині W_{C_α} можна знайти як $\alpha = \mathbb{P}\left\{\vec{\xi} \in W_{C_\alpha}/H_0\right\} = \int_{W_{C_\alpha}} \mathcal{L}_{H_0}(\vec{x})d\vec{x}$. Нехай W — деяка інша критична множина рівня значущості α, $\mathbb{P}\left\{\vec{\xi} \in W/H_0\right\} \leq \alpha$. Позначимо $I_W(\vec{x})$ та $I_{W_{C_\alpha}}(\vec{x})$ індикатори цих множин і розглянемо функцію

$$f(\vec{x}) = \left(I_{W_{C_\alpha}}(\vec{x}) - I_W(\vec{x})\right) \cdot \left(\mathcal{L}_{H_1}(\vec{x}) - C_\alpha \cdot \mathcal{L}_{H_0}(\vec{x})\right)$$

Якщо $\vec{x} \in W_{C_\alpha}$, то $I_{W_{C_\alpha}}(\vec{x}) - I_W(\vec{x}) = 1$ і $\mathcal{L}_{H_1}(\vec{x}) - C_\alpha \cdot \mathcal{L}_{H_0}(\vec{x}) \geq 0$ за побудовою. Якщо $\vec{x} \in W$, то $I_{W_{C_\alpha}}(\vec{x}) - I_W(\vec{x}) = -1$ і $\mathcal{L}_{H_1}(\vec{x}) - C_\alpha \cdot \mathcal{L}_{H_0}(\vec{x}) \leq 0$ за побудовою. В інших випадках ($\vec{x}$ належить обом множинам одночасно або не належить жодній) перший множник дорівнює 0. Отже, $f(\vec{x}) \geq 0$ для всіх $\vec{x} \in \mathbb{R}^n$.

$$\begin{aligned}
0 \leq \int_{\mathbb{R}^n} f(\vec{x})d\vec{x} &= \int_{\mathbb{R}^n} \left(I_{W_{C_\alpha}}(\vec{x}) - I_W(\vec{x})\right) \cdot \left(\mathcal{L}_{H_1}(\vec{x}) - C_\alpha \cdot \mathcal{L}_{H_0}(\vec{x})\right) d\vec{x} = \\
&= \int_{W_{C_\alpha}} \mathcal{L}_{H_1}(\vec{x})d\vec{x} - \int_W \mathcal{L}_{H_1}(\vec{x})d\vec{x} - C_\alpha \cdot \int_{W_{C_\alpha}} \mathcal{L}_{H_0}(\vec{x})d\vec{x} + C_\alpha \cdot \int_W \mathcal{L}_{H_0}(\vec{x})d\vec{x} = \\
&= \mathbb{P}\left\{\vec{\xi} \in W_{C_\alpha}/H_1\right\} - \mathbb{P}\left\{\vec{\xi} \in W/H_1\right\} - C_\alpha \cdot \left(\mathbb{P}\left\{\vec{\xi} \in W_{C_\alpha}/H_0\right\} - \mathbb{P}\left\{\vec{\xi} \in W/H_0\right\}\right)
\end{aligned}$$

$\mathbb{P}\left\{\vec{\xi} \in W_{C_\alpha}/H_0\right\} - \mathbb{P}\left\{\vec{\xi} \in W/H_0\right\} \geq 0$, тому $\mathbb{P}\left\{\vec{\xi} \in W_{C_\alpha}/H_1\right\} - \mathbb{P}\left\{\vec{\xi} \in W/H_1\right\} \geq 0$. Це й означає, що вибір критичної множини W_{C_α} дає більшу потужність критерію, ніж W. ▲

Твердження леми Неймана-Пірсона можна застосовувати для знаходження потужності критерію та обсягу вибірки, що забезпечує заданий рівень значущості та ймовірність помилки другого роду. Знаючи C_α, можна знайти ймовірність помилки другого роду β як

$$\mathbb{P}\left\{ \ln \frac{\mathcal{L}_{H_1}(\vec{x})}{\mathcal{L}_{H_0}(\vec{x})} < \ln C_\alpha \Big/ H_1 \right\} = \beta$$

Для визначення мінімального обсягу вибірки, при якому досягаються наперед задані ймовірності помилок першого α та другого роду β, треба розв'язати систему

$$\begin{cases} \mathbb{P}\left\{ \ln \frac{\mathcal{L}_{H_1}(\vec{x})}{\mathcal{L}_{H_0}(\vec{x})} \geq \ln C_\alpha \Big/ H_0 \right\} = \alpha \\ \mathbb{P}\left\{ \ln \frac{\mathcal{L}_{H_1}(\vec{x})}{\mathcal{L}_{H_0}(\vec{x})} < \ln C_\alpha \Big/ H_1 \right\} = \beta \end{cases}$$

Відповідні приклади будуть розглянуті нижче (ст. 103).

9.3.2 Приклади побудови найкращої критичної області

Приклад 1. Побудувати найкращу критичну область за лемою Неймана-Пірсона для параметра a нормально розподіленої ГС з відомою дисперсією σ^2. Розглянемо дві прості гіпотези: $H_0 : \mathbb{E}\xi = a_0$ та $H_1 : \mathbb{E}\xi = a_1$, де $a_0 < a_1$. Функція правдоподібності має вигляд

$$\mathcal{L}(x_1, ..., x_n, a) = \frac{1}{(2\pi)^{\frac{n}{2}}\sigma^n} \cdot \exp\left\{-\frac{1}{2\sigma^2}\sum_{k=1}^{n}(x_k - a)^2\right\}$$

Тому звичайне та логарифмічне відношення правдоподібності матимуть вигляд

$$\frac{\mathcal{L}_{H_1}(\vec{x})}{\mathcal{L}_{H_0}(\vec{x})} = \exp\left\{-\frac{1}{2\sigma^2}\sum_{k=1}^{n}(x_k - a_1)^2 + \frac{1}{2\sigma^2}\sum_{k=1}^{n}(x_k - a_0)^2\right\} =$$

$$= \exp\left\{-\frac{1}{2\sigma^2}\sum_{k=1}^{n}\left(x_k^2 - 2x_k a_1 + a_1^2 - x_k^2 + 2x_k a_0 - a_0^2\right)\right\} =$$

$$= \exp\left\{-\frac{1}{2\sigma^2}\left(2(a_0 - a_1)\sum_{k=1}^{n}x_k + n(a_1^2 - a_0^2)\right)\right\} = \exp\left\{\frac{a_1 - a_0}{\sigma^2}\sum_{k=1}^{n}x_k\right\}\exp\left\{-\frac{n(a_1^2 - a_0^2)}{2\sigma^2}\right\}$$

$$\ln\frac{\mathcal{L}_{H_1}(\vec{x})}{\mathcal{L}_{H_0}(\vec{x})} = \frac{a_1 - a_0}{\sigma^2}\sum_{k=1}^{n}x_k - \frac{n(a_1^2 - a_0^2)}{2\sigma^2} = \frac{n(a_1 - a_0)}{\sigma^2} \cdot \left(\underbrace{\frac{1}{n}\sum_{k=1}^{n}x_k}_{\overline{x}} - \frac{a_0 + a_1}{2}\right)$$

Розв'яжемо відносно $\overline{x}$ нерівність $\ln\frac{\mathcal{L}_{H_1}(\vec{x})}{\mathcal{L}_{H_0}(\vec{x})} \geq \ln C_\alpha$:

$$\frac{n(a_1 - a_0)}{\sigma^2} \cdot \left(\overline{x} - \frac{a_0 + a_1}{2}\right) \geq \ln C_\alpha \Leftrightarrow \overline{x} \geq \frac{\sigma^2 \ln C_\alpha}{n(a_1 - a_0)} + \frac{a_0 + a_1}{2} = t_{\text{кр}}$$

Зауважимо, що знак нерівності тут не змінюється, оскільки $a_1 > a_0$. Знайдемо t з рівності $\mathbb{P}\left\{\overline{\xi} \geq t/H_0\right\} = \alpha$, користуючись тим, що $\overline{\xi} \sim \mathrm{N}\left(a_0, \frac{\sigma^2}{n}\right)$, якщо H_0 справджується:

$$\mathbb{P}\left\{\overline{\xi} \geq t/H_0\right\} = \mathbb{P}\left\{\frac{(\overline{\xi} - a_0)\sqrt{n}}{\sigma} \geq \frac{(t - a_0)\sqrt{n}}{\sigma}\Big/\overline{\xi} \sim \mathrm{N}\left(a_0, \frac{\sigma^2}{n}\right)\right\} = 0.5 - \Phi\left(\frac{(t - a_0)\sqrt{n}}{\sigma}\right) = \alpha$$

Знайдене значення t буде межею критичної області, яка в цьому випадку буде *правосторонньою.* Якщо значення вибіркового середнього, обчисленого за конкретною реалізацією, буде більше за t, то гіпотезу $H_0 : a = a_0$ треба відхиляти, бо дослідні дані суперечать цій гіпотезі на рівні значущості α. В іншому випадку гіпотезу H_0 приймаємо.

У випадку $a_1 < a_0$ при розв'язанні нерівності відносно $\overline{x}$ зміниться її знак, тому критична область буде *лівосторонньою.*

Зауваження. Якщо дисперсія ГС є невідомою, то треба користуватися статистикою $\eta = \frac{(\overline{\xi} - a_0)\sqrt{n}}{\sqrt{\mathbb{D}^{**}\xi}}$, що має розподіл St_{n-1}, якщо гіпотеза H_0 справджується. В цьому випадку критична область також буде *правосторонньою* при $a_1 > a_0$, а при $a_1 < a_0$ — *лівосторонньою.*

Продемонструємо застосування знайденої НКО на конкретній задачі.

Задача 1. Стверджується, що кульки, вироблені верстатом, мають середній діаметр 10 мм. Перевірити цю гіпотезу на рівні значущості $\alpha = 0.05$, якщо після перевірки 16 кульок середнє значення діаметра дорівнювало 10.3 мм. Вважати, що результати вимірювання ξ нормально розподілені з $\sigma = 1$. Що зміниться, якщо дисперсія є невідомою, але $(\mathbb{D}^{**}\xi)_{\text{зн}} = 1$?

З умови задачі висуваємо гіпотези $H_0 : \mathbb{E}\xi = 10$ та $H_1 : \mathbb{E}\xi = a > 10$, бо $\overline{x} > 10$. Обчислимо межу критичної області з рівності $0.05 = 0.5 - \Phi\left(\frac{(t-10)\sqrt{16}}{1}\right) \Leftrightarrow \Phi\left(4(t - 10)\right) = 0.45$. З таблиці значень функції Лапласа $4(t-10) = 1.65 \Leftrightarrow t = 10.4125$. Отже, $\overline{x} = 10.3 < t$ і на рівні значущості 0.05 основна гіпотеза приймається. Якщо дисперсія є невідомою, то межу критичної області знайдемо з $\mathbb{P}\left\{\eta > t_{\text{кр}}\right\} = \alpha$ для $\eta \sim \mathrm{St}_{15}$, вона дорівнює 1.753. $\eta_{\text{зн}} = \frac{(10.3-10)\cdot 4}{1} = 1.2 < t_{\text{кр}}$. Отже, і в цьому випадку дані не суперечать основній гіпотезі на рівні значущості 0.05.

В умовах цієї задачі знайдемо також ймовірність помилки другого роду з рівності $\beta = \mathbb{P}\left\{\ln\frac{\mathcal{L}_{H_1}(\vec{x})}{\mathcal{L}_{H_0}(\vec{x})} < \ln C_\alpha \middle/ H_1\right\}$. Повторюючи міркування при знаходженні межі критичної області, отримаємо $\beta = \mathbb{P}\left\{\overline{\xi} < t_{\text{кр}}/H_1\right\} = 0.5 + \Phi\left(\frac{(t_{\text{кр}}-10.3)\sqrt{16}}{1}\right) = 0.5 + \Phi\left(\frac{(10.41-10.3)\cdot 4}{1}\right) \approx 0.617$. Відповідно, потужність критерію дорівнює $1-\beta \approx 0.383$.

Якщо ймовірність помилки другого роду при заданому рівні значущості треба зменшити, необхідно збільшити обсяг вибірки. Нехай $\beta = 0.1$, а $\alpha = 0.05$, як і було в задачі. Розв'яжемо систему:

$$\begin{cases}\mathbb{P}\left\{\ln\frac{\mathcal{L}_{H_1}(\vec{x})}{\mathcal{L}_{H_0}(\vec{x})} \geq \ln C_\alpha \middle/ H_0\right\} = \alpha \\ \mathbb{P}\left\{\ln\frac{\mathcal{L}_{H_1}(\vec{x})}{\mathcal{L}_{H_0}(\vec{x})} < \ln C_\alpha \middle/ H_1\right\} = \beta\end{cases} \Leftrightarrow \begin{cases}0.5 - \Phi\left(\frac{(t-10)\sqrt{n}}{1}\right) = 0.05 \\ 0.5 + \Phi\left(\frac{(t-10.3)\sqrt{n}}{1}\right) = 0.1\end{cases} \Leftrightarrow$$

$$\Leftrightarrow \begin{cases}\Phi\left((t-10)\sqrt{n}\right) = 0.45 \\ \Phi\left((t-10.3)\sqrt{n}\right) = -0.4\end{cases} \Leftrightarrow \begin{cases}(t-10)\sqrt{n} \approx 1.64 \\ (t-10.3)\sqrt{n} \approx -1.28\end{cases} \Rightarrow \begin{cases}\sqrt{n} \approx 9.733 \\ t \approx 10.17\end{cases}$$

Отже, обсяг вибірки $n \geq 95$ забезпечить ймовірність помилки другого роду, рівну 0.1.

Приклад 2. Побудувати найкращу критичну область за лемою Неймана-Пірсона для параметра σ^2 нормального закону розподілу з відомим математичним сподіванням a. Введемо дві прості гіпотези (основну та альтернативну): $H_0 : \mathbb{D}\xi = \sigma_0^2$ та $H_1 : \mathbb{D}\xi = \sigma_1^2$, де $\sigma_0^2 > \sigma_1^2$. Як і у прикладі 1, запишемо логарифмічне відношення правдоподібності:

$$\ln\frac{\mathcal{L}_{H_1}(\vec{x})}{\mathcal{L}_{H_0}(\vec{x})} = \ln\left(\frac{\sigma_0}{\sigma_1}\right)^n - \frac{1}{2\sigma_1^2}\sum_{k=1}^n (x_k - a)^2 + \frac{1}{2\sigma_0^2}\sum_{k=1}^n (x_k-a)^2 =$$

$$= n\cdot\ln\frac{\sigma_0}{\sigma_1} + \frac{n}{2}\cdot\frac{1}{n}\cdot\sum_{k=1}^n (x_k-a)^2\cdot\frac{\sigma_1^2-\sigma_0^2}{\sigma_1^2\cdot\sigma_0^2} = n\cdot\ln\frac{\sigma_0}{\sigma_1} + \frac{n}{2}\cdot\left(\mathbb{D}^*\xi\right)_{\text{зн}}\cdot\frac{\sigma_1^2-\sigma_0^2}{\sigma_1^2\cdot\sigma_0^2}$$

Розв'яжемо відносно $\left(\mathbb{D}^*\xi\right)_{\text{зн}}$ нерівність $\ln\frac{\mathcal{L}_{H_1}(\vec{x})}{\mathcal{L}_{H_0}(\vec{x})} \geq \ln C_\alpha$:

$$\frac{n}{2}\cdot\left(\mathbb{D}^*\xi\right)_{\text{зн}}\cdot\frac{\sigma_1^2-\sigma_0^2}{\sigma_1^2\cdot\sigma_0^2} \geq \ln C_\alpha - n\cdot\ln\frac{\sigma_0}{\sigma_1} \Leftrightarrow \left(\mathbb{D}^*\xi\right)_{\text{зн}} \leq \frac{\ln C_\alpha - n\cdot\ln\frac{\sigma_0}{\sigma_1}}{\frac{n}{2}\cdot\frac{\sigma_1^2-\sigma_0^2}{\sigma_1^2\cdot\sigma_0^2}}$$

Зауважимо, що знак нерівності тут змінюється, оскільки $\sigma_1^2 < \sigma_0^2$. Знайдемо t з рівності $\mathbb{P}\left\{\frac{n\mathbb{D}^*\xi}{\sigma_0^2} \leq t/H_0\right\} = \alpha$, користуючись тим, що $\eta = \frac{n\mathbb{D}^*\xi}{\sigma_0^2} \sim \chi_n^2$, якщо H_0 справджується. Знайдене значення t буде межею критичної області, яка в цьому випадку буде *лівосторонньою*. Якщо значення статистики η, обчисленої за конкретною реалізацією, буде менше за t, то гіпотезу $H_0 : \mathbb{D}\xi = \sigma_0^2$ треба відхиляти, бо дослідні дані суперечать цій гіпотезі на рівні значущості α. В іншому випадку гіпотезу H_0 приймаємо.

У випадку $\sigma_1^2 > \sigma_0^2$ при розв'язанні нерівності відносно $\left(\mathbb{D}^*\xi\right)_{\text{зн}}$ її знак не зміниться, тому критична область буде *правосторонньою*.

Зауваження. Якщо математичне сподівання ГС є невідомим, то треба користуватися статистикою $\eta = \frac{(n-1)\mathbb{D}^{**}\xi}{\sigma_0^2}$, що має розподіл χ_{n-1}^2, якщо гіпотеза H_0 справджується. В цьому випадку критична область також буде *правосторонньою* при $\sigma_1^2 > \sigma_0^2$, а при $\sigma_1^2 < \sigma_0^2$ — *лівосторонньою*.

Продемонструємо застосування знайденої НКО на конкретній задачі.

Задача 2. Точність деякого верстата характеризується дисперсією довжини виготовлених деталей: якщо вона більша за 400 мкм^2, то верстат треба переналаштувати. Після перевірки $n = 15$ деталей виявилося, що $\left(\mathbb{D}^{**}\xi\right)_{\text{зн}} = 680$ мкм^2. Чи потрібно переналаштувати верстат?

З умови задачі висуваємо основну гіпотезу $H_0 : \sigma^2 = \sigma_0^2 = 400$ і альтернативну $H_1 : \sigma^2 = \sigma_1^2$, де $\sigma_1^2 > 400$. Оберемо рівень значущості $\alpha = 0.01$. Оскільки математичне сподівання невідоме, скористаємося статистикою $\eta = \frac{(n-1)\mathbb{D}^{**}\xi}{\sigma_0^2}$, що має розподіл χ_{n-1}^2, якщо гіпотеза H_0 справджується. Критична область буде правосторонньою, за таблицею знайдемо її межу $t_{\text{кр}} = 29.14$. Оскільки $\eta_{\text{зн}} = \frac{14\cdot 680}{400} = 23.8 < t_{\text{кр}}$, гіпотеза H_0 приймається — верстат не треба переналаштувати.

Приклад 3. Побудувати найкращу критичну область для перевірки простої гіпотези про значення параметра a пуассонівської ГС при великому обсягу вибірки. Маємо основну гіпотезу $H_0 : a = a_0$ та альтернативну $H_1 : a = a_1$, нехай $a_1 > a_0$. Логарифмічна функція правдоподібності має вигляд

$$\ln \mathcal{L}(x_1, ..., x_n, a) = -na + \ln a \cdot \sum_{k=1}^{n} x_k - \sum_{k=1}^{n} \ln x_k!$$

Запишемо логарифмічне відношення правдоподібності

$$\ln \frac{\mathcal{L}_{H_1}(\vec{x})}{\mathcal{L}_{H_0}(\vec{x})} = \ln \mathcal{L}_{H_1}(\vec{x}) - \ln \mathcal{L}_{H_0}(\vec{x}) = -n \cdot (a_1 - a_0) + \ln \left(\frac{a_1}{a_0}\right) \cdot \sum_{k=1}^{n} x_k =$$

$$= -n \cdot (a_1 - a_0) + n \ln \left(\frac{a_1}{a_0}\right) \cdot \frac{1}{n} \sum_{k=1}^{n} x_k = -n \cdot (a_1 - a_0) + n \ln \left(\frac{a_1}{a_0}\right) \cdot \overline{x}$$

Розв'яжемо відносно $\overline{x}$ нерівність $\ln \frac{\mathcal{L}_{H_1}(\vec{x})}{\mathcal{L}_{H_0}(\vec{x})} \geq \ln C_\alpha$:

$$-n \cdot (a_1 - a_0) + n \ln \left(\frac{a_1}{a_0}\right) \cdot \overline{x} \geq \ln C_\alpha \Leftrightarrow \overline{x} \geq \frac{\ln C_\alpha + n \cdot (a_1 - a_0)}{n \ln \left(\frac{a_1}{a_0}\right)}$$

Зауважимо, що знак нерівності тут не змінюється, оскільки $a_1 > a_0$ і тому $\ln \left(\frac{a_1}{a_0}\right) > 0$. Якщо обсяг вибірки великий, то $\overline{\xi}$ наближено має розподіл $\mathrm{N}\left(a_0, \frac{a_0}{n}\right)$, якщо H_0 справджується:

$$\mathbb{P}\left\{\overline{\xi} \geq t / H_0\right\} = \alpha \Rightarrow \alpha \approx 0.5 - \Phi\left(\frac{(t - a_0)\sqrt{n}}{\sqrt{a_0}}\right)$$

Знайдене значення t буде межею критичної області, яка в цьому випадку буде *правосторонньою.* Якщо значення вибіркового середнього, обчисленого за конкретною реалізацією, буде більше за t, то гіпотезу $H_0 : a = a_0$ треба відхиляти, бо дослідні дані суперечать цій гіпотезі на рівні значущості α. В іншому випадку гіпотезу H_0 приймаємо.

У випадку $a_1 < a_0$ при розв'язанні нерівності відносно $\overline{x}$ зміниться її знак, тому критична область буде *лівосторонньою.*

Зрозуміло, що аналогічні міркування можна застосовувати і для інших розподілів ГС, якщо у логарифмічному відношенні правдоподібності вдається виділити вибіркове середнє чи іншу статистику, для якої справджується ЦГТ.

Вправа. Побудувати за лемою Неймана-Пірсона найкращу критичну область для параметра p ГС $\xi \sim \mathrm{Bin}(N, p)$, вважаючи N відомим, а обсяг вибірки великим. Розглянути випадки, коли альтернативне значення p_1 більше та менше основного p_0.

9.3.3 Перевірка складних параметричних гіпотез

Розглянемо задачу перевірки *простої* основної параметричної гіпотези $H_0 : \theta = \theta_0$ проти *складної* альтернативної гіпотези, яка записується у вигляді $H_1 : \theta > \theta_0$ або $H_1 : \theta < \theta_0$. В першому випадку задачу називають *перевіркою простої основної гіпотези проти правосторонньої альтернативи*, а в другому випадку — *перевіркою простої основної гіпотези проти лівосторонньої альтернативи.* Має місце теорема, яку іноді називають теоремою Неймана-Пірсона.

Теорема. *Побудований на основі відношення правдоподібності критерій перевірки простої параметричної гіпотези (за лемою Неймана-Пірсона) буде оптимальним критерієм і для перевірки простої параметричної гіпотези проти наведених вище складних альтернативних параметричних гіпотез.*

Зробимо пояснення щодо справедливості цієї теореми. Розглянемо просту основну параметричну гіпотезу $H_0 : \theta = \theta_0$ проти складної альтернативної гіпотези $H_1 : \theta > \theta_0$. Візьмемо довільне $\theta_1 > \theta_0$ та розглянемо альтернативну гіпотезу $\widetilde{H}_1 : \theta = \theta_1$, яка вже є простою. Для неї за лемою Неймана-Пірсона можна побудувати найкращу критичну область W_{C_α} для заданого рівня значущості α. Оскільки $\theta_1 > \theta_0$, критична область є правосторонньою. Як було показано в прикладах, межа критичної області від θ_1 не залежить, тому логічно припустити, що знайдена критична область буде найкращою і для інших значень θ_1 — це і є твердженням теореми.

Зауваження. На практиці іноді користуються альтернативною гіпотезою у вигляді $H_1 : \theta \neq \theta_0$ — так звану двосторонню альтернативу. Оптимального критерію в такому випадку не існує, а для пошуку лівої та правої межі критичної області за заданим рівнем значущості α беруть рівні $\alpha_1 = \alpha_2 = \frac{\alpha}{2}$.

9.3.4 Перевірка гіпотез про рівність параметрів двох ГС

На практиці часто виникає потреба порівняти результат однієї серії випробувань із результатом іншої серії. Виникає питання: чи можна розбіжність, що виникла пояснити випадковою помилкою експерименту, чи ця розбіжність невипадкова?

Перевірка гіпотези про рівність математичних сподівань двох гауссівських ГС при відомих дисперсіях

Дано дві реалізації обсягами n_1 та n_2 відповідно двох гауссівських ГС $\xi \sim \mathrm{N}\left(\mathbb{E}\xi, \sigma_1^2\right)$ та $\eta \sim \mathrm{N}\left(\mathbb{E}\eta, \sigma_2^2\right)$. Дисперсії цих ГС відомі, а самі ГС — незалежні. На рівні значущості α треба перевірити просту основну гіпотезу $H_0 : \mathbb{E}\xi = \mathbb{E}\eta$ проти однієї з альтернативних: $H_1 : \mathbb{E}\xi > \mathbb{E}\eta$ або $H_2 : \mathbb{E}\xi < \mathbb{E}\eta$. Статистикою критерію є $\zeta = \frac{\overline{\xi}-\overline{\eta}}{\sqrt{\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}}}$. Вона має гауссівський розподіл з $\mathbb{D}\zeta = 1$, причому, якщо H_0 справджується, то $\mathbb{E}\zeta = 0$. Критична область є правосторонньою у випадку $H_1 : \mathbb{E}\xi > \mathbb{E}\eta$ та лівосторонньою у випадку $H_2 : \mathbb{E}\xi < \mathbb{E}\eta$, її межа знаходиться з таблиці значень функції Лапласа як квантиль $\mathrm{N}(0, 1)$ рівня α для лівосторонньої критичної області та $1 - \alpha$ для правосторонньої.

Приклад. За вибіркою обсягом $n_1 = 14$ знайдемо середній розмір $\overline{x} = 182$ мм діаметра деталей, які виготовлено першим верстатом, а за вибіркою $n_2 = 9$ — середній розмір $\overline{y} = 185$ мм деталей, які виготовлено другим верстатом. Розмір діаметра деталі має гауссівський розподіл з дисперсією $\sigma_1^2 = 5$ мм^2 для першого верстату та $\sigma_2^2 = 7$ мм^2 для другого. Чи можна на рівні значущості 0.05 пояснити різницю вибіркових середніх випадковою помилкою?

Висуваємо основну гіпотезу $H_0 : \mathbb{E}\xi = \mathbb{E}\eta$ проти альтернативної $H_1 : \mathbb{E}\xi < \mathbb{E}\eta$, оскільки $\overline{x} < \overline{y}$. Критична область буде лівосторонньою. Знайдемо значення статистики критерію:

$$\eta_{\text{зн}} = \frac{182 - 185}{\sqrt{\frac{5}{14} + \frac{7}{9}}} \approx -2.81$$

Оскільки $t_{\text{кр}} = -1.64 > \eta_{\text{зн}}$, то на рівні значущості 0.05 дані суперечать гіпотезі про рівність математичних сподівань двох ГС: тобто, різницю між вибірковими середніми не можна вважати випадковою.

Перевірка гіпотези про рівність дисперсій двох гауссівських ГС при відомих або невідомих математичних сподіваннях

Дано дві реалізації обсягами n_1 та n_2 відповідно двох гауссівських ГС $\xi \sim \mathrm{N}\left(\mathbb{E}\xi, \sigma_1^2\right)$ та $\eta \sim \mathrm{N}\left(\mathbb{E}\eta, \sigma_2^2\right)$. Дисперсії цих ГС невідомі, а самі ГС — незалежні. На рівні значущості α треба перевірити просту основну гіпотезу $H_0 : \mathbb{D}\xi = \mathbb{D}\eta$ проти однієї з альтернативних: $H_1 : \mathbb{D}\xi > \mathbb{D}\eta$ або $H_2 : \mathbb{D}\xi < \mathbb{D}\eta$.

Припустимо, що математичні сподівання $\mathbb{E}\xi$ та $\mathbb{E}\eta$ відомі. Розглянемо дві статистики:

$$h_1 = \frac{n_1 \mathbb{D}^* \xi}{\sigma_1^2} \sim \chi^2_{n_1},\ h_2 = \frac{n_2 \mathbb{D}^* \eta}{\sigma_2^2} \sim \chi^2_{n_2}$$

Оскільки ξ та η незалежні, то h_1 та h_2 — також. Якщо H_0 справджується, то $\gamma = \frac{h_1/n_1}{h_2/n_2} = \frac{\mathbb{D}^*\xi}{\mathbb{D}^*\eta} \sim \mathrm{F}(n_1, n_2)$.

Якщо ж математичні сподівання $\mathbb{E}\xi$ та $\mathbb{E}\eta$ невідомі, то

$$h_1 = \frac{(n_1 - 1)\mathbb{D}^{**}\xi}{\sigma_1^2} \sim \chi^2_{n_1-1},\ h_2 = \frac{(n_2 - 1)\mathbb{D}^{**}\eta}{\sigma_2^2} \sim \chi^2_{n_2-1}$$

Якщо H_0 справджується, то $\gamma = \frac{h_1/(n_1-1)}{h_2/(n_2-1)} = \frac{\mathbb{D}^{**}\xi}{\mathbb{D}^{**}\eta} \sim \mathrm{F}(n_1-1, n_2-1)$.

В обох випадках критична область є правосторонньою чи лівосторонньою в залежності від альтернативної гіпотези, як і у попередніх прикладах. Межа критичної області береться з таблиці квантилів розподілу Фішера-Снедекора. Нагадаємо, що за властивістю цього розподілу $\frac{1}{\gamma} \sim \mathrm{F}(n_2, n_1)$ (або, відповідно, $\mathrm{F}(n_2-1, n_1-1)$).

Приклад. На двох токарних верстатах обробляються деталі. Відібрані дві проби: з деталей, що зроблені на першому верстаті, $n_1 = 15$, на другому верстаті — $n_2 = 18$. За обома вибірками обчислені значення виправлених вибіркових дисперсій розміру деталей: $(\mathbb{D}^{**}\xi)_{\text{зн}} = 8.5$ мм^2 для першого верстата та $(\mathbb{D}^{**}\eta)_{\text{зн}} = 6.3$ мм^2 для другого. Розмір деталі, що виготовляється верстатами, має гауссівський розподіл. На рівні значущості 0.05 з'ясувати, чи можна вважати, що верстати мають різну точність.

Маємо нульову гіпотезу $H_0 : \mathbb{D}\xi = \mathbb{D}\eta$ — дисперсії розміру деталі на різних верстатах є рівними. В якості альтернативної гіпотези варто брати $H_1 : \mathbb{D}\xi > \mathbb{D}\eta$, бо значення виправленої вибіркової дисперсії для першого верстата більше. Статистика критерію γ, якщо H_0 справджується, має розподіл $\mathrm{F}(n_1-1, n_2-1) = \mathrm{F}(14, 17)$. Значення статистики $\gamma_{\text{зн}} = \frac{(\mathbb{D}^{**}\xi)_{\text{зн}}}{(\mathbb{D}^{**}\eta)_{\text{зн}}} \approx 1.37$. За таблицею квантилів розподілу Фішера-Снедекора на рівні значущості 0.05 при кількості ступенів вільності $n_1 - 1 = 14$ та $n_2 - 1 = 17$ знаходимо $t_{\text{кр}} = 2.33$. Згідно з альтернативною гіпотезою, критична область знаходиться справа від $t_{\text{кр}}$ (тобто, є правосторонньою),тому робимо висновок, що дослідні дані на рівні значущості 0.05 не суперечать основній гіпотезі, оскільки $\gamma_{\text{зн}} < t_{\text{кр}}$. Це означає, що дослідні дані не дозволяють вважати, що верстати мають різну точність.

Перевірка гіпотези про рівність математичних сподівань двох гауссівських ГС при невідомих, але рівних дисперсіях

Дано дві реалізації обсягами n_1 та n_2 відповідно двох гауссівських ГС $\xi \sim \mathrm{N}\left(\mathbb{E}\xi, \sigma_1^2\right)$ та $\eta \sim \mathrm{N}\left(\mathbb{E}\eta, \sigma_2^2\right)$. Дисперсії цих ГС невідомі, але рівні ($\sigma_1 = \sigma_2 = \sigma$), а самі ГС — незалежні. На рівні значущості α потрібно перевірити просту основну гіпотезу $H_0 : \mathbb{E}\xi = \mathbb{E}\eta$ проти однієї з альтернативних гіпотез $H_1 : \mathbb{E}\xi > \mathbb{E}\eta$ або $H_2 : \mathbb{E}\xi < \mathbb{E}\eta$.

Оскільки дисперсії невідомі, то ми можемо побудувати тільки їх точкові оцінки $\mathbb{D}^{**}\xi$ та $\mathbb{D}^{**}\eta$, тоді

$$h_1 = \frac{(n_1-1)\mathbb{D}^{**}\xi}{\sigma^2} \sim \chi^2_{n_1-1},\ h_2 = \frac{(n_2-1)\mathbb{D}^{**}\eta}{\sigma^2} \sim \chi^2_{n_2-1}$$

Тоді $h = h_1 + h_2 = \frac{1}{\sigma^2}\left((n_1-1)\mathbb{D}^{**}\xi + (n_2-1)\mathbb{D}^{**}\eta\right) \sim \chi^2_{n_1+n_2-2}$ за властивостями розподілу хі-квадрат. Якщо $H_0 : \mathbb{E}\xi = \mathbb{E}\eta$ справджується, то $\zeta = \frac{\bar{\xi}-\bar{\eta}}{\sqrt{\frac{\sigma^2}{n_1}+\frac{\sigma^2}{n_2}}} = \frac{\bar{\xi}-\bar{\eta}}{\sigma\sqrt{\frac{n_1+n_2}{n_1 n_2}}} \sim \mathrm{N}(0,1)$.

В якості статистики критерію береться

$$\gamma = \frac{\zeta}{\sqrt{h/(n_1+n_2-2)}} = \frac{\bar{\xi}-\bar{\eta}}{\sqrt{\frac{n_1+n_2}{n_1 n_2}}\sqrt{\frac{(n_1-1)\mathbb{D}^{**}\xi+(n_2-1)\mathbb{D}^{**}\eta}{n_1+n_2-2}}} \sim St_{n_1+n_2-2}$$

Критична область є правосторонньою чи лівосторонньою в залежності від альтернативної гіпотези. Межа критичної області береться з таблиці квантилів розподілу Стьюдента.

Зауваження. Якщо про рівність дисперсій заздалегідь нічого не відомо, треба перевіряти гіпотезу про їх рівність на тому ж самому рівні значущості.

Приклад. Є дані про витрати сировини на одиницю продукції за старою технологією та за новою:

	стара			нова		
Витрати сировини	304	307	308	303	304	306
Кількість виробів	1	4	4	2	6	4

Припустивши, що відповідні генеральні сукупності мають гауссівський розподіл, на рівні значущості 0.05 відповісти на питання: чи дає нова технологія економію середньої витрати сировини на один виріб?

Потрібно перевірити гіпотезу про рівність математичних сподівань. Оскільки дисперсії невідомі, перевірити також треба гіпотезу про їх рівність. Знайдемо значення вибіркових середніх та виправлених вибіркових дисперсій, позначивши ξ ГС, що відповідає старій технології, а η — новій:

$$\left(\overline{\xi}\right)_{\text{зн}} = \overline{x} = \frac{1}{9}\left(304 \cdot 1 + 307 \cdot 4 + 308 \cdot 4\right) \approx 307.11$$

$$(\mathbb{D}^{**}\xi)_{\text{зн}} = \frac{1}{9-1}\left((304 - 307.11)^2 \cdot 1 + (307 - 307.11)^2 \cdot 4 + (308 - 307.11)^2 \cdot 4\right) \approx 1.611$$

$$\left(\overline{\eta}\right)_{\text{зн}} = \overline{y} \approx 304.5,\ (\mathbb{D}^{**}\eta)_{\text{зн}} \approx 1.364$$

Розглянемо основну гіпотезу $H_0 : \mathbb{D}\xi = \mathbb{D}\eta$ та альтернативну $H_1 : \mathbb{D}\xi > \mathbb{D}\eta$, бо $(\mathbb{D}^{**}\xi)_{\text{зн}} > (\mathbb{D}^{**}\eta)_{\text{зн}}$. Критична область правостороння, за таблицею для $\mathrm{F}(8, 11)$ знаходимо її межу — $t_{\text{кр}} = 2.95$, тобто, критична область — $(2.95; +\infty)$. Значення статистики критерію для перевірки гіпотези про рівність дисперсій $\frac{(\mathbb{D}^{**}\xi)_{\text{зн}}}{(\mathbb{D}^{**}\eta)_{\text{зн}}} = \frac{1.611}{1.364} \approx 1.181 < t_{\text{кр}}$. Отже, гіпотеза про рівність дисперсій не відхиляється на рівні значущості 0.05.

Вправа. Перевірити, що якби у вибірці, що відповідає новій технології, був один виріб з витратою сировини 308, то гіпотезу про рівність дисперсій довелося б відхилити.

Перейдемо до перевірки гіпотези про рівність математичних сподівань $H_0 : \mathbb{E}\xi = \mathbb{E}\eta$ проти альтернативної $H_1 : \mathbb{E}\xi > \mathbb{E}\eta$, бо $\overline{x} > \overline{y}$. Статистикою для перевірки буде

$$\gamma = \frac{\overline{\xi} - \overline{\eta}}{\sqrt{\frac{n_1+n_2}{n_1 n_2}}\sqrt{\frac{(n_1-1)\mathbb{D}^{**}\xi + (n_2-1)\mathbb{D}^{**}\eta}{n_1+n_2-2}}} = \frac{\overline{\xi} - \overline{\eta}}{\sqrt{\frac{9+12}{9\cdot 12}}\sqrt{\frac{(9-1)\mathbb{D}^{**}\xi + (12-1)\mathbb{D}^{**}\eta}{9+12-2}}} \sim St_{n_1+n_2-2} = St_{19}$$

$\gamma_{\text{зн}} \approx 4.885$. За таблицею квантилів розподілу Стьюдента знайдемо $t_{\text{кр}} = 1.729$, причому критична область є правосторонньою. Отже, на рівні значущості 0.05 гіпотеза про рівність математичних сподівань відхиляється, тому можна вважати, що нова технологія дійсно дає економію середньої кількості сировини для виготовлення одного виробу.

Зауваження. Якщо дисперсії двох ГС невідомі і немає підстав вважати їх рівними, має місце *проблема Беренса-Фішера.* В такому випадку також застосовується статистика, що має розподіл Стьюдента, але з кількістю ступенів вільності (з округленням до найближчого цілого)

$$k = (n_1 + n_2 - 2) \cdot \left(\frac{1}{2} + \frac{(\mathbb{D}^{**}\xi)_{\text{зн}}(\mathbb{D}^{**}\eta)_{\text{зн}}}{(\mathbb{D}^{**}\xi)^2_{\text{зн}} + (\mathbb{D}^{**}\eta)^2_{\text{зн}}}\right)$$

Розглянемо другий множник. Оскільки $\frac{(\mathbb{D}^{**}\xi)_{\text{зн}}(\mathbb{D}^{**}\eta)_{\text{зн}}}{(\mathbb{D}^{**}\xi)^2_{\text{зн}} + (\mathbb{D}^{**}\eta)^2_{\text{зн}}} \le \frac{1}{2}$ та рівність досягається при $(\mathbb{D}^{**}\xi)_{\text{зн}} = (\mathbb{D}^{**}\eta)_{\text{зн}}$, то зі збільшенням різниці між значенням виправлених вибіркових дисперсій цей доданок зменшується. Якщо ці значення розрізняються досить суттєво, то кількість степенів свободи зменшується майже вдвічі порівняно з $n_1 + n_2 - 2$. Це, в свою чергу, звужує критичну область (та, відповідно, ймовірність потрапити туди, оскільки вона розміщена на «хвостах» розподілу). До того ж, розкид можливих значень $\overline{x} - \overline{y}$ буде визначатися в основному тією ГС, яка має більшу дисперсію, тому дані з ГС з меншою дисперсією «губляться». Ці два фактори разом призводять до більшої недовіри висновку щодо основної гіпотези, ніж у випадку відомих дисперсій.

Перевірка гіпотези про рівність ймовірностей біномних генеральних сукупностей при великому обсязі вибірки

Дано дві реалізації обсягами n_1 та n_2 відповідно двох незалежних ГС ξ та η, що мають розподіл Бернуллі, тобто $\xi \sim \mathrm{Bin}(1, p_1)$, $\eta \sim \mathrm{Bin}(1, p_2)$, причому обсяги цих вибірок достатньо великі, щоб можна було застосувати асимптотичну нормальність оцінки ймовірності «успіху» в одному випробуванні. На рівні значущості α потрібно перевірити основну гіпотезу $H_0 : p_1 = p_2$ проти альтернативної $H_1 : p_1 > p_2$ або $H_2 : p_1 < p_2$. Нехай m_1 та m_2 — кількості успішних випробувань у кожній з реалізацій. Тоді за ЦГТ $\frac{m_1}{n_1} = p_1^*$ та $\frac{m_2}{n_2} = p_2^*$ наближено мають розподіли $\mathrm{N}\left(p_1, \frac{p_1 q_1}{n_1}\right)$ та $\mathrm{N}\left(p_2, \frac{p_2 q_2}{n_2}\right)$ відповідно, а $p_1^* - p_2^*$, якщо H_0 справджується — розподіл $\mathrm{N}\left(0, \frac{p_1 q_1}{n_1} + \frac{p_2 q_2}{n_2}\right)$. Також, якщо H_0 справджується, то можна переписати $\frac{p_1 q_1}{n_1} + \frac{p_2 q_2}{n_2}$

як $pq\left(\frac{1}{n_1}+\frac{1}{n_2}\right)$, тому в якості статистики для перевірки основної гіпотези беруть

$$\gamma=\frac{p_1^*-p_2^*}{\sqrt{p^*(1-p^*)\left(\frac{1}{n_1}+\frac{1}{n_2}\right)}},\text{ де } p^*=\frac{m_1+m_2}{n_1+n_2}$$

Оскільки вона має наближено розподіл $\mathrm{N}(0,1)$, то критичне значення знаходять за таблицею значень функції Лапласа як квантиль рівня α для лівосторонньої критичної області та рівня $1-\alpha$ для правосторонньої. Критична область береться правосторонньою чи лівосторонньою в залежності від альтернативної гіпотези.

Приклад. 1. Порівняти частку браку на рівні значущості 0.1 для двох партій виробів:

номер партії	обсяг вибірки	к-сть бракованих	частка бракованих
1	$n_1=200$	$m_1=5$	$(p_1^*)_{\text{зн}}=\frac{m_1}{n_1}=\frac{1}{40}$
2	$n_2=300$	$m_2=10$	$(p_2^*)_{\text{зн}}=\frac{m_2}{n_2}=\frac{1}{30}$

Потрібно перевірити гіпотезу $H_0: p_1=p_2$ проти альтернативної $H_1: p_1<p_2$. Обсяги обох вибірок великі, тому розподіли p_1^* та p_2^* можна наближувати нормальним. $(p^*)_{\text{зн}}=\frac{5+10}{200+300}=\frac{3}{100}$, тому $\gamma_{\text{зн}}=\frac{\frac{1}{40}-\frac{1}{30}}{\sqrt{\frac{3}{100}\frac{97}{100}\left(\frac{1}{200}+\frac{1}{300}\right)}}\approx-0.535$. Критичне значення — $t_{\text{кр}}=-1.28$. Оскільки критична область є лівосторонньою, то на рівні значущості 0.1 гіпотеза про однакову частку бракованих виробів не відхиляється.

2. Контрольну роботу з теорії ймовірностей виконували студенти двох груп другого курсу ІПСА. Першій групі викладач запропонував 105 задач, з яких правильно було розв'язано 60, а другій — 140 задач, з яких правильно розв'язаних вийшло 69. На рівні значущості 0.02 перевірити гіпотезу про відсутність різниці в засвоєнні дисципліни «Теорія ймовірностей» студентами різних груп.

Нехай p_1 та p_2 — ймовірності розв'язання задачі студентами першої та другої групи відповідно. Потрібно перевірити гіпотезу $H_0: p_1=p_2$ проти альтернативної $H_1: p_1>p_2$, оскільки $(p_1^*)_{\text{зн}}=\frac{m_1}{n_1}=\frac{60}{105}\approx0.571$, $(p_2^*)_{\text{зн}}=\frac{m_2}{n_2}=\frac{69}{140}\approx0.493$. Також, $(p^*)_{\text{зн}}=\frac{60+69}{105+140}\approx 0.527$. Тому $\gamma_{\text{зн}}=\frac{0.571-0.493}{\sqrt{0.527\cdot(1-0.527)\cdot\left(\frac{1}{105}+\frac{1}{140}\right)}}\approx1.21$. Критична область є правосторонньою, її межа $t_{\text{кр}}=2.05>\gamma_{\text{зн}}$. Отже, на рівні значущості 0.02 дослідні дані не суперечать гіпотезі про те, що різниці в засвоєнні дисципліни студентами немає.

Розділ 10

Елементи регресійного аналізу

Часто за дослідними даними треба визначити, як залежить випадкова величина, яку ми спостерігаємо, від однієї чи кількох інших випадкових величин. Найзагальніший випадок — *статистична залежність*, як-от $\eta = \xi_1 + \xi_2$. Є сенс розглядати математичне сподівання однієї величини при фіксованому значенні іншої (чи інших). Наприклад, вартість квартири залежить від площі, поверху, району та багатьох інших параметрів, але не є функцію (в звичному розумінні) від цих параметрів: якби функціональна залежність була, то кожному набору цих параметрів відповідало б одне значення вартості — у реальному світі такого не відбувається. До того ж, врахувати всі можливі впливи на вартість квартири неможливо: площу чи поверх дізнатися просто, але може статися, що продавець встановлює вартість на основі якихось власних міркувань, які «зі сторони» визначити чи оцінити неможливо. Натомість, можна розглядати функціональну залежність середнього значення (математичного сподівання) вартості від цих параметрів. Такий опис вже є більш реалістичним: хоча саме середнє значення спостерігати неможливо, але можна збирати інформацію про значення вартості при різних значеннях деякого фіксованого набору параметрів.

10.1 Математична модель лінійної регресії

Нехай спостерігається деяка випадкова величина η, яка залежить від іншої випадкової величини ξ (або вектору $\vec{\xi}$), значення яких теж спостерігаються.

Означення 10.1.1. Функція $f(x) = \mathbb{E}(\eta/\xi = x)$ (або $f(\vec{x}) = \mathbb{E}(\eta/\vec{\xi} = \vec{x})$) називається *лінією регресії* η на ξ (або $\vec{\xi}$).

Термін «регресія» вперше з'явився в роботі Френсіса Гальтона (англійського антрополога та статистика), який у 1885 році досліджував зв'язок між зростом дітей та їх батьків. Виявилося, що в середньому діти високих батьків не були такими вже й високими, і навпаки. Він назвав це «регресією (поверненням) до посереднього». Зокрема, Гальтон побудував таку лінію регресії (для зросту в дюймах):

$$\mathbb{E}\left(\text{зріст дитини}/\xi_1 = x \text{ (зріст батька)}, \xi_2 = y \text{ (зріст матері)}\right) = 22.3 + 0.38x + 0.28y$$

10.1.1 Однофакторна лінійна регресія

Для початку будемо вважати, що η, яку називають *вихідною величиною* або *відкликом* залежить лише від однієї випадкової величини ξ, яку називають *вхідною* або *фактором* чи *предиктором*. Припустимо, що було проведено n експериментів зі значеннями фактору ξ $x_1, x_2, ..., x_n$ та відповідними значенням відклику η $\eta_1, \eta_2, ..., \eta_n$. Позначимо $\varepsilon_i = \eta_i - \mathbb{E}(\eta/\xi = x_i)$ — це похибки спостережень, що дорівнюють різниці між отриманим значенням та усередненим значенням η при цьому значенні ξ. Про розподіл $\vec{\varepsilon} = (\varepsilon_1, \varepsilon_2, ..., \varepsilon_n)^T$, як правило, мало що відомо: припускається, що його координати незалежні, однаково розподілені, їх дисперсія невідома, але математичне сподівання нульове: $\mathbb{E}\varepsilon_i = \mathbb{E}\eta_i - \mathbb{E}(\mathbb{E}(\eta/\xi = x_i)) = \mathbb{E}(\eta/\xi = x_i) - \mathbb{E}(\eta/\xi = x_i) = 0$.

Необхідно за значеннями $x_1, x_2, ..., x_n$ та $\eta_1, \eta_2, ..., \eta_n$ оцінити $f(x)$. Величини $x_1, x_2, ..., x_n$ вважають невипадковими, тому вся випадковість зосереджена в $\varepsilon_1, \varepsilon_2, ..., \varepsilon_n$ та $\eta_1, \eta_2, ..., \eta_n$.

Спочатку треба визначити вигляд цієї функції. Часто це випливає з постановки задачі або з'ясовується після візуалізації отриманих даних. Якщо мова йде про *лінійну регресію*, то вважають, що $f(x) = \sum\limits_{k=1}^{m} \beta_k \psi_k(x)$, де $\psi_k(x)$ — деякі функції, що називають *базисними*. Окремим часто застосованим випадком є *поліноміальна регресія* з $f(x) = \sum\limits_{k=0}^{m} \beta_k x^k$. Задача регресійного аналізу полягає в оцінці невідомих параметрів моделей наведеного виду.

Зауваження. Назва «лінійна регресія» походить від того, що невідомі параметри $\beta_1, \beta_2, ..., \beta_n$ входять до $f(x)$ лінійно. Іноді (не зовсім коректно) вважається, що лінійною називають тільки ту модель, яка також є лінійною за x.

10.1.2 Метод найменших квадратів

Оцінки $\vec{\beta} = (\beta_1, \beta_2, ..., \beta_m)^T$ можна знайти за допомогою методу максимальної правдоподібності, застосованого до вектору $\eta = (\eta_1, \eta_2, ..., \eta_n)^T$. Оскільки $\eta_i = \varepsilon_i + f(x_i)$, то розподіл η_i буде таким самим, як ε_i, але з математичним сподіванням $f(x_i)$.

Припустимо, що $\vec{\varepsilon} \sim \mathrm{N}(\vec{0}, \sigma^2 \mathbb{I})$, де $\mathbb{I}$ — одинична матриця $n \times n$. Тоді

$$\mathcal{L}\left(\eta_1, \eta_2, ..., \eta_n, \vec{\beta}\right) = \frac{1}{\left(\sqrt{2\pi}\sigma\right)^n} \exp\left\{-\frac{1}{2\sigma^2} \sum_{i=1}^{n} (\eta_i - f(x_i))^2\right\}$$

При фіксованому значенні σ максимум функції правдоподібності по $\vec{\beta}$ досягається при найменшому значенні $\sum\limits_{i=1}^{n} (\eta_i - f(x_i))^2 = \sum\limits_{i=1}^{n} \varepsilon_i^2$. Цей окремий випадок ММП називається *методом найменших квадратів (МНК)*.

Означення 10.1.2. *Оцінкою методу найменших квадратів (ОМНК)* для параметрів $\vec{\beta} = (\beta_1, \beta_2, ..., \beta_m)^T$ називається такий набір цих параметрів, при яких досягається мінімум $\sum\limits_{i=1}^{n} \varepsilon_i^2$.

Якщо ми знайдемо оцінки $\vec{\beta}^* = (\beta_1^*, \beta_2^*, ..., \beta_m^*)^T$, то також знайдемо оцінку функції $f^*(x)$.

Зауваження. На практиці іноді можуть припускати, що координати $\vec{\varepsilon}$ мають розподіл Лапласа зі щільністю $f_\varepsilon(x) = \frac{\alpha}{2} e^{-\alpha|x|}$. В такому випадку знаходження максимуму функції правдоподібності зводиться до мінімізації суми $\sum\limits_{i=1}^{n} |\eta_i - f(x_i)|$ — це вже не буде методом найменших квадратів.

Приклад. Нехай $f(x) = \beta$, тобто ми хочемо оцінити параметр лінії регресії, що є константою. Маємо $\eta_i = \beta + \varepsilon_i,\ i = 1, ..., n$. По суті, ця задача є задачею оцінки невідомого математичного сподівання за вибіркою з незалежними й однаково розподіленими гауссівськими випадковими величинами $\eta_1, \eta_2, ..., \eta_n$. Знайдемо розв'язок задачі мінімізації $\sum\limits_{i=1}^{n} \varepsilon_i^2 = \sum\limits_{i=1}^{n} (\eta_i - \beta)^2 \to \min$:

$$\frac{d}{d\beta}\left(\sum_{i=1}^{n} (\eta_i - \beta)^2\right) = -2\sum_{i=1}^{n} (\eta_i - \beta) = 0 \Rightarrow \beta = \frac{1}{n}\sum_{i=1}^{n} \eta_i$$

$$\frac{d^2}{d\beta^2}\left(\sum_{i=1}^{n} (\eta_i - \beta)^2\right) = 2n > 0$$

Отже, $\beta^* = \frac{1}{n} \sum\limits_{i=1}^{n} \eta_i = \overline{\eta}$. Такий само результат було отримано і при розв'язанні цієї задачі як задачі оцінки математичного сподівання.

Вправа. Показати, що якщо координати вектора похибок ε_i мають розподіл Лапласа, то максимізація такої функції правдоподібності буде еквівалентна мінімізації суми $\sum\limits_{i=1}^{n} |\eta_i - \beta|$, а розв'язком цієї задачі буде вибіркова медіана.

10.1.3 Багатофакторна (множинна) лінійна регресія

Тепер розглянемо випадок, коли η залежить від m факторів $\vec{\xi} = (\xi_1, \xi_2, ..., \xi_m)^T$.

Означення 10.1.3. *Простою лінійною багатофакторною регресійною моделлю* називається функція регресії вигляду $f(\vec{x}) = \mathbb{E}(\eta/\vec{\xi} = \vec{x}) = \beta_0 + \sum\limits_{i=1}^{m} \beta_i x_i$.

Для спрощення цю модель іноді записують як $f(\vec{x}) = \sum\limits_{i=0}^{m} \beta_i x_i$, де фактор x_0 завжди дорівнює 1. Надалі без втрати загальності будемо вважати, що розглядається модель $f(\vec{x}) = \sum\limits_{i=1}^{m} \beta_i x_i$, в якій константного фактору або немає, або він присутній як один з факторів, що завжди дорівнює 1.

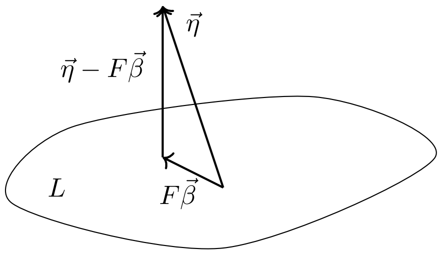
Нехай експеримент проводився n разів, а на j-тому проведенні фактори набули значень $\vec{x}^{(j)} = \left(x_1^{(j)}, x_2^{(j)}, ..., x_m^{(j)}\right)^T$, $j = 1, ..., n$. Як правило, $n > m$. Також, після проведення n експериментів маємо вектор відкликів $\vec{\eta} = (\eta_1, \eta_2, ..., \eta_n)^T$, де $\eta_j = f\left(\vec{x}^{(j)}\right) + \varepsilon_j = \sum\limits_{i=1}^{m} \beta_i x_i^{(j)} + \varepsilon_j$, $j = 1, ..., n$. Цей вектор можна записати як $\vec{\eta} = F\vec{\beta} + \vec{\varepsilon}$, де

$$F = \begin{pmatrix} x_1^{(1)} & x_2^{(1)} & \cdots & x_m^{(1)} \\ x_1^{(2)} & x_2^{(2)} & \cdots & x_m^{(2)} \\ \vdots & \vdots & \ddots & \vdots \\ x_1^{(n)} & x_2^{(n)} & \cdots & x_m^{(n)} \end{pmatrix} \tag{1}$$

Ця матриця називається *матрицею плану.* Щоб знайти оцінку $\vec{\beta}$ за допомогою МНК, зробимо **припущення**: $\mathrm{rang} F = m$ та $\vec{\varepsilon} \sim \mathrm{N}(\vec{0}, \sigma^2 \mathbb{I})$. Перше припущення фактично означає, що фактори є лінійно незалежними, а друге — що можна вимагати як незалежності похибок, так і їх некорельованості. ОМНК для $\vec{\beta}$ буде розв'язком задачі мінімізації

$$S(\beta) = \sum_{i=1}^{n} \varepsilon_i^2 = \|\vec{\varepsilon}\|^2 = \|\vec{\eta} - F\vec{\beta}\|^2 \to \min$$

$S(\beta)$ — це квадрат відстані від точки $\vec{\eta} \in \mathbb{R}^n$ (коректніше було б писати $\vec{\eta}_{\text{зн}}$, але це ускладнить позначення) до точки $F\vec{\beta}$, що належить підпростору $L = \{F\vec{x} : \vec{x} \in \mathbb{R}^m\}$. Мінімальною ця відстань буде тоді, коли вектор $\vec{\eta} - F\vec{\beta}$ буде ортогональним до усіх векторів з L.



Отже, $\forall\ \vec{x} \in \mathbb{R}^m : \left(F\vec{x}, \vec{\eta} - F\vec{\beta}\right) = 0$. Оскільки $\left(F\vec{x}, \vec{\eta} - F\vec{\beta}\right) = \left(\vec{x}, F^T(\vec{\eta} - F\vec{\beta})\right)$, то, з умови довільності $\vec{x}$ можна взяти, наприклад, вектори стандартного базису $\mathbb{R}^m$ та отримати, що $F^T(\vec{\eta} - F\vec{\beta}) = 0 \Leftrightarrow F^T\vec{\eta} = F^T F\vec{\beta}$. Умова $\mathrm{rang} F = m$ забезпечує існування $(F^T F)^{-1}$: нехай $\vec{x} \in \mathbb{R}^m$ та $F^T F\vec{x} = \vec{0}$, тоді $\vec{x}^T F^T F\vec{x} = (F\vec{x})^T F\vec{x} = \vec{0}$, звідки $F\vec{x} = \vec{0}$, а з умови $\mathrm{rang} F = m$ маємо наслідок $\vec{x} = \vec{0}$, тому $\mathrm{Ker} F^T F = \{\vec{0}\}$. Таким чином, отримуємо ОМНК для $\vec{\beta}$:

$$\vec{\beta}^* = (F^T F)^{-1} F^T \vec{\eta} \tag{2}$$

Цю рівність іноді називають *нормальним рівнянням.*

Означення 10.1.4. Матрицю $A = F^T F$ називають *інформаційною матрицею*, а $A^{-1} = (F^T F)^{-1}$ — *дисперсійною матрицею (Фішера).*

Розглянемо властивості інформаційної матриці A:

1. F — матриця розміру $n \times m$, F^T — розміру $m \times n$, тому A — квадратна матриця розміру $m \times m$.
2. A — симетрична: $A^T = (F^T F)^T = F^T (F^T)^T = F^T F = A$.
3. A — невід'ємно визначена: $\forall\, \vec{x} \in \mathbb{R}^m : (A\vec{x}, \vec{x}) = \left(F^T F\vec{x}, \vec{x}\right) = (F\vec{x}, F\vec{x}) = \|F\vec{x}\|^2 \geq 0$.
4. Оскільки A — невід'ємно визначена та невироджена, то вона додатно визначена ($A > 0$) і тому $\exists!\; B > 0 : B^2 = A$, тобто, існує і єдиний квадратний корінь $\sqrt{A}$.

Зауваження. Корінь із невід'ємно визначеної матриці будується за загальним принципом побудови функцій від симетричних матриць. Нехай M — невід'ємно визначена ($M \geq 0$) матриця $n \times n$, тоді її власні числа $\lambda_1 \geq \lambda_2 \geq ... \geq \lambda_n \geq 0$. Існує ортогональна матриця U, для якої $M = U \cdot \mathrm{diag}(\lambda_1, ..., \lambda_n) \cdot U^{-1}$. Для $S = U \cdot \mathrm{diag}(\sqrt{\lambda_1}, ..., \sqrt{\lambda_n}) \cdot U^{-1}$ маємо $S^2 = \left(\mathrm{diag}(\sqrt{\lambda_1}, ..., \sqrt{\lambda_n})\right)^2 = \mathrm{diag}(\lambda_1, ..., \lambda_n) = M$, причому $\sqrt{\lambda_1} \geq \sqrt{\lambda_2} \geq ... \geq \sqrt{\lambda_n} \geq 0$, тому $S \geq 0$. У випадку $M > 0$ мали б $\lambda_n > 0$, звідки $\sqrt{\lambda_n} > 0$ і $S > 0$. Зрозуміло, що $S = \sqrt{M}$ теж є симетричною матрицею. Також за побудовою $\sqrt{M}$ комутує з M та M^{-1} (якщо вона існує): $\sqrt{M} \cdot M = M \cdot \sqrt{M}$, $\sqrt{M} \cdot M^{-1} = M^{-1} \cdot \sqrt{M}$.

Приклад. Знайти оцінки параметрів простої однофакторної лінійної регресійної моделі, що має вигляд $f(x) = \beta_0 + \beta_1 x$. Нехай є n спостережень відклику $\vec{\eta}_{\text{зн}} = (y_1, y_2, ..., y_n)^T$ та фактора — $\vec{\xi}_{\text{зн}} = (x_1, x_2, ..., x_n)$. Матрицю плану для спрощення можна записати з індексами x_i замість $x^{(i)}$ як $F = \begin{pmatrix} 1 & x_1 \\ \vdots & \vdots \\ 1 & x_n \end{pmatrix}$. Знайдемо дисперсійну матрицю:

$$A = F^T F = \begin{pmatrix} 1 & \cdots & 1 \\ x_1 & \cdots & x_n \end{pmatrix} \cdot \begin{pmatrix} 1 & x_1 \\ \vdots & \vdots \\ 1 & x_n \end{pmatrix} = \begin{pmatrix} n & \sum\limits_{k=1}^{n} x_k \\ \sum\limits_{k=1}^{n} x_k & \sum\limits_{k=1}^{n} x_k^2 \end{pmatrix}$$

$$\det A = n \cdot \sum_{k=1}^{n} x_k^2 - \left(\sum_{k=1}^{n} x_k\right)^2 = n^2 \cdot \left(\frac{1}{n}\sum_{k=1}^{n} x_k^2 - \left(\frac{1}{n}\sum_{k=1}^{n} x_k\right)^2\right) = n^2 \cdot (\mathbb{D}^*\xi)_{\text{зн}}$$

$$A^{-1} = (F^T F)^{-1} = \frac{1}{\det A}\begin{pmatrix} \sum\limits_{k=1}^{n} x_k^2 & -\sum\limits_{k=1}^{n} x_k \\ -\sum\limits_{k=1}^{n} x_k & n \end{pmatrix}$$

$$F^T \vec{\eta}_{\text{зн}} = \begin{pmatrix} 1 & \cdots & 1 \\ x_1 & \cdots & x_n \end{pmatrix} \cdot \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix} = \begin{pmatrix} \sum\limits_{k=1}^{n} y_k \\ \sum\limits_{k=1}^{n} x_k y_k \end{pmatrix}$$

$$\vec{\beta}^*_{\text{зн}} = (F^T F)^{-1} F^T \vec{\eta}_{\text{зн}} = \frac{1}{\det A}\begin{pmatrix} \sum\limits_{k=1}^{n} y_k \cdot \sum\limits_{k=1}^{n} x_k^2 - \sum\limits_{k=1}^{n} x_k \cdot \sum\limits_{k=1}^{n} x_k y_k \\ -\sum\limits_{k=1}^{n} y_k \cdot \sum\limits_{k=1}^{n} x_k + n \cdot \sum\limits_{k=1}^{n} x_k y_k \end{pmatrix}$$

Якщо позначити $\sum\limits_{k=1}^{n} x_k = n \cdot \overline{x}$, $\sum\limits_{k=1}^{n} y_k = n \cdot \overline{y}$, $\sum\limits_{k=1}^{n} x_k y_k = n \cdot \overline{xy}$, $\sum\limits_{k=1}^{n} x_k^2 = n \cdot \overline{x^2}$, то

$$\vec{\beta}^*_{\text{зн}} = \frac{1}{n^2 \cdot (\mathbb{D}^*\xi)_{\text{зн}}}\begin{pmatrix} n^2 \cdot \overline{y} \cdot \overline{x^2} - n^2 \cdot \overline{x} \cdot \overline{xy} \\ -n^2 \overline{y} \cdot \overline{x} + n^2 \cdot \overline{xy} \end{pmatrix} = \frac{1}{\overline{x^2} - (\overline{x})^2}\begin{pmatrix} \overline{x^2} \cdot \overline{y} - \overline{x} \cdot \overline{xy} \\ \overline{xy} - \overline{x} \cdot \overline{y} \end{pmatrix}$$

Зауважимо, що $\overline{xy} - \overline{x} \cdot \overline{y}$ — це значення *вибіркового аналогу коваріації* $cov(\xi, \eta)$, який ще можна позначити $\mathbb{K}^*\xi\eta$. Цей результат доволі очікуваний: якби статистична залежність між ξ та η описувалася рівнянням $\eta = \beta_0 + \beta_1\xi$, то $cov(\xi, \eta) = cov(\xi, \beta_0 + \beta_1\xi) = \beta_1 cov(\xi, \xi) = \beta_1 \mathbb{D}\xi$.

Зауваження. До простої лінійної багатофакторної моделі можна звести дослідження багатьох моделей більш загального виду. Наприклад, однофакторну модель виду $f(x) = \sum\limits_{k=1}^{m} \beta_k \psi_k(x)$ можна розглядати як багатофакторну модель з факторами $x_k = \psi_k(x)$ (зокрема, це стосується поліноміальної регресії).

10.2 Статистичний аналіз лінійної регресійної моделі

10.2.1 Властивості оцінок параметрів

Будемо досліджувати властивості оцінок параметрів $\vec{\beta}$, отриманих за допомогою МНК за формулою (2): $\vec{\beta}^* = (F^T F)^{-1} F^T \vec{\eta}$.

Твердження 1. *Вектор $\vec{\beta}^*$ має гауссівський розподіл та є незміщеною оцінкою $\vec{\beta}$.*

Доведення. $\vec{\beta}^* - \vec{\beta} = (F^T F)^{-1} F^T \vec{\eta} - \vec{\beta} = \left[\vec{\eta} = F\vec{\beta} + \vec{\varepsilon}\right] = (F^T F)^{-1} F^T (F\vec{\beta} + \vec{\varepsilon}) - \vec{\beta} =$ $= (F^T F)^{-1} F^T F \vec{\beta} + (F^T F)^{-1} F^T \vec{\varepsilon} - \vec{\beta} = \vec{\beta} + (F^T F)^{-1} F^T \vec{\varepsilon} - \vec{\beta} = (F^T F)^{-1} F^T \vec{\varepsilon}$. Оскільки, за припущенням, $\vec{\varepsilon} \sim \mathrm{N}(\vec{0}, \sigma^2 \mathbb{I})$, то $\vec{\beta}^* = (F^T F)^{-1} F^T \vec{\varepsilon} + \vec{\beta}$ також має гауссівський розподіл. Знайдемо параметри цього розподілу:

$$\mathbb{E}\vec{\beta}^* = \mathbb{E}(F^T F)^{-1} F^T \vec{\varepsilon} + \vec{\beta} = (F^T F)^{-1} F^T \mathbb{E}\vec{\varepsilon} + \vec{\beta} = \vec{\beta}$$

$$\mathbb{K}\vec{\beta}^* = (F^T F)^{-1} F^T \mathbb{K}\vec{\varepsilon} \left((F^T F)^{-1} F^T\right)^T = (F^T F)^{-1} F^T \cdot \sigma^2 \mathbb{I} \cdot F (F^T F)^{-1} = \sigma^2 (F^T F)^{-1}$$

Отже, $\vec{\beta}^* \sim \mathrm{N}(\vec{\beta}, \sigma^2 A^{-1})$, де $A = F^T F$. Це, зокрема, пояснює назву матриці $(F^T F)^{-1}$ — «дисперсійна». Зауважимо, що значення σ^2 невідоме. Пізніше буде досліджено його оцінку. ▲

Твердження 2 (теорема Гаусса-Маркова)**.** *Незміщена оцінка $\vec{\beta}^*$, що знайдена за формулою (2), є ефективною в класі незміщених лінійних оцінок.*

Доведення. Нехай існує інша незміщена оцінка $\vec{\beta}^{**}$. Оскільки розглядається клас лінійних оцінок, то $\vec{\beta}^{**} = C\vec{\eta}$ для деякої матриці C. Розпишемо параметри розподілу цієї оцінки.

$$\vec{\beta} = \mathbb{E}\vec{\beta}^{**} = \mathbb{E}(C(F\vec{\beta} + \vec{\varepsilon})) = CF\vec{\beta} + C\mathbb{E}\vec{\varepsilon} = CF\vec{\beta}$$

звідки $CF = \mathbb{I}$ — одинична матриця.

$$\mathbb{K}\vec{\beta}^{**} = \mathbb{E}(\vec{\beta}^{**} - \mathbb{E}\vec{\beta}^{**})(\vec{\beta}^{**} - \mathbb{E}\vec{\beta}^{**})^T = \mathbb{E}(C\vec{\eta} - \vec{\beta})(C\vec{\eta} - \vec{\beta})^T =$$
$$= \mathbb{E}(C\vec{\eta} - CF\vec{\beta})(C\vec{\eta} - CF\vec{\beta})^T = C\mathbb{E}(\vec{\eta} - F\vec{\beta})(\vec{\eta} - F\vec{\beta})^T C^T = C\mathbb{E}\vec{\varepsilon}\vec{\varepsilon}^T C^T = \sigma^2 CC^T$$

Нехай $D = C - A^{-1}F^T = C - (F^T F)^{-1} F^T$, тоді $C = A^{-1}F^T + D$. Маємо $\mathbb{I} = CF = A^{-1}F^T F + DF = (F^T F)^{-1} F^T F + DF = \mathbb{I} + DF$, звідки $DF = 0$. Підставимо цей вираз для C в вираз для $\mathbb{K}\vec{\beta}^{**}$:

$$\mathbb{K}\vec{\beta}^{**} = \sigma^2 CC^T = \sigma^2 (A^{-1}F^T + D)(A^{-1}F^T + D)^T =$$
$$= \sigma^2 \left((A^{-1}F^T)(A^{-1}F^T)^T + D(A^{-1}F^T)^T + (A^{-1}F^T)D^T + DD^T\right) =$$
$$= \sigma^2 \left(A^{-1}F^T F (A^T)^{-1} + DF(A^T)^{-1} + A^{-1}(DF)^T + DD^T\right) =$$
$$= \sigma^2 \left(A^{-1} + DD^T\right) = \mathbb{K}\vec{\beta}^* + \sigma^2 DD^T$$

Звідси $\mathbb{K}\vec{\beta}^{**} - \mathbb{K}\vec{\beta}^* = \sigma^2 DD^T$. Оскільки $\forall\ \vec{x} : \left(DD^T\vec{x}, \vec{x}\right) = \left(D^T\vec{x}, D^T\vec{x}\right) = \|D^T\vec{x}\|^2 \geq 0$, то матриця $\sigma^2 DD^T \geq 0$ (невід'ємно визначена). Отже, $\mathbb{K}\vec{\beta}^{**} - \mathbb{K}\vec{\beta}^* \geq 0$, що доводить ефективність $\vec{\beta}^*$. ▲

Зауваження. Дослідження ефективності лише в класі незміщених лінійних оцінок інтуїтивно пояснюється тим, що якщо $\vec{\eta}$ лінійно залежить від $\vec{\beta}$, то й усі «хороші» оцінки $\vec{\beta}$ будуть лінійно залежати від $\vec{\eta}$.

Твердження 3. *Незміщена оцінка $\vec{\beta}^*$, що знайдена за формулою (2), є консистентною.*

Доведення. В твердженні 1 було показано, що $\vec{\beta}^* = (F^T F)^{-1} F^T \vec{\varepsilon} + \vec{\beta}$ і $\vec{\beta}^* - \vec{\beta} \sim \mathrm{N}(\vec{0}, \sigma^2 A^{-1})$. Матриця A^{-1} залежить від n, тому застосувати закон великих чисел для випадкових векторів в тому формулюванні, яке було наведено (ст. 76), не вдасться. Натомість, буде достатньо довести, що для $i = 1, ..., m$ $\mathbb{D}\left(\beta_i^* - \beta_i\right) = \mathbb{D}\beta_i^* \to 0, n \to \infty$.

Без втрати загальності можна вважати, що $\sigma = 1$, тоді на діагоналі матриці A^{-1} будуть знаходитися дисперсії $\mathbb{D}\beta_i^*$. $\forall\ i = 1, ..., m : \mathbb{D}\beta_i^* \leq \sum\limits_{k=1}^{m} \mathbb{D}\beta_k^* = \mathrm{tr}A^{-1}$. Нехай $\lambda_1 \geq ... \geq \lambda_m > 0$

— власні числа матриці A, тоді $\text{tr}A^{-1} = \frac{1}{\lambda_1} + ... + \frac{1}{\lambda_m} = \frac{\frac{\lambda_1\lambda_2...\lambda_m}{\lambda_1} + \frac{\lambda_1\lambda_2...\lambda_m}{\lambda_2} + ... + \frac{\lambda_1\lambda_2...\lambda_m}{\lambda_m}}{\lambda_1\lambda_2...\lambda_m}$. З лінійної алгебри відомо, що $\lambda_1\lambda_2...\lambda_m = \det A$, а доданки $\frac{\lambda_1\lambda_2...\lambda_m}{\lambda_j}$ в чисельнику — визначники мінорів матриці A, що отримані викреслюванням j-того стовпця та рядка — це випливає з формули для характеристичного полінома $\det(A - \lambda\mathbb{I})$ та фактів, що його коефіцієнти не залежать від базису, а в деякому базисі матриця A є діагональною. Тепер можна оцінити $\frac{\lambda_1\lambda_2...\lambda_m}{\lambda_1} + \frac{\lambda_1\lambda_2...\lambda_m}{\lambda_2} + ... + \frac{\lambda_1\lambda_2...\lambda_m}{\lambda_m} \le m \cdot \max\limits_{j=1,...,m} \frac{\lambda_1\lambda_2...\lambda_m}{\lambda_j}$.

Повернемося до матриці A та розглянемо її структуру:

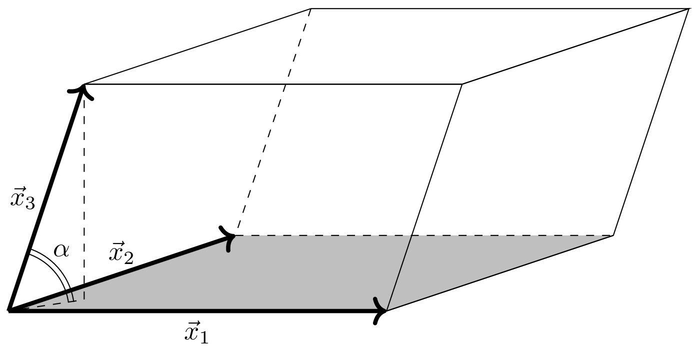
$$A = F^T F = \begin{pmatrix} \sum\limits_{k=1}^{n} \left(x_1^{(k)}\right)^2 & \sum\limits_{k=1}^{n} x_1^{(k)} x_2^{(k)} & \cdots & \sum\limits_{k=1}^{n} x_1^{(k)} x_m^{(k)} \\ \sum\limits_{k=1}^{n} x_2^{(k)} x_1^{(k)} & \sum\limits_{k=1}^{n} \left(x_2^{(k)}\right)^2 & \cdots & \sum\limits_{k=1}^{n} x_2^{(k)} x_m^{(k)} \\ \vdots & \vdots & \ddots & \vdots \\ \sum\limits_{k=1}^{n} x_m^{(k)} x_1^{(k)} & \sum\limits_{k=1}^{n} x_m^{(k)} x_2^{(k)} & \cdots & \sum\limits_{k=1}^{n} \left(x_m^{(k)}\right)^2 \end{pmatrix}$$

Позначимо $\vec{x}_i = \left(x_i^{(1)}, x_i^{(2)}, ..., x_i^{(n)}\right)^T$ — спостереження i-того фактору за n проведень експерименту. Тоді можна сказати, що A — це матриця Грама для цих векторів:

$$A = \begin{pmatrix} (\vec{x}_1, \vec{x}_1) & (\vec{x}_1, \vec{x}_2) & \cdots & (\vec{x}_1, \vec{x}_m) \\ (\vec{x}_2, \vec{x}_1) & (\vec{x}_2, \vec{x}_2) & \cdots & (\vec{x}_2, \vec{x}_m) \\ \vdots & \vdots & \ddots & \vdots \\ (\vec{x}_m, \vec{x}_1) & (\vec{x}_m, \vec{x}_2) & \cdots & (\vec{x}_m, \vec{x}_m) \end{pmatrix}$$

Знову ж з лінійної алгебри відомо, що в такому випадку $\det A$ дорівнює квадрату об'єму m-вимірного паралелепіпеда Π_m, побудованого на векторах $\vec{x}_1, ..., \vec{x}_m$. Аналогічно, якщо викреслити j-тий стовпець та рядок, то визначник такої матриці буде дорівнювати квадрату об'єму вже $m-1$-вимірного паралелепіпеда $\Pi_{m\setminus j}$, побудованого на векторах $\vec{x}_1, ..., \vec{x}_{j-1}, \vec{x}_{j+1}, ..., \vec{x}_m$.

Позначимо $\Delta = \max\limits_{j=1,...,m} \frac{\lambda_1\lambda_2...\lambda_m}{\lambda_j}$, а j — той індекс з $1, ..., m$, на якому цей максимум досягається. З геометричних міркувань $\det A = \Delta \cdot \|\vec{x}_j\|^2 \sin^2\alpha$, де α — кут між $\vec{x}_j$ та $\Pi_{m\setminus j}$. Покажемо це на прикладі, коли $\Pi_{m\setminus j}$ побудований на $\vec{x}_1$ та $\vec{x}_2$, а Π_m — на них та $\vec{x}_3$:



За відомими формулами, об'єм Π_m в цьому випадку дорівнює добутку площі основи на висоту паралелепіпеда. Основа — це паралелепіпед $\Pi_{m\setminus j}$, а висота дорівнює $\|\vec{x}_3\| \sin\alpha$.

Повернемося до $\text{tr}A^{-1} \le \frac{m\cdot\Delta}{\det A} = \frac{m\cdot\Delta}{\Delta\cdot\|\vec{x}_j\|^2\sin^2\alpha} = \frac{m}{\|\vec{x}_j\|^2\sin^2\alpha}$. Без втрати загальності можна вважати, що $\exists\, \delta > 0$ таке, що $\left|x_i^{(j)}\right| \ge \delta$ для всіх i та j. Якби такого δ не існувало, можна було б «зсунути» всі значення факторів так, щоб воно існувало. Це не надто вплине на модель, оскільки усі зсуви врахуються у вільному коефіцієнті. З цього випливає, що діагональні елементи A необмежено зростають, а отже — зростає їх сума, що дорівнює сумі власних чисел. Оскільки власні числа завжди додатні, то зростає і їх добуток, тому $\det A \to \infty$, $n \to \infty$. З цього, зокрема, випливає, що $\sin\alpha$ теж завжди залишається обмеженим знизу.

Отже, $\frac{m}{\|\vec{x}_j\|^2\sin^2\alpha} \to 0, n \to \infty$, а тому й $\sum\limits_{k=1}^{m} \mathbb{D}\beta_k^* = \text{tr}A^{-1} \to 0, n \to \infty$. Це, нарешті, доводить консистентність оцінок $\beta_1^*, ..., \beta_m^*$. ▲

10.2.2 Оцінка дисперсії похибок спостережень

Нагадаємо, що дисперсії похибок спостережень σ^2 вважаються однаковими, але невідомими. В такому випадку для аналізу регресійної моделі необхідно знайти оцінку $(\sigma^2)^*$. Зробимо це за допомогою методу максимальної правдоподібності: $\vec{\varepsilon} \sim \mathrm{N}(\vec{0}, \sigma^2 \mathbb{I}_n)$, тому

$$\mathcal{L}(\vec{\eta}, \vec{\beta}, \sigma^2) = \frac{1}{(\sqrt{2\pi}\sigma)^n} \exp\left\{-\frac{1}{2\sigma^2}\sum_{i=1}^{n}\left(\eta_i - f\left(\vec{x}^{(i)}\right)\right)^2\right\}$$

$$\frac{\partial \ln \mathcal{L}}{\partial(\sigma^2)}(\vec{\eta}, \vec{\beta}, \sigma^2) = -\frac{n}{2\sigma^2} + \frac{1}{2\sigma^4}\sum_{i=1}^{n}\left(\eta_i - f\left(\vec{x}^{(i)}\right)\right)^2 = 0 \Rightarrow$$

$$\Rightarrow (\sigma^2)^* = \frac{1}{n}\sum_{i=1}^{n}\left(\eta_i - f^*\left(\vec{x}^{(i)}\right)\right)^2 = \frac{1}{n}\|\vec{\eta} - F\vec{\beta}^*\|^2 = \frac{1}{n}\|\vec{\varepsilon}^*\|^2$$

Означення 10.2.1. Вираз $\frac{1}{n}\|\vec{\varepsilon}^*\|^2$ називається *залишковою оцінкою дисперсії* або *залишковою сумою квадратів.*

Перед знаходженням розподілу $(\sigma^2)^*$ розглянемо m-вимірний вектор $\sqrt{A}(\vec{\beta}^* - \vec{\beta})$ та його числові характеристики:

$$\mathbb{E}\left(\sqrt{A}(\vec{\beta}^* - \vec{\beta})\right) = \sqrt{A}\mathbb{E}\left(\vec{\beta}^* - \vec{\beta}\right) = \vec{0}$$

$$\mathbb{K}\left(\sqrt{A}(\vec{\beta}^* - \vec{\beta})\right) = \sqrt{A} \cdot \mathbb{K}\left(\vec{\beta}^* - \vec{\beta}\right) \cdot (\sqrt{A})^T = \sqrt{A} \cdot \sigma^2 A^{-1} \cdot \sqrt{A} = \sigma^2 \cdot \mathbb{I}_m$$

Отже, $\frac{1}{\sigma}\sqrt{A}(\vec{\beta}^* - \vec{\beta}) \sim \mathrm{N}(\vec{0}, \mathbb{I}_m)$. Повернемося до пошуку розподілу $(\sigma^2)^*$. За побудовою вектор $\vec{\eta} - F\vec{\beta}^*$ є ортогональним до векторів виду $F\vec{x}$, $\vec{x} \in \mathbb{R}^m$, а тому й буде ортогональним до $F(\vec{\beta}^* - \vec{\beta})$. Тоді за теоремою Піфагора

$$\left\|\vec{\eta} - F\vec{\beta}^*\right\|^2 + \left\|F(\vec{\beta}^* - \vec{\beta})\right\|^2 = \left\|\vec{\eta} - F\vec{\beta}^* + F(\vec{\beta}^* - \vec{\beta})\right\|^2 = \left\|\vec{\eta} - F\vec{\beta}\right\|^2 = \|\vec{\varepsilon}\|^2$$

Звідси $\|\vec{\eta} - F\vec{\beta}^*\|^2 = \|\vec{\varepsilon}\|^2 - \|F(\vec{\beta}^* - \vec{\beta})\|^2$. Розглянемо другий доданок цієї різниці:

$$\left\|F(\vec{\beta}^* - \vec{\beta})\right\|^2 = \left(F(\vec{\beta}^* - \vec{\beta})\right)^T\left(F(\vec{\beta}^* - \vec{\beta})\right) = (\vec{\beta}^* - \vec{\beta})^T F^T F(\vec{\beta}^* - \vec{\beta}) =$$

$$= (\vec{\beta}^* - \vec{\beta})^T\sqrt{A} \cdot \sqrt{A}(\vec{\beta}^* - \vec{\beta}) = \left(\sqrt{A}(\vec{\beta}^* - \vec{\beta})\right)^T\left(\sqrt{A}(\vec{\beta}^* - \vec{\beta})\right) = \left\|\sqrt{A}(\vec{\beta}^* - \vec{\beta})\right\|^2 =$$

$$= \left[\vec{\beta}^* - \vec{\beta} = A^{-1}F^T\vec{\varepsilon}\right] = \left\|(\sqrt{A})^{-1}F^T\vec{\varepsilon}\right\|^2$$

Оскільки $\frac{1}{\sigma}\sqrt{A}(\vec{\beta}^* - \vec{\beta}) \sim \mathrm{N}(\vec{0}, \mathbb{I}_m)$, то й $\frac{1}{\sigma}(\sqrt{A})^{-1}F^T\vec{\varepsilon} \sim \mathrm{N}(\vec{0}, \mathbb{I}_m)$, тому $\left\|\frac{1}{\sigma}(\sqrt{A})^{-1}F^T\vec{\varepsilon}\right\|^2 \sim \chi^2_m$. Аналогічно, $\frac{1}{\sigma}\vec{\varepsilon} \sim \mathrm{N}(\vec{0}, \mathbb{I}_n)$ та $\left\|\frac{1}{\sigma}\vec{\varepsilon}\right\|^2 \sim \chi^2_n$. Отже, $\frac{n(\sigma^2)^*}{\sigma^2} = \left\|\frac{1}{\sigma}\vec{\varepsilon}\right\|^2 - \left\|\frac{1}{\sigma}(\sqrt{A})^{-1}F^T\vec{\varepsilon}\right\|^2 \sim \chi^2_{n-m}$. За властивостями розподілу хі-квадрат, $\mathbb{E}\left(\frac{n(\sigma^2)^*}{\sigma^2}\right) = n - m$, тому $\mathbb{E}(\sigma^2)^* = \frac{n-m}{n}\sigma^2$ і треба вводити виправлену незміщену оцінку $(\sigma^2)^{**} = \frac{n}{n-m}(\sigma^2)^*$. Таким чином, маємо незміщену оцінку дисперсії σ^2:

$$(\sigma^2)^{**} = \frac{1}{n-m}\left\|\vec{\eta} - F\vec{\beta}^*\right\|^2 \tag{1}$$

Надалі для дослідження побудованої регресійної моделі знадобиться статистика

$$\vec{\gamma} = \frac{\frac{1}{\sigma}\sqrt{A}(\vec{\beta}^* - \vec{\beta})}{\sqrt{\frac{1}{n-m}\left(\frac{n(\sigma^2)^*}{\sigma^2}\right)}} = \frac{\sqrt{A}(\vec{\beta}^* - \vec{\beta})}{\sqrt{(\sigma^2)^{**}}}$$

Оскільки координати $\frac{1}{\sigma}\sqrt{A}(\vec{\beta}^* - \vec{\beta})$ є стандартними гауссівськими величинами, то за аналогією з теоремою Фішера (ст. 90), $\gamma_i \sim \mathrm{St}_{n-m}$, $i = 1, ..., m$.

10.2.3 Перевірка адекватності моделі

Після побудови регресійної моделі її перевіряють на *адекватність*. Це означає, що значення відклику, що прогнозуються моделлю, узгоджуються з результатами спостережень.

Одним із способів такої перевірки є так званий F-критерій: порівняння *залишкової оцінки дисперсії* $(\sigma^2)^{**}$, що обумовлена випадковими помилками вимірів та, можливо, неврахованими факторами, з *незміщеною оцінкою дисперсії* $\mathbb{D}^{**}\eta$. Оскільки

$$\frac{n-m}{\sigma^2}(\sigma^2)^{**} = \frac{n}{\sigma^2}(\sigma^2)^{*} = \frac{1}{\sigma^2}\left\|\vec{\eta} - F\vec{\beta}^*\right\|^2 \sim \chi^2_{n-m}$$
$$\frac{n-1}{\sigma^2}\mathbb{D}^{**}\eta = \frac{n}{\sigma^2}\mathbb{D}^{*}\eta = \frac{1}{\sigma^2}\sum_{k=1}^{n}(\eta_k - \overline{\eta})^2 \sim \chi^2_{n-1}$$

то розглядається статистика

$$\zeta = \frac{\frac{1}{n-1}\sum\limits_{k=1}^{n}(\eta_k - \overline{\eta})^2}{\frac{1}{n-m}\left\|\vec{\eta} - F\vec{\beta}^*\right\|^2} = \frac{\frac{1}{n-1}\sum\limits_{k=1}^{n}(\eta_k - \overline{\eta})^2}{\frac{1}{n-m}\sum\limits_{k=1}^{n}\left(\eta_k - f^*\left(\vec{x}^{(k)}\right)\right)^2} \sim \mathrm{F}(n-1, n-m) \tag{2}$$

У прикладі на ст. 111 було показано, що якщо не включати в регресійну модель жодних факторів, окрім константи, то оцінкою цього єдиного коефіцієнта буде $\overline{\eta}$. Таким чином, F-критерій перевіряє, чи є побудована регресійна модель кращою за найпростішу — константну. *Основною гіпотезою* є те, що константа та побудована моделі не відрізняються (в тому сенсі, що дисперсії їх похибок однакові), а *альтернативною* — що побудована модель є кращою за константну. За заданим рівнем значущості треба знайти межу критичної області $t_{\text{кр}}$. Критична область є правосторонною: при $\zeta_{\text{зн}} > t_{\text{кр}}$ основна гіпотеза відхиляється і модель вважається адекватною: це означає, що її дисперсія значно менше, ніж дисперсія моделі-константи. Якщо основна гіпотеза не відхиляється, то треба спробувати додати в модель інші фактори.

Після перевірки на адекватність корегувати побудовану модель можна за допомогою перевірки *гіпотези про значущість параметрів*. Якщо значення якоїсь з оцінок β_j^* виявилася близькою до нуля, це може означати, що відповідний фактор не вносить внеску до моделі. Висувається основна гіпотеза $H_0 : \beta_j = 0$ та альтернативна $H_1 : \beta_j < 0$ чи $H_2 : \beta_j > 0$ в залежності від знаку отриманої оцінки β_j^*. Оскільки $\vec{\beta}^* \sim \mathrm{N}(\vec{\beta}, \sigma^2 A^{-1})$, то $\beta_j^* \sim \mathrm{N}(\beta_j, \sigma^2 a_{jj})$, де a_{jj} — елемент A^{-1}. Відповідно, статистикою для перевірки гіпотези є

$$\gamma = \frac{\beta_j^*}{\sqrt{(\sigma^2)^{**} \cdot a_{jj}}} \sim \mathrm{St}_{n-m} \tag{3}$$

За заданим рівнем значущості треба знайти межу критичної області, яка буле правосторонньою чи лівосторонньою в залежності від альтернативної гіпотези. Якщо основна гіпотеза $H_0 : \beta_j = 0$ приймається, то відповідний фактор прибирається з моделі, але нову модель потрібно знову перевіряти на адекватність.

Таким чином, перевірка адекватності моделі за F-критерієм показує, чи не треба додати більше факторів, а перевірка значущості близьких до нуля параметрів — чи не треба зменшити кількість факторів.

10.2.4 Аналіз точності прогнозів моделі

Нехай побудована регресійна модель перевірена на адекватність та значущість параметрів. Однією з задач, для яких її можна застосувати — *прогнозування значення відклику* з тими значеннями факторів, з якими експеримент не проводився. Аналіз точності отриманих прогнозів відбувається за допомогою побудови довірчих інтервалів для самого значення відклику та його середнього значення.

Позначимо $f^*(\vec{x}) = \sum\limits_{i=0}^{m}\beta_i^* x_i$. $\mathbb{E}f^*(\vec{x}) = \sum\limits_{i=0}^{m}\beta_i^* x_i = \sum\limits_{i=0}^{m}\beta_i x_i = f(\vec{x})$ називається *середнім значенням відклику*. Можна записати $f^*(\vec{x}) = \vec{x}^T\vec{\beta}^*$ та $f(\vec{x}) = \vec{x}^T\vec{\beta}$, тоді

$$\mathbb{D}f^*(\vec{x}) = \mathbb{E}\left(\vec{x}^T(\vec{\beta}^* - \vec{\beta})\right)^2 = \mathbb{E}\left(\vec{x}^T(\vec{\beta}^* - \vec{\beta})\right)\left(\vec{x}^T(\vec{\beta}^* - \vec{\beta})\right)^T = \mathbb{E}\left(\vec{x}^T(\vec{\beta}^* - \vec{\beta})(\vec{\beta}^* - \vec{\beta})^T\vec{x}\right) =$$
$$= \vec{x}^T\mathbb{E}(\vec{\beta}^* - \vec{\beta})(\vec{\beta}^* - \vec{\beta})^T\vec{x} = \vec{x}^T\mathbb{K}\vec{\beta}^*\vec{x} = \sigma^2\vec{x}^T A^{-1}\vec{x}$$

Отже, маємо статистику $\frac{f^*(\vec{x})-f(\vec{x})}{\sigma\sqrt{\vec{x}^TA^{-1}\vec{x}}} \sim \mathrm{N}(0,1)$. Оскільки $\frac{n-m}{\sigma^2}(\sigma^2)^{**} \sim \chi^2_{n-m}$, то

$$\zeta = \frac{\frac{f^*(\vec{x})-f(\vec{x})}{\sigma\sqrt{\vec{x}^TA^{-1}\vec{x}}}}{\sqrt{\frac{1}{n-m}\cdot\frac{n-m}{\sigma^2}(\sigma^2)^{**}}} = \frac{f^*(\vec{x})-f(\vec{x})}{\sqrt{(\sigma^2)^{**}\vec{x}^TA^{-1}\vec{x}}} \sim \mathrm{St}_{n-m}$$

Для побудови довірчого інтервалу з рівнем надійності γ за таблицею знаходять таке t, що $\mathbb{P}\left\{|\zeta| < t\right\} = \gamma$. Довірчий інтервал для середнього значення відклику має вигляд

$$f(\vec{x}) \in \left(f^*(\vec{x}) - t\sqrt{(\sigma^2)^{**}\vec{x}^TA^{-1}\vec{x}}, f^*(\vec{x}) + t\sqrt{(\sigma^2)^{**}\vec{x}^TA^{-1}\vec{x}}\right) \tag{4}$$

Для побудови довірчого інтервалу для самого значення відклику скористаємося тим, що $\mathbb{E}\left(\eta - f^*(\vec{x})\right) = 0$, а $\mathbb{D}\left(\eta - f^*(\vec{x})\right) = \mathbb{D}\eta + \mathbb{D}f^*(x) = \sigma^2\left(1 + \vec{x}^TA^{-1}\vec{x}\right)$. Аналогічними міркуваннями отримаємо

$$\frac{\eta - f^*(\vec{x})}{\sqrt{(\sigma^2)^{**}\left(1 + \vec{x}^TA^{-1}\vec{x}\right)}} \sim \mathrm{St}_{n-m}$$

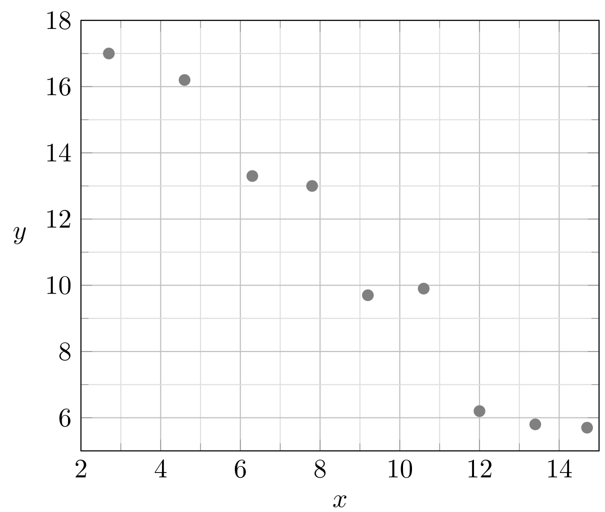
тому довірчий інтервал для значення відклику матиме вигляд

$$\eta \in \left(f^*(\vec{x}) - t\sqrt{(\sigma^2)^{**}\left(1 + \vec{x}^TA^{-1}\vec{x}\right)}, f^*(\vec{x}) + t\sqrt{(\sigma^2)^{**}\left(1 + \vec{x}^TA^{-1}\vec{x}\right)}\right) \tag{5}$$

10.2.5 Приклад побудови та аналізу моделі

За заданими значенням y_i та x_i знайти оцінки параметрів простої однофакторної лінійної регресійної моделі $y = f(x) = \beta_0 + \beta_1 x$. Перевірити побудовану модель на адекватність та на значущість параметрів на рівні значущості $\alpha = 0.05$. Побудувати довірчі інтервали для середнього значення відклику та значення відклику в точці $x_0 = 10$ з рівнем надійності 0.99.

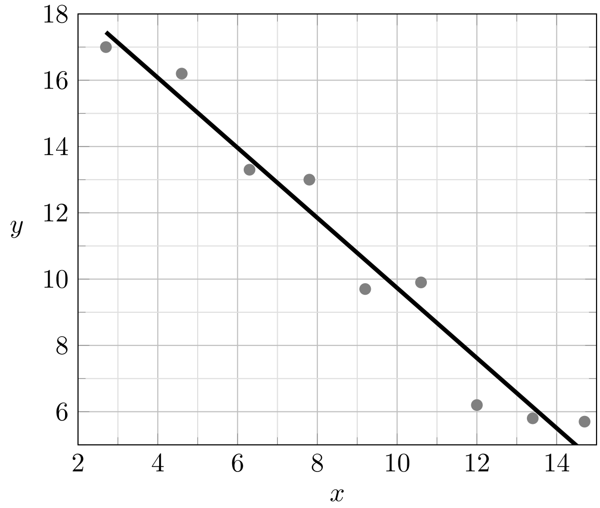
x_i	2.7	4.6	6.3	7.8	9.2	10.6	12.0	13.4	14.7
y_i	17.0	16.2	13.3	13.0	9.7	9.9	6.2	5.8	5.7



Складемо матрицю плану $F = \begin{pmatrix} 1 & 1 & \cdots & 1 \\ 2.7 & 4.6 & \cdots & 14.7 \end{pmatrix}^T$. Інформаційна та дисперсійна матриці дорівнюють, відповідно, $A = F^TF = \begin{pmatrix} 9 & 81.3 \\ 81.3 & 865.63 \end{pmatrix}$, $A^{-1} = \begin{pmatrix} 0.73298 & -0.06884 \\ -0.06884 & 0.00762 \end{pmatrix}$. Вектор значень відкликів $\vec{\eta}_{\text{зн}} = \begin{pmatrix} 17.0 & 16.2 & \cdots & 5.7 \end{pmatrix}^T$, тому можемо знайти значення оцінок параметрів моделі:

$$\vec{\beta}^*_{\text{зн}} = A^{-1}F^T\vec{\eta}_{\text{зн}} \approx \begin{pmatrix} 20.30566 \\ -1.05721 \end{pmatrix}$$

Отже, маємо модель $y = f^*(x) = 20.30566 - 1.05721x$, графік якої зображено нижче.



Перевіримо побудовану модель на адекватність. Обчислимо значення виправленої вибіркової дисперсії для η та залишкову оцінку дисперсії: $(\mathbb{D}^{**}\eta)_{\text{зн}} = \frac{1}{8}\sum_{i=1}^{9}(y_i - \overline{y})^2 \approx 19.1$, $(\sigma^2)^{**}_{\text{зн}} = \frac{1}{7}\sum_{i=1}^{9}(y_i - f^*(x_i))^2 \approx 0.88564$. Значення статистики F-критерію (2) $\zeta_{\text{зн}} \approx 21.5752$. За таблицею знайдемо межу критичної області для $\mathrm{F}(8,7)$: $t_{\text{кр}} = 3.73$. Отже, $\zeta_{\text{зн}} > t_{\text{кр}}$ і модель можна вважати адекватною на рівні значущості 0.05.

Перевіримо значущість параметра β_2. Основною гіпотезою є $H_0 : \beta_2 = 0$, альтернативною — $H_1 : \beta_2 < 0$. Знайдемо значення відповідної статистики (3): $\gamma_{\text{зн}} = \frac{-1.05721}{\sqrt{0.88564\cdot 0.00762}} \approx -12.87$. За таблицею знайдемо межу критичної області для St_7: $t_{\text{кр}} = -1.895$. Отже, $\gamma_{\text{зн}} > t_{\text{кр}}$ і основна гіпотеза відхиляється, тому параметр β_2 є значущим.

Для знаходження обох довірчих інтервалів знайдемо $\vec{x}^T A^{-1}\vec{x}$ для $\vec{x} = \begin{pmatrix}1\\10\end{pmatrix}$ та $f^*(x_0)$. Отримані значення — 0.11823 та 9.7336 відповідно. За таблицею знайдемо для St_7 значення $t = 3.499$, тому за формулою (4) отримаємо довірчий інтервал для середнього значення відклику в точці $x_0 = 10$:

$$f^*(\vec{x}) \pm t\sqrt{(\sigma^2)^{**}\vec{x}^T A^{-1}\vec{x}} = 9.7336 \pm 3.499 \cdot \sqrt{0.88564 \cdot 0.11823} \approx 9.7336 \pm 1.13223$$
$$f(\vec{x}) \in (8.60137, 10.86583) \approx (8.6, 10.87)$$

За формулою 5 отримаємо довірчий інтервал для значення відклику в точці $x_0 = 10$:

$$f^*(\vec{x}) \pm t\sqrt{(\sigma^2)^{**}\left(1 + \vec{x}^T A^{-1}\vec{x}\right)} = 9.7336 \pm 3.499 \cdot \sqrt{0.88564 \cdot 1.11823} \approx 9.7336 \pm 3.482$$
$$\eta_{\text{зн}} \in (6.2516, 13.2156) \approx (6.25, 13.22)$$

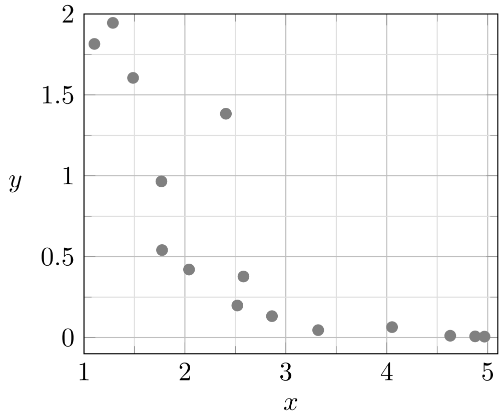
Такі широкі довірчі інтервали пояснюються, зокрема, малим обсягом вибірки.

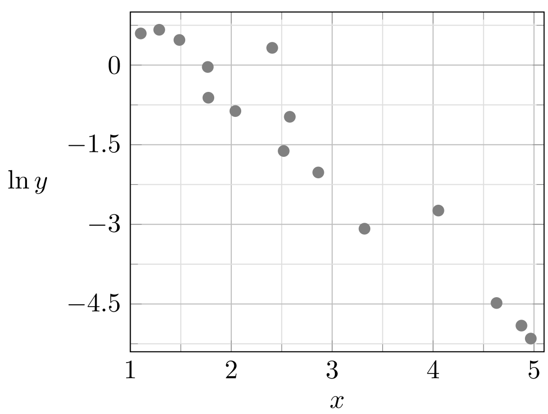
10.2.6 Деякі труднощі застосування лінійної регресійної моделі

Одне з основних припущень для застосування лінійної регресійної моделі $\vec{\varepsilon} \sim \mathrm{N}(\vec{x}, \sigma^2\mathbb{I})$ є доволі «жорстким» і не завжди виправданим: наприклад, дисперсії похибок можуть змінюватися в залежності від значень факторів.

Також, складним іноді може бути вибір кількості факторів. З одного боку, включення до моделі багатьох факторів може як підвищити її точність, так і погіршити оцінки параметрів, оскільки деякі фактори можуть бути лінійно залежними (або близькими до таких), що призведе до виродженості чи поганої обумовленості матриці A — таке явище називається *мультиколінеарністю* факторів. На практиці ще перед побудовою моделі часто досліджують зміст самих факторів (наприклад, щоб не включати два фактори, один з яких показуватиме вимірювання деякої величини в метрах, а інший — в сантиметрах), а також вибіркові коефіцієнти кореляції між факторами: якщо значення цього коефіцієнту для якихось двох факторів буде близьким за модулем до 1, то, скоріш за все, вони є лінійно залежними, тому один із них не варто включати до моделі.

Іншим припущенням при побудові лінійної регресійної моделі є лінійна за параметрами залежність середнього значення відклику від значень факторів. Якщо все ж таки нелінійність очевидна (наприклад, з погляду на *діаграму розсіювання*), то треба застосовувати якесь перетворення змінних. Наприклад, розглянемо дві діаграми розсіювання:





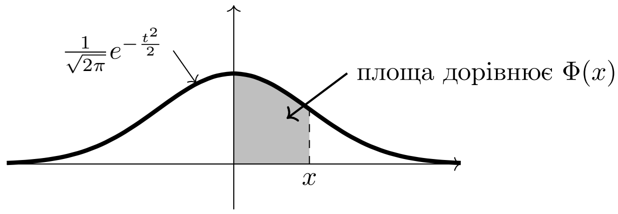
Видно, що логарифмування змінної y дало залежність, що схожа на лінійну. В такому випадку будується модель виду $z = \ln y = \beta_0 + \beta_1 x$, звідки потім отримується $y = e^{\beta_0 + \beta_1 x}$.

Насамкінець, важливо зауважити, що методи регресійного аналізу дозволяють з'ясувати *лише числові залежності* — як значення деяких величин впливає на середнє значення іншої. Причинно-наслідкові зв'язки цими методами дослідити неможливо. Наприклад, між віком будинку та ціною квартир в ньому можна знайти деякий кореляційний зв'язок. Чи можна з цього зробити висновок, що чим дорожча квартира, тим новіше будинок, в якому вона знаходиться? Ні. Але можна перевірити, як вік будинку впливає на середню ціну квартир — на це питання регресійний аналіз вже може дати відповідь.

Таблиці значень деяких функцій

Таблиця значень функції Лапласа

$$\Phi(x) = \frac{1}{\sqrt{2\pi}} \int\limits_0^x e^{-\frac{t^2}{2}}\, dt$$



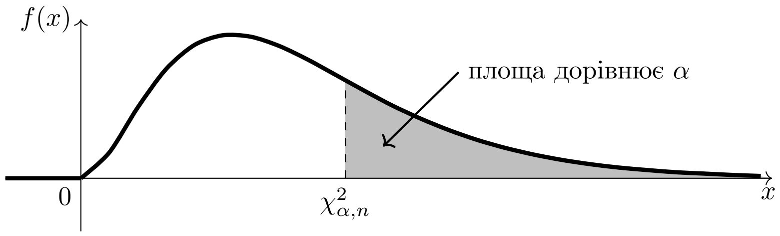
В таблиці наведено лише дробову частину усіх значень, оскільки усі цілі частини рівні 0: наприклад, .48745 треба читати як 0.48745.

x	**0.00**	**0.01**	**0.02**	**0.03**	**0.04**	**0.05**	**0.06**	**0.07**	**0.08**	**0.09**
0.00	.00000	.00399	.00798	.01197	.01595	.01994	.02392	.02790	.03188	.03586
0.10	.03983	.04380	.04776	.05172	.05567	.05962	.06356	.06749	.07142	.07535
0.20	.07926	.08317	.08706	.09095	.09483	.09871	.10257	.10642	.11026	.11409
0.30	.11791	.12172	.12552	.12930	.13307	.13683	.14058	.14431	.14803	.15173
0.40	.15542	.15910	.16276	.16640	.17003	.17364	.17724	.18082	.18439	.18793
0.50	.19146	.19497	.19847	.20194	.20540	.20884	.21226	.21566	.21904	.22240
0.60	.22575	.22907	.23237	.23565	.23891	.24215	.24537	.24857	.25175	.25490
0.70	.25804	.26115	.26424	.26730	.27035	.27337	.27637	.27935	.28230	.28524
0.80	.28814	.29103	.29389	.29673	.29955	.30234	.30511	.30785	.31057	.31327
0.90	.31594	.31859	.32121	.32381	.32639	.32894	.33147	.33398	.33646	.33891
1.00	.34134	.34375	.34614	.34849	.35083	.35314	.35543	.35769	.35993	.36214
1.10	.36433	.36650	.36864	.37076	.37286	.37493	.37698	.37900	.38100	.38298
1.20	.38493	.38686	.38877	.39065	.39251	.39435	.39617	.39796	.39973	.40147
1.30	.40320	.40490	.40658	.40824	.40988	.41149	.41309	.41466	.41621	.41774
1.40	.41924	.42073	.42220	.42364	.42507	.42647	.42785	.42922	.43056	.43189
1.50	.43319	.43448	.43574	.43699	.43822	.43943	.44062	.44179	.44295	.44408
1.60	.44520	.44630	.44738	.44845	.44950	.45053	.45154	.45254	.45352	.45449
1.70	.45543	.45637	.45728	.45818	.45907	.45994	.46080	.46164	.46246	.46327
1.80	.46407	.46485	.46562	.46638	.46712	.46784	.46856	.46926	.46995	.47062
1.90	.47128	.47193	.47257	.47320	.47381	.47441	.47500	.47558	.47615	.47670
2.00	.47725	.47778	.47831	.47882	.47932	.47982	.48030	.48077	.48124	.48169
2.10	.48214	.48257	.48300	.48341	.48382	.48422	.48461	.48500	.48537	.48574
2.20	.48610	.48645	.48679	.48713	.48745	.48778	.48809	.48840	.48870	.48899
2.30	.48928	.48956	.48983	.49010	.49036	.49061	.49086	.49111	.49134	.49158
2.40	.49180	.49202	.49224	.49245	.49266	.49286	.49305	.49324	.49343	.49361
2.50	.49379	.49396	.49413	.49430	.49446	.49461	.49477	.49492	.49506	.49520
2.60	.49534	.49547	.49560	.49573	.49585	.49598	.49609	.49621	.49632	.49643
2.70	.49653	.49664	.49674	.49683	.49693	.49702	.49711	.49720	.49728	.49736
2.80	.49744	.49752	.49760	.49767	.49774	.49781	.49788	.49795	.49801	.49807
2.90	.49813	.49819	.49825	.49831	.49836	.49841	.49846	.49851	.49856	.49861

x	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
3.00	.49865	.49869	.49874	.49878	.49882	.49886	.49889	.49893	.49896	.49900
3.10	.49903	.49906	.49910	.49913	.49916	.49918	.49921	.49924	.49926	.49929
3.20	.49931	.49934	.49936	.49938	.49940	.49942	.49944	.49946	.49948	.49950
3.30	.49952	.49953	.49955	.49957	.49958	.49960	.49961	.49962	.49964	.49965
3.40	.49966	.49968	.49969	.49970	.49971	.49972	.49973	.49974	.49975	.49976
3.50	.49977	.49978	.49978	.49979	.49980	.49981	.49981	.49982	.49983	.49983
3.60	.49984	.49985	.49985	.49986	.49986	.49987	.49987	.49988	.49988	.49989
3.70	.49989	.49990	.49990	.49990	.49991	.49991	.49992	.49992	.49992	.49992
3.80	.49993	.49993	.49993	.49994	.49994	.49994	.49994	.49995	.49995	.49995
3.90	.49995	.49995	.49996	.49996	.49996	.49996	.49996	.49996	.49997	.49997
4.00	.49997	.49997	.49997	.49997	.49997	.49997	.49998	.49998	.49998	.49998
4.10	.49998	.49998	.49998	.49998	.49998	.49998	.49998	.49998	.49999	.49999
4.20	.49999	.49999	.49999	.49999	.49999	.49999	.49999	.49999	.49999	.49999
4.30	.49999	.49999	.49999	.49999	.49999	.49999	.49999	.49999	.49999	.49999
4.40	.49999	.49999	.50000	.50000	.50000	.50000	.50000	.50000	.50000	.50000
4.50	.50000	.50000	.50000	.50000	.50000	.50000	.50000	.50000	.50000	.50000
4.60	.50000	.50000	.50000	.50000	.50000	.50000	.50000	.50000	.50000	.50000
4.70	.50000	.50000	.50000	.50000	.50000	.50000	.50000	.50000	.50000	.50000
4.80	.50000	.50000	.50000	.50000	.50000	.50000	.50000	.50000	.50000	.50000
4.90	.50000	.50000	.50000	.50000	.50000	.50000	.50000	.50000	.50000	.50000

Деякі квантилі розподілу χ^2_n

Значення функції $\chi^2_{\alpha,n}$ визначаються з рівняння $\int\limits_{\chi^2_{\alpha,n}}^{+\infty} f(x)dx = \alpha$ де $f(x)$ — щільність розподілу χ^2_n (хі-квадрат з n ступенями вільності). Це значення є квантилем рівня $q = 1 - \alpha$.

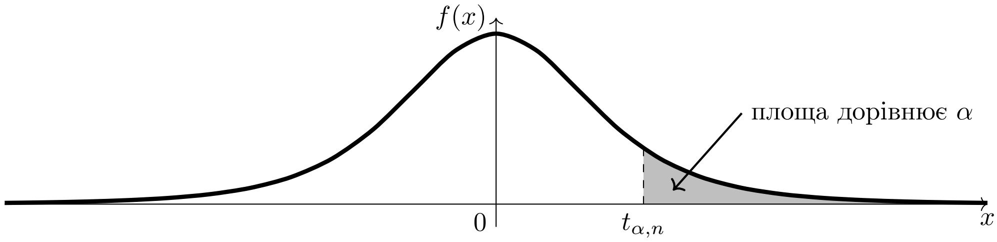


n \ α	**0.990**	**0.975**	**0.950**	**0.900**	**0.100**	**0.050**	**0.025**	**0.010**
1	0.00	0.00	0.00	0.02	2.71	3.84	5.02	6.63
2	0.02	0.05	0.10	0.21	4.61	5.99	7.38	9.21
3	0.11	0.22	0.35	0.58	6.25	7.81	9.35	11.34
4	0.30	0.48	0.71	1.06	7.78	9.49	11.14	13.28
5	0.55	0.83	1.15	1.61	9.24	11.07	12.83	15.09
6	0.87	1.24	1.64	2.20	10.64	12.59	14.45	16.81
7	1.24	1.69	2.17	2.83	12.02	14.07	16.01	18.48
8	1.65	2.18	2.73	3.49	13.36	15.51	17.53	20.09
9	2.09	2.70	3.33	4.17	14.68	16.92	19.02	21.67
10	2.56	3.25	3.94	4.87	15.99	18.31	20.48	23.21
11	3.05	3.82	4.57	5.58	17.28	19.68	21.92	24.72
12	3.57	4.40	5.23	6.30	18.55	21.03	23.34	26.22
13	4.11	5.01	5.89	7.04	19.81	22.36	24.74	27.69
14	4.66	5.63	6.57	7.79	21.06	23.68	26.12	29.14
15	5.23	6.26	7.26	8.55	22.31	25.00	27.49	30.58
16	5.81	6.91	7.96	9.31	23.54	26.30	28.85	32.00
17	6.41	7.56	8.67	10.09	24.77	27.59	30.19	33.41
18	7.01	8.23	9.39	10.86	25.99	28.87	31.53	34.81
19	7.63	8.91	10.12	11.65	27.20	30.14	32.85	36.19
20	8.26	9.59	10.85	12.44	28.41	31.41	34.17	37.57
21	8.90	10.28	11.59	13.24	29.62	32.67	35.48	38.93
22	9.54	10.98	12.34	14.04	30.81	33.92	36.78	40.29
23	10.20	11.69	13.09	14.85	32.01	35.17	38.08	41.64
24	10.86	12.40	13.85	15.66	33.20	36.42	39.36	42.98
25	11.52	13.12	14.61	16.47	34.38	37.65	40.65	44.31
26	12.20	13.84	15.38	17.29	35.56	38.89	41.92	45.64
27	12.88	14.57	16.15	18.11	36.74	40.11	43.19	46.96
28	13.56	15.31	16.93	18.94	37.92	41.34	44.46	48.28
29	14.26	16.05	17.71	19.77	39.09	42.56	45.72	49.59
30	14.95	16.79	18.49	20.60	40.26	43.77	46.98	50.89
40	22.16	24.43	26.51	29.05	51.81	55.76	59.34	63.69
50	29.71	32.36	34.76	37.69	63.17	67.50	71.42	76.15
60	37.48	40.48	43.19	46.46	74.40	79.08	83.30	88.38
70	45.44	48.76	51.74	55.33	85.53	90.53	95.02	100.43
80	53.54	57.15	60.39	64.28	96.58	101.88	106.63	112.33
90	61.75	65.65	69.13	73.29	107.57	113.15	118.14	124.12
100	70.06	74.22	77.93	82.36	118.50	124.34	129.56	135.81

Деякі квантилі розподілу St_n

Значення функції $t_{\alpha,n}$ визначаються з рівняння $\int\limits_{t_{\alpha,n}}^{+\infty} f(x)dx = \alpha$ де $f(x)$ — щільність розподілу St_n (Стьюдента з n ступенями вільності). Це значення є квантилем рівня $q = 1 - \alpha$.

Якщо для $\xi \sim \mathrm{St}_n$ необхідно знайти таке значення t, що $\mathbb{P}\{|\xi| < t\} = \beta$, то, внаслідок симетричності розподілу відносно нуля, можна шукати за таблицею значення $t = t_{(1-\beta)/2,n}$.

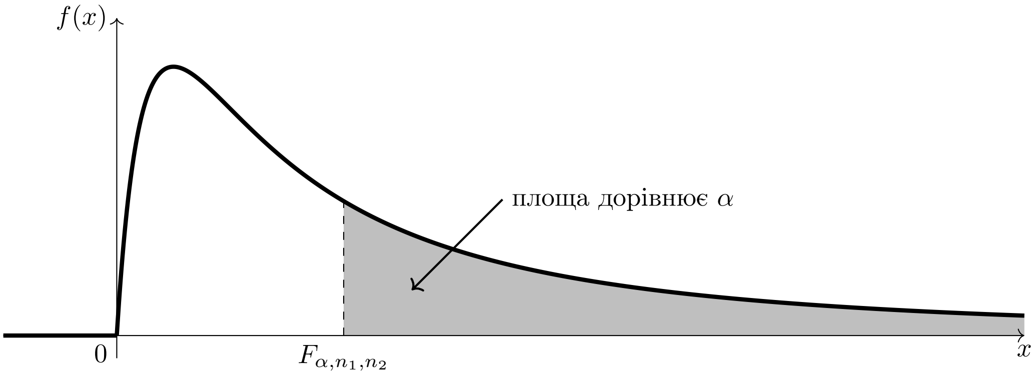


n \ α	**0.050**	**0.025**	**0.010**	**0.005**
1	6.314	12.706	31.821	63.657
2	2.920	4.303	6.965	9.925
3	2.353	3.182	4.541	5.841
4	2.132	2.776	3.747	4.604
5	2.015	2.571	3.365	4.032
6	1.943	2.447	3.143	3.707
7	1.895	2.365	2.998	3.499
8	1.860	2.306	2.896	3.355
9	1.833	2.262	2.821	3.250
10	1.812	2.228	2.764	3.169
11	1.796	2.201	2.718	3.106
12	1.782	2.179	2.681	3.055
13	1.771	2.160	2.650	3.012
14	1.761	2.145	2.624	2.977
15	1.753	2.131	2.602	2.947
16	1.746	2.120	2.583	2.921
17	1.740	2.110	2.567	2.898
18	1.734	2.101	2.552	2.878
19	1.729	2.093	2.539	2.861
20	1.725	2.086	2.528	2.845

n \ α	**0.050**	**0.025**	**0.010**	**0.005**
21	1.721	2.080	2.518	2.831
22	1.717	2.074	2.508	2.819
23	1.714	2.069	2.500	2.807
24	1.711	2.064	2.492	2.797
25	1.708	2.060	2.485	2.787
26	1.706	2.056	2.479	2.779
27	1.703	2.052	2.473	2.771
28	1.701	2.048	2.467	2.763
29	1.699	2.045	2.462	2.756
30	1.697	2.042	2.457	2.750
40	1.684	2.021	2.423	2.704
50	1.676	2.009	2.403	2.678
60	1.671	2.000	2.390	2.660
70	1.667	1.994	2.381	2.648
70	1.667	1.994	2.381	2.648
90	1.662	1.987	2.368	2.632
100	1.660	1.984	2.364	2.626
120	1.658	1.980	2.358	2.617
150	1.655	1.976	2.351	2.609
∞	1.645	1.960	2.326	2.576

Квантилі рівня 0.95 розподілу $\mathrm{F}(n_1, n_2)$

Значення функції F_{α,n_1,n_2} визначаються з рівняння $\int\limits_{F_{\alpha,n_1,n_2}}^{+\infty} f(x)dx = \alpha$ де $f(x)$ — щільність розподілу $\mathrm{F}(n_1, n_2)$ (Фішера-Снедекора з n_1, n_2 ступенями вільності). Це значення є квантилем рівня $q = 1 - \alpha$. Нагадаємо, що даний розподіл — це розподіл частки $\frac{\chi^2_{n_1}/n_1}{\chi^2_{n_2}/n_2}$ з незалежними чисельником і знаменником.

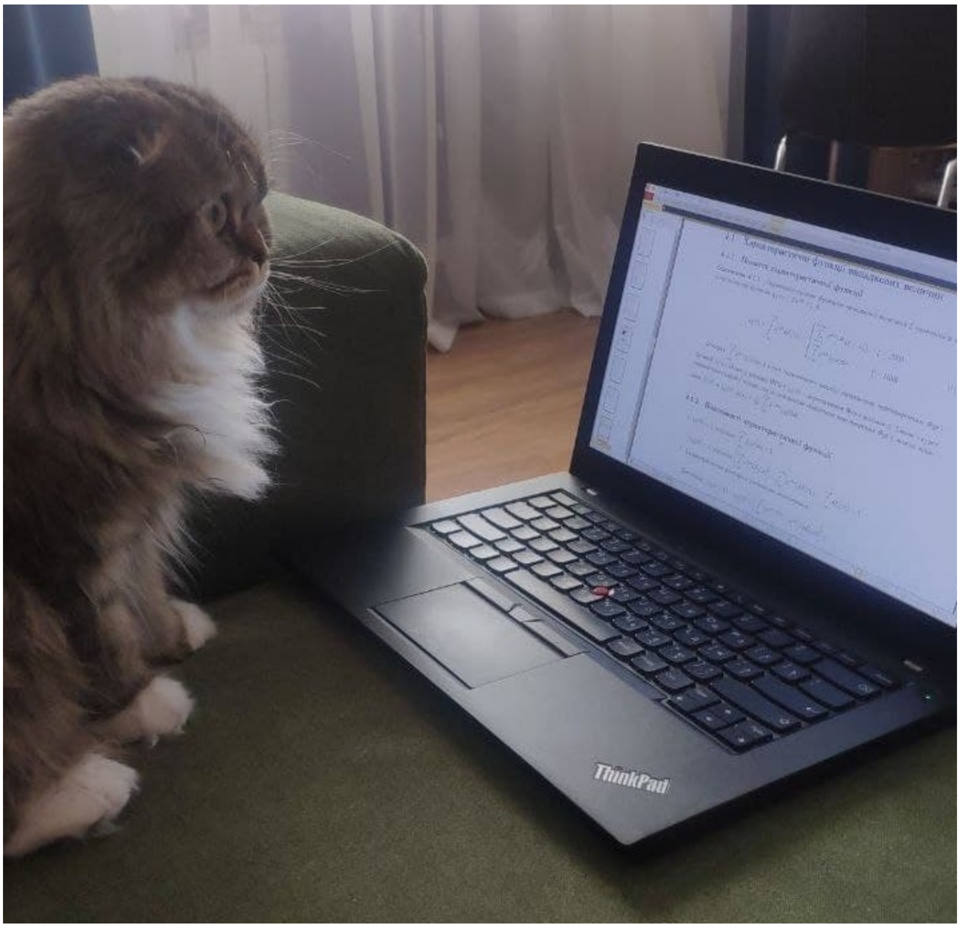


Для наведеної таблиці $\alpha = 0.05$.

n_2 \ n_1	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
3	10.1	9.55	9.28	9.12	9.01	8.94	8.89	8.85	8.81	8.79
4	7.71	6.94	6.59	6.39	6.26	6.16	6.09	6.04	6.00	5.96
5	6.61	5.79	5.41	5.19	5.05	4.95	4.88	4.82	4.77	4.74
6	5.99	5.14	4.76	4.53	4.39	4.28	4.21	4.15	4.10	4.06
7	5.59	4.74	4.35	4.12	3.97	3.87	3.79	3.73	3.68	3.64
8	5.32	4.46	4.07	3.84	3.69	3.58	3.50	3.44	3.39	3.35
9	5.12	4.26	3.86	3.63	3.48	3.37	3.29	3.23	3.18	3.14
10	4.96	4.10	3.71	3.48	3.33	3.22	3.14	3.07	3.02	2.98
11	4.84	3.98	3.59	3.36	3.20	3.09	3.01	2.95	2.90	2.85
12	4.75	3.89	3.49	3.26	3.11	3.00	2.91	2.85	2.80	2.75
13	4.67	3.81	3.41	3.18	3.03	2.92	2.83	2.77	2.71	2.67
14	4.60	3.74	3.34	3.11	2.96	2.85	2.76	2.70	2.65	2.60
15	4.54	3.68	3.29	3.06	2.90	2.79	2.71	2.64	2.59	2.54
16	4.49	3.63	3.24	3.01	2.85	2.74	2.66	2.59	2.54	2.49
17	4.45	3.59	3.20	2.96	2.81	2.70	2.61	2.55	2.49	2.45
18	4.41	3.55	3.16	2.93	2.77	2.66	2.58	2.51	2.46	2.41
19	4.38	3.52	3.13	2.90	2.74	2.63	2.54	2.48	2.42	2.38
20	4.35	3.49	3.10	2.87	2.71	2.60	2.51	2.45	2.39	2.35
22	4.30	3.44	3.05	2.82	2.66	2.55	2.46	2.40	2.34	2.30
24	4.26	3.40	3.01	2.78	2.62	2.51	2.42	2.36	2.30	2.25
26	4.23	3.37	2.98	2.74	2.59	2.47	2.39	2.32	2.27	2.22
28	4.20	3.34	2.95	2.71	2.56	2.45	2.36	2.29	2.24	2.19
30	4.17	3.32	2.92	2.69	2.53	2.42	2.33	2.27	2.21	2.16
40	4.08	3.23	2.84	2.61	2.45	2.34	2.25	2.18	2.12	2.08
50	4.03	3.18	2.79	2.56	2.40	2.29	2.20	2.13	2.07	2.03
60	4.00	3.15	2.76	2.53	2.37	2.25	2.17	2.10	2.04	1.99
100	3.94	3.09	2.70	2.46	2.31	2.19	2.10	2.03	1.97	1.93

n_2 \ n_1	**12**	**14**	**16**	**18**	**20**	**30**	**40**	**50**	**60**	**100**
3	8.74	8.71	8.69	8.67	8.66	8.62	8.59	8.58	8.57	8.55
4	5.91	5.87	5.84	5.82	5.80	5.75	5.72	5.70	5.69	5.66
5	4.68	4.64	4.60	4.58	4.56	4.50	4.46	4.44	4.43	4.41
6	4.00	3.96	3.92	3.90	3.87	3.81	3.77	3.75	3.74	3.71
7	3.57	3.53	3.49	3.47	3.44	3.38	3.34	3.32	3.30	3.27
8	3.28	3.24	3.20	3.17	3.15	3.08	3.04	3.02	3.01	2.97
9	3.07	3.03	2.99	2.96	2.94	2.86	2.83	2.80	2.79	2.76
10	2.91	2.86	2.83	2.80	2.77	2.70	2.66	2.64	2.62	2.59
11	2.79	2.74	2.70	2.67	2.65	2.57	2.53	2.51	2.49	2.46
12	2.69	2.64	2.60	2.57	2.54	2.47	2.43	2.40	2.38	2.35
13	2.60	2.55	2.51	2.48	2.46	2.38	2.34	2.31	2.30	2.26
14	2.53	2.48	2.44	2.41	2.39	2.31	2.27	2.24	2.22	2.19
15	2.48	2.42	2.38	2.35	2.33	2.25	2.20	2.18	2.16	2.12
16	2.42	2.37	2.33	2.30	2.28	2.19	2.15	2.12	2.11	2.07
17	2.38	2.33	2.29	2.26	2.23	2.15	2.10	2.08	2.06	2.02
18	2.34	2.29	2.25	2.22	2.19	2.11	2.06	2.04	2.02	1.98
19	2.31	2.26	2.21	2.18	2.16	2.07	2.03	2.00	1.98	1.94
20	2.28	2.22	2.18	2.15	2.12	2.04	1.99	1.97	1.95	1.91
22	2.23	2.17	2.13	2.10	2.07	1.98	1.94	1.91	1.89	1.85
24	2.18	2.13	2.09	2.05	2.03	1.94	1.89	1.86	1.84	1.80
26	2.15	2.09	2.05	2.02	1.99	1.90	1.85	1.82	1.80	1.76
28	2.12	2.06	2.02	1.99	1.96	1.87	1.82	1.79	1.77	1.73
30	2.09	2.04	1.99	1.96	1.93	1.84	1.79	1.76	1.74	1.70
40	2.00	1.95	1.90	1.87	1.84	1.74	1.69	1.66	1.64	1.59
50	1.95	1.89	1.85	1.81	1.78	1.69	1.63	1.60	1.58	1.52
60	1.92	1.86	1.82	1.78	1.75	1.65	1.59	1.56	1.53	1.48
100	1.85	1.79	1.75	1.71	1.68	1.57	1.52	1.48	1.45	1.39

Творці цього конспекту Олексій Галганов та Нікіта Фордуй вдячні Ірині Юріївні за викладання, а студентами КА-91 та КА-92 — за знайдені помилки.



Таємний рецензент — кіт Гаррі.