

Міністерство освіти і науки, молоді та спорту України
Львівський національний університет імені Івана Франка

Ю. В. Жерновий

**ТЕОРІЯ ЙМОВІРНОСТЕЙ
ТА
МАТЕМАТИЧНА СТАТИСТИКА**

Тексти лекцій
для студентів нематематичних спеціальностей

Львів
2012

Лекція 1. ПРЕДМЕТ ТЕОРІЇ ЙМОВІРНОСТЕЙ

У повсякденному житті нам часто доводиться зустрічатися з різними явищами і фактами, які ми називаємо *випадковими*. Зокрема, інформація, на основі якої розв'язуються практичні задачі в економіці, зазвичай носить *наближений, неточний, випадковий* характер. Наприклад, власник магазину не знає, скільки буде покупців, бізнесмен – яким буде завтра курс гривні, банкір – чи повернуть йому позику. Але й у випадкових фактах за певних умов можуть бути виявлені певні закономірності. Ці закономірності вивчає теорія ймовірностей. Для розв'язання задач, пов'язаних з аналізом економічної інформації, використовують ймовірнісні та статистичні методи, оскільки характерною особливістю теорії ймовірностей є те, що вона розглядає явища, в яких в тій чи іншій формі присутня невизначеність.

Широко розповсюджене уявлення пов'язує невизначеність і, отже, ймовірність з такими ситуаціями як гра в кості, в рулетку, витягування карт з колоди і т.п. Саме потреба у розв'язуванні практичних задач, пов'язаних з азартними іграми, а також з питаннями страхування і демографії, якими в *середині 17 ст.* займалися такі відомі вчені як *Гюйгенс, Паскаль, Ферма і Яків Бернуллі*, призвела до виникнення теорії ймовірностей як самостійної науки.

Як і в кожній математичній дисципліні, в теорії ймовірностей існують деякі *початкові, первісні* поняття, які покладені в її основу.

Першим таким поняттям є поняття *випадкової події*. До нього приходимо так.

По-перше, під *подією* розуміємо таку *дію*, про яку можна сказати, що вона відбулась, або відбувається, або може відбутись, або неможлива.

Приклади: 1) 1 вересня почалось навчання в ЛНУ ім. І.Франка (подія відбулась). 2) При киданні монети герб випаде зверху (може відбутись, може не відбутись). 3) Витягнути зелену кульку зі скриньки, яка містить 1 білу і 1 чорну кульку (неможлива подія).

По-друге, поняття випадкової події пов'язане з заданням *певного комплексу умов*.

Приклади комплексів умов:

1) На *тверду плоску рівну* поверхню кидається кубик правильної форми, виготовлений з однорідного матеріалу, з гранями, позначеними одним, двома, ..., шістьма очками.

2) Ведеться спостереження за певною частиною неба в зоряну ніч протягом деякого проміжку часу.

3) Серед населення якого-небудь населеного пункту вибирають навмання 100 чол.

Процес реалізації певного комплексу умов називається *експериментом*.

Тепер можемо дати означення *випадкової події*. **Випадкова подія** – це подія, яка може *відбутись або не відбутись* в результаті здійснення деякого експерименту, тобто в результаті реалізації певного комплексу умов.

Приклади випадкових подій (в зв'язку з вищевказаними комплексами умов):

- 1) Випадання грані кубика з парною кількістю очок.
- 2) Поява комети.
- 3) Наявність хоча б 10-ти чоловік з вищою освітою серед вибраних ста чоловік.

Якщо під час кожної реалізації заданого комплексу умов подія *обов'язково* відбувається, то вона називається **вірогідною**.

Якщо ж в результаті експерименту подія *обов'язково* не відбудеться, то це – **неможлива** подія.

Очевидно, що після *одноразового* здійснення експерименту, ми не виявимо закономірностей, які властиві для конкретної випадкової події. Однак закономірності можна виявити, якщо здійснювати експеримент *багаторазово в однакових умовах*.

Наприклад, дані реєстрації народжувань дітей в невеликому населеному пункті за невеликий період часу не дають *стійких* співвідношень між кількістю народжених хлопчиків і дівчаток. Однак якщо зібрати статистичні дані *по всій країні за тривалий період* (кілька десятиріч) і проаналізувати їх, то виявиться певна закономірність: на кожну 1000 народжених припадає в *середньому* 515 хлопчиків.

Отже, другим з початкових понять теорії ймовірностей є поняття *масовості* явищ (подій).

Масові явища розглядаються як протилежність до одиничних і масовість може проявлятися: 1) в часі; 2) в просторі; 3) і в часі, і в просторі.

Тепер ми можемо остаточно сформулювати, що є *предметом* теорії ймовірностей.

Предметом теорії ймовірностей є вивчення кількісних закономірностей, характерних для масових однорідних випадкових подій.

Теорія ймовірностей є теоретичною базою для математичної статистики.

Предмет математичної статистики – дослідження закономірностей, яким підпорядковані масові випадкові явища, на підставі статистичних даних – результатів спостережень. Ці закономірності вивчають за допомогою методів теорії ймовірностей.

Статистичне означення ймовірності

Нехай деякий експеримент здійснюється n разів і при цьому в результаті t випробувань відбувається подія A . Відношення t/n називається *відносною частотою* появи події A в серії з n випробувань. Теорія ймовірностей вивчає лише такі події, для яких має місце властивість *стійкості частот*. Вона полягає в тому, що частота появи події A при *великій кількості* випробувань мало відрізняється від деякого числа.

Наприклад, якщо багато разів підкидати монету, то частота випадання герба буде мало відрізнятися від $\frac{1}{2}$. Частота народження хлопчика для великої кількості проаналізованих даних буде мало відрізнятися від 0,515.

Звідси, логічно міркуючи, отримуємо так зване *статистичне означення ймовірності*. *Ймовірністю* події називається *об'єктивно існуюча величина*, навколо якої групуються відносні частоти цієї події при великій кількості випробувань. Позначаємо $P(A)$ – ймовірність події A .

Легко зрозуміти, що за значеннями відносних частот можна отримати лише наближене значення ймовірності, тому з точки зору математики таке означення є недосконалим.

Якщо формулювати досконале статистичне означення з точки зору математики, то можна сказати, що *ймовірністю події є границя*, до якої прямує відносна частота появи події при необмеженому зростанні кількості випробувань: $P(A) = \lim_{n \rightarrow \infty} \frac{m(n)}{n}$. Це означення належить німецькому математику Ріхарду Мізесу.

Лекція 2. ЕЛЕМЕНТИ КОМБІНАТОРИКИ

Комбінаторика – розділ математики, предметом якого є теорія скінченних множин.

Множина – сукупність об'єктів довільної природи, які володіють спільною для всіх них характеристичною властивістю. **Приклади:** множина всіх дійсних чисел; множина цілих чисел – нескінченні множини; множина цілих чисел від 1 до 100 – скінченна множина.

A, B, C, \dots – множини; a, b, c – їх елементи. Нагадаємо, що $A \cup B, A \cap B, A \setminus B, \emptyset$ – відповідно об'єднання, перетин, різниця множин і порожня множина. Позначимо: $N(A)$ – кількість елементів множини A . Якщо $N(A)=n$, то казатимемо, що A – n -множина.

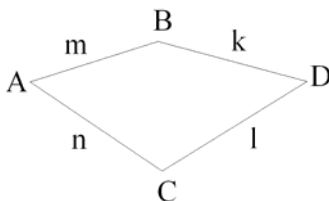
В основі комбінаторики лежать два елементарні правила – суми і добутку.

Сформулюємо **правило суми**.

$$N(A)=m, N(B)=n, A \cap B = \emptyset \Rightarrow N(A \cup B) = m + n.$$

Правило добутку (основне правило комбінаторики): якщо I-у дію можна здійснити n_1 способами, II-у, яка не залежить від першої, – n_2 способами, ..., k -ту, яка не залежить від усіх попередніх, – n_k способами, то першу, другу, ..., k -ту дії *поспідовно* можна здійснити $n_1 n_2 \dots n_k$ способами.

Приклад. Скількома способами можна потрапити з п.А до п.Д, якщо з А до В веде m доріг, з В до D – k доріг, з А до С – n доріг і з С до D – l доріг?



Розв'язання. За правилом добутку рух шляхом ABD можна здійснити mk способами, а шляхом ACD – nl способами. Згідно з правилом суми з А до D можна потрапити $mk + nl$ способами.

Впорядковані множини

Озн. Множина M називається *впорядкованою*, якщо в ній встановлено відношення порядку, що має такі властивості:

- 1) $\forall a, b \in M$ або $a \prec b$, або $b \prec a$;
- 2) якщо $a \prec b, b \prec c$, то $a \prec c$.

Для впорядкування n -множини досить кожному з її елементів приписати один з номерів $1, 2, \dots, n$, або просто записати її елементи в певному порядку.

Ту саму множину, очевидно, можна впорядкувати по-різному. Наприклад, для множини $M = \{1, 2, 3\}$ можливі такі впорядкування:

$$(1, 2, 3), (1, 3, 2), (2, 1, 3), (2, 3, 1), (3, 1, 2), (3, 2, 1).$$

Розміщення, перестановки, комбінації

Озн. Нехай M – n -множина; $k \leq n$, $k \in N$. **Розміщенням** з n елементів по k називають будь-яку *впорядковану* k -підмножину множини M .

Кількість розміщень з n елементів по k обчислюють за формулою:

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

Приклад. $M = \{1, 2, 3\}$; $n=3$, $k=2$; $A_3^2 = \frac{3!}{1!} = 6$. Справді, можливі розміщення

такі:

(1;2), (1;3), (2;1), (2;3), (3;1), (3;2) – всього шість.

Озн. **Перестановкою** з n елементів називається розміщення з n елементів по n , тобто будь-яке впорядкування n -множини, яка складається з різних елементів.

Щоб задати перестановку з n елементів, досить якимось чином впорядкувати n -множину.

Кількість перестановок з n елементів обчислюють за формулою:

$$P_n = n! \quad \left(P_n = A_n^n = \frac{n!}{0!} = n! \right).$$

Зокрема, $P_3 = 3! = 6$, в чому можна переконавшись, повернувшись до прикладу про можливі впорядкування множини $M = \{1, 2, 3\}$.

Озн. Нехай M – n -множина; $k \leq n$, $k \in N$. **Комбінацією** з n елементів по k називають будь-яку k -підмножину множини M .

Кількість комбінацій з n елементів по k обчислюють за формулою:

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

Приклад. $M = \{1, 2, 3\}$; $n=3$, $k=2$; $C_3^2 = \frac{3!}{2!1!} = 3$. Справді, можливі комбінації

такі: $\{1;2\}$, $\{1;3\}$, $\{2;3\}$ – всього три.

Числа C_n^k називають біномними коефіцієнтами, оскільки вони фігурують у відомій формулі біному Ньютона $(a+b)^n = \sum_{k=0}^n C_n^k a^k b^{n-k}$. Поклавши у ній $a = b = 1$,

одержимо: $2^n = \sum_{k=0}^n C_n^k$ – кількість підмножин n -множини.

Розміщення з повтореннями

Озн. Нехай M – n -множина; $k \in N$ – довільне натуральне число. **Розміщенням з повтореннями** з n елементів по k називають будь-яку *впорядковану* множину вигляду (a_1, a_2, \dots, a_k) , де $a_i, i = \overline{1, k}$ – елементи множини M , не обов'язково різні.

Кількість розміщень з повтореннями з n елементів по k обчислюють за формулою: $\overline{A}_n^k = n^k$.

Приклади. 1) $M=\{1, 2\}$; $n=2$, $k=3$; $\overline{A_2^3} = 2^3 = 8$. Справді, можливі розміщення такі:

$(2,1,1), (1;2,1), (1,1,2), (1,2,2), (2,1,2), (2;2,1), (1,1,1), (2,2,2)$ – всього вісім.

2) $M=\{1, 2, 3\}$; $n=3$, $k=2$; $\overline{A_3^2} = 3^2 = 9$. Справді, можливі розміщення такі:

$(1;2), (1;3), (2;1), (2;3), (3;1), (3;2), (1,1), (2,2), (3,3)$ – всього дев'ять.

Перестановки з повтореннями

Озн. *Перестановкою з повтореннями* з n елементів називається будь-яке впорядкування n -множини, серед елементів якої є однакові.

Якщо серед елементів n -множини є n_1 елементів першого типу, n_2 елементів другого типу, ..., n_k елементів k -ого типу, причому $n_1 + n_2 + \dots + n_k = n$, то кількість перестановок з повтореннями з n елементів обчислюють за формулою: $P_n(n_1, n_2, \dots, n_k) = \frac{n!}{n_1! n_2! \dots n_k!}$. Остання формула задає також кількість

способів розбиття n -множини на k підмножин, що містять відповідно n_1, n_2, \dots, n_k елементів.

Приклад. $M=\{1, 2, 2\}$; $n=3$, $n_1=1$, $n_2=2$, $k=2$; $P_3(1,2) = \frac{3!}{1!2!} = 3$. Справді, можливі перестановки такі: $(1,2,2), (2,1,2), (2,2,1)$ – всього три.

Приклад розбиття n -множини. $M=\{1, 2, 3\}$; $n=3$, $n_1=1$, $n_2=2$, $k=2$; $P_3(1,2) = 3$. Справді, можливі розбиття такі: $\{1\}, \{2,3\}$; $\{2\}, \{1,3\}$; $\{3\}, \{1,2\}$ – всього три.

Комбінації з повтореннями

Озн. Нехай M – n -множина; $k \in N$ – довільне натуральне число. *Комбінацією з повтореннями* з n елементів по k називають будь-яку k -множину вигляду $\{a_1, a_2, \dots, a_k\}$, де $a_i, i = \overline{1, k}$ – елементи множини M , не обов'язково різні.

Кількість комбінацій з повтореннями з n елементів по k обчислюють за формулою: $\overline{C_n^k} = C_{n+k-1}^k = \frac{(n+k-1)!}{(n-1)!k!}$.

Приклади. 1) $M=\{1, 2\}$; $n=2$, $k=3$; $\overline{C_2^3} = C_4^3 = \frac{4!}{1!3!} = 4$. Справді, можливі комбінації такі: $\{2,1,1\}, \{1,2,2\}, \{1,1,1\}, \{2,2,2\}$ – всього чотири.

2) $M=\{1, 2, 3\}$; $n=3$, $k=2$; $\overline{C_3^2} = C_4^2 = \frac{4!}{2!2!} = 6$. Справді, можливі комбінації такі: $\{1;2\}, \{1;3\}, \{2;3\}, \{1,1\}, \{2,2\}, \{3,3\}$ – всього шість.

Лекція 3

Простір елементарних подій. Операції над випадковими подіями

Простором елементарних подій (Ω) назвемо множину, елементами якої є всі елементарні події, пов'язані з даним випробуванням.

Під *елементарними* подіями (ω) ми розуміємо всі нерозкладні результати даного випробування, які взаємно виключають один одного.

Приклади. 1) Підкидання монети 1 раз. Можливі результати у цьому експерименті: випадання герба (елементарна подія G або ω_1); випадання цифри (елементарна подія C або ω_2). $\Omega = \{G, C\}$ або $\Omega = \{\omega_1, \omega_2\}$.

2) Підкидання грального кубика 1 раз. Тоді $\Omega = \{\omega_1, \omega_2, \omega_3, \omega_4, \omega_5, \omega_6\}$, де ω_k – випадання k очок.

3) Підкидання монети 2 рази. $\Omega = \{GG, CC, GC, CG\}$.

4) Стрільба по плоскій мішені. Введемо в площині мішені прямокутну систему координат xOy і влучання в певну точку площини поставимо у відповідність координати цієї точки. Тоді простором елементарних подій є вся площина, тобто множина всіх впорядкованих пар дійсних чисел:

$$\Omega = \{(x, y) : -\infty < x < \infty, -\infty < y < \infty\}.$$

Нехай Ω – довільний простір елементарних подій. *Випадковими подіями* або просто *подіями* назвемо підмножини A множини Ω . Отже, ми поняття *подія* ототожнюємо з поняттям *множина*. Для прикладу 2 (підкидання кубика) подіями є: $A = \{\omega_1, \omega_3, \omega_5\}$ – випадання непарної кількості очок; $B = \{\omega_2, \omega_4, \omega_6\}$ – випадання парної кількості очок і т.п. Для прикладу 4 (стрільба) подією є будь-яка область A в площині xOy . Подія A відбувається, якщо відбувається влучання в точку $(x, y) \in A$.

Подію Ω називатимемо *вірогідною*, вона обов'язково відбудеться в результаті випробування, а подію \emptyset – *неможливою*, вона обов'язково не відбудеться в результаті випробування.

Розглянемо *відношення*, в яких можуть перебувати події одна відносно одної, і *операції над подіями*.

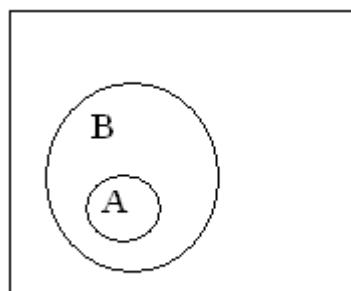
Означення. Кажуть, що подія A є *окремим випадком* події B (або подія A тягне за собою B , або B є наслідком A), якщо множина A є підмножиною множини B .

Позначають ці відношення так само як для множин: $A \subset B$ або $B \supset A$.

Відношення $A \subset B$ означає, що всі елементарні події, які входять до складу A , належать також і до B , тобто кожного разу, коли відбувається подія A , відбувається також і подія B .

Приклади. 1) $A = \{\omega_2, \omega_6\}$, $B = \{\omega_2, \omega_4, \omega_6\} \Rightarrow A \subset B$.

2) Подія A – влучання в область A ,
подія B – влучання в область B
 $\Rightarrow A \subset B$.



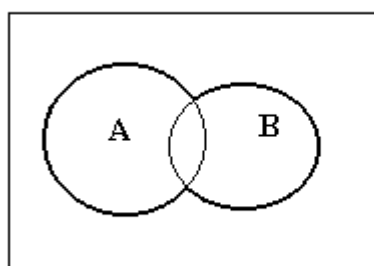
Означення. Події A і B називаються **рівносильними**, якщо $A \subset B$ і $B \subset A$. Рівносильність подій позначають так: $A=B$.

Означення. **Сумою** двох подій A і B ($A+B$ або $A \cup B$) називається така подія, яка відбудеться тоді і лише тоді, коли відбудеться хоча б одна з подій A або B (або лише A , або лише B , або і A , і B).

Згідно з цим означенням, якщо події A відповідає підмножина A простору Ω , а події B – підмножина B , то події $A+B$ відповідає підмножина $A \cup B$, яка складається з усіх елементарних подій, які належать об'єднанню множин A і B .

Приклади. 1) $A=\{\omega_2, \omega_4\}$, $B=\{\omega_4, \omega_6\} \Rightarrow A+B=\{\omega_2, \omega_4, \omega_6\}$.

2) Подія A – влучання в область A , подія B – влучання в область $B \Rightarrow A+B$ – влучання хоча б в одну з областей A або B .

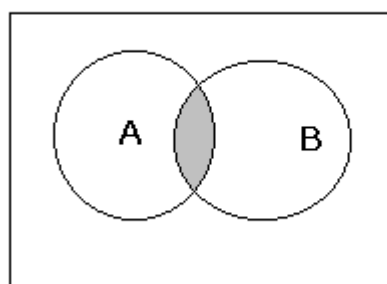


Означення. **Добутком** двох подій A і B (AB або $A \cap B$) називається така подія, яка відбудеться тоді і лише тоді, коли відбудеться як подія A , так і подія B .

Події AB відповідає перетин множин A і B , тобто вона складається з елементарних подій, які входять до складу обидвох множин A і B .

Приклади. 1) $A=\{\omega_2, \omega_4\}$, $B=\{\omega_4, \omega_6\} \Rightarrow A \cap B=\{\omega_4\}$.

2) Подія A – влучання в область A , подія B – влучання в область $B \Rightarrow A \cap B$ – влучання в перетин цих областей.

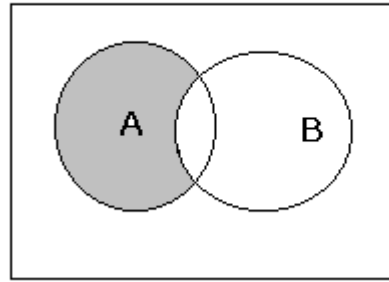


Означення. **Різницею** двох подій A і B ($A-B$ або $A \setminus B$) називається така подія, яка відбудеться тоді і лише тоді, коли відбудеться подія A , але не відбудеться подія B .

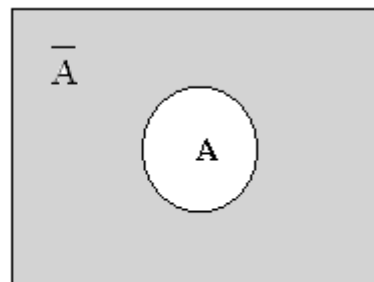
Події $A-B$ відповідає різниця множин A і B , тобто вона складається з елементарних подій, які входять до A , але не входять до B .

Приклади. 1) $A=\{\omega_2, \omega_4\}$, $B=\{\omega_4, \omega_6\} \Rightarrow A-B=\{\omega_2\}$.

2) Подія A – влучання в область A , подія B – влучання в область $B \Rightarrow A-B$ – влучання в різницю цих областей.



Означення. *Протилежною подією \bar{A} до події A називається подія $\Omega-A$, вона означає, що подія A не відбулася.*



Приклади. 1) $A = \{\omega_2, \omega_4\} \Rightarrow \bar{A} = \{\omega_1, \omega_3, \omega_5, \omega_6\}$.

2) Подія A – влучання в область A , подія \bar{A} – невлучання в область A .

Означення. *Події A і B називаються несумісними, якщо їх добуток є неможлива подія, тобто $AB = \emptyset$.*

Несумісність подій A і B означає, що поява події A виключає можливість появи події B , і навпаки.

Приклад. $A = \{\omega_2, \omega_4\}, B = \{\omega_3, \omega_5\} \Rightarrow A$ і B – несумісні.

Поняття суми і добутку подій поширюються на випадок будь-якої скінченної, а також зліченної кількості подій. Зокрема, подія $\bigcup_{i=1}^{\infty} A_i = A_1 + A_2 + \dots$ відбувається тоді і лише тоді, коли відбувається принаймні одна з подій $A_i, i = 1, 2, \dots$, а подія $\bigcap_{i=1}^{\infty} A_i = A_1 A_2 \dots$ відбувається тоді і лише тоді, коли відбуваються всі події $A_i, i = 1, 2, \dots$

Означення. *Події A_1, A_2, \dots, A_n утворюють повну групу подій, якщо в результаті виконання експерименту принаймні одна з цих подій обов'язково відбудеться, тобто $A_1 + A_2 + \dots + A_n = \Omega$.*

Особливо істотними для нас надалі будуть повні групи попарно несумісних подій. Такою, зокрема, є сукупність усіх елементарних подій у випадку, коли Ω є скінченна множина. Повну групу попарно несумісних подій

завжди утворюють будь-яка подія A та протилежна до неї подія \bar{A} , оскільки $A\bar{A} = \emptyset$, $A + \bar{A} = \Omega$.

Для довільних подій безпосередньо з означень операцій над подіями випливають співвідношення $AA = A$, $A + A = A$, $(A + B)C = AC + BC$, а також закони де Моргана $\overline{A + B} = \bar{A}\bar{B}$, $\overline{AB} = \bar{A} + \bar{B}$.

Алгебра і σ -алгебра подій. Аксиоматичне означення ймовірності

Ймовірність – це функція від випадкової події. Функції дійсної змінної, як ми знаємо, зазвичай визначені не для всіх дійсних чисел, а лише для деякої підмножини множини \mathbb{R} , яка називається областю визначення. Ймовірність також не завжди вдається визначити для будь-яких підмножин (випадкових подій) множини Ω . Доводиться обмежуватися деяким класом підмножин. Від цього класу природно вимагати, щоб він був замкненим відносно операцій, введених для подій вище.

Означення. Нехай Ω – довільний простір елементарних подій, а S – деяка сукупність випадкових подій. Сукупність подій S називається **алгеброю подій**, якщо виконуються умови:

- 1) $\Omega \in S$;
- 2) якщо $A \in S$, $B \in S$, то $A + B \in S$, $A - B \in S$.

Неможлива подія завжди входить до алгебри подій, оскільки $\emptyset = \Omega - \Omega$. Якщо $A \in S$, то $\bar{A} \in S$, оскільки $\bar{A} = \Omega - A$. Нарешті, якщо $A \in S$, $B \in S$, то $AB \in S$, тому що $AB = \overline{\bar{A} + \bar{B}}$ (це випливає з другого закону де Моргана).

Найменшою сукупністю подій, яка є алгеброю, очевидно, є множина $S = \{\Omega, \emptyset\}$, оскільки $\Omega \pm \emptyset = \Omega$, $\emptyset - \Omega = \emptyset$.

Означення. Алгебра подій \mathcal{F} називається **σ -алгеброю** або борелівською алгеброю, якщо з того, що $A_i \in \mathcal{F}$, $i=1,2,\dots$, випливає, що $\bigcup_{i=1}^{\infty} A_i \in \mathcal{F}$.

Зауважимо, що у випадку скінченної множини Ω будь-яка алгебра подій є σ -алгеброю.

Тепер ми можемо ввести поняття ймовірності події.

Означення. Числова функція P , визначена на σ -алгебрі подій \mathcal{F} , називається **ймовірністю**, якщо

- 1) $P(A) \geq 0$;
- 2) $P(\Omega) = 1$;
- 3а) $AB = \emptyset \Rightarrow P(A + B) = P(A) + P(B)$ (скінченна адитивність);
- 3б) якщо в послідовності подій A_1, A_2, \dots всі події попарно несумісні, то

$$P\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} P(A_i) \text{ (зліченна адитивність);}$$

4) Нехай для зростаючої послідовності подій $A_1 \subset A_2 \subset \dots \subset A_n \subset A_{n+1} \subset \dots$
 і $A = \bigcup_{n=1}^{\infty} A_n$ або для спадної послідовності подій $A_1 \supset A_2 \supset \dots \supset A_n \supset A_{n+1} \supset \dots$ і
 $A = \bigcap_{n=1}^{\infty} A_n$, тоді $P(A) = \lim_{n \rightarrow \infty} P(A_n)$ (аксіома неперервності).

Трійку (Ω, \mathcal{F}, P) , де \mathcal{F} – σ -алгебра, а P – ймовірність, визначена вище, називатимемо ймовірнісним простором.

Наслідки з аксіом

$$1) P(\bar{A}) = 1 - P(A).$$

Доведення. Оскільки $A + \bar{A} = \Omega$, то $1 = P(\Omega) = P(A + \bar{A}) = P(A) + P(\bar{A})$.

$$2) P(\emptyset) = 0.$$

Доведення. Оскільки $\bar{\Omega} = \Omega - \Omega = \emptyset$, то $P(\emptyset) = P(\bar{\Omega}) \stackrel{H1}{=} 1 - P(\Omega) \stackrel{A2}{=} 1 - 1 = 0$.

3) $\forall A, B \quad P(A + B) = P(A) + P(B) - P(AB)$ – теорема додавання ймовірностей.

Доведення. Оскільки $A + B = A + B\bar{A}$, де доданки у правій частині – несумісні події, то $P(A + B) \stackrel{A3}{=} P(A) + P(B\bar{A})$. Але $B = BA + B\bar{A}$, тому, аналогічно, $P(B) \stackrel{A3}{=} P(BA) + P(B\bar{A})$, звідки $P(B\bar{A}) = P(B) - P(BA)$, тому

$$P(A + B) = P(A) + P(B) - P(AB).$$

$$4) P(A + B) \leq P(A) + P(B); \quad P(A_1 + \dots + A_n) \leq P(A_1) + \dots + P(A_n).$$

Доведення. $P(A + B) = P(A) + P(B) - P(AB) \leq P(A) + P(B)$, оскільки $P(AB) \stackrel{A1}{\geq} 0$.
 Для випадку n доданків нерівність можна довести, використовуючи метод математичної індукції.

$$5) A \subset B \Rightarrow P(A) \leq P(B).$$

Доведення. Якщо $A \subset B$, то $B = BA + B\bar{A} = A + B\bar{A}$, тому

$$P(B) \stackrel{A3}{=} P(A) + P(B\bar{A}) \stackrel{A1}{\geq} P(A).$$

$$6) \forall A \quad 0 \leq P(A) \leq 1.$$

Доведення. $A \subset \Omega \stackrel{H5}{\Rightarrow} P(A) \leq P(\Omega) = 1$.

Лекція 4

Класичне означення ймовірності

Нехай простір елементарних подій є скінченною множиною

$$\Omega = \{\omega_1, \omega_2, \dots, \omega_N\}, \quad N(\Omega) = N,$$

тобто є лише N можливих результатів випробування, отже, множина Ω є повною групою всіх попарно несумісних результатів випробування.

Вважатимемо додатково, що всі елементарні події *рівноможливі*, тобто з міркувань симетрії або якихось інших впливає, що нема об'єктивних причин вважати одну з елементарних подій більш імовірною порівняно з іншими. Наприклад, якщо експеримент полягає в одноразовому киданні грального кубика правильної форми, виготовленого з однорідного матеріалу, то всі 6 результатів цього випробування природно вважати рівноможливими.

Класичне означення ймовірності формулюють для подій, які є підмножинами множини Ω . Якщо $A = \{\omega_{i_1}, \omega_{i_2}, \dots, \omega_{i_{N(A)}}\}$, де

$$1 \leq i_1 < i_2 < \dots < i_{N(A)} \leq N, \quad N(A) = 0, 1, 2, \dots, N,$$

то ймовірність події A визначають за формулою:

$$P(A) = \frac{N(A)}{N(\Omega)}, \quad (1)$$

де $N(A)$ – кількість елементів множини A .

Отже, **ймовірністю** події A називається відношення кількості сприятливих для події A елементарних подій до кількості всіх рівноможливих і попарно несумісних результатів випробування.

Визначена за формулою (1) функція $P(A)$ задовольняє всі аксіоми теорії ймовірностей. З формули (1) випливає, що ймовірність кожної елементарної події ω_i рівна

$$P(\{\omega_i\}) = \frac{1}{N}, \quad i = \overline{1, N}. \quad (2)$$

Отож класична схема означення ймовірності може служити моделлю тих випадкових явищ, для яких є природним припущення (2).

Приклад. Знайти ймовірність того, що кількість очок, яка випаде при одноразовому киданні кубика, буде а) парною (подія A); б) кратною трьом (подія B).

Розв'язання.

$$\Omega = \{\omega_1, \omega_2, \dots, \omega_6\}, \quad N(\Omega) = 6, \quad A = \{\omega_2, \omega_4, \omega_6\}, \quad B = \{\omega_3, \omega_6\}, \quad N(A) = 3, \quad N(B) = 2,$$

$$P(A) = \frac{3}{6} = \frac{1}{2}, \quad P(B) = \frac{2}{6} = \frac{1}{3}.$$

Геометричні ймовірності

Класичне означення ймовірності не можна застосувати до випробування, для якого множина Ω є незліченною множиною елементарних подій.

Нехай простір елементарних подій Ω – це відрізок числової прямої або область на площині, або в просторі, а елементарні події ω – окремі точки в межах цієї області. Припустимо, що область Ω має скінченну міру $\mu(\Omega)$ (на прямій – довжину, на площині – площу, у просторі – об’єм). Розглянемо систему \mathcal{F} підмножин простору Ω , які мають міру. Відомо, що вони утворюють σ -алгебру. Множини з \mathcal{F} назвемо випадковими подіями. Якщо експеримент має властивість симетрії щодо елементарних результатів (наприклад, деякий „точковий” об’єкт “навмання” кидаємо в межах області), то всі елементарні події „рівноправні”, тож природно припустити, що ймовірність попадання елементарної події ω у будь-яку частину Ω пропорційна мірі цієї частини і не залежить від її розташування і форми. Тоді ймовірність будь-якої події $A \in \mathcal{F}$ можна обчислити, користуючись таким означенням.

Означення. *Геометричною ймовірністю події A називається відношення міри $\mu(A)$ до міри $\mu(\Omega)$, тобто $P(A) = \frac{\mu(A)}{\mu(\Omega)}$.*

Можна показати, що геометрична ймовірність задовольняє всі аксіоми теорії ймовірностей.

Зауваження. За класичного означення ймовірності $P(A) = 0$ лише для $A = \emptyset$. За геометричного це не так. Справді, нехай Ω – плоска область, A – точка або лінія, розміщена в Ω . Тоді за формулою геометричної ймовірності $P(A) = 0$, хоча подія A є можливою – точка в разі її „кидання” на Ω може потрапити на A .

Лекція 5.

Умовні ймовірності. Теорема множення ймовірностей. Незалежність подій

Пояснимо спочатку на *прикладі* суть умовної ймовірності. Нехай двічі кидають гральний кубик. Простір елементарних подій складається з 36 елементів: $\Omega = \{(m, n) : m, n = \overline{1, 6}\}$. Розглянемо подію A – „сума очок після двох кидань рівна 5”, тоді $A = \{(1, 4), (2, 3), (3, 2), (4, 1)\}$, $P(A) = \frac{N(A)}{N(\Omega)} = \frac{4}{36} = \frac{1}{9}$.

Припустимо, що результат першого кидання – випадання 3 очок (подія $B = \{(3, 1), (3, 2), (3, 3), (3, 4), (3, 5), (3, 6)\}$). В такій ситуації ймовірність події A зміниться: сумарна кількість очок буде рівною 5 лише тоді, коли за другим разом випаде „2”, тобто в одному випадку з шести. Отже, ймовірність події A за умови, що відбулась подія B , $P(A/B) = 1/6$.

В загальному випадку (в умовах класичної схеми) міркуємо так. Якщо відбулась подія B , то здійснився один з $N(B)$ елементарних результатів. Серед них подія A з’являється $N(AB)$ разів. Тому

$$P(A/B) = \frac{N(AB)}{N(B)} = \frac{N(AB)/N(\Omega)}{N(B)/N(\Omega)} = \frac{P(AB)}{P(B)}.$$

У випадку загальної ймовірнісної моделі рівність

$$P(A/B) = \frac{P(AB)}{P(B)}, \quad P(B) > 0 \quad (1)$$

приймається за означення **умовної ймовірності**.

З цього означення випливають такі властивості умовної ймовірності:

- 1) $P(A/B) \geq 0$;
- 2) $P(\Omega/B) = 1$;
- 3) $P(B/B) = 1$;
- 4) $A_1 A_2 = \emptyset \Rightarrow P(A_1 + A_2 / B) = P(A_1 / B) + P(A_2 / B)$.

Доведення.

$$2) P(\Omega/B) = \frac{P(\Omega B)}{P(B)} = \frac{P(B)}{P(B)} = 1;$$

$$3) P(B/B) = \frac{P(BB)}{P(B)} = \frac{P(B)}{P(B)} = 1;$$

$$4) P(A_1 + A_2 / B) = \frac{P((A_1 + A_2)B)}{P(B)} = \frac{P(A_1 B)}{P(B)} + \frac{P(A_2 B)}{P(B)} = P(A_1 / B) + P(A_2 / B). \quad \square$$

З рівності (1) безпосередньо випливає **теорема множення ймовірностей**: якщо $P(A) > 0$, $P(B) > 0$, то

$$P(AB) = P(A)P(B/A) = P(B)P(A/B). \quad (2)$$

Приклад. Обчислити ймовірність $P(AB)$ у прикладі про кидання кубика, розглянутому на початку параграфа.

Розв'язання. $P(A/B) = 1/6$, $P(B) = 6/36 = 1/6 \Rightarrow P(AB) = 1/36$.

Справді, $AB = \{3, 2\}$, і за класичним означенням $P(AB) = 1/36$. \square

Формулу множення ймовірностей (2) можна узагальнити для випадку будь-якої скінченної кількості подій. Нехай A_1, A_2, \dots, A_n – випадкові події такі, що $P(A_1 A_2 \dots A_{n-1}) > 0$. Тоді

$$P(A_1 A_2 \dots A_n) = P(A_1)P(A_2 / A_1)P(A_3 / A_1 A_2) \dots P(A_n / A_1 A_2 \dots A_{n-1}). \quad (3)$$

Доведемо формулу (3) за допомогою методу математичної індукції. Для $n=2$ справедливість (3) випливає безпосередньо з (2). Припустимо, що формула (3) виконується для $n-1$ множника і введемо позначення: $A_1 A_2 \dots A_{n-1} = A$, $A_n = B$. Тоді

$$\begin{aligned} P(A_1 A_2 \dots A_n) &= P((A_1 A_2 \dots A_{n-1}) A_n) = P(AB) = P(A)P(B/A) = \\ &= P(A_1 A_2 \dots A_{n-1})P(A_n / A_1 A_2 \dots A_{n-1}) = \\ &= P(A_1)P(A_2 / A_1)P(A_3 / A_1 A_2) \dots P(A_n / A_1 A_2 \dots A_{n-1}). \quad \square \end{aligned}$$

Введемо поняття **незалежності** випадкових подій, яке в теорії ймовірностей відіграє дуже важливу роль. Нехай Ω – простір елементарних подій, $A \subset \Omega$, $B \subset \Omega$ – випадкові події.

Означення. Події A і B називаються **незалежними**, якщо $P(AB) = P(A)P(B)$.

Приклад. Монету кидають двічі. Нехай A – подія, яка полягає в тому, що за першим разом випав герб, B – подія, яка полягає в тому, що за другим разом випав герб. З'ясувати, чи будуть незалежними події A і B .

Розв'язання. Простором елементарних подій даного експерименту є множина $\Omega = \{ГГ, ЦЦ, ГЦ, ЦГ\}$. Тоді

$$A = \{ГГ, ГЦ\}, \quad B = \{ГГ, ЦГ\}, \quad AB = \{ГГ\}$$

і $P(A) = P(B) = \frac{2}{4} = \frac{1}{2}$, $P(AB) = \frac{1}{4}$. Отже, $P(AB) = P(A)P(B)$, тому випадкові події A і B – незалежні. \square

Розглянемо найважливіші **властивості** ймовірностей для незалежних подій.

1. Якщо події A і B незалежні, то

$$P(A/B) = P(A) \quad (P(B) > 0), \quad P(B/A) = P(B) \quad (P(A) > 0).$$

Доведення випливає безпосередньо з теореми множення.

2. Якщо події A і B незалежні, то незалежні також A і \bar{B} , \bar{A} і B , \bar{A} і \bar{B} .

Доведення. Для обґрунтування цієї властивості досить довести, що події A і \bar{B} – незалежні. Оскільки $A = AB + A\bar{B}$, причому доданки цієї суми – несумісні події, то $P(A) = P(AB) + P(A\bar{B})$. Звідси

$$P(A\bar{B}) = P(A) - P(AB) = P(A) - P(A)P(B) = P(A)(1 - P(B)) = P(A)P(\bar{B}),$$

отже, A і \bar{B} – незалежні. \square

Якщо маємо сукупність подій, яка містить більше, ніж дві події, то формула для обчислення ймовірності їх добутку також спрощується, якщо ці події незалежні в сукупності.

Означення. *Випадкові події A_1, A_2, \dots, A_n називаються незалежними в сукупності, якщо*

$$P(A_{i_1} A_{i_2} \cdot \dots \cdot A_{i_k}) = P(A_{i_1}) P(A_{i_2}) \cdot \dots \cdot P(A_{i_k}) \quad (4)$$

для будь-яких $1 \leq i_1 < i_2 < \dots < i_k \leq n$ ($2 \leq k \leq n$).

Для незалежних у сукупності подій A_1, A_2, \dots, A_n імовірність їх добутку дорівнює добуткові ймовірностей цих подій, тобто

$$P(A_1 A_2 \cdot \dots \cdot A_n) = P(A_1) P(A_2) \cdot \dots \cdot P(A_n).$$

Якщо умова (4) виконується лише для $k=2$, то події називаються *попарно незалежними*. Виявляється, що попарно незалежні події можуть не бути незалежними в сукупності. Про це свідчить приклад, наведений С.Н.Бернштейном.

Приклад. На площину кидають тетраедр, три грані якого пофарбовані відповідно в червоний, синій, жовтий кольори, а на четверту грань нанесено всі три кольори. Розглянемо випадкові події: A_1 – випаде грань із червоним кольором; A_2 – випаде грань із синім кольором; A_3 – випаде грань із жовтим кольором. З'ясувати, чи будуть ці події незалежними в сукупності.

Розв'язання. Очевидно, що $P(A_1) = P(A_2) = P(A_3) = \frac{2}{4} = \frac{1}{2}$;

$$P(A_1 A_2) = \frac{1}{4} = P(A_1) P(A_2); \quad P(A_1 A_3) = \frac{1}{4} = P(A_1) P(A_3); \quad P(A_2 A_3) = \frac{1}{4} = P(A_2) P(A_3).$$

Отже, події A_1, A_2, A_3 – попарно незалежні. Проте ці події не є незалежними в сукупності, бо

$$P(A_1 A_2 A_3) = \frac{1}{4} \neq P(A_1) P(A_2) P(A_3) = \frac{1}{8}. \quad \square$$

Задача. Нехай A_1, A_2, \dots, A_n – випадкові події, незалежні в сукупності, і їхні ймовірності $P(A_1), P(A_2), \dots, P(A_n)$ є відомі. Знайти ймовірність події A , яка полягає в тому, що в результаті виконання експерименту відбудеться хоча б одна з подій A_1, A_2, \dots, A_n .

Розв'язання. З умови задачі випливає, що $A = \bigcup_{i=1}^n A_i$. Розглянемо протилежну подію \bar{A} . Ця подія полягає в тому, що в результаті виконання експерименту не відбудеться жодної з подій A_1, A_2, \dots, A_n , тобто одночасно відбудуться події $\bar{A}_1, \bar{A}_2, \dots, \bar{A}_n$. Тому $\bar{A} = \bigcap_{i=1}^n \bar{A}_i$, і внаслідок незалежності подій $\bar{A}_1, \bar{A}_2, \dots, \bar{A}_n$

$$P(\bar{A}) = P(\bar{A}_1) P(\bar{A}_2) \cdot \dots \cdot P(\bar{A}_n).$$

Отже,

$$P(A) = 1 - P(\bar{A}) = 1 - P(\bar{A}_1)P(\bar{A}_2) \cdot \dots \cdot P(\bar{A}_n),$$

тобто ймовірність появи хоча б однієї з незалежних подій A_1, A_2, \dots, A_n дорівнює різниці між одиницею і добутком імовірностей протилежних до них подій.

Лекція 6

Формула повної ймовірності. Формули Байєса

Припустимо, що подія A може відбуватися в різних умовах, яким відповідають попарно несумісні події H_1, H_2, \dots, H_n , які утворюють повну групу і які назвемо *гіпотезами*. Нехай відомі ймовірності $P(H_1), P(H_2), \dots, P(H_n)$ та умовні ймовірності $P(A/H_1), P(A/H_2), \dots, P(A/H_n)$. Як знайти $P(A)$ в даній ситуації?

Теорема. Якщо H_1, H_2, \dots, H_n – повна група попарно несумісних подій і $P(H_i) > 0$, $i = \overline{1, n}$, то для будь-якої події A справедлива рівність:

$$P(A) = \sum_{i=1}^n P(H_i)P(A/H_i). \quad (1)$$

Доведення. Оскільки події H_1, H_2, \dots, H_n утворюють повну групу подій, то $H_1 + H_2 + \dots + H_n = \Omega$, і подію A можна записати так:

$$A = A\Omega = A(H_1 + H_2 + \dots + H_n) = H_1A + H_2A + \dots + H_nA,$$

де всі доданки у правій частині – попарно несумісні події. Використовуючи аксіому адитивності та теорему множення ймовірностей, маємо:

$$P(A) = P(H_1A) + P(H_2A) + \dots + P(H_nA) = \sum_{i=1}^n P(H_iA) = \sum_{i=1}^n P(H_i)P(A/H_i),$$

що й треба було довести. \square

Рівність (1) називають **формулою повної ймовірності**. Вона виражає ймовірність події A за умови, що відбулася одна і тільки одна з попарно несумісних подій H_1, H_2, \dots, H_n .

Якщо додатково припустити, що $P(A) > 0$, то за теоремою множення ймовірностей можемо записати:

$$P(AH_k) = P(A)P(H_k/A),$$

звідки

$$P(H_k/A) = \frac{P(AH_k)}{P(A)} = \frac{P(H_k)P(A/H_k)}{P(A)}$$

або

$$P(H_k/A) = \frac{P(H_k)P(A/H_k)}{\sum_{i=1}^n P(H_i)P(A/H_i)}, \quad k = \overline{1, n}. \quad (2)$$

Формули (2) називаються **формулами Байєса** (формулами ймовірностей гіпотез). Їм можна дати таке тлумачення. Нехай подія A може відбуватись в різних умовах, щодо характеру яких можна зробити n припущень (гіпотез) H_1, H_2, \dots, H_n . Ймовірності цих гіпотез $P(H_k)$ нам відомі, як і умовні ймовірності $P(A/H_k)$ події A за умови здійснення кожної гіпотези. Якщо в результаті експерименту подія A відбулась, то за формулами Байєса ми можемо *переоцінити* ймовірності кожної з гіпотез, знайшовши $P(H_k/A)$.

Приклад. Зі скриньки, яка містить 5 білих і 3 чорних кулі, одна куля невідомого кольору загублена. Яка ймовірність витягнути навмання зі скриньки білу кулю (подія A)? Яка ймовірність того, що загублено чорну кулю, якщо витягнута навмання куля виявилась білою?

Розв'язання. Тут можливі дві події-гіпотези: H_k – загублено k білих куль ($k=0; 1$). Очевидно, події H_0, H_1 – несумісні й утворюють повну групу, а їхні ймовірності становлять: $P(H_0) = \frac{3}{8}$, $P(H_1) = \frac{5}{8}$. Відповідні умовні ймовірності

події $A = AH_0 + AH_1$ становлять: $P(A/H_0) = \frac{5}{7}$, $P(A/H_1) = \frac{4}{7}$. За формулою повної ймовірності

$$P(A) = \sum_{i=0}^1 P(H_i)P(A/H_i) = \frac{3}{8} \cdot \frac{5}{7} + \frac{5}{8} \cdot \frac{4}{7} = \frac{5}{8}.$$

Ймовірність $P(H_0/A)$ обчислимо за формулою Байєса:

$$P(H_0/A) = \frac{P(H_0)P(A/H_0)}{P(A)} = \frac{\frac{3}{8} \cdot \frac{5}{7}}{\frac{5}{8}} = \frac{3}{7}.$$

Лекція 7

Послідовність незалежних випробувань за схемою Бернуллі. Біномна формула. Біномний розподіл

Припустимо, що проводиться певна кількість однакових випробувань, у кожному з яких можливі лише два несумісні результати: деяка подія A може відбутися або не відбутися. Наприклад, коли підкидаємо 10 разів монету, то за кожного підкидання монети випаде або герб (подія A), або цифра (подія \bar{A}).

Означення. Випробування називаються *незалежними* стосовно деякої події A , якщо ймовірність цієї події в кожному випробуванні не залежить від результатів інших випробувань.

Означення. Серія повторних незалежних випробувань з одним із можливих результатів A або \bar{A} , у кожному з яких подія A має одну і ту саму ймовірність $P(A) = p$, називається *схемою Бернуллі*.

Отже, якщо випробування проводяться за схемою Бернуллі, то в кожному з них можливий тільки один з двох результатів: A (успіх) або \bar{A} (невдача), до того ж ймовірності $P(A) = p$ і $P(\bar{A}) = 1 - p = q$ є однаковими в кожному випробуванні.

Позначимо через μ_n кількість успіхів (кількість появ події A) в серії із n послідовних незалежних випробувань за схемою Бернуллі і обчислимо ймовірність того, що кількість успіхів в n випробуваннях рівна k $P\{\mu_n = k\} = P_n(k)$, за умови, що ймовірність успіху в кожному окремому випробуванні $P(A) = p$. Введемо події: A_i – в i -му випробуванні відбулась подія A , \bar{A}_i – в i -му випробуванні не відбулась подія A . Подію $\{\mu_n = k\}$ можна записати у вигляді

$$\begin{aligned} \{\mu_n = k\} = & A_1 A_2 \dots A_k \bar{A}_{k+1} \bar{A}_{k+2} \dots \bar{A}_n + \\ & + \bar{A}_1 A_2 \dots A_k A_{k+1} \bar{A}_{k+2} \bar{A}_{k+3} \dots \bar{A}_n + \dots + \bar{A}_1 \bar{A}_2 \dots \bar{A}_{n-k} A_{n-k+1} A_{n-k+2} \dots A_n. \end{aligned}$$

Беручи до уваги незалежність випробувань, маємо, що ймовірність кожної з подій-доданків у правій частині рівна $p^k q^{n-k}$. Всі ці доданки – попарно несумісні події, а їхня кількість рівна кількості k -підмножин з n -множини, тобто C_n^k . Застосовуючи аксіому адитивності, одержимо:

$$P_n(k) = C_n^k p^k q^{n-k}. \quad (1)$$

Формула (1) називається **біномною формулою** або *формулою Бернуллі*. Вона виражає ймовірність того, що кількість успіхів в серії з n послідовних незалежних випробувань за схемою Бернуллі дорівнює k .

Набір чисел $P_n(k)$ ($k = 0, 1, \dots, n$) називають **біномним розподілом**.

Події $\{\mu_n = 0\}, \{\mu_n = 1\}, \dots, \{\mu_n = n\}$ утворюють повну групу попарно несумісних подій, отже, $\{\mu_n = 0\} + \{\mu_n = 1\} + \dots + \{\mu_n = n\} = \Omega$, тому

$$\sum_{k=0}^n P_n(k) = P(\Omega) = 1.$$

Цю формулу можна також отримати й безпосереднім обчисленням

$$\sum_{k=0}^n P_n(k) = \sum_{k=0}^n C_n^k p^k q^{n-k} = (p+q)^n = 1^n = 1.$$

Приклад. Кубик кидають 5 разів. Обчислити ймовірність випадання чотирьох шісток.

Розв'язання. Тут $n=5, k=4, p=1/6, q=5/6$. За допомогою біномної формули отримуємо: $P_5(4) = C_5^4 \cdot \frac{1}{6^4} \cdot \frac{5}{6} = \frac{5^2}{6^5} = \frac{25}{7776}$.

Найімовірніша кількість успіхів у випробуваннях за схемою Бернуллі

Теорема. Найімовірніша кількість успіхів k_0 в n випробуваннях за схемою Бернуллі задовольняє умову $np - q \leq k_0 \leq np + p$. Якщо $np - q$ – неціле, то є одне таке значення k_0 , якщо $np - q$ – ціле, то таких значень два.

Доведення. Розглянемо відношення

$$\frac{P_n(k+1)}{P_n(k)} = \frac{C_n^{k+1} p^{k+1} q^{n-k-1}}{C_n^k p^k q^{n-k}} = \frac{\frac{n!}{(n-k-1)!(k+1)!} p}{\frac{n!}{(n-k)!k!} q} = \frac{(n-k)p}{(k+1)q}.$$

Отже, $P_n(k+1) > P_n(k)$, якщо $(n-k)p > (k+1)q$, тобто $np - q > k(p+q) = k$. Тому

$$\begin{cases} P_n(k+1) > P_n(k), & \text{якщо } k < np - q; \\ P_n(k+1) = P_n(k), & \text{якщо } k = np - q; \\ P_n(k+1) < P_n(k), & \text{якщо } k > np - q. \end{cases} \quad (2)$$

Це означає, що зі зростанням k величина $P_n(k)$ спочатку зростає до певного максимуму, а потім спадає. Знайдемо, при якому k досягається максимум. Можливі два випадки.

1) $np - q$ – неціле число. Тоді число $np - q + 1 = np + p$ також неціле і існує єдине ціле число k_0 , що задовольняє умову $np - q < k_0 < np + p$. Покажемо, що k_0 – найімовірніша кількість успіхів, тобто що $P_n(k)$ досягає найбільшого значення при $k = k_0$. Справді, оскільки $k_0 - 1 < np - q$, то згідно з першим співвідношенням (2) $P_n(k_0) > P_n(k_0 - 1)$, а оскільки $k_0 > np - q$, то із третього співвідношення (2) випливає, що $P_n(k_0 + 1) < P_n(k_0)$. Звідси дістаємо, що $P_n(k_0) > P_n(k)$ для всіх $k \neq k_0$, тобто k_0 – найімовірніша кількість успіхів.

2) $np - q$ – ціле число. Покладемо $k_1 = np - q$, тоді $k_1 + 1 = np - q + 1 = np + p$. Згідно з другим співвідношенням (2) $P_n(k_1 + 1) = P_n(k_1)$. Через те, що $k_1 - 1 < np - q$ і $k_1 + 1 > np - q$, то відповідно з першого і третього співвідношень (2) випливає, що $P_n(k_1) > P_n(k_1 - 1)$ і $P_n(k_1 + 2) < P_n(k_1 + 1)$. Отож у цьому випадку є два найімовірніші значення: $k = k_1$ і $k = k_1 + 1$.

Отже, в загальному випадку $np - q \leq k_0 \leq np + p$. \square

Приклад. Кубик кидають 5 разів. Яка найімовірніша кількість випадань шістки?

Розв'язання. Тут $n = 5$, $p = 1/6$, $q = 5/6$, $np = 5/6$, $np - q = 0$, $np + p = 1$.
Отже, $0 \leq k_0 \leq 1$, і найімовірнішими є два значення кількості випадань шістки:
 $k_0 = 0$ і $k_0 = 1$.

Лекція 8

Граничні теореми для схеми Бернуллі

Якщо проводяться випробування за схемою Бернуллі і числа n і k – великі, то обчислення ймовірностей $P_n(k)$ за біномною формулою викликає певні труднощі. У такому разі для обчислення цих ймовірностей застосовують асимптотичні (наближені) формули, які випливають із локальної та інтегральної теорем Муавра-Лапласа і граничної теореми Пуассона. Назва „гранична” в обох випадках пов’язана з тим, що згадані теореми встановлюють поведінку ймовірностей $P_n(k)$ або $P_n(k_1 \leq k \leq k_2)$ за певних умов, до яких обов’язково входить умова $n \rightarrow \infty$.

Гранична теорема Пуассона. *Якщо ймовірність успіху в кожному з n незалежних випробувань за схемою Бернуллі рівна p і якщо для $n \rightarrow \infty$ $np \rightarrow \lambda$ ($0 < \lambda < \infty$), то*

$$\lim_{n \rightarrow \infty} P_n(k) = \frac{\lambda^k}{k!} e^{-\lambda}$$

для будь-якого $k = 0, 1, 2, \dots$, де $P_n(k)$ – ймовірність появи k успіхів в n випробуваннях.

Доведення. Поклавши $np = \lambda_n$ (отже, $\lambda_n \rightarrow \lambda$), запишемо

$$\begin{aligned} P_n(k) &= C_n^k p^k q^{n-k} = \frac{n!}{(n-k)!k!} p^k q^{n-k} = \frac{n(n-1)\cdots(n-k+1)}{k!} \cdot \left(\frac{\lambda_n}{n}\right)^k \left(1 - \frac{\lambda_n}{n}\right)^{n-k} = \\ &= \frac{\lambda_n^k}{k!} \left(1 - \frac{\lambda_n}{n}\right)^n \left(1 - \frac{1}{n}\right) \left(1 - \frac{2}{n}\right) \cdots \left(1 - \frac{k-1}{n}\right) \left(1 - \frac{\lambda_n}{n}\right)^{-k} \xrightarrow{n \rightarrow \infty} \frac{\lambda^k}{k!} \cdot \lim_{n \rightarrow \infty} \left(1 - \frac{\lambda_n}{n}\right)^n = \\ &= \frac{\lambda^k}{k!} \cdot \lim_{n \rightarrow \infty} \left[\left(1 + \left(-\frac{\lambda}{n}\right)\right)^{\frac{n}{\lambda}} \right]^{-\lambda} = \frac{\lambda^k}{k!} e^{-\lambda}, \end{aligned}$$

що й треба було довести. \square

Отже, при великих n ($n > 100$) і малих p ($np < 30$) ми можемо користуватися наближеними формулами:

$$P_n(k) \approx \frac{\lambda^k}{k!} e^{-\lambda}; \quad (1)$$

$$P_n(k_1 \leq k \leq k_2) = \sum_{k=k_1}^{k_2} P_n(k) \approx \sum_{k=k_1}^{k_2} \frac{\lambda^k}{k!} e^{-\lambda}, \quad \lambda = np. \quad (2)$$

Формули (1) і (2) називаються *асимптотичними формулами Пуассона*. Друга з них дає наближений вираз для ймовірності того, що кількість успіхів в n випробуваннях міститься між заданими числами k_1 і k_2 . Дослідження питання про точність формул (1) і (2) ми не розглядаємо. Обмежимося лише тим, що прийемо без доведення нерівність

$$\left| \sum_{k \in M} P_n(k) - \sum_{k \in M} \frac{\lambda^k}{k!} e^{-\lambda} \right| < np^2,$$

яка є правильною для будь-якої множини $M \subset \{0, 1, 2, \dots\}$. Зокрема, якщо M складається з одного числа k , то

$$\left| P_n(k) - \frac{\lambda^k}{k!} e^{-\lambda} \right| < np^2. \quad (3)$$

Для виразу $\frac{\lambda^k}{k!} e^{-\lambda}$, який розглядається як функція двох змінних k і λ , складено таблицю значень.

Введемо позначення $P(k) = \frac{\lambda^k}{k!} e^{-\lambda}$. Сукупність значень $\{P(k) : k = 0, 1, 2, \dots\}$

називається *розподілом Пуассона з параметром $\lambda > 0$* .

Приклад. У фірмі працює 500 співробітників. Знайти ймовірність того, що у двох співробітників день народження припаде на новий рік, вважаючи, що ймовірність народитися у фіксований день становить $1/365$.

Розв'язання. Маємо: $n=500$, $p=1/365$, $\lambda = np = \frac{500}{365} = \frac{100}{73}$, $k=2$;

$$P_{500}(2) \approx \frac{\left(\frac{100}{73}\right)^2}{2} e^{-\frac{100}{73}} = 0,2384517.$$

Обчислення ж за біномною формулою дають $P_{500}(2) = 0,2388347$. Бачимо, що

$$|P_{500}(2) - 0,2384517| = 0,000383 < 0,003753 = np^2.$$

Отже, оцінка (3) в даному випадку правильна. \square

Локальна формула Муавра-Лапласа. Нехай імовірність успіху в кожному з n незалежних випробувань за схемою Бернуллі рівна p , $0 < p < 1$. Тоді для великих значень n імовірність $P_n(k)$ появи k успіхів в n випробуваннях обчислюється за наближеною формулою:

$$P_n(k) \approx \frac{\varphi(x_0)}{\sqrt{npq}}, \quad \text{якщо} \quad x_0 = \frac{k - np}{\sqrt{npq}}, \quad (4)$$

де $\varphi(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$ – функція Гаусса.

При цьому встановлено, що відносна похибка формули (4) наближається до нуля, коли $n \rightarrow \infty$.

Функція Гаусса табульована. В більшості підручників наведено значення $\varphi(x)$ для $0 \leq x \leq 3,99$. Для обчислення значень $\varphi(x)$ при від'ємних значеннях $-3,99 \leq x \leq 0$ використовуємо парність функції $\varphi(x)$, а для $|x| > 3,99$ приймаємо, що $\varphi(x) = 0$.

Інтегральна теорема Муавра-Лапласа. Нехай ймовірність успіху в кожному з n незалежних випробувань за схемою Бернуллі рівна p , $0 < p < 1$, а μ_n – кількість успіхів в n випробуваннях. Тоді

$$\lim_{n \rightarrow \infty} P \left\{ a \leq \frac{\mu_n - np}{\sqrt{npq}} \leq b \right\} = \frac{1}{\sqrt{2\pi}} \int_a^b e^{-\frac{x^2}{2}} dx$$

для всіх a, b ($-\infty \leq a \leq b \leq \infty$).

Спираючись на цю теорему, для великих значень n записують наближену формулу для ймовірності $P\{k_1 \leq \mu_n \leq k_2\}$. Щоб отримати її, введемо позначення

$$k_1 = a\sqrt{npq} + np, \quad k_2 = b\sqrt{npq} + np.$$

Тоді

$$\begin{aligned} P \left\{ a \leq \frac{\mu_n - np}{\sqrt{npq}} \leq b \right\} &= P\{a\sqrt{npq} + np \leq \mu_n \leq b\sqrt{npq} + np\} = P\{k_1 \leq \mu_n \leq k_2\} \approx \\ &\approx \frac{1}{\sqrt{2\pi}} \int_a^b e^{-\frac{x^2}{2}} dx = \frac{1}{\sqrt{2\pi}} \int_{x_1}^{x_2} e^{-\frac{x^2}{2}} dx, \end{aligned}$$

де

$$x_1 = \frac{k_1 - np}{\sqrt{npq}}, \quad x_2 = \frac{k_2 - np}{\sqrt{npq}}.$$

Для простішого запису отриманої наближеної формули вводять **функцію Лапласа**

$$\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_0^x e^{-\frac{t^2}{2}} dt.$$

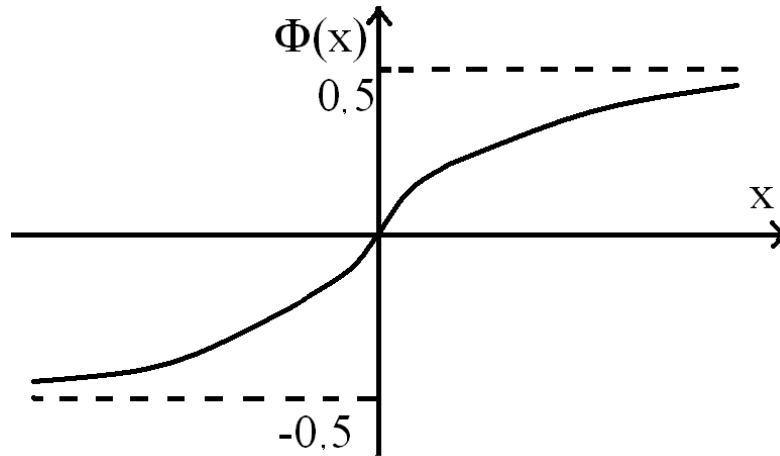
Оскільки

$$\int_{x_1}^{x_2} e^{-\frac{x^2}{2}} dx = \int_0^{x_2} e^{-\frac{x^2}{2}} dx - \int_0^{x_1} e^{-\frac{x^2}{2}} dx,$$

то остаточно отримаємо

$$P\{k_1 \leq \mu_n \leq k_2\} \approx \Phi(x_2) - \Phi(x_1) = \Phi\left(\frac{k_2 - np}{\sqrt{npq}}\right) - \Phi\left(\frac{k_1 - np}{\sqrt{npq}}\right). \quad (5)$$

Функцію Лапласа часто використовують у теорії ймовірностей і математичній статистиці, тому опишемо її найпростіші властивості.



Графік функції Лапласа

- $\Phi(x)$ – непарна функція: $\Phi(-x) = -\Phi(x)$.

Доведення. В інтегралі $\Phi(-x) = \frac{1}{\sqrt{2\pi}} \int_0^{-x} e^{-\frac{t^2}{2}} dt$ зробимо заміну $t = -u$. Тоді

$$dt = -du, \quad \Phi(-x) = -\frac{1}{\sqrt{2\pi}} \int_0^x e^{-\frac{u^2}{2}} du = -\Phi(x).$$

- $\Phi(0) = 0$; $\Phi(+\infty) = 0,5$ і $\Phi(-\infty) = -0,5 \Rightarrow$ прямі $y = 0,5$ і $y = -0,5$ - асимптоти графіка функції при $x \rightarrow +\infty$ і $x \rightarrow -\infty$ відповідно.

Доведення. В інтегралі $\Phi(+\infty) = \frac{1}{\sqrt{2\pi}} \int_0^{\infty} e^{-\frac{t^2}{2}} dt$ зробимо заміну $t = u\sqrt{2}$. Тоді

$$dt = \sqrt{2} du, \quad \Phi(+\infty) = \frac{\sqrt{2}}{\sqrt{2\pi}} \int_0^{\infty} e^{-u^2} du = \frac{\sqrt{2}}{\sqrt{2\pi}} \cdot \frac{\sqrt{\pi}}{2} = \frac{1}{2}.$$

- $\Phi(x)$ – зростаюча функція.

Функція Лапласа $\Phi(x)$ табульована для $0 \leq x \leq 5$. Для обчислення значень $\Phi(x)$ при від'ємних значеннях $-5 \leq x \leq 0$ використовуємо непарність функції.

Приклад. Знайти ймовірність того, що при 600 киданнях кубика „шістка” випаде 100 разів. Яка ймовірність того, що кількість випадань „шістки” є в межах від 90 до 110?

Розв'язання. Маємо: $n=600$; $k=100$; $p=1/6$; $np = 100$; $npq = 100 \cdot \frac{5}{6} = 83,3$;

$$\sqrt{npq} = 9,1287; \quad x_0 = \frac{k - np}{\sqrt{npq}} = 0; \quad P_{600}(100) \approx \frac{\varphi(0)}{\sqrt{npq}} = \frac{0,3989}{9,1287} = 0,0437;$$

$$\begin{aligned} P\{90 \leq \mu_{600} \leq 110\} &\approx \Phi\left(\frac{110 - np}{\sqrt{npq}}\right) - \Phi\left(\frac{90 - np}{\sqrt{npq}}\right) = \Phi(1,095) - \Phi(-1,095) = \\ &= 2\Phi(1,095) = 2 \cdot 0,36324 = 0,72648. \end{aligned}$$

Зауваження. Точність наближених формул (4) і (5) істотно залежить від взаємовідношення величин n і p . Зокрема, добрі наближення ці формули дають

при $p=q=1/2$, їх часто використовують, коли $npq \geq 10$. Звідси, до речі, видно: що ближче одне з чисел p або q до нуля, то більшим слід вибирати n . Тому в разі близькості однієї з величин p або q до нуля формулами (4) і (5) зазвичай не користуються; для цього випадку значно точнішими є наближені формули Пуассона.

Теорема Бернуллі про стійкість відносних частот

Нехай імовірність появи випадкової події A в кожному з n незалежних випробувань за схемою Бернуллі становить p . Для довільного числа $\varepsilon > 0$ визначимо за допомогою інтегральної теореми Муавра-Лапласа ймовірність події $\left| \frac{\mu_n}{n} - p \right| \leq \varepsilon$ при $n \rightarrow \infty$, де $\frac{\mu_n}{n}$ – відносна частота події, тобто ймовірність того, що відхилення відносної частоти події від її ймовірності за абсолютним значенням не перевищує числа $\varepsilon > 0$.

Можемо записати

$$P \left\{ \left| \frac{\mu_n}{n} - p \right| \leq \varepsilon \right\} = P \left\{ -\varepsilon \leq \frac{\mu_n - np}{n} \leq \varepsilon \right\} = P \left\{ -\varepsilon \sqrt{\frac{n}{pq}} \leq \frac{\mu_n - np}{\sqrt{npq}} \leq \varepsilon \sqrt{\frac{n}{pq}} \right\} \xrightarrow{n \rightarrow \infty} \\ \rightarrow \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-\frac{x^2}{2}} dx = \Phi(+\infty) - \Phi(-\infty) = 2\Phi(+\infty) = 1.$$

Отже,

$$\forall \varepsilon > 0 \quad \lim_{n \rightarrow \infty} P \left\{ \left| \frac{\mu_n}{n} - p \right| \leq \varepsilon \right\} = 1.$$

Ця формула виражає теорему Бернуллі.

Теорема Бернуллі. Якщо в кожному з серії незалежних випробувань випадкова подія настає з однією і тою самою ймовірністю p , то за достатньо великої кількості випробувань з **імовірністю як завгодно близькою до 1**, відхилення відносної частоти появи цієї події від її ймовірності p не перевищуватиме як завгодно малого наперед заданого числа ε .

Оскільки для великих n

$$P \left\{ -\varepsilon \sqrt{\frac{n}{pq}} \leq \frac{\mu_n - np}{\sqrt{npq}} \leq \varepsilon \sqrt{\frac{n}{pq}} \right\} \approx \frac{1}{\sqrt{2\pi}} \int_{-\varepsilon \sqrt{\frac{n}{pq}}}^{\varepsilon \sqrt{\frac{n}{pq}}} e^{-\frac{x^2}{2}} dx = 2\Phi \left(\varepsilon \sqrt{\frac{n}{pq}} \right),$$

то для великих n використовують наближену формулу

$$P \left\{ \left| \frac{\mu_n}{n} - p \right| \leq \varepsilon \right\} \approx 2\Phi \left(\varepsilon \sqrt{\frac{n}{pq}} \right). \quad (6)$$

Приклад. Проводиться 100 випробувань за схемою Бернуллі з імовірністю $p=0,5$ появи події A в кожному окремому випробуванні. Знайти межі, в яких міститься частота події A з імовірністю 0,9545.

Розв'язання. За умовою задачі $n=100$, $p=q=0,5$ і

$$P\left\{\left|\frac{\mu_n}{n} - 0,5\right| \leq \varepsilon\right\} = 0,9545.$$

За формулою (6) маємо, що $0,9545 = 2\Phi\left(\varepsilon\sqrt{\frac{100}{0,25}}\right) \Leftrightarrow \Phi(20\varepsilon) = 0,47725$. За таблицею значень функції Лапласа знаходимо, що $20\varepsilon = 2 \Rightarrow \varepsilon = 0,1$. Із нерівності $\left|\frac{\mu_n}{n} - 0,5\right| \leq 0,1 \Leftrightarrow -0,1 \leq \frac{\mu_n}{n} - 0,5 \leq 0,1 \Leftrightarrow 0,4 \leq \frac{\mu_n}{n} \leq 0,6$ випливає, що $40 \leq \mu_n \leq 60$, тобто з імовірністю 0,9545 подія A може з'явитися від 40 до 60 разів.

Лекція 9

Поняття випадкової величини та функції розподілу

Оскільки результат експерименту може змінюватися від випадку до випадку, то кількісна ознака, яка в ньому розглядається, взагалі кажучи, є змінною величиною, до того ж випадковою. Отже, **випадкова величина** – це величина, яка в результаті експерименту з випадковим результатом набуває того чи іншого числового значення.

Прикладами випадкових величин, що набувають різних числових значень під впливом багатьох випадкових факторів, можуть бути:

- а) кількість очок, яка випадає на верхній грані за одне кидання грального кубика;
- б) кількість бракованих виробів серед n навмання вибраних;
- в) кількість кидань монети до першої появи герба;
- г) кількість викликів, які надходять на телефонну станцію протягом деякого проміжку часу;
- д) тривалість часу обслуговування покупця;
- е) час виконання деякого завдання і т. д.

Випадкові величини позначатимемо великими літерами X, Y, Z, \dots , а їхні можливі значення – малими літерами x, y, z, \dots латинського алфавіту.

У наведених прикладах траплялися два типи випадкових величин: **дискретні** величини, множини можливих значень яких скінченні або зліченні, - приклади а) – г) і **неперервні** величини, множини можливих значень яких суцільно заповнюють деякий інтервал, – приклади д), е).

Зазначимо, що за теоретико-множинним трактуванням основних понять теорії ймовірностей випадкова величина X є функція елементарної події: $X = X(\omega)$, де ω – елементарна подія, яка належить простору Ω ($\omega \in \Omega$). При цьому множина можливих значень випадкової величини X складається з усіх значень, яких набуває функція $X(\omega)$. Якщо ця множина скінченна або зліченна, то випадкова величина X називається *дискретною*, якщо незліченна – *неперервною*.

Наведемо приклади дискретної і неперервної випадкових величин.

1. Симетричну монету кидають двічі. Нехай випадкова величина X – кількість появ герба. Простір елементарних подій складається з чотирьох елементів:

$$\Omega = \{\omega_1 = (ЦЦ), \omega_2 = (ЦГ), \omega_3 = (ГЦ), \omega_4 = (ГГ)\}.$$

Таблиця значень випадкової величини має такий вигляд:

ω_i	ω_1	ω_2	ω_3	ω_4
$X(\omega_i)$	0	1	1	2

2. Нехай випадкова величина Y є час очікування трамвая на зупинці. Якщо відомо, що проміжок часу між прибуттям трамваїв не перевищує T , то

значення Y належать відрізку $[0, T]$.

Для того, щоб описати випадкову величину, необхідно вказати не тільки множину її можливих значень, а й охарактеризувати ймовірності всіх можливих подій, пов'язаних із випадковою величиною (наприклад, ймовірність того, що вона набуде того чи іншого значення або потрапить у деякий інтервал). Такий повний опис випадкової величини називається її *законом розподілу*.

У випадку довільного ймовірнісного простору (Ω, \mathcal{F}, P) не будь-які функції, визначені на Ω , можна розглядати як випадкові величини. Вивчаючи закони розподілу випадкових величин, часто доводиться відповідати на питання: яка ймовірність того, що значення випадкової величини $X(\omega)$ належать до тієї чи іншої множини. Отже, для достатньо широкого класу множин $\{B\}$ на числовій прямій повинна бути впевненість, що множина $\{\omega: X(\omega) \in B\}$ належить σ -алгебрі подій \mathcal{F} , і тому можна розглядати ймовірність $P\{\omega: X(\omega) \in B\}$. Виявляється, достатньо припустити, що для кожного інтервалу $(-\infty, x)$ множина $\{\omega: X(\omega) \in (-\infty, x)\} = \{\omega: X(\omega) < x\}$ належить σ -алгебрі подій \mathcal{F} , і тоді для кожної множини дійсних чисел B , яка зображається як об'єднання або перетин скінченної або зліченної кількості проміжків, отримаємо, що $\{\omega: X(\omega) \in B\} \in \mathcal{F}$.

Означення. Нехай (Ω, \mathcal{F}, P) – ймовірнісний простір. *Випадковою величиною* назвемо дійсну функцію $X = X(\omega)$, визначену на Ω і таку, що для кожного дійсного числа x виконується співвідношення:

$$\{\omega: X(\omega) < x\} \in \mathcal{F}.$$

Означення. Функція дійсної змінної x , $x \in \mathbb{R} = (-\infty, +\infty)$, визначена рівністю

$$F(x) = P\{\omega: X(\omega) < x\} = P\{X < x\},$$

називається *функцією розподілу* випадкової величини $X = X(\omega)$.

Функція розподілу є найбільш загальною формою закону розподілу, придатною для характеристики всіх випадкових величин (як дискретних, так і неперервних). Знаючи функцію розподілу $F(x)$ випадкової величини X , можна обчислити ймовірності будь-яких подій, які з нею пов'язані.

Властивості функції розподілу

- *Ймовірність того, що випадкова величина X набуде значення з проміжку $[a, b)$ дорівнює приросту її функції розподілу на цьому проміжку, тобто*

$$P\{a \leq X < b\} = F(b) - F(a).$$

Доведення. Оскільки $\{X < b\} = \{X < a\} + \{a \leq X < b\}$, то за аксіомою адитивності (події-доданки – несумісні):

$$F(b) = P\{X < b\} = P\{X < a\} + P\{a \leq X < b\} = F(a) + P\{a \leq X < b\},$$

звідки $P\{a \leq x < b\} = F(b) - F(a)$.

- Значення функції розподілу належать відрізку $[0, 1]$, тобто $0 \leq F(x) \leq 1$.

- Якщо $x_1 \leq x_2$, то $F(x_1) \leq F(x_2)$, тобто $F(x)$ – неспадна функція.

Доведення. Якщо $x_1 \leq x_2$, то $A = \{X < x_1\} \subset B = \{X < x_2\}$, тому $P(A) \leq P(B)$, тобто $F(x_1) \leq F(x_2)$.

- $\lim_{x \rightarrow \infty} F(x) = 1$ і $\lim_{x \rightarrow -\infty} F(x) = 0$.

- $\lim_{x \rightarrow x_0-0} F(x) = F(x_0)$, тобто функція $F(x)$ – неперервна зліва.

Доведення. Розглянемо довільну зростаючу числову послідовність

$$x_1 < x_2 < \dots < x_n < \dots, \quad \lim_{n \rightarrow \infty} x_n = x_0$$

і позначимо $A_n = \{x_n \leq X < x_0\}$. Тоді послідовність подій

$$A_1 \supset A_2 \supset \dots \supset A_n \supset \dots -$$

спадна, причому $\bigcap_{n=1}^{\infty} A_n = \emptyset$. За аксіомою неперервності $\lim_{n \rightarrow \infty} P(A_n) = P(\emptyset) = 0$. З

другого боку,

$$\lim_{n \rightarrow \infty} P(A_n) = \lim_{n \rightarrow \infty} P\{x_n \leq X < x_0\} = \lim_{n \rightarrow \infty} (F(x_0) - F(x_n)) = F(x_0) - F(x_0 - 0) = 0.$$

- $P\{X = x\} = F(x + 0) - F(x)$.

Доведення. $A = \{X = x\} = \bigcap_{n=1}^{\infty} A_n$, де $A_n = \left\{x \leq X < x + \frac{1}{n}\right\}$. Тоді

послідовність подій $A_1 \supset A_2 \supset \dots \supset A_n \supset \dots -$ спадна. За аксіомою неперервності

$$P(A) = \lim_{n \rightarrow \infty} P(A_n) \Rightarrow P\{X = x\} = \lim_{n \rightarrow \infty} \left(F\left(x + \frac{1}{n}\right) - F(x) \right) = F(x + 0) - F(x).$$

Дискретні випадкові величини. Закон розподілу

Означення. Випадкова величина називається **дискретною**, якщо множина її можливих значень є скінченною або зліченною.

Нехай X – дискретна випадкова величина, можливими і єдиною можливими значеннями якої є числа x_1, x_2, \dots, x_n . Через $p_k = P\{X = x_k\}$ позначимо ймовірність того, що випадкова величина X набуває значення x_k , $p_k \geq 0$.

Події $\{X = x_k\}$, $k = \overline{1, n}$ утворюють повну групу попарно несумісних подій, тому $\sum_{k=1}^n p_k = p_1 + p_2 + \dots + p_n = 1$.

Означення. **Законом розподілу ймовірностей** (законом розподілу) дискретної випадкової величини називається відповідність між усіма її можливими значеннями та їхніми ймовірностями.

Табличний запис закону розподілу – це таблиця значень x_k випадкової величини та відповідних їхніх ймовірностей p_k :

x_k	x_1	x_2	\dots	x_n
p_k	p_1	p_2	\dots	p_n

За допомогою табличного запису закону розподілу можна визначити функцію розподілу $F(x)$ випадкової величини X за формулою:

$$F(x) = P\{X < x\} = \sum_{k: x_k < x} p_k,$$

у якій сумування проводиться за усіма індексами k , для яких $x_k < x$.

У випадку, коли множина різних значень x_k випадкової величини X є нескінченною і зліченною, то її закон розподілу також можна записати у формі таблиці, яка складатиметься з двох нескінченних рядків:

$$x_k : x_1, x_2, \dots, x_n, \dots$$

$$p_k : p_1 = P\{X = x_1\}, p_2 = P\{X = x_2\}, \dots, p_n = P\{X = x_n\}, \dots,$$

до того ж $\sum_{k=1}^{\infty} p_k = 1$.

Приклад. У грошовій лотереї розігрується 2 виграші по 1000 грн, 10 виграшів по 100 грн і 100 виграшів по 10 грн за загальної кількості білетів 10000. Визначити закон розподілу випадкової величини X – виграшу власника одного лотерейного білета.

Розв'язання. Можливими значеннями дискретної випадкової величини X є числа $x_1=1000$, $x_2=100$, $x_3=10$, $x_4=0$. Відповідні їхні ймовірності обчислюємо за формулою: $p_k = \frac{n_k}{n}$, де n_k – кількість виграшних білетів на відповідну суму гривень, n – кількість всіх білетів лотереї. Одержимо:

$$p_1 = \frac{2}{10000} = 0,0002; \quad p_2 = \frac{10}{10000} = 0,001; \quad p_3 = \frac{100}{10000} = 0,01;$$

$$p_4 = 1 - (p_1 + p_2 + p_3) = 1 - (0,0002 + 0,001 + 0,01) = 0,9888.$$

Закон розподілу випадкової величини X запишемо у вигляді таблиці:

$X = x_k$	1000	100	10	0
$p = p_k$	0,0002	0,001	0,01	0,9888

Приклад. Дискретна випадкова величина X має закон розподілу ймовірностей, що заданий таблицею:

$X = x_k$	-2	1	4	6
$p = p_k$	0,2	0,1	0,3	0,4

Знайти функцію розподілу випадкової величини X .

Розв'язання. Якщо $x \leq -2$, то $F(x) = P\{X < x\} = 0$, бо подія $\{X < x\}$ неможлива.

Якщо $-2 < x \leq 1$, то $F(x) = P\{X < x\} = 0,2$, бо подія $\{X < x\}$ рівносильна події $\{X = -2\}$, яка має ймовірність 0,2.

Якщо $1 < x \leq 4$, то $F(x) = P\{X < x\} = 0,2 + 0,1 = 0,3$, бо подія $\{X < x\}$ є сумою двох несумісних подій: $\{X = -2\}$, яка має ймовірність 0,2, і $\{X = 1\}$, яка має ймовірність 0,1.

Якщо $4 < x \leq 6$, то $F(x) = P\{X < x\} = 0,2 + 0,1 + 0,3 = 0,6$, бо подія $\{X < x\}$ є сумою трьох несумісних подій: $\{X = -2\}$, яка має ймовірність 0,2, $\{X = 1\}$, яка має ймовірність 0,1 і $\{X = 4\}$, яка має ймовірність 0,3.

Якщо $x > 6$, то $F(x) = P\{X < x\} = 1$, бо подія $\{X < x\}$ є вірогідною.

Отже, функція розподілу заданої дискретної випадкової величини має такий аналітичний вигляд:

$$F(x) = \begin{cases} 0, & x \leq -2; \\ 0,2, & -2 < x \leq 1; \\ 0,3, & 1 < x \leq 4; \\ 0,6, & 4 < x \leq 6; \\ 1, & x > 6. \end{cases}$$

Графік функції розподілу дискретної випадкової величини має „східчастий” характер.

Основні закони розподілу дискретних випадкових величин

Біномний закон розподілу. Нехай проводиться n незалежних випробувань за схемою Бернуллі і $p = P(A)$ – ймовірність появи події A в кожному окремому випробуванні. Сформулюємо задачу: написати закон розподілу дискретної випадкової величини X – кількості появ події A в цих n випробуваннях.

Випадкова величина X може набути значень

$$x_0 = 0, x_1 = 1, x_2 = 2, \dots, x_n = n.$$

Ймовірності можливих значень x_k випадкової величини X обчислимо за біномною формулою:

$$p_k = P_n(k) = C_n^k p^k q^{n-k}, \quad q = 1 - p$$

і одержимо закон розподілу описаної випадкової величини X , який називається **біномним**

$X = x_k$	0	1	2	...	n
$p = p_k$	q^n	$C_n^1 p q^{n-1}$	$C_n^2 p^2 q^{n-2}$...	p^n

Раніше ми вже переконалися, що

$$\sum_{k=0}^n p_k = \sum_{k=0}^n P_n(k) = \sum_{k=0}^n C_n^k p^k q^{n-k} = (p + q)^n = 1^n = 1.$$

Приклад. Прилад складається з чотирьох елементів і ймовірність наявності технічних неполадок у кожному з них становить 0,5. Написати закон розподілу випадкової величини X – кількості елементів приладу, в яких наявні

технічні неполадки.

Розв'язання. Можливими значеннями дискретної випадкової величини X є числа $x_0=0$, $x_1=1$, $x_2=2$, $x_3=3$, $x_4=4$. За біномною формулою обчислимо відповідні ймовірності цих значень, знаючи, що $p = q = \frac{1}{2}$:

$$p_0 = C_4^0 p^0 q^4 = 1 \cdot 1 \cdot \left(\frac{1}{2}\right)^4 = \frac{1}{16}; \quad p_1 = C_4^1 p q^3 = 4 \cdot \frac{1}{2} \cdot \left(\frac{1}{2}\right)^3 = \frac{1}{4};$$

$$p_2 = C_4^2 p^2 q^2 = 6 \cdot \left(\frac{1}{2}\right)^2 \cdot \left(\frac{1}{2}\right)^2 = \frac{3}{8}; \quad p_3 = C_4^3 \cdot \left(\frac{1}{2}\right)^3 \cdot \frac{1}{2} = \frac{1}{4}; \quad p_4 = C_4^4 \cdot \left(\frac{1}{2}\right)^4 = \frac{1}{16}.$$

Зробимо перевірку: $p_0 + p_1 + p_2 + p_3 + p_4 = \frac{1}{16} + \frac{1}{4} + \frac{3}{8} + \frac{1}{4} + \frac{1}{16} = 1$. Закон розподілу даної випадкової величини має форму таблиці:

$X = x_k$	0	1	2	3	4
$p = p_k$	1/16	1/4	3/8	1/4	1/16

З таблиці видно, що найімовірніша кількість елементів приладу, в яких є технічні неполадки, $k_0 = 2$.

Розподіл Пуассона. Розподіл імовірностей дискретної випадкової величини X , яка набуває значень

$$x_k : 0, 1, 2, \dots, n, \dots$$

з імовірностями

$$p_k = P\{X = x_k\} = \frac{\lambda^k}{k!} e^{-\lambda}, \quad k = 0, 1, 2, \dots, \quad (1)$$

називається **законом розподілу Пуассона**, що залежить від параметра λ , $\lambda > 0$.

Розподіл Пуассона записують у формі таблиці:

$X = x_k$	0	1	2	...	n	...
$p = p_k$	$e^{-\lambda}$	$\frac{\lambda}{1!} e^{-\lambda}$	$\frac{\lambda^2}{2!} e^{-\lambda}$...	$\frac{\lambda^n}{n!} e^{-\lambda}$...

Сумуючи всі ймовірності розподілу Пуассона і використовуючи рівність

$$e^\lambda = \sum_{k=0}^{\infty} \frac{\lambda^k}{k!},$$

отримуємо підтвердження основної властивості розподілу:

$$\sum_{k=0}^{\infty} p_k = e^{-\lambda} \sum_{k=0}^{\infty} \frac{\lambda^k}{k!} = e^{-\lambda} \cdot e^\lambda = 1.$$

Під час вивчення схеми Бернуллі було зауважено, що при великих n для обчислення ймовірностей $P_n(k)$ доцільно використовувати асимптотичні формули Пуассона, які полегшують ці обчислення. Зокрема, з асимптотичної формули Пуассона випливає, що за допомогою розподілу Пуассона можна

апроксимувати біномний закон розподілу, коли кількість експериментів n необмежено зростає ($n \rightarrow \infty$) й одночасно ймовірність успіху в одному експерименті необмежено зменшується ($p \rightarrow 0$) так, що їх добуток np наближається до числа λ : $\lim_{n \rightarrow \infty} np = \lambda$.

Приклад. Електронна пошта банку підтримує зв'язки із сотнею абонентів. Імовірність того, що за одиницю часу на електронну пошту надійде повідомлення від абонента, становить 0,02. Написати закон розподілу випадкової величини X – кількості повідомлень від абонентів за одиницю часу.

Розв'язання. У даному випадку проводиться $n=100$ випробувань за схемою Бернуллі, і випадкова величина X може набувати значень $x_0=0$, $x_1=1$, $x_2=2$, $x_3=3$, $x_4=4$, ..., $x_{100}=100$. Імовірність події A – надходження повідомлення від одного абонента є мала, а число $n=100$ – велике і $\lambda = 100 \cdot 0,02 = 2$, тому відповідні ймовірності обчислюємо за формулою (1):

$$p_0 = P_{100}(0) \approx e^{-2} \approx 0,1353; \quad p_1 = P_{100}(1) \approx 2e^{-2} \approx 0,2707; \quad p_2 = P_{100}(2) \approx 2e^{-2} \approx 0,2707;$$

$$p_3 = P_{100}(3) \approx \frac{4}{3}e^{-2} \approx 0,1804; \quad p_4 = P_{100}(4) \approx \frac{2}{3}e^{-2} \approx 0,0902;$$

.....

$$p_{100} = P_{100}(100) = 1 - \sum_{k=0}^{99} P_{100}(k).$$

Закон розподілу описаної в задачі випадкової величини X записуємо у формі таблиці:

$X=x_k$	0	1	2	3	4	...
$p=p_k$	0,1353	0,2707	0,2707	0,1804	0,0902	...

З таблиці видно, що найімовірніша кількість повідомлень від абонентів за одиницю часу – одне або два.

Неперервні випадкові величини. Щільність розподілу

Вище *неперервною* випадковою величиною ми назвали випадкову величину X , можливі значення якої суцільно заповнюють деякий скінченний або нескінченний проміжок на числовій прямій. Це означення ми далі уточнимо.

Оскільки множина можливих значень неперервної випадкової величини незліченна, то для неї непридатна характеристика розподілу у формі переліку її значень та відповідних ймовірностей хоча б з тієї причини, що значення цієї множини неможливо записати як послідовність. Тому характеристика розподілу такої величини базується на понятті функції розподілу.

Означення. Випадкову величину X називають *неперервною* (або

абсолютно неперервною), якщо існує невід'ємна функція $p(x)$ така, що для всіх x функція розподілу випадкової величини X визначається у вигляді

$$F(x) = P\{X < x\} = \int_{-\infty}^x p(u) du, \quad (2)$$

до того ж функція $p(x)$ неперервна всюди, крім, можливо, скінченної кількості точок.

З означення (2) випливає, що **функція розподілу неперервної випадкової величини неперервна** (інтеграл – неперервна функція верхньої межі інтегрування). Функція $p(x)$ називається **щільністю розподілу ймовірностей** випадкової величини X . В точках своєї неперервності функцію $p(x)$ можна визначити як похідну функції розподілу: $p(x) = F'(x)$.

Спираючись на співвідношення (2) та властивості функції розподілу, можемо також записати:

$$P\{a \leq X < b\} = F(b) - F(a) = \int_{-\infty}^b p(u) du - \int_{-\infty}^a p(u) du = \int_a^b p(x) dx.$$

З неперервності функції розподілу неперервної випадкової величини виводимо, що $P\{X = x\} = F(x+0) - F(x) = 0$. Тому для неперервної випадкової величини X

$$P\{a \leq X \leq b\} = P\{a < X \leq b\} = P\{a \leq X < b\} = P\{a < X < b\} = \int_a^b p(x) dx.$$

Використовуючи, що $F(+\infty) = 1$, з (2) отримуємо основну властивість щільності розподілу

$$\int_{-\infty}^{\infty} p(x) dx = 1.$$

Основні закони розподілу неперервних випадкових величин

А. Рівномірний розподіл. Нехай на проміжок $[a, b]$ навмання кидають точку, отже, ймовірність потрапляння точки на деяку частину проміжка пропорційна довжині цієї частини проміжка. Випадкову величину визначимо як координату тієї точки відрізка, в яку влучила кинута точка:

$$X(\omega) = \omega, \quad \omega \in [a, b].$$

Визначимо функцію розподілу випадкової величини X .

Якщо $x \leq a$, то $F(x) = P\{X < x\} = P(\emptyset) = 0$. Якщо $x \in (a, b]$, то $F(x) = P\{X < x\} = c(x - a)$ (імовірність пропорційна довжині проміжка $[a, x]$).
Тоді

$$F(b) = P\{X < b\} = P(\Omega) = 1 = c(b - a) \quad \Rightarrow \quad c = \frac{1}{b - a}.$$

Отже, рівномірний розподіл задають функцією

$$F(x) = \begin{cases} 0, & x \leq a; \\ \frac{x-a}{b-a}, & a \leq x \leq b; \\ 1, & x \geq b, \end{cases}$$

яка, очевидно, є неперервною. Відповідну щільність розподілу ймовірностей можна визначити як похідну функції розподілу:

$$p(x) = \begin{cases} \frac{1}{b-a}, & x \in [a, b]; \\ 0, & x \notin [a, b]. \end{cases}$$

Перевіримо основну властивість щільності розподілу

$$\int_{-\infty}^{\infty} p(x) dx = \int_a^b p(x) dx = \int_a^b \frac{dx}{b-a} = 1.$$

Б. Нормальний розподіл. Надзвичайно важливу роль у теорії ймовірностей відіграє нормальний розподіл (закон Гаусса). Він проявляється у всіх випадках, коли випадкова величина X є наслідком дії великої кількості різних випадкових факторів, кожний з яких окремо має на величину X незначний вплив.

Означення. Неперервна випадкова величина X називається **розподіленою за нормальним законом**, або **нормально розподіленою**, з параметрами $-\infty < a < +\infty$ і $\sigma > 0$, якщо її щільність розподілу має вигляд:

$$p(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-a)^2}{2\sigma^2}}, \quad -\infty < x < \infty.$$

Перевіримо для нормального розподілу основну властивість щільності розподілу. Для цього в інтегралі

$$\int_{-\infty}^{\infty} p(x) dx = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-\frac{(x-a)^2}{2\sigma^2}} dx$$

зробимо заміну $t = \frac{x-a}{\sigma\sqrt{2}}$. Тоді $x = \sigma\sqrt{2}t + a$; $dx = \sigma\sqrt{2} dt$ і враховуючи, що

$$\int_{-\infty}^{\infty} e^{-t^2} dt = \sqrt{\pi},$$

одержимо:

$$\int_{-\infty}^{\infty} p(x) dx = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-\frac{(x-a)^2}{2\sigma^2}} dx = \frac{\sigma\sqrt{2}}{\sigma\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-t^2} dt = \frac{1}{\sqrt{\pi}} \cdot \sqrt{\pi} = 1.$$

Обчислимо ймовірність попадання значень нормально розподіленої випадкової величини у заданий інтервал. Для цього в інтегралі

$$P\{\alpha < X < \beta\} = \int_{\alpha}^{\beta} p(x) dx = \frac{1}{\sigma\sqrt{2\pi}} \int_{\alpha}^{\beta} e^{-\frac{(x-a)^2}{2\sigma^2}} dx$$

зробимо заміну $t = \frac{x-a}{\sigma}$. Тоді $x = \sigma t + a$; $dx = \sigma dt$ і враховуючи, що

$$\frac{1}{\sqrt{2\pi}} \int_0^x e^{-\frac{t^2}{2}} dt = \Phi(x), \quad \frac{1}{\sqrt{2\pi}} \int_a^b e^{-\frac{t^2}{2}} dt = \frac{1}{\sqrt{2\pi}} \left(\int_0^b e^{-\frac{t^2}{2}} dt - \int_0^a e^{-\frac{t^2}{2}} dt \right) = \Phi(b) - \Phi(a),$$

одержимо:

$$P\{\alpha < X < \beta\} = \frac{1}{\sigma\sqrt{2\pi}} \int_{\alpha}^{\beta} e^{-\frac{(x-a)^2}{2\sigma^2}} dx = \frac{1}{\sqrt{2\pi}} \int_{\frac{\alpha-a}{\sigma}}^{\frac{\beta-a}{\sigma}} e^{-\frac{t^2}{2}} dt = \Phi\left(\frac{\beta-a}{\sigma}\right) - \Phi\left(\frac{\alpha-a}{\sigma}\right).$$

Отже, імовірність попадання значень нормально розподіленої випадкової величини в інтервал (α, β) дорівнює різниці значень функції Лапласа $\Phi(x)$,

якщо $x = \frac{\beta-a}{\sigma}$ і $x = \frac{\alpha-a}{\sigma}$:

$$P\{\alpha < X < \beta\} = \Phi\left(\frac{\beta-a}{\sigma}\right) - \Phi\left(\frac{\alpha-a}{\sigma}\right). \quad (3)$$

Оскільки функція $\Phi(x)$ табульована, то формула (3) є зручною для обчислень.

Якщо взяти $\alpha = a - \varepsilon$, $\beta = a + \varepsilon$, то з (3) отримаємо:

$$P\{a - \varepsilon < X < a + \varepsilon\} = P\{|X - a| < \varepsilon\} = \Phi\left(\frac{\varepsilon}{\sigma}\right) - \Phi\left(\frac{-\varepsilon}{\sigma}\right) = 2\Phi\left(\frac{\varepsilon}{\sigma}\right),$$

тобто

$$P\{|X - a| < \varepsilon\} = 2\Phi\left(\frac{\varepsilon}{\sigma}\right). \quad (4)$$

Приймемо у формулі (4) $\varepsilon = 3\sigma$. Одержимо:

$$P\{|X - a| < 3\sigma\} = 2\Phi(3) = 2 \cdot 0,49865 = 0,9973,$$

тобто ймовірність події $\{|X - a| < 3\sigma\}$ близька до одиниці, а це означає, що ця подія майже вірогідна. Звідси маємо так зване **правило „трьох сигм”**: якщо випадкова величина нормально розподілена, то практично вірогідно, що абсолютна величина її відхилення від параметра a не перевищує потроєного параметра σ .

В. Показниковий розподіл. У практичних застосуваннях теорії ймовірностей часто трапляються випадкові величини, які мають показниковий розподіл.

Означення. Неперервна випадкова величина X називається **розподіленою за показниковим законом** або **показниково розподіленою**, з параметром α , якщо щільність розподілу її ймовірностей має вигляд:

$$p(x) = \begin{cases} 0, & x < 0; \\ \alpha e^{-\alpha x}, & x \geq 0, \end{cases}$$

де $\alpha > 0$ – параметр розподілу.

Перевіримо для нормального розподілу основну властивість щільності розподілу:

$$\int_{-\infty}^{\infty} p(x) dx = \int_0^{\infty} \alpha e^{-\alpha x} dx = -e^{-\alpha x} \Big|_0^{\infty} = 0 - (-1) = 1.$$

Оскільки для $x > 0$

$$\int_{-\infty}^x p(u) du = \int_0^x \alpha e^{-\alpha u} du = -e^{-\alpha u} \Big|_0^x = -e^{-\alpha x} - (-1) = 1 - e^{-\alpha x},$$

то відповідна функція розподілу має вигляд:

$$F(x) = \begin{cases} 0, & x < 0; \\ 1 - e^{-\alpha x}, & x \geq 0. \end{cases}$$

Обчислимо ймовірність попадання значень показниково розподіленої випадкової величини у заданий інтервал:

$$P\{a < X < b\} = \int_a^b p(x) dx = \int_a^b \alpha e^{-\alpha x} dx = -e^{-\alpha x} \Big|_a^b = e^{-\alpha a} - e^{-\alpha b}. \quad (5)$$

Серед усіх законів розподілу неперервних випадкових величин лише показниковому притаманна властивість *відсутності післядії*, а саме: якщо випадкову величину пов'язати з часом, то для показникового закону розподілу минуле не впливає на передбачення подій у майбутньому. Наприклад, якщо випадкова величина T – тривалість безвідмовної роботи приладу має показниковий розподіл, то час роботи приладу впродовж інтервалу часу $(0, t_0)$ не впливає на величину ймовірності його безвідмовної роботи впродовж наступного інтервалу часу $(t_0, t_0 + t)$, а залежить лише від довжини t цього інтервалу.

Приклад. Час безвідмовної роботи деякого приладу – показниково розподілена випадкова величина T з параметром $\alpha = 0,02$. Знайти ймовірність того, що прилад працюватиме безвідмовно не менше ніж 100 год.

Розв'язання. За формулою (5) обчислюємо:

$$P\{T \geq 100\} = P\{100 \leq T < \infty\} = P\{100 < T < \infty\} = e^{-0,02 \cdot 100} - e^{-\infty} = e^{-2} \approx 0,1353.$$

Випадкові вектори та їхні функції розподілу

Дуже часто в імовірнісних моделях доводиться розглядати відразу кілька випадкових величин. Наприклад, кількість очок, яка випаде за одночасного кидання двох гральних кубиків, є можливим значенням системи двох випадкових величин X і Y , де X – кількість очок, яка випаде на першому кубіку, Y – кількість очок, яка випаде на другому кубіку. У математичній моделі в таких випадках на ймовірнісному просторі (Ω, \mathcal{F}, P) визначені кілька випадкових величин X_1, X_2, \dots, X_n , які іноді зручно розглядати як координати випадкового вектора $X = (X_1, X_2, \dots, X_n)$ із n -вимірнього простору \mathbb{R}^n . При цьому випадкові величини X_1, X_2, \dots, X_n можуть бути як дискретними, так і неперервними. Закон

розподілу випадкового вектора $X = (X_1, X_2, \dots, X_n)$ у загальному випадку визначається функцією розподілу.

Означення. *Функцією розподілу n -вимірної випадкової величини (X_1, X_2, \dots, X_n) називається ймовірність сумісного виконання нерівностей:*

$$X_1 < x_1, X_2 < x_2, \dots, X_n < x_n; \quad x_i \in \mathbb{R}, \quad i = \overline{1, n},$$

тобто

$$F(x_1, x_2, \dots, x_n) = P\{X_1 < x_1, X_2 < x_2, \dots, X_n < x_n\}.$$

Властивості функції розподілу розглянемо для випадку $n=2$, тобто для двовимірної випадкової величини (X, Y) . Функція розподілу $F(x, y)$ має такі властивості:

- $F(x, y)$ – неспадна функція за кожним аргументом;
- для функції $F(x, y)$ виконуються такі граничні співвідношення:
 $F(-\infty, y) = 0, \quad F(x, -\infty) = 0, \quad F(-\infty, -\infty) = 0, \quad F(+\infty, +\infty) = 1;$
- $\lim_{y \rightarrow \infty} F(x, y) = F(x, \infty) = F_X(x); \quad \lim_{x \rightarrow \infty} F(x, y) = F(\infty, y) = F_Y(y),$
де $F_X(x)$ – функція розподілу випадкової величини X ; $F_Y(y)$ – функція розподілу випадкової величини Y .

Двовимірні випадкові величини

Двовимірну випадкову величину (X, Y) назвемо *дискретною*, якщо її складові X і Y є дискретними одновимірними випадковими величинами, і *неперервною*, якщо її складові X і Y є неперервними одновимірними випадковими величинами. Складові X і Y двовимірної випадкової величини (X, Y) називають ще її *компонентами*.

А. Випадок дискретної величини. Для задання дискретної випадкової величини (X, Y) досить задати її можливі значення (x_i, y_k) та ймовірності кожного з них:

$$p_{ik} = P\{X = x_i, Y = y_k\}, \quad i, k = 1, 2, \dots$$

Для одновимірних випадкових величин X і Y введемо позначення:

$$p_i = P\{X = x_i\}, \quad i = 1, 2, \dots; \quad q_k = P\{Y = y_k\}, \quad k = 1, 2, \dots$$

Знайдемо зв'язок між ймовірностями p_{ik} та p_i і q_k . Для цього достатньо зауважити, що подію $\{X = x_i\}$ можна представити як суму попарно несумісних подій:

$$\{X = x_i\} = \{X = x_i, Y = y_1\} + \{X = x_i, Y = y_2\} + \dots,$$

звідки за аксіомою адитивності $p_i = \sum_{k=1}^{\infty} p_{ik}$. Аналогічно можна отримати

формулу $q_k = \sum_{i=1}^{\infty} p_{ik}$. Нарешті,

$$\sum_{k=1}^{\infty} \sum_{i=1}^{\infty} p_{ik} = \sum_{k=1}^{\infty} q_k = \sum_{i=1}^{\infty} p_i = 1.$$

Б. *Випадок неперервної величини.* Для двовимірної неперервної випадкової величини (X, Y) існує невід'ємна функція $p(x, y)$ така, що функція розподілу випадкового вектора (X, Y) визначається у вигляді

$$F(x, y) = \int_{-\infty}^x \int_{-\infty}^y p(u, v) dudv, \quad (6)$$

до того ж функція $p(x, y)$ неперервна всюди, крім, можливо, скінченної кількості точок і називається **щільністю розподілу ймовірностей** випадкового вектора (X, Y) .

З (6) випливає, що в точках неперервності щільність розподілу можна визначити як другу мішану похідну функції розподілу:

$$p(x, y) = \frac{\partial^2 F}{\partial x \partial y}.$$

Використовуючи, що $F(+\infty, +\infty) = 1$, з (6) отримуємо основну властивість щільності розподілу випадкового вектора (X, Y) :

$$\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} p(x, y) dx dy = 1.$$

Якщо можливі значення двовимірної неперервної випадкової величини (X, Y) розміщені в області $G \subset \mathbb{R}^2$, то формула (6) набуває вигляду:

$$P\{(x, y) \in G\} = \iint_G p(x, y) dx dy.$$

Знаючи щільність розподілу $p(x, y)$ випадкового вектора (X, Y) , можна знайти щільності розподілу його компонент $p_X(x)$ та $p_Y(y)$. Справді,

$$F_X(x) = \lim_{y \rightarrow \infty} F(x, y) = \int_{-\infty}^x \int_{-\infty}^{\infty} p(u, v) dudv,$$

звідки

$$p_X(x) = \frac{dF_X(x)}{dx} = \int_{-\infty}^{\infty} p(x, v) dv.$$

Аналогічно можна отримати формулу для $p_Y(y)$:

$$p_Y(y) = \int_{-\infty}^{\infty} p(u, y) du.$$

Незалежність випадкових величин

У попередньому пункті ми показали, як, знаючи закон розподілу двовимірної випадкової величини (X, Y) , знайти закон розподілу окремих величин X і Y . Виникає природне питання: чи не можна, знаючи закони розподілу окремих випадкових величин X і Y , знайти закон розподілу випадкового вектора (X, Y) ? Виявляється, що це можна зробити лише в одному

частковому випадку, коли випадкові величини X і Y є незалежні.

Означення. Випадкові величини X_1, X_2, \dots, X_n називаються незалежними, якщо для будь-яких дійсних чисел x_1, x_2, \dots, x_n

$$F(x_1, x_2, \dots, x_n) = F_1(x_1)F_2(x_2) \cdot \dots \cdot F_n(x_n),$$

де $F(x_1, x_2, \dots, x_n)$ – функція розподілу випадкового вектора (X_1, X_2, \dots, X_n) ,

$F_k(x_k)$ – функція розподілу випадкової величини X_k , $k = \overline{1, n}$.

Розглянемо випадок $n=2$. Якщо X, Y – дискретні випадкові величини, то їхня незалежність означає, що для будь-яких i, k події $\{X = x_i\}$ і $\{Y = y_k\}$ є незалежні, тому

$$p_{ik} = P\{X = x_i, Y = y_k\} = P\{X = x_i\} \cdot P\{Y = y_k\} = p_i q_k, \quad i, k = 1, 2, \dots$$

Нехай X, Y – неперервні випадкові величини з функціями розподілу $F_X(x)$ і $F_Y(y)$ і щільностями розподілу ймовірностей $p_X(x)$ та $p_Y(y)$ відповідно. Нехай $F(x, y)$ і $p(x, y)$ – функція розподілу і щільність розподілу двовимірної випадкової величини (X, Y) . Тоді за означенням незалежності X і Y маємо: $F(x, y) = F_X(x) \cdot F_Y(y)$. Диференціюючи двічі (по x і по y), одержимо:

$$\frac{\partial^2 F}{\partial x \partial y} = F'_X(x) \cdot F'_Y(y), \text{ тобто}$$

$$p(x, y) = p_X(x) \cdot p_Y(y). \quad (7)$$

Можна показати, що умова (7) є не лише необхідною, а й достатньою для незалежності неперервних випадкових величин X і Y . Справді, якщо (7) виконується, то

$$\begin{aligned} F(x, y) &= \int_{-\infty}^x \int_{-\infty}^y p(u, v) du dv = \int_{-\infty}^x \int_{-\infty}^y p_X(u) p_Y(v) du dv = \\ &= \int_{-\infty}^x p_X(u) du \int_{-\infty}^y p_Y(v) dv = F_X(x) \cdot F_Y(y), \end{aligned}$$

а це означає, що X і Y – незалежні.

Лекція 10

Математичне сподівання і його властивості

Найповнішу інформацію про випадкову величину дає її функція розподілу. Проте іноді навіть зручніше користуватися числами, які описують випадкову величину сумарно і називаються *числовими характеристиками* цієї величини. Розглянемо основні з них.

Нехай X – дискретна випадкова величина, яка може набувати значень x_1, x_2, \dots відповідно з імовірностями p_1, p_2, \dots .

Означення. *Математичним сподіванням дискретної випадкової величини X називається число*

$$E(X) = \sum_{k=1}^{\infty} x_k p_k = x_1 p_1 + x_2 p_2 + \dots, \quad (1)$$

якщо ряд у правій частині абсолютно збіжний.

Якщо ряд (1) не збіжний абсолютно, то кажуть, що математичне сподівання не існує. Зауважимо, що у випадку скінченної кількості можливих значень дискретної випадкової величини її математичне сподівання

$$E(X) = \sum_{k=1}^n x_k p_k = x_1 p_1 + x_2 p_2 + \dots + x_n p_n$$

завжди існує.

Щоб пояснити ймовірнісний зміст математичного сподівання, розглянемо приклад. Нехай проводиться n незалежних випробувань, в кожному з яких випадкова величина X може набувати деякого значення з множини $\{x_1, x_2, \dots, x_k\}$. Припустимо, що значення x_1 набуло n_1 разів, значення x_2 набуло n_2 разів, ..., значення x_k набуло n_k разів. Тоді середнє арифметичне значення набутих випадковою величиною X значень в n випробуваннях обчислимо за формулою:

$$\bar{x} = \frac{1}{n} (x_1 n_1 + x_2 n_2 + \dots + x_k n_k) = x_1 \cdot \frac{n_1}{n} + x_2 \cdot \frac{n_2}{n} + \dots + x_k \cdot \frac{n_k}{n}.$$

Відношення $\frac{n_i}{n}$ – це відносна частота події $\{X = x_i\}$. Якщо кількість

випробувань n – число досить велике, то $\frac{n_i}{n} \approx p_i = P\{X = x_i\}$. Тому і середнє арифметичне \bar{x} буде в цьому випадку наближатися до математичного сподівання $E(X)$:

$$\bar{x} \approx x_1 p_1 + x_2 p_2 + \dots + x_k p_k = E(X).$$

Отже, *ймовірнісний зміст математичного сподівання* полягає в тому, що *математичне сподівання наближено дорівнює (тим точніше, чим більша кількість випробувань n) середньому арифметичному спостережуваних значень випадкової величини X .*

Нехай X – неперервна випадкова величина з щільністю розподілу ймовірностей $p(x)$.

Означення. Математичним сподіванням неперервної випадкової величини X називається число

$$E(X) = \int_{-\infty}^{\infty} xp(x) dx, \quad (2)$$

якщо інтеграл у правій частині абсолютно збіжний.

Якщо інтеграл (1) не збіжний абсолютно, то кажуть, що математичне сподівання не існує. Зауважимо, що у випадку, коли можливі значення неперервної випадкової величини зосереджені на проміжку (a, b) , її математичне сподівання

$$E(X) = \int_a^b xp(x) dx,$$

завжди існує.

Оскільки в точках неперервності $p(x)$ маємо: $p(x) = \frac{dF(x)}{dx}$, то формулу (2) можна записати також у вигляді

$$E(X) = \int_{-\infty}^{\infty} x dF(x). \quad (3)$$

Приклади. Для випадкової величини X , розподіленої за законом Пуассона,

$$p_k = P\{X = x_k\} = \frac{\lambda^k}{k!} e^{-\lambda}, \quad k = 0, 1, 2, \dots,$$

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

Отже, математичне сподівання випадкової величини X , розподіленої за законом Пуассона, рівне параметрові цього розподілу λ .

Для випадкової величини X , рівномірно розподіленої на проміжку $[a, b]$,

$$p(x) = \begin{cases} \frac{1}{b-a}, & x \in [a, b]; \\ 0, & x \notin [a, b]; \end{cases}$$

$$E(X) = \int_a^b xp(x) dx = \int_a^b \frac{x dx}{b-a} = \frac{b^2 - a^2}{b-a} = \frac{a+b}{2}.$$

Це означає, що математичне сподівання випадкової величини X , рівномірно розподіленої на проміжку $[a, b]$, – це середина цього відрізка.

Властивості математичного сподівання

- Якщо X – неперервна випадкова величина з щільністю розподілу ймовірностей $p(x)$, а $f(x)$ – неперервна функція на множині можливих значень X , то

$$E(f(X)) = \int_{-\infty}^{\infty} f(x)p(x)dx. \quad (4)$$

Якщо X – дискретна випадкова величина з законом розподілу

$$P\{X = x_k\} = p_k, \quad k = 1, 2, \dots,$$

то

$$E(f(X)) = \sum_{k=1}^{\infty} f(x_k)p_k. \quad (5)$$

Приклад. Перевіримо формулу (4) для випадкової величини $Y = f(X) = X^3$. Маємо:

$$F_Y(x) = P\{Y < x\} = P\{X^3 < x\} = P\{X < \sqrt[3]{x}\} = F_X(\sqrt[3]{x}).$$

Тоді за формулою (3)

$$E(X^3) = E(Y) = \int_{-\infty}^{\infty} x dF_Y(x) = \int_{-\infty}^{\infty} x dF_X(\sqrt[3]{x}),$$

або після заміни $y = \sqrt[3]{x}$, $x = y^3$:

$$E(X^3) = \int_{-\infty}^{\infty} y^3 dF_X(y) = \int_{-\infty}^{\infty} y^3 p(y) dy = \int_{-\infty}^{\infty} x^3 p(x) dx.$$

• Якщо X, Y – неперервні випадкові величини, $p(x, y)$ – щільність розподілу ймовірностей випадкового вектора (X, Y) , а $f(x, y)$ – неперервна функція на множині можливих значень випадкового вектора (X, Y) , то

$$E(f(X, Y)) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(x, y)p(x, y) dx dy. \quad (6)$$

Якщо X, Y – дискретні випадкові величини із законами розподілу

$$P\{X = x_i\} = p_i, \quad i = 1, 2, \dots, \quad P\{Y = y_k\} = q_k, \quad k = 1, 2, \dots,$$

то

$$E(f(X, Y)) = \sum_{i=1}^{\infty} \sum_{k=1}^{\infty} f(x_i, y_k)p_{ik}, \quad (7)$$

де $p_{ik} = P\{X = x_i, Y = y_k\}$, $i, k = 1, 2, \dots$ – закон розподілу випадкового вектора (X, Y) .

• Математичне сподівання сталої величини дорівнює цій величині, тобто якщо $C = \text{const}$, то $E(C) = C$.

Доведення. Сталу C можна розглядати як дискретну випадкову величину, яка набуває лише одного значення C з імовірністю 1. Тому $E(C) = C \cdot 1 = C$.

• Сталій множник виноситься за знак математичного сподівання, тобто якщо $C = \text{const}$, то $E(CX) = CE(X)$.

Доведення є безпосереднім наслідком співвідношень (4) і (5).

• Математичне сподівання алгебраїчної суми двох (або кількох) випадкових величин дорівнює алгебраїчній сумі математичних сподівань цих

величин, тобто $E(X \pm Y) = E(X) \pm E(Y)$.

Доведення. Якщо X, Y – дискретні випадкові величини, то застосовуючи співвідношення (7), одержимо:

$$\begin{aligned} E(X + Y) &= \sum_{i=1}^{\infty} \sum_{k=1}^{\infty} (x_i + y_k) p_{ik} = \sum_{i=1}^{\infty} \sum_{k=1}^{\infty} x_i p_{ik} + \sum_{i=1}^{\infty} \sum_{k=1}^{\infty} y_k p_{ik} = \\ &= \sum_{i=1}^{\infty} x_i \sum_{k=1}^{\infty} p_{ik} + \sum_{k=1}^{\infty} y_k \sum_{i=1}^{\infty} p_{ik} = \sum_{i=1}^{\infty} x_i p_i + \sum_{k=1}^{\infty} y_k q_k = E(X) + E(Y). \end{aligned}$$

Для неперервних випадкових величин використовуємо формулу (6):

$$\begin{aligned} E(X + Y) &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} (x + y) p(x, y) dx dy = \int_{-\infty}^{\infty} x dx \int_{-\infty}^{\infty} p(x, y) dy + \int_{-\infty}^{\infty} y dy \int_{-\infty}^{\infty} p(x, y) dx = \\ &= \int_{-\infty}^{\infty} x p_X(x) dx + \int_{-\infty}^{\infty} y p_Y(y) dy = E(X) + E(Y). \end{aligned}$$

Приклад. Знайти математичне сподівання кількості успіхів μ та частоти успіхів μ/n в n випробуваннях за схемою Бернуллі, якщо ймовірність успіху в одному випробуванні дорівнює p і $q = 1 - p$.

Розв'язання. Позначимо через μ_k кількість успіхів у випробуванні під номером k . Тоді $\mu = \mu_1 + \mu_2 + \dots + \mu_n$, $E(\mu_k) = 1 \cdot p + 0 \cdot q = p$, оскільки кількість успіхів в одному випробуванні рівна 1 з імовірністю p або 0 з імовірністю q . Отже,

$$E(\mu) = E(\mu_1 + \mu_2 + \dots + \mu_n) = E(\mu_1) + E(\mu_2) + \dots + E(\mu_n) = np;$$

$$E\left(\frac{\mu}{n}\right) = \frac{1}{n} \cdot E(\mu) = \frac{np}{n} = p.$$

• Якщо випадкові величини X і Y – незалежні, то $E(XY) = E(X) \cdot E(Y)$.

Доведення. Якщо X і Y – дискретні випадкові величини і

$$p_i = P\{X = x_i\}, \quad q_k = P\{Y = y_k\}, \quad p_{ik} = P\{X = x_i, Y = y_k\},$$

то з їхньої незалежності випливає, що $p_{ik} = p_i q_k$, тому

$$\begin{aligned} E(XY) &= \sum_{i=1}^{\infty} \sum_{k=1}^{\infty} (x_i y_k) p_{ik} = \sum_{i=1}^{\infty} \sum_{k=1}^{\infty} (x_i y_k) p_i q_k = \\ &= \sum_{i=1}^{\infty} x_i p_i \sum_{k=1}^{\infty} y_k q_k = E(X) \cdot E(Y). \end{aligned}$$

Для незалежних неперервних випадкових величин X і Y з щільностями розподілу відповідно $p_X(x)$ і $p_Y(y)$ двовимірною випадковою величиною (X, Y) має щільність розподілу $p(x, y) = p_X(x) p_Y(y)$ і за формулою (6) отримаємо:

$$E(XY) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} (x \cdot y) p(x, y) dx dy = \int_{-\infty}^{\infty} x p_X(x) dx \cdot \int_{-\infty}^{\infty} y p_Y(y) dy = E(X) \cdot E(Y).$$

Лекція 11

Дисперсія та її властивості

Дисперсія випадкової величини характеризує відхилення випадкової величини від її математичного сподівання.

Для характеристики цього відхилення неможливо використати його математичне сподівання, оскільки воно дорівнює нулю:

$$E(X - E(X)) = E(X) - E(E(X)) = E(X) - E(X) = 0.$$

Цей результат цілком природний. Він показує, що випадкова величина „однаково часто” відхиляється від свого математичного сподівання як у більшу, так і в меншу сторону. Щоб запобігти взаємному знищенню додатних і від’ємних значень відхилення $X - E(X)$, можна замість самого відхилення розглядати його квадрат $(X - E(X))^2$.

Означення. *Дисперсією* випадкової величини X називається математичне сподівання квадрата відхилення цієї випадкової величини від її математичного сподівання, тобто

$$D(X) = E(X - E(X))^2.$$

З властивостей математичного сподівання випливає, що дисперсію дискретної і нерерервної випадкових величин можна обчислити відповідно за формулами:

$$D(X) = \sum_{k=1}^{\infty} (x_k - E(X))^2 p_k; \quad D(X) = \int_{-\infty}^{\infty} (x - E(X))^2 p(x) dx.$$

Використовуючи дисперсію для характеристики розсіювання випадкової величини, стикаємося з незручністю: якщо випадкова величина вимірюється в деяких одиницях, то дисперсія вимірюється у квадратах цих одиниць. Тому доцільно мати характеристику розсіювання значень випадкової величини тієї ж вимірності, що й сама величина. Такою характеристикою є середнє квадратичне відхилення.

Означення. *Середнім квадратичним відхиленням* випадкової величини X називається корінь квадратний із дисперсії $D(X)$, тобто

$$\sigma(X) = \sqrt{D(X)}.$$

Дисперсія має такі властивості.

- Дисперсія будь-якої випадкової величини невід’ємна: $D(X) \geq 0$.

Доведення. Дисперсія – це математичне сподівання (середнє значення) невід’ємної випадкової величини $(X - E(X))^2$.

- Дисперсія сталої величини дорівнює нулю, тобто якщо $C = \text{const}$, то $D(C) = 0$.

Доведення. $D(C) = E(C - E(C))^2 = E(C - C)^2 = E(0) = 0$.

- Сталій множник виноситься у квадраті за знак дисперсії, тобто якщо $C = \text{const}$, то $D(CX) = C^2 D(X)$.

Доведення.

$$D(CX) = E(CX - E(CX))^2 = E(CX - CE(X))^2 = C^2 E(X - E(X))^2 = C^2 D(X).$$

• Дисперсія випадкової величини дорівнює різниці між математичним сподіванням квадрата цієї величини і квадратом її математичного сподівання, тобто $D(X) = E(X^2) - (E(X))^2$.

Доведення.

$$\begin{aligned} D(X) &= E(X - E(X))^2 = E(X^2 - 2X \cdot E(X) + (E(X))^2) = \\ &= E(X^2) - 2E(X) \cdot E(X) + (E(X))^2 = E(X^2) - (E(X))^2. \end{aligned}$$

• Дисперсія суми двох незалежних випадкових величин дорівнює сумі дисперсій цих величин, тобто $D(X + Y) = D(X) + D(Y)$.

Доведення. Для незалежних випадкових величин X і Y : $E(XY) = E(X) \cdot E(Y)$, тому

$$\begin{aligned} D(X + Y) &= E(X + Y)^2 - (E(X + Y))^2 = E(X^2 + 2XY + Y^2) - (E(X) + E(Y))^2 = \\ &= E(X^2) + 2E(XY) + E(Y^2) - (E(X))^2 - 2E(X) \cdot E(Y) - (E(Y))^2 = \\ &= E(X^2) - (E(X))^2 + E(Y^2) - (E(Y))^2 = D(X) + D(Y). \end{aligned}$$

• Дисперсія різниці двох незалежних випадкових величин дорівнює сумі дисперсій цих величин, тобто $D(X - Y) = D(X) + D(Y)$.

• Якщо кожна з випадкових величин X_1, X_2, \dots, X_n не залежить від суми попередніх, то дисперсія суми цих величин дорівнює сумі їхніх дисперсій, тобто $D(X_1 + X_2 + \dots + X_n) = D(X_1) + D(X_2) + \dots + D(X_n)$.

Доведення. Досить застосувати метод математичної індукції і використати той факт, що дисперсія суми двох незалежних випадкових величин дорівнює сумі дисперсій цих величин.

• Якщо випадкові величини X_1, X_2, \dots, X_n – попарно незалежні, то дисперсія суми цих величин дорівнює сумі їхніх дисперсій:

$$D(X_1 + X_2 + \dots + X_n) = D(X_1) + D(X_2) + \dots + D(X_n).$$

Доведення. Досить повторити міркування, викладені у доведенні властивості $D(X + Y) = D(X) + D(Y)$, замінивши у ньому формули квадрата суми двох доданків на формули квадрата суми n доданків.

Дисперсія випадкової величини, розподіленої за законом Пуассона. Для випадкової величини X , розподіленої за законом Пуассона,

$$p_k = P\{X = x_k\} = \frac{\lambda^k}{k!} e^{-\lambda}, \quad k = 0, 1, 2, \dots, \quad E(X) = \lambda,$$

$$\begin{aligned}
E(X^2) &= \sum_{k=0}^{\infty} k^2 p_k = \sum_{k=0}^{\infty} k^2 \frac{\lambda^k}{k!} e^{-\lambda} = e^{-\lambda} \sum_{k=1}^{\infty} \frac{k \lambda^k}{(k-1)!} = e^{-\lambda} \sum_{k=1}^{\infty} \frac{[(k-1)+1] \lambda^k}{(k-1)!} = \\
&= e^{-\lambda} \sum_{k=2}^{\infty} \frac{\lambda^k}{(k-2)!} + e^{-\lambda} \sum_{k=1}^{\infty} \frac{\lambda^k}{(k-1)!} = \lambda^2 e^{-\lambda} \sum_{k=2}^{\infty} \frac{\lambda^{k-2}}{(k-2)!} + \lambda e^{-\lambda} \sum_{k=1}^{\infty} \frac{\lambda^{k-1}}{(k-1)!} = \\
&= \lambda^2 e^{-\lambda} e^{\lambda} + \lambda e^{-\lambda} e^{\lambda} = \lambda^2 + \lambda.
\end{aligned}$$

Отже, $D(X) = E(X^2) - (E(X))^2 = \lambda^2 + \lambda - \lambda^2 = \lambda$, тобто як і математичне сподівання, дисперсія випадкової величини X , розподіленої за законом Пуассона, рівна параметрові цього розподілу λ .

Дисперсія рівномірно розподіленої випадкової величини. Для випадкової величини X , рівномірно розподіленої на проміжку $[a, b]$,

$$p(x) = \begin{cases} \frac{1}{b-a}, & x \in [a, b]; \\ 0, & x \notin [a, b]; \end{cases}$$

$$E(X) = \frac{a+b}{2}; \quad E(X^2) = \int_a^b x^2 p(x) dx = \int_a^b \frac{x^2 dx}{b-a} = \frac{b^3 - a^3}{3(b-a)} = \frac{a^2 + ab + b^2}{3};$$

$$D(X) = E(X^2) - (E(X))^2 = \frac{a^2 + ab + b^2}{3} - \frac{(a+b)^2}{4} = \frac{(a-b)^2}{12}.$$

Дисперсія випадкової величини, розподіленої за біномним законом. Так розподілена кількість успіхів μ в n випробуваннях за схемою Бернуллі, якщо ймовірність успіху в одному випробуванні дорівнює p і $q = 1 - p$. Позначимо через μ_k кількість успіхів у випробуванні під номером k . Тоді $\mu = \mu_1 + \mu_2 + \dots + \mu_n$, $E(\mu_k) = p$, $\mu_k^2 = \mu_k$, оскільки кількість успіхів в одному випробуванні може набувати лише значень 1 або 0. Отже,

$$E(\mu_k^2) = E(\mu_k) = p, \quad D(\mu_k) = E(\mu_k^2) - (E(\mu_k))^2 = p - p^2 = p(1-p) = pq.$$

Випадкові величини $\mu_1, \mu_2, \dots, \mu_n$ — попарно незалежні (бо випробування за схемою Бернуллі є незалежними), тому

$$D(\mu) = D(\mu_1 + \mu_2 + \dots + \mu_n) = D(\mu_1) + D(\mu_2) + \dots + D(\mu_n) = npq.$$

Математичне сподівання та дисперсія нормально розподіленої випадкової величини. Нагадаємо, що неперервна випадкова величина X називається розподіленою за нормальним законом з параметрами $-\infty < a < +\infty$ і $\sigma > 0$, якщо її щільність розподілу має вигляд:

$$p(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-a)^2}{2\sigma^2}}, \quad -\infty < x < \infty.$$

Обчислимо математичне сподівання випадкової величини X . Для цього в інтегралі

$$E(X) = \int_{-\infty}^{\infty} xp(x) dx = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{\infty} xe^{-\frac{(x-a)^2}{2\sigma^2}} dx$$

зробимо заміну $t = \frac{x-a}{\sigma\sqrt{2}}$. Тоді $x = \sigma\sqrt{2}t + a$; $dx = \sigma\sqrt{2} dt$ і враховуючи, що

$$\int_{-\infty}^{\infty} e^{-t^2} dt = \sqrt{\pi}, \quad \int_{-\infty}^{\infty} te^{-t^2} dt = 0,$$

одержимо:

$$E(X) = \frac{\sigma\sqrt{2}}{\sigma\sqrt{2\pi}} \int_{-\infty}^{\infty} (\sigma\sqrt{2}t + a)e^{-t^2} dt = \frac{\sigma\sqrt{2}}{\sqrt{\pi}} \int_{-\infty}^{\infty} te^{-t^2} dt + \frac{a}{\sqrt{\pi}} \int_{-\infty}^{\infty} e^{-t^2} dt = 0 + \frac{a}{\sqrt{\pi}} \cdot \sqrt{\pi} = a.$$

Дисперсію обчислимо за формулою:

$$D(X) = \int_{-\infty}^{\infty} (x - E(X))^2 p(x) dx = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{\infty} (x - a)^2 e^{-\frac{(x-a)^2}{2\sigma^2}} dx.$$

Знову зробимо заміну $t = \frac{x-a}{\sigma\sqrt{2}}$, і отриманий інтеграл проінтегруємо частинами,

поклавши $t = u$, $2te^{-t^2} = dv$:

$$D(X) = \frac{\sigma^2}{\sqrt{\pi}} \int_{-\infty}^{\infty} 2t^2 e^{-t^2} dt = \frac{\sigma^2}{\sqrt{\pi}} \left(-te^{-t^2} \Big|_{-\infty}^{\infty} + \int_{-\infty}^{\infty} e^{-t^2} dt \right) = \frac{\sigma^2}{\sqrt{\pi}} (0 + \sqrt{\pi}) = \sigma^2.$$

Отже, математичне сподівання і дисперсія нормально розподіленої випадкової величини відповідно дорівнюють a і σ^2 , тобто виражаються через параметри цього розподілу. Середнє квадратичне відхилення нормально розподіленої випадкової величини рівне параметрові σ цього розподілу:

$$\sigma(X) = \sqrt{D(X)} = \sigma.$$

Математичне сподівання двовимірної випадкової величини

Нехай задано n -вимірну випадкову величину (X_1, X_2, \dots, X_n) . Її математичним сподіванням називається n -вимірний вектор

$$(E(X_1), E(X_2), \dots, E(X_n)),$$

де $E(X_k)$ – математичне сподівання випадкової величини X_k .

Розглянемо, зокрема, двовимірну неперервну випадкову величину (X, Y) з щільністю розподілу ймовірностей $p(x, y)$. Якщо $p_X(x)$, $p_Y(y)$ – щільності розподілу випадкових величин X і Y , то згідно з означенням математичного сподівання одновимірної випадкової величини

$$E(X) = \int_{-\infty}^{\infty} xp_X(x) dx, \quad E(Y) = \int_{-\infty}^{\infty} yp_Y(y) dy.$$

Однак, $p_X(x) = \int_{-\infty}^{\infty} p(x, y) dy$, $p_Y(y) = \int_{-\infty}^{\infty} p(x, y) dx$, тому

$$E(X) = \int_{-\infty}^{\infty} x \left(\int_{-\infty}^{\infty} p(x, y) dy \right) dx = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} xp(x, y) dx dy.$$

Аналогічно отримаємо формулу для $E(Y)$:

$$E(Y) = \int_{-\infty}^{\infty} y \left(\int_{-\infty}^{\infty} p(x, y) dx \right) dy = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} yp(x, y) dx dy.$$

Дисперсія двовимірної випадкової величини. Коваріація, коефіцієнт кореляції та його властивості

Означення. *Дисперсією (або дисперсійною матрицею) n -вимірної випадкової величини (X_1, X_2, \dots, X_n) називається сукупність n^2 чисел, які визначаються формулами*

$$b_{ik} = E((X_i - E(X_i))(X_k - E(X_k))), \quad i, k = \overline{1, n}.$$

Очевидно, що дисперсійна матриця симетрична: $b_{ik} = b_{ki}$.

Для двовимірної випадкової величини (X, Y) дисперсією є сукупність 3-х чисел: b_{11}, b_{22} і $b_{12} = b_{21}$. Перші два числа є дисперсіями компонент цієї випадкової величини: $b_{11} = E((X - E(X))(X - E(X))) = E(X - E(X))^2 = D(X)$, $b_{22} = E(Y - E(Y))^2 = D(Y)$, а третя називається **коваріацією** випадкових величин X і Y :

$$\text{cov}(X, Y) = E((X - E(X))(Y - E(Y))),$$

або після розкриття дужок:

$$\begin{aligned} \text{cov}(X, Y) &= E(XY) + E(X(-E(Y))) + E(-E(X) \cdot Y) + E(E(X)E(Y)) = \\ &= E(XY) - E(X)E(Y) - E(X)E(Y) + E(X)E(Y) = E(XY) - E(X)E(Y). \end{aligned}$$

З отриманої формули випливає, що **коваріація незалежних випадкових величин дорівнює нулю**.

Зв'язок між дисперсією та коваріацією встановлюють формули:

$$\text{cov}(X, X) = D(X); \quad D(X + Y) = D(X) + 2\text{cov}(X, Y) + D(Y).$$

Справді,

$$\begin{aligned} D(X + Y) &= E((X + Y) - E(X + Y))^2 = E((X - E(X)) + (Y - E(Y)))^2 = \\ &= E(X - E(X))^2 + 2E((X - E(X))(Y - E(Y))) + E(Y - E(Y))^2 = \\ &= D(X) + 2\text{cov}(X, Y) + D(Y). \end{aligned}$$

Використовуючи означення математичного сподівання, одержимо:

- для дискретного розподілу

$$\text{cov}(X, Y) = \sum_{i, k} (x_i - E(X))(y_k - E(Y)) p_{ik};$$

- для неперервного розподілу

$$\text{cov}(X, Y) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} (x - E(X))(y - E(Y)) p(x, y) dx dy.$$

Коваріація двох випадкових величин X і Y характеризує не тільки ступінь

взаємозалежності цих випадкових величин, а й також їхнє розсіювання навколо точки $(E(X), E(Y))$ на площині.

Розмірність коваріації $\text{cov}(X, Y)$ дорівнює добуткові розмірностей випадкових величин X і Y . Для того, щоб отримати безрозмірну величину, до того ж таку, яка характеризує тільки залежність між випадковими величинами, а не їхнє розсіювання, вводиться поняття коефіцієнта кореляції.

Означення. *Коефіцієнтом кореляції між випадковими величинами X і Y називається відношення коваріації до добутку середніх квадратичних відхилень цих величин:*

$$\rho(X, Y) = \frac{\text{cov}(X, Y)}{\sqrt{D(X) \cdot D(Y)}} = \frac{\text{cov}(X, Y)}{\sigma(X) \cdot \sigma(Y)}.$$

Розглянемо властивості коефіцієнта кореляції.

- *Коефіцієнт кореляції незалежних випадкових величин дорівнює нулю.*
Доведення. Для незалежних випадкових величин X і Y $\text{cov}(X, Y) = 0$.

- *Для будь-яких випадкових величин $|\rho(X, Y)| \leq 1$.*

Доведення. Досить довести, що $\rho^2(X, Y) \leq 1$, тобто

$$\text{cov}^2(X, Y) \leq D(X)D(Y). \quad (1)$$

Оскільки математичне сподівання невід'ємної випадкової величини невід'ємне, то для кожного дійсного числа t маємо $E(t(X - E(X)) - (Y - E(Y)))^2 \geq 0$, тобто

$$t^2 \cdot D(X) - 2t \text{cov}(X, Y) + D(Y) \geq 0. \quad (2)$$

У лівій частині цієї нерівності – квадратний тричлен відносно t з додатним коефіцієнтом при t^2 . Цей тричлен буде невід'ємним для всіх дійсних значень t тоді і лише тоді, коли його дискримінант недодатний:

$$4\text{cov}^2(X, Y) - 4D(X)D(Y) \leq 0,$$

отже, виконується нерівність (1).

- $|\rho(X, Y)| = 1$ тоді і лише тоді, коли випадкові величини зв'язані лінійною залежністю $Y = \alpha X + \beta$. Тоді

$$\rho(X, Y) = \begin{cases} 1, & \text{якщо } \alpha > 0; \\ -1, & \text{якщо } \alpha < 0. \end{cases} \quad (3)$$

Доведення. (\Rightarrow) Нехай $|\rho(X, Y)| = 1$, тобто $\text{cov}^2(X, Y) = D(X)D(Y)$, тому дискримінант квадратного тричлена у лівій частині нерівності (2) рівний нулю, і цей тричлен має двократний дійсний корінь

$$t = \alpha = \frac{\text{cov}(X, Y)}{D(X)}. \quad (4)$$

Підставивши $t = \alpha$ у рівність $E(t(X - E(X)) - (Y - E(Y)))^2 = 0$, одержимо

$$E(\alpha X - Y - E(\alpha X - Y))^2 = 0,$$

тобто $D(\alpha X - Y) = 0$. Це означає, що випадкова величина $\alpha X - Y$ є сталою, позначимо її через $-\beta$. Отже, $\alpha X - Y = -\beta$, тому $Y = \alpha X + \beta$. В (4) $D(X) > 0$,

тому числа α і $\text{cov}(X, Y)$ мають однакові знаки. Але ж знак $\rho(X, Y)$ визначається знаком $\text{cov}(X, Y)$, а, отже, й знаком α , тобто виконується (3).

(\Leftarrow) Нехай випадкові величини зв'язані лінійною залежністю $Y = \alpha X + \beta$. Доведемо, що $|\rho(X, Y)| = 1$, тобто $\text{cov}^2(X, Y) = D(X)D(Y)$. Оскільки

$E(Y) = E(\alpha X + \beta) = \alpha E(X) + \beta$, $Y - E(Y) = \alpha X + \beta - \alpha E(X) - \beta = \alpha(X - E(X))$, то

$$\text{cov}(X, Y) = E((X - E(X))(Y - E(Y))) = E(\alpha(X - E(X))^2) = \alpha D(X).$$

З другого боку, $D(Y) = D(\alpha X + \beta) = D(\alpha X) + D(\beta) = \alpha^2 D(X)$. Отже,

$$\text{cov}^2(X, Y) = \alpha^2 (D(X))^2, \quad D(X)D(Y) = D(X) \cdot \alpha^2 D(X) = \alpha^2 (D(X))^2,$$

тобто $\text{cov}^2(X, Y) = D(X)D(Y)$. Теорему доведено.

Модуль коефіцієнта кореляції випадкових величин X і Y характеризує *ступінь тісноти лінійної залежності між ними*. Якщо лінійної залежності немає, то $\rho(X, Y) = 0$. Якщо між випадковими величинами існує лінійна функціональна залежність $Y = \alpha X + \beta$, то $\rho(X, Y) = 1$ при $\alpha > 0$ і $\rho(X, Y) = -1$ при $\alpha < 0$.

Означення. Дві випадкові величини X і Y називаються **корельованими**, якщо $\rho(X, Y) \neq 0$, і **некорельованими**, якщо $\rho(X, Y) = 0$.

Легко переконатися, що дві корельовані випадкові величини є також залежні (якщо припустимо від супротивного, що вони незалежні, то прийдемо до суперечності $\rho(X, Y) = 0$). Обернене твердження правильне не завжди, тобто якщо дві випадкові величини залежні, то вони можуть бути як корельованими, так і некорельованими. Покажемо, зокрема, на прикладі, що може бути $\rho(X, Y) = 0$ навіть для випадкових величин, зв'язаних функціональною залежністю, тобто з некорельованості випадкових величин не завжди випливає їхня незалежність.

Приклад. Нехай випадкова величина X рівномірно розподілена на проміжку $[-1, 1]$, $Y = X^2$. Тоді

$$p_X(x) = \begin{cases} \frac{1}{2}, & x \in [-1, 1]; \\ 0, & x \notin [-1, 1]; \end{cases} \quad E(X) = 0;$$

$$\text{cov}(X, Y) = E(XY) - E(X)E(Y) = E(XY) = E(X^3) = \frac{1}{2} \int_{-1}^1 x^3 dx = 0 \Rightarrow \rho(X, Y) = 0.$$

Можна довести, що для *нормально розподілених випадкових величин X і Y некорельованість рівносильна незалежності*.

Лекція 12

Вивчення зв'язку між випадковими величинами за допомогою прямих регресії

Якщо $\rho(X, Y) \neq 0$, то X і Y – корельовані випадкові величини і з наближенням величини $|\rho(X, Y)|$ до одиниці залежність між цими випадковими величинами наближається до лінійної залежності вигляду $Y = \alpha X + \beta$. Тому цілком природним є питання про наближене зображення зв'язку між X і Y у вигляді лінійної функції.

Означення. Пряма $y = \alpha x + \beta$ називається **прямою регресії** (або прямою середньої квадратичної регресії) **Y на X** , якщо коефіцієнти α , β вибрані так, що середнє квадратичне відхилення $\alpha X + \beta$ від Y є мінімальним: $E(Y - \alpha X - \beta)^2 = \min$.

Щоб знайти значення α , β , про які сказано в означенні, здійснимо перетворення величини $E(Y - \alpha X - \beta)^2$, яку нам треба мінімізувати:

$$\begin{aligned} E(Y - \alpha X - \beta)^2 &= E(Y - E(Y) - \alpha(X - E(X)) + E(Y) - \alpha E(X) - \beta)^2 = \\ &= E(Y - E(Y))^2 - 2\alpha E((X - E(X))(Y - E(Y))) + \alpha^2 E(X - E(X))^2 + \\ &\quad + (E(Y) - \alpha E(X) - \beta)^2, \end{aligned}$$

оскільки

$$\begin{aligned} E((Y - E(Y))(E(Y) - \alpha E(X) - \beta)) &= (E(Y) - \alpha E(X) - \beta)(E(Y) - E(Y)) = 0; \\ E((X - E(X))(E(Y) - \alpha E(X) - \beta)) &= 0. \end{aligned}$$

Отже,

$$\begin{aligned} E(Y - \alpha X - \beta)^2 &= \sigma^2(Y) - 2\alpha \operatorname{cov}(X, Y) + \alpha^2 \sigma^2(X) + (E(Y) - \alpha E(X) - \beta)^2 = \\ &= \sigma^2(Y) - 2\alpha \rho(X, Y) \sigma(X) \sigma(Y) + \alpha^2 \sigma^2(X) + (E(Y) - \alpha E(X) - \beta)^2 = \\ &= \rho^2(X, Y) \sigma^2(Y) - 2\alpha \rho(X, Y) \sigma(X) \sigma(Y) + \alpha^2 \sigma^2(X) + \sigma^2(Y) - \rho^2(X, Y) \sigma^2(Y) + \\ &\quad + (E(Y) - \alpha E(X) - \beta)^2 = (\rho(X, Y) \sigma(Y) - \alpha \sigma(X))^2 + (1 - \rho^2(X, Y)) \sigma^2(Y) + \\ &\quad + (E(Y) - \alpha E(X) - \beta)^2. \end{aligned}$$

Тепер видно, що величина $E(Y - \alpha X - \beta)^2$ набуватиме найменшого значення тоді і лише тоді, коли виконуватимуться рівності:

$$\rho(X, Y) \sigma(Y) - \alpha \sigma(X) = 0, \quad E(Y) - \alpha E(X) - \beta = 0,$$

звідки можна визначити відповідні значення сталих α , β :

$$\alpha = \rho(X, Y) \frac{\sigma(Y)}{\sigma(X)}, \quad \beta = E(Y) - \alpha E(X).$$

Отже, рівняння прямої регресії Y на X має вигляд:

$$y = \rho(X, Y) \frac{\sigma(Y)}{\sigma(X)} x + E(Y) - \alpha E(X),$$

або

$$y - E(Y) = \rho(X, Y) \frac{\sigma(Y)}{\sigma(X)} (x - E(X)).$$

Приклад. Монету кинули двічі. Нехай X – кількість випадань герба, Y – кількість випадань цифри. Записати рівняння прямої регресії Y на X .

Розв'язання. Маємо: кількість випробувань $n=2$, ймовірність випадання герба в одному випробуванні $p=0,5$, цифри $q=0,5$. Тоді

$$E(X) = np = 1, \quad E(Y) = nq = 1; \quad D(X) = D(Y) = npq = 0,5; \quad \sigma(X) = \sigma(Y) = \frac{1}{\sqrt{2}}.$$

Закон розподілу випадкової величини XY задається таблицею:

$XY=k$	0	1
$P\{XY=k\}$	0,5	0,5

Отже,

$$E(XY) = 0,5; \quad \text{cov}(X, Y) = E(XY) - E(X) \cdot E(Y) = 0,5 - 1 = -0,5;$$

$$\rho(X, Y) = \frac{\text{cov}(X, Y)}{\sigma(X)\sigma(Y)} = \frac{-0,5}{0,5} = -1; \quad \alpha = \rho(X, Y) \frac{\sigma(Y)}{\sigma(X)} = -1,$$

$$\beta = E(Y) - \alpha E(X) = 1 - \alpha = 2,$$

і рівняння прямої регресії Y на X запишеться у вигляді $y = 2 - x$.

Початкові і центральні моменти випадкових величин. Асиметрія та ексцес

Математичне сподівання і дисперсія характеризують найважливіші параметри розподілу випадкових величин: середнє значення та ступінь розсіювання можливих значень величини навколо середнього значення. Крім цих характеристик, у теорії ймовірностей застосовуються деякі інші числові характеристики відповідного призначення, кожна з яких характеризує випадкову величину з позиції тих чи інших особливостей її розподілу. Серед них особливе значення мають початкові і центральні моменти, а також асиметрія та ексцес.

Означення. *Початковим моментом k -го порядку випадкової величини X називають математичне сподівання величини X^k і позначають v_k :*

$$v_k = E(X^k).$$

Зокрема, $v_1 = E(X)$, $v_2 = E(X^2)$, і для обчислення дисперсії маємо таку формулу: $D(X) = v_2 - v_1^2$.

Означення. *Центральним моментом k -го порядку випадкової величини X називають математичне сподівання величини $(X - E(X))^k$ і позначають μ_k :*

$$\mu_k = E(X - E(X))^k.$$

Зокрема,

$$\mu_1 = E(X - E(X)) = E(X) - E(X) = 0, \quad \mu_2 = E(X - E(X))^2 = D(X).$$

Між центральними і початковими моментами існує зв'язок:

$$\mu_k = E(X - E(X))^k = E(X - \nu_1)^k;$$

$$\mu_2 = E(X - \nu_1)^2 = E(X^2 - 2\nu_1 X + \nu_1^2) = \nu_2 - 2\nu_1^2 + \nu_1^2 = \nu_2 - \nu_1^2;$$

$$\begin{aligned} \mu_3 &= E(X - \nu_1)^3 = E(X^3 - 3\nu_1 X^2 + 3\nu_1^2 X - \nu_1^3) = \\ &= \nu_3 - 3\nu_1 \nu_2 + 3\nu_1^3 - \nu_1^3 = \nu_3 - 3\nu_1 \nu_2 + 2\nu_1^3. \end{aligned}$$

Для дискретної випадкової величини X , яка набуває значень x_i з імовірностями p_i ,

$$\nu_k = \sum_i x_i^k p_i, \quad \mu_k = \sum_i (x_i - E(X))^k p_i.$$

Для неперервної випадкової величини X зі щільністю розподілу $p(x)$

$$\nu_k = \int_{-\infty}^{\infty} x^k p(x) dx, \quad \mu_k = \int_{-\infty}^{\infty} (x - E(X))^k p(x) dx.$$

Означення. *Асиметрією* випадкової величини X називається число A , яке обчислюється за формулою:

$$A = \frac{\mu_3}{\sigma^3(X)}.$$

З цієї формули випливає, що напрям асиметрії випадкової величини визначається за допомогою центрального моменту μ_3 , який має розмірність куба випадкової величини. Якщо $\mu_3 = 0$, то випадкова величина розподілена симетрично відносно математичного сподівання $E(X)$. Коли $\mu_3 < 0$ ($\mu_3 > 0$), то розподіл випадкової величини X має від'ємну (додатну) асиметрію. Коефіцієнт асиметрії A є безрозмірною величиною. Його використовують для оцінки як напрямку, так і сили асиметрії розподілу випадкової величини.

Означення. *Ексцесом* випадкової величини X називається число E_S , яке визначається рівністю:

$$E_S = \frac{\mu_4}{\sigma^4(X)} - 3.$$

Ексцес E_S , як і коефіцієнт асиметрії A , є безрозмірною величиною. Він характеризує „гостровершинність” графіка щільності розподілу розглядуваної випадкової величини у порівнянні з кривою нормального розподілу

$$p(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-a)^2}{2\sigma^2}}, \quad -\infty < x < \infty.$$

Для нормального розподілу ексцес $E_S = 0$; криві, які є більш гостровершинні, порівнюючи з нормальною, мають ексцес $E_S > 0$, менш гостровершинні – ексцес $E_S < 0$.

Лекція 13

Нерівність Чебишова

I форма нерівності Чебишова. Якщо випадкова величина X може набувати лише невід'ємні значення ($X \geq 0$) і має скінченне математичне сподівання, то

$$P\{X \geq 1\} \leq E(X).$$

Доведення. Нехай X – дискретна випадкова величина, яка набуває значення $x_i \geq 0$ з імовірністю p_i ($i=1,2,\dots$), тоді згідно з аксіомою адитивності:

$$P\{X \geq 1\} = \sum_{i: x_i \geq 1} p_i \leq \sum_{i: x_i \geq 1} x_i p_i \leq \sum_i x_i p_i = E(X).$$

Для неперервної випадкової величини X зі щільністю розподілу $p(x)$ маємо:

$$P\{X \geq 1\} = P\{1 \leq X < \infty\} = \int_1^{\infty} p(x) dx \leq \int_1^{\infty} xp(x) dx \leq \int_0^{\infty} xp(x) dx = \int_{-\infty}^{\infty} xp(x) dx = E(X).$$

Теорему доведено.

II форма нерівності Чебишова. Якщо випадкова величина X має скінченні математичне сподівання $E(X)$ та дисперсію $D(X)$, то для будь-якого числа $\varepsilon > 0$ виконується нерівність

$$P\{|X - E(X)| \geq \varepsilon\} \leq \frac{D(X)}{\varepsilon^2}, \quad (1)$$

і, крім цього,

$$P\{|X - E(X)| < \varepsilon\} \geq 1 - \frac{D(X)}{\varepsilon^2}.$$

Доведення. Для будь-якого числа $\varepsilon > 0$ розглянемо випадкову величину $Y = \frac{(X - E(X))^2}{\varepsilon^2}$. Вона може набувати лише невід'ємних значень, тому

$$P\{Y \geq 1\} \leq E(Y),$$

тобто

$$P\left\{\frac{(X - E(X))^2}{\varepsilon^2} \geq 1\right\} \leq E\left(\frac{(X - E(X))^2}{\varepsilon^2}\right) = \frac{1}{\varepsilon^2} E(X - E(X))^2 = \frac{D(X)}{\varepsilon^2}.$$

Але ж нерівності $\frac{(X - E(X))^2}{\varepsilon^2} \geq 1$ і $|X - E(X)| \geq \varepsilon$ еквівалентні, тому нерівність

(1) доведено.

Оскільки подія $\{|X - E(X)| < \varepsilon\}$ протилежна до події $\{|X - E(X)| \geq \varepsilon\}$, то

$$P\{|X - E(X)| < \varepsilon\} = 1 - P\{|X - E(X)| \geq \varepsilon\} \geq 1 - \frac{D(X)}{\varepsilon^2}.$$

Теорему доведено.

Збіжність послідовності випадкових величин за ймовірністю

Нехай задана послідовність випадкових величин $X_1, X_2, \dots, X_n, \dots$

Кажуть, що ця послідовність збігається за ймовірністю до числа a , якщо за необмеженого збільшення n імовірність події $\{|X_n - a| < \varepsilon\}$ (де $\varepsilon > 0$ – як завгодно мале фіксоване число) наближається до одиниці, тобто

$$\lim_{n \rightarrow \infty} P\{|X_n - a| < \varepsilon\} = 1.$$

Збіжність за ймовірністю будемо позначати так:

$$X_n \xrightarrow{\text{йм}} a.$$

Закон великих чисел

Суть закону великих чисел полягає в тому, що в разі дуже великої кількості випадкових явищ усереднений їхній результат практично перестає бути випадковим і може бути передбачений із великою часткою вірогідності.

Із деякими окремими випадками закону великих чисел ми вже стикалися раніше. Зокрема, це стосується *стійкості відносної частоти* події за необмеженого збільшення кількості випробувань.

Інший вид статистичної закономірності – *стійкість середнього значення*. Логічною основою для неї є такий хід міркувань. Під час кожного окремого випробування його результат, що характеризується деяким показником, під впливом випадкових факторів буде відхилятися від деякого сталого значення в той чи інший бік. Тому середнє значення показника за достатньо великої кількості випробувань унаслідок взаємного погашення індивідуальних відхилень стає стійким, наближаючись до деякого сталого числа, тобто практично не залежить від випадку.

Отже, закон великих чисел приводить до встановлення детермінованих закономірностей у поведінці відносної частоти, середнього значення або інших показників, що характеризують результат достатньо великої кількості випробувань в умовах непередбачуваності кожної спроби окремо.

Теорема 1 (теорема Маркова, закон великих чисел). *Якщо випадкові величини $X_1, X_2, \dots, X_n, \dots$ попарно незалежні, мають скінченні математичні сподівання і для їхніх дисперсій виконується рівність*

$$\lim_{n \rightarrow \infty} \frac{1}{n^2} \sum_{k=1}^n D(X_k) = 0, \quad (2)$$

то для будь-якого числа $\varepsilon > 0$ виконується граничне співвідношення

$$\lim_{n \rightarrow \infty} P\left\{\left|\frac{1}{n} \sum_{k=1}^n X_k - \frac{1}{n} \sum_{k=1}^n E(X_k)\right| < \varepsilon\right\} = 1, \quad (3)$$

яке означає, що послідовність середніх арифметичних цих випадкових величин збігається за ймовірністю до середнього арифметичного їхніх математичних сподівань.

Доведення. Покладемо $Y_n = \frac{1}{n} \sum_{k=1}^n X_k$. Твердження теореми рівносильне тому, що $\lim_{n \rightarrow \infty} P\{|Y_n - E(Y_n)| \geq \varepsilon\} = 0$ (ймовірність протилежної події).

Оскільки випадкові величини X_1, X_2, \dots, X_n попарно незалежні, то дисперсія їхньої суми рівна сумі дисперсій, тому

$$D(Y_n) = \sum_{k=1}^n D\left(\frac{1}{n} X_k\right) = \frac{1}{n^2} \sum_{k=1}^n D(X_k) \xrightarrow{n \rightarrow \infty} 0$$

згідно зі співвідношенням (2). За нерівністю Чебишова (1)

$$P\{|Y_n - E(Y_n)| \geq \varepsilon\} \leq \frac{D(Y_n)}{\varepsilon^2} \xrightarrow{n \rightarrow \infty} 0.$$

Отже, ймовірність протилежної події при $n \rightarrow \infty$ наближається до одиниці, тобто виконується граничне співвідношення (3).

Теорема 2 (теорема Чебишова). Якщо випадкові величини $X_1, X_2, \dots, X_n, \dots$ попарно незалежні, мають скінченні математичні сподівання та обмежені в сукупності дисперсії

$$D(X_k) \leq C, \quad k = 1, 2, \dots, \quad (4)$$

де $C = \text{const} > 0$, то для будь-якого числа $\varepsilon > 0$ виконується граничне співвідношення (3).

Доведення. Використовуючи нерівності (4), одержимо:

$$\frac{1}{n^2} \sum_{k=1}^n D(X_k) \leq \frac{nC}{n^2} = \frac{C}{n} \xrightarrow{n \rightarrow \infty} 0.$$

Отже, виконується умова (2), тобто виконуються всі умови теореми Маркова, тому виконується граничне співвідношення (3). \square

Суть теорем Чебишова і Маркова полягає в тому, що хоч окремі випадкові величини X_i ($i = 1, 2, \dots$) можуть набувати значень, досить віддалених від своїх математичних сподівань, зате середнє арифметичне великої кількості цих випадкових величин з імовірністю, близькою до одиниці, набуває значення, яке близьке до середнього арифметичного їхніх математичних сподівань. Тобто середнє арифметичне великої кількості незалежних випадкових величин втрачає випадковий характер і має властивість стійкості. Наприклад, індивідуальні ціни окремих товаровиробників на той самий товар можуть значно відрізнятись від його вартості, але середня суспільна ціна наближається до вартості товару.

Теорема 3. Якщо випадкові величини $X_1, X_2, \dots, X_n, \dots$ однаково розподілені, попарно незалежні, мають скінченні математичні сподівання $E(X_k) = a$ та дисперсії $D(X_k) = C$, $k = 1, 2, \dots$, де $C = \text{const} > 0$, то для будь-якого числа $\varepsilon > 0$ виконується граничне співвідношення:

$$\lim_{n \rightarrow \infty} P\left\{\left|\frac{1}{n} \sum_{k=1}^n X_k - a\right| < \varepsilon\right\} = 1, \quad (5)$$

тобто середнє арифметичне випадкових величин збігається за ймовірністю до

математичного сподівання a .

Доведення. З того, що $D(X_k) = C$, $k = 1, 2, \dots$, випливає, що виконуються умови теореми 2, тому виконується граничне співвідношення (3), в якому

$$\frac{1}{n} \sum_{k=1}^n E(X_k) = \frac{na}{n} = a.$$

Теорема 4 (теорема Бернуллі про стійкість відносних частот). *Якщо в кожному з серії незалежних випробувань випадкова подія настає з однією і тою самою ймовірністю p , то за необмеженого зростання кількості випробувань відносна частота появи цієї події збігається за ймовірністю до її ймовірності p , тобто для будь-якого наперед заданого числа $\varepsilon > 0$ виконується граничне співвідношення*

$$\lim_{n \rightarrow \infty} P \left\{ \left| \frac{\mu_n}{n} - p \right| < \varepsilon \right\} = 1, \quad (6)$$

де μ_n – кількість появ події A в n випробуваннях.

Доведення. Через X_k позначимо кількість появ події A у випробуванні за номером k , тоді випадкові величини $X_1, X_2, \dots, X_n, \dots$ попарно незалежні і

$$\mu_n = \sum_{k=1}^n X_k, \quad E(X_k) = p, \quad D(X_k) = pq = p(1-p).$$

Отже, виконуються умови теореми 3 і з граничного співвідношення (5) випливає (6).

Поняття про центральну граничну теорему

Центральна гранична теорема встановлює умови, за яких розподіл суми випадкових величин наближається до нормального закону.

Нехай μ_n – кількість появ події A в n незалежних випробуваннях, в кожному з яких $P(A) = p$. Згідно з інтегральною теоремою Муавра-Лапласа

$$\lim_{n \rightarrow \infty} P \left\{ a < \frac{\mu_n - np}{\sqrt{npq}} < b \right\} = \frac{1}{\sqrt{2\pi}} \int_a^b e^{-\frac{x^2}{2}} dx$$

для будь-яких a, b ($-\infty \leq a \leq b \leq \infty$). Покладемо $a = -\infty$, $b = x$, тоді

$$\lim_{n \rightarrow \infty} P \left\{ \frac{\mu_n - np}{\sqrt{npq}} < x \right\} = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^x e^{-\frac{t^2}{2}} dt = F(x), \quad (7)$$

де $F(x)$ – функція розподілу випадкової величини X , щільність розподілу

ймовірностей якої $p(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$. Граничне співвідношення (7) означає, що

функції розподілу випадкових величин $Z_n = \frac{\mu_n - np}{\sqrt{npq}}$ за необмеженого зростання n наближаються до функції розподілу $F(x)$ нормального закону з

параметрами $a = 0$, $\sigma = 1$.

Надамо співвідношенню (7) іншу форму. Для цього позначимо через X_k кількість появ події A у випробуванні за номером k , тоді випадкові величини $X_1, X_2, \dots, X_n, \dots$ незалежні,

$$\mu_n = \sum_{k=1}^n X_k, \quad E(\mu_n) = \sum_{k=1}^n E(X_k) = np, \quad D(\mu_n) = \sum_{k=1}^n D(X_k) = npq,$$

і твердження, отримане з (7), можна сформулювати так: якщо кожна з незалежних випадкових величин $X_1, X_2, \dots, X_n, \dots$ може набувати лише два значення 1 і 0 з відповідними ймовірностями p і $q = 1 - p$ ($0 < p < 1$), то при $n \rightarrow \infty$ функція розподілу випадкової величини

$$Y_n = \left(\sum_{k=1}^n X_k - \sum_{k=1}^n E(X_k) \right) / \sqrt{\sum_{k=1}^n D(X_k)} \quad (8)$$

для будь-якого дійсного числа x прямує до функції розподілу нормального закону з параметрами $a = 0$, $\sigma = 1$:

$$\lim_{n \rightarrow \infty} F_{Y_n}(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^x e^{-\frac{t^2}{2}} dt. \quad (9)$$

О.М.Ляпунов отримав твердження, яке виражається граничним співвідношенням (9), для послідовності випадкових величин, на які накладаються слабші умови, ніж для випадкових величин $X_1, X_2, \dots, X_n, \dots$ зі схеми Бернуллі.

Теорема Ляпунова (центральна гранична теорема). Нехай $X_1, X_2, \dots, X_n, \dots$ – послідовність незалежних випадкових величин зі скінченними математичними сподіваннями і дисперсіями. Якщо виконана умова

$$\lim_{n \rightarrow \infty} \frac{\sum_{k=1}^n E|X_k - E(X_k)|^3}{\left(\sqrt{\sum_{k=1}^n D(X_k)} \right)^3} = 0, \quad (10)$$

то для будь-якого дійсного числа x для послідовності випадкових величин (8) виконується гранична рівність (9).

Суть умови (10) полягає в тому, що кожна з випадкових величин $X_1, X_2, \dots, X_n, \dots$ має на випадкову величину, яка є їхньою сумою, незначний вплив.

У випадку, коли випадкові величини $X_1, X_2, \dots, X_n, \dots$ – однаково розподілені, твердження центральної граничної теореми справедливе без умови (10).

Наслідок з центральної граничної теореми (теорема Ліндеберга-Леві). Нехай $X_1, X_2, \dots, X_n, \dots$ – послідовність незалежних випадкових величин зі скінченними математичними сподіваннями $E(X_k) = a$ і дисперсіями

$D(X_k) = \sigma^2$. Тоді для будь-якого дійсного числа x виконується гранична рівність

$$\lim_{n \rightarrow \infty} P \left\{ \frac{\sum_{k=1}^n X_k - na}{\sigma \sqrt{n}} < x \right\} = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^x e^{-\frac{t^2}{2}} dt.$$

Отже, суть центральної граничної теореми полягає в тому, що розподіл випадкової величини, яка формується як результат дії багатьох незалежних випадкових факторів, кожний з яких має на неї незначний вплив, мало відрізняється від нормального закону. Оскільки ці умови на практиці досить часто виконуються, то нормальний закон розподілу є найпоширенішим із законів розподілу, які зустрічаються у випадкових явищах. Розглянемо лише деякі приклади застосування центральної граничної теореми.

Зокрема, в більшості випадків похибки, які виникають під час вимірювання фізичних величин, розподілені саме за нормальним законом, оскільки виникають як результат багатьох незалежних елементарних помилок, породжених різними причинами: станом приладу, атмосферними умовами, фізичним і психічним станом дослідника.

Якщо випадкова величина X – відхилення курсу гривні від фіксованого за 1 рік, то $X = \sum_{k=1}^{365} X_k$, де X_k – відхилення курсу за k -ий день року. Тому, очевидно, випадкова величина X матиме розподіл, близький до нормального.

Лекція 14

Предмет математичної статистики

Часто під час вивчення випадкових явищ не вдається перерахувати випадки, на які розпадається подія A , тому неможливо точно обчислити ймовірність цієї події користуючись класичним означенням імовірності.

Нехай, наприклад, нас цікавить імовірність того, що з навмання взятої ділянки землі буде зібрано урожай пшениці в межах від M до N центнерів з 1 га. Очевидно, ця подія є випадковою, однак, не знаючи характеристик вибраної ділянки землі, неможливо визначити ймовірність цієї події до випробування навіть наближено.

Очевидно, для наближеного визначення ймовірності події, що вивчається, необхідно здійснити серію випробувань і проаналізувавши результати цих випробувань, взяти за наближене значення ймовірності події A відносну частоту цієї події. В таких дослідженнях виникають питання: скільки треба провести випробувань? Як обробити результати спостережень і зробити обґрунтовані висновки?

Крім наближеного обчислення ймовірності, дослідника можуть цікавити й інші питання, наприклад, залежність урожайності від кількості внесених добрив і якості обробки ґрунту. Для з'ясування цієї залежності необхідно зібрати дані про урожайність, кількість внесених добрив і якість обробки ґрунту, певним чином систематизувати і проаналізувати їх.

Математична статистика – це розділ математики, в якому вивчаються методи збору, систематизації і аналізу результатів спостережень масових випадкових явищ з метою виявлення існуючих закономірностей.

З'ясування закономірностей, яким підпорядковані масові випадкові явища, здійснюється за допомогою методів теорії ймовірностей.

Аніліз статистичних даних може здійснюватися з метою:

а) оцінки невідомої ймовірності події; оцінки невідомої функції розподілу; оцінки невідомих параметрів розподілу, загальний вигляд якого відомий; оцінки залежності випадкової величини від одної або кількох випадкових величин;

б) перевірки статистичних гіпотез про вигляд невідомого розподілу або про величину параметрів розподілу, вигляд якого відомий.

Генеральна сукупність. Вибірка

Вихідними поняттями математичної статистики є поняття *генеральної сукупності* і *вибірки*. Під *генеральною сукупністю* розуміють множину всіх реально існуючих або навіть тільки умовно можливих однорідних об'єктів, які вивчають під кутом зору їхнього розподілу за деякою ознакою. Наприклад, це можуть бути множини людей за віком, множини тварин певного виду за вагою, множини орних земель за врожайністю, множини виробів певного найменування за якістю, множини акціонерних банків України за прибутком

і т. д.

Оскільки практично будь-яка ознака генеральної сукупності допускає кількісну оцінку, то замість того, щоб говорити про розподіл одиниць сукупності за ознакою, можна говорити про розподіл деякої випадкової величини X . Експеримент, з яким пов'язана випадкова величина X , полягає у виборі одного представника даної сукупності, а значення x , яке приймає X , є значенням ознаки для цього представника.

Отже, з теоретико-ймовірнісного погляду *генеральна сукупність – це випадкова величина $X(\omega)$, задана на просторі елементарних подій Ω* .

Зрозуміло, що повний опис закону розподілу випадкової величини X можна отримати, з'ясувавши значення ознаки для *всіх без винятку* представників даної сукупності. В окремих ситуаціях так і роблять: наприклад, дані про розподіл жителів тієї чи іншої країни щодо статі, віку, освіти і т. д. отримують у результаті загальних переписів населення, які проводяться один раз на кілька десятиліть. Однак такий спосіб суцільного обстеження всієї досліджуваної сукупності пов'язаний із низкою труднощів. Одна з них – це *великий обсяг* сукупності. У деяких випадках є ще й *трудність* принципового характеру, яка полягає в тому, що сукупність, яку ми розглядаємо, не існує в готовому вигляді, а є лише визначеною в уяві. Наприклад, якщо нас цікавить розподіл похибки, яку допускає вимірювальний прилад, то досліджувана сукупність становитиме *перелік усіх можливих вимірювань*, які можна здійснити за допомогою даного приладу. Зрозуміло, що обстежити всі елементи такої сукупності неможливо. В такому випадку кажуть, що генеральна сукупність є нескінченною.

Щоб подолати або обійти вказані труднощі, найчастіше чинять так: обстеження всієї сукупності замінюють обстеженням невеликої (до того ж вибраної навмання) її частини. Таку частину генеральної сукупності називають *вибіркою*.

Із теоретико-ймовірнісного погляду, *вибірка з даної генеральної сукупності – це результати обмеженого ряду спостережень x_1, x_2, \dots, x_n випадкової величини X* .

Число n , яке відповідає кількості спостережень, що утворюють вибірку, називають *обсягом (або об'ємом) вибірки*, а числа x_1, x_2, \dots, x_n – *елементами або варіантами вибірки*.

У статистиці інтерпретація вибірки та її окремих елементів допускає, залежно від контексту, два різних варіанти.

У першому (практичному) варіанті інтерпретації вибірки під x_1, x_2, \dots, x_n розуміють фактично спостережувані в даному конкретному n -кратному експерименті значення досліджуваної випадкової величини X , тобто конкретні числа.

Згідно з другим (теоретичним) варіантом інтерпретації під вибіркою x_1, x_2, \dots, x_n розуміють послідовність випадкових величин, i -тий член якої x_i лише означає результат спостереження, який ми могли б отримати на i -му

кроці n -кратного експерименту, пов'язаного зі спостереженням досліджуваної випадкової величини X .

Якщо умови експерименту не змінюються від спостереження до спостереження і якщо n -кратний експеримент організований у такий спосіб, що результати спостереження на кожному (i -му) кроці ніяк не залежать від попередніх і не впливають на майбутні результати спостережень, то очевидно, що ймовірнісні закономірності поведінки i -го спостереження теоретичної вибірки залишаються одними і тими ж для всіх $i=1,2,\dots,n$ і цілком визначаються законом розподілу ймовірностей спостережуваної випадкової величини, тобто

$$P\{x_i < x\} = P\{X < x\} = F(x).$$

При цьому із взаємної незалежності спостережень вибірки впливає незалежність випадкових величин x_1, x_2, \dots, x_n .

Якщо в межах теоретичного варіанту інтерпретації вибірки ряд спостережень x_1, x_2, \dots, x_n утворює послідовність незалежних і однаково розподілених випадкових величин, то вибірка називається випадковою.

Надалі, використовуючи теоретичний варіант інтерпретації вибірки, завжди будемо вважати, що ця вибірка є випадковою.

Статистичний розподіл вибірки

Припустимо, що вивчають деяку генеральну випадкову величину X . Для цього проводять низку незалежних дослідів або спостережень, у кожному з яких величина X набуває того чи іншого значення. Сукупність отриманих значень x_1, x_2, \dots, x_n величини X (де n – кількість дослідів) і є утворена нами вибірка. Цю сукупність часто називають *статистичним рядом*; він відіграє роль вихідного числового матеріалу, що підлягає подальшій обробці та аналізу.

Перший етап обробки статистичного ряду – побудова так званого *простого варіаційного ряду*. Його отримують з елементів наявної вибірки, розмістивши x_i ($i=1,2,\dots,n$) у порядку зростання (неспадання) їхніх значень. Позначимо члени нового ряду через $x_{(i)}$, щоб відрізнити його від x_i . Тоді простий варіаційний ряд буде поданий як неспадна послідовність:

$$x_{(1)}, x_{(2)}, \dots, x_{(n)},$$

де $x_{(1)} \leq x_{(2)} \leq \dots \leq x_{(n)}$.

Наступний етап обробки вихідного статистичного ряду – побудова статистичного (емпіричного) закону розподілу. Форма його запису залежить від характеру досліджуваної випадкової величини X .

Нехай X – дискретна випадкова величина. Тоді найбільш природна форма статистичного закону розподілу вибірки описується за допомогою так званого *згрупованого варіаційного ряду*. Його отримують у такий спосіб: серед чисел простого варіаційного ряду відбирають усі різні і розміщують їх у порядку зростання:

$$x_1, x_2, \dots, x_k,$$

де $x_1 < x_2 < \dots < x_k$ ($k \leq n$). При цьому для виділених варіант x_i ($i = 1, 2, \dots, k$) одночасно обчислюють частоти n_i , що їм відповідають або відповідні відносні частоти w_i ; частота n_i дорівнює кількості спостережень, в яких випадкова величина X набула значення x_i , а відносна частота $w_i = \frac{n_i}{n}$ ($i = 1, 2, \dots, k$).

Очевидно, що $\sum_{i=1}^k n_i = n$, $\sum_{i=1}^k w_i = 1$.

Означення. *Дискретним статистичним розподілом вибірки називається відповідність між варіантами та їхніми частотами або відносними частотами.*

Дискретний статистичний розподіл вибірки можна подати у формі таблиць:

- *дискретний статистичний розподіл частот:*

x_i	x_1	x_2	...	x_k
n_i	n_1	n_2	...	n_k

- *дискретний статистичний розподіл відносних частот:*

x_i	x_1	x_2	...	x_k
w_i	w_1	w_2	...	w_k

Розглянемо тепер випадок, коли випадкова величина X – неперервна. Характерною рисою неперервного розподілу є, як ми знаємо, той факт, що ймовірність кожного окремого значення дорівнює нулю. Отже, у вихідному статистичному ряді, як правило, не буде повторів, і тому дискретний розподіл виявиться малопридатним для подальшого аналізу. У такому разі (а також у випадку, коли випадкова величина дискретна і обсяг вибірки $n \geq 30$) статистичний закон розподілу вибірки записують як інтервальний варіаційний ряд – частот або відносних частот. Для цього весь діапазон зміни ознаки від найменшого (x_{\min}) до найбільшого (x_{\max}) значення розбивають на певну кількість проміжків (частіше однакової довжини) $[z_0, z_1), [z_1, z_2), \dots, [z_{m-1}, z_m)$ і збирають в окремі групи елементи сукупності, для яких величина X набуває значень у кожному з проміжків.

Для визначення приблизної кількості інтервалів можна використати таку таблицю:

Обсяг вибірки	Кількість інтервалів
30-60	6-8
60-100	7-10
100-200	9-12
200-500	11-17

500-1500	16-25
----------	-------

Межі проміжків або беруть „круглими” числами, або визначають з точністю до $\alpha/2$, де α – точність, з якою визначені значення вибірки. Це означає, що початок першого проміжка z_0 і кінець останнього z_m не обов’язково збігаються відповідно з x_{\min} і x_{\max} . Вимагається лише, щоб $z_0 \leq x_{\min}$ і $z_m \geq x_{\max}$. Якщо $z_m = x_{\max}$, то будемо вважати, що в інтервальному варіаційному ряді останній проміжок є відрізком.

Частота n_i події $X \in [z_{i-1}, z_i)$ обчислюється як кількість випробувань, в яких значення випадкової величини X потрапило в i -ий проміжок, а відносна частота цієї події $w_i = \frac{n_i}{n}$ ($i = 1, 2, \dots, k$).

Означення. *Інтервальним статистичним розподілом вибірки називається відповідність між інтервалами варіаційного ряду та їхніми частотами або відносними частотами.*

Інтервальний статистичний розподіл вибірки, як і дискретний, записують у формі таблиць:

- *інтервальний статистичний розподіл частот:*

$[z_{i-1}, z_i)$	$[z_0, z_1)$	$[z_1, z_2)$...	$[z_{m-1}, z_m)$
n_i	n_1	n_2	...	n_m

$$\sum_{i=1}^m n_i = n.$$

- *інтервальний статистичний розподіл відносних частот:*

$[z_{i-1}, z_i)$	$[z_0, z_1)$	$[z_1, z_2)$...	$[z_{m-1}, z_m)$
w_i	w_1	w_2	...	w_m

$$\sum_{i=1}^m w_i = 1.$$

Інтервальний статистичний розподіл вибірки за необхідності можна замінити дискретним. Для цього досить частинні інтервали $[z_{i-1}, z_i)$ замінити числами – серединами цих інтервалів (тобто прийняти $x_i = \frac{z_{i-1} + z_i}{2}$), а відповідні значення частот (відносних частот) залишити без змін.

Полігон та гістограма частот

Статистичний розподіл вибірки можна задати графічно полігоном або гістограмою частот (відносних частот).

Полігон розподілу вибірки використовується для зображення як дискретних, так і інтервальних варіаційних рядів, а гістограма – лише для

інтервальних рядів.

Означення. *Полігоном частот* називають ламану, відрізки якої послідовно з'єднують точки $(x_1; n_1)$, $(x_2; n_2)$, ..., $(x_k; n_k)$ координатної площини.

Щоб побудувати полігон частот, на осі абсцис відкладають варіанти x_i , а на осі ординат – відповідні частоти n_i . Далі, точки $(x_i; n_i)$ з'єднують відрізками прямих і отримують полігон частот.

Означення. *Полігоном відносних частот* називають ламану, відрізки якої послідовно з'єднують точки $(x_1; w_1)$, $(x_2; w_2)$, ..., $(x_k; w_k)$ координатної площини.

Щоб побудувати полігон частот, на осі абсцис відкладають варіанти x_i , а на осі ординат – відповідні відносні частоти w_i . Далі, точки $(x_i; w_i)$ з'єднують відрізками прямих і отримують полігон відносних частот.

Означення. *Гістограмою частот* називається східчаста фігура, яка складена з прямокутників, основами яких є частинні інтервали $[z_{i-1}, z_i)$,

$i = 1, 2, \dots, t$, а їхні висоти $h_i = \frac{n_i}{z_i - z_{i-1}}$.

Щоб побудувати гістограму частот, на осі абсцис відкладають частинні інтервали $[z_{i-1}, z_i)$ і на них, як на основі, будують прямокутники з висотами h_i . Площа кожного такого прямокутника дорівнює n_i .

Означення. *Гістограмою відносних частот* називається східчаста фігура, яка складена з прямокутників, основами яких є частинні інтервали

$[z_{i-1}, z_i)$, $i = 1, 2, \dots, t$, а їхні висоти $\tilde{h}_i = \frac{w_i}{z_i - z_{i-1}}$.

Щоб побудувати гістограму частот, на осі абсцис відкладають частинні інтервали $[z_{i-1}, z_i)$ і на них, як на основі, будують прямокутники з висотами \tilde{h}_i . Площа кожного такого прямокутника дорівнює w_i .

Емпірична функція розподілу

Нагадаємо, що функція розподілу $F(x)$ випадкової величини X визначається рівністю

$$F(x) = P\{X < x\}.$$

Її називають ще теоретичною функцією розподілу випадкової величини X або функцією розподілу генеральної сукупності.

Зафіксуємо довільне дійсне число x і позначимо через n_x кількість елементів вибірки, що менші за x . Якщо n – обсяг вибірки, то $\frac{n_x}{n}$ – відносна частота події $\{X < x\}$.

Означення. *Емпіричною функцією розподілу* випадкової величини X (функцією розподілу вибірки) називають функцію $F_n(x)$, що визначає для будь-

якого дійсного числа x відносно частоту події $\{X < x\}$, тобто

$$F_n(x) = \frac{n_x}{n}. \quad (1)$$

Коли вихідні статистичні дані згруповані в дискретний варіаційний ряд, то емпірична функція розподілу записується у вигляді:

$$F_n(x) = \begin{cases} 0, & x \leq x_1; \\ \frac{n_1}{n}, & x_1 < x \leq x_2; \\ \frac{n_1 + n_2}{n}, & x_2 < x \leq x_3; \\ \dots\dots\dots \\ \frac{n_1 + n_2 + \dots + n_{k-1}}{n}, & x_{k-1} < x \leq x_k; \\ 1, & x > x_k. \end{cases}$$

Якщо ми маємо у своєму розпорядженні згрупований інтервальний статистичний ряд, то можемо лише наближено побудувати емпіричну функцію розподілу $F_n(x)$. За значення x , для яких $F_n(x)$ обчислюється за формулою (1), природно взяти межі частинних проміжків. У всіх інших точках $F_n(x)$ визначається за допомогою лінійного інтерполювання. Це означає, що геометрично $F_n(x)$ буде зображена неперервною ламаною лінією, яка з'єднає послідовно точки (z_i, ω_i) , де

$$\omega_0 = 0, \quad \omega_i = \frac{n_1 + n_2 + \dots + n_i}{n}, \quad i = 1, 2, \dots, m.$$

Зв'язок між функціями $F(x)$ і $F_n(x)$ встановлює така теорема.

Теорема Глівенка. Для будь-якого дійсного числа x за необмеженого зростання обсягу вибірки n емпірична функція розподілу $F_n(x)$ збігається за ймовірністю до теоретичної функції розподілу $F(x)$, тобто

$$\forall \varepsilon > 0 \quad \forall x \in \mathbb{R} \quad \lim_{n \rightarrow \infty} P\{|F_n(x) - F(x)| < \varepsilon\} = 1.$$

Доведення. Зафіксуємо довільне дійсне число x і розглянемо послідовність n незалежних експериментів, результатами яких будуть значення x_1, x_2, \dots, x_n випадкової величини X . Якщо подія $A = \{X < x\}$, то в кожному з експериментів $P(A) = P\{X < x\} = F(x)$, а $F_n(x) = \frac{n_x}{n}$ – це відносна частота появ події A в n випробуваннях. Згідно з теоремою Бернуллі про стійкість відносних частот

$$\forall \varepsilon > 0 \quad \lim_{n \rightarrow \infty} P\left\{\left|\frac{n_x}{n} - F(x)\right| < \varepsilon\right\} = \lim_{n \rightarrow \infty} P\{|F_n(x) - F(x)| < \varepsilon\} = 1,$$

що й треба було довести.

Лекція 15

Числові характеристики статистичного розподілу вибірки

У практичних задачах часто замість повного вивчення даних вибірки буває достатньо обмежитися знаходженням їхніх числових характеристик. Далі наведемо основні з них, припускаючи, що статистичні дані згруповано в дискретний варіаційний ряд.

Означення. *Вибірковим середнім \bar{x} статистичного розподілу вибірки називається середнє арифметичне значення її варіант x_i з урахуванням їхніх частот, тобто:*

$$\bar{x} = \frac{1}{n} \sum_{i=1}^k n_i x_i = \frac{n_1 x_1 + n_2 x_2 + \dots + n_k x_k}{n}. \quad (1)$$

Вибіркове середнє \bar{x} є основною характеристикою статистичного розподілу вибірки. Його узагальненням є поняття початкового емпіричного моменту.

Означення. *Початковим емпіричним моментом s -го порядку M_s статистичного розподілу вибірки називається середнє арифметичне значення степенів порядку s варіант x_i , тобто:*

$$M_s = \frac{1}{n} \sum_{i=1}^k n_i x_i^s. \quad (2)$$

Якщо $s=1$, то $M_1 = \bar{x}$ – вибірковому середньому.

Переходимо до означення основних характеристик розсіювання значень випадкової величини навколо її середнього значення, які також розраховуються на основі вибірки.

Означення. *Розмахом вибірки R називають різницю між найбільшим і найменшим значеннями її варіант, тобто:*

$$R = x_k - x_1.$$

Означення. *Вибірковою дисперсією D_B статистичного розподілу вибірки називають середнє арифметичне значення квадратів відхилень варіант x_i від вибіркового середнього \bar{x} , тобто:*

$$D_B = \frac{1}{n} \sum_{i=1}^k (x_i - \bar{x})^2 n_i. \quad (3)$$

Для обчислення вибіркової дисперсії часто зручніше використовувати іншу формулу:

$$D_B = \frac{1}{n} \sum_{i=1}^k n_i x_i^2 - (\bar{x})^2. \quad (3')$$

Розмірність дисперсії дорівнює квадрату розмірності значень випадкової величини, що створює незручність у дослідженнях. Щоб її усунути, за характеристику розсіювання значень випадкової величини за результатами вибірки приймають *вибіркове середнє квадратичне відхилення σ_B* , яке визначається рівністю:

$$\sigma_B = \sqrt{D_B} = \sqrt{\frac{1}{n} \sum_{i=1}^k (x_i - \bar{x})^2 n_i}, \quad (4)$$

або

$$\sigma_B = \sqrt{D_B} = \sqrt{\frac{1}{n} \sum_{i=1}^k n_i x_i^2 - (\bar{x})^2}. \quad (4')$$

Означення. *Коефіцієнтом варіації V статистичного розподілу вибірки називається відношення вибіркового середнього квадратичного відхилення до вибіркового середнього, тобто:*

$$V = \frac{\sigma_B}{\bar{x}} \cdot 100\%.$$

Означення. *Центральним емпіричним моментом s -го порядку m_s статистичного розподілу вибірки називається середнє арифметичне значення степенів порядку s відхилень варіант x_i від середнього вибіркового значення \bar{x} , тобто:*

$$m_s = \frac{1}{n} \sum_{i=1}^k (x_i - \bar{x})^s n_i. \quad (5)$$

Зокрема, $m_1 = 0$, $m_2 = D_B$.

Для оцінки відхилення статистичного розподілу вибірки від нормального розподілу використовують числові характеристики – асиметрію та ексцес.

Означення. *Асиметрією (коефіцієнтом асиметрії) A_B називають число, яке обчислюється за формулою:*

$$A_B = \frac{m_3}{\sigma_B^3},$$

де m_3 – центральний емпіричний момент 3-го порядку, σ_B – середнє квадратичне відхилення статистичного розподілу вибірки.

Означення. *Ексцесом E_B статистичного розподілу вибірки називається число, яке обчислюється за формулою:*

$$E_B = \frac{m_4}{\sigma_B^4} - 3,$$

де m_4 – центральний емпіричний момент 4-го порядку, σ_B – середнє квадратичне відхилення статистичного розподілу вибірки.

Якщо випадкова величина X розподілена за нормальним законом, то її асиметрія та ексцес дорівнюють нулю. Тому, що більше віддалені від нуля асиметрія та ексцес, то менше підстав сподіватися, що вибірка, з якої утворено варіаційний ряд, одержана з нормально розподіленої генеральної сукупності.

Співвідношення (1)-(5) можуть бути також використані при обчисленні відповідних числових характеристик вибірки для випадків, коли емпіричні дані записані як вихідний статистичний ряд або згруповані за допомогою інтервального варіаційного ряду. Зокрема, у першому з них у згаданих співвідношеннях слід покласти $k = n$ та $n_i = 1$. У другому випадку ці формули

залишаються без змін, якщо вважати, що в них x_i – середини частинних проміжків $[z_{i-1}, z_i)$, $i = 1, 2, \dots, m$.

Точкові оцінки параметрів розподілу (незміщені, слушні, ефективні)

Відомо, що повною характеристикою випадкової величини – ознаки генеральної сукупності є її закон розподілу. Для його встановлення дослідними методами потрібні певні затрати ресурсів і часу. Однак часто є підстави вважати, що той чи інший закон розподілу є відомий (наприклад: пуассонівський, нормальний, показниковий тощо). Але щоб конкретизувати цей закон, потрібно знати його параметри, які називають *параметрами розподілу*. Зокрема для нормального закону розподілу параметрами є a і σ , для пуассонівського – λ , для показникового – α і т. д.

Отже, вивчаючи певну ознаку X генеральної сукупності, ми можемо знати характер закону розподілу випадкової величини X , але параметри цього закону залишаються невідомими. Тоді виникає *задача: на основі одержаної вибірки з генеральної сукупності визначити наближені числові значення невідомих параметрів розподілу*. Такі наближені числові значення параметрів розподілу називають їхніми *точковими статистичними оцінками*, або скорочено – *точковими оцінками*.

Нехай ми вивчаємо випадкову величину X , закон розподілу якої відомий, але містить невідомий параметр θ . Потрібно знайти точкову статистичну оцінку параметра θ за результатами n незалежних випробувань, у кожному з яких випадкова величина X набуває значень x_1, x_2, \dots, x_n (вибірка обсягу n).

Означення. Будь-яку однозначну функцію $\theta_n^* = \theta_n^*(x_1, x_2, \dots, x_n)$, за допомогою якої знаходять наближене значення параметра θ розподілу випадкової величини, називають **точковою оцінкою** цього параметра.

Для того, щоб оцінка θ_n^* була в певному сенсі „найкращою” для параметра θ , тобто мала практичну цінність, вона мусить задовольняти певні умови.

Здійснивши вибірку обсягу n і знайшовши за нею точкову оцінку параметра θ , отримаємо наближене значення $\theta_{n(1)}^*$. Оцінку, отриману за ще одною вибіркою такого самого обсягу позначимо через $\theta_{n(2)}^*$. Отримані після k таких експериментів, взагалі кажучи, різні числа $\theta_{n(1)}^*, \theta_{n(2)}^*, \dots, \theta_{n(k)}^*$ будуть значеннями випадкової величини θ_n^* . Отже, точкова оцінка θ_n^* сама є випадковою величиною.

Якщо оцінка θ_n^* дає значення оцінюваного параметра з надлишком, то $\theta_{n(i)}^* > \theta$ для всіх $i = \overline{1, k}$. Якщо ж оцінка θ_n^* дає значення оцінюваного параметра з недостатчею, то $\theta_{n(i)}^* < \theta$ для всіх $i = \overline{1, k}$. В обидвох випадках, оцінюючи параметр θ , ми робимо систематичні помилки (помилки одного знаку). Якщо

накладемо вимогу, щоб $E(\theta_n^*) = \theta$, то доб'ємося того, що оцінка θ_n^* буде „застрахована” від накопичення систематичних помилок.

Означення. Точкова оцінка $\theta_n^* = \theta_n^*(x_1, x_2, \dots, x_n)$ параметра розподілу θ випадкової величини X називається **незмщеною**, якщо її математичне сподівання дорівнює точному значенню цього параметра.

Але навіть використовуючи незмщену оцінку, дисперсія якої значна, ми не застраховані від великих помилок при знаходженні наближеного значення параметра θ . Якщо ж вимагати, щоб дисперсія $D(\theta_n^*)$ була малою, то можливість допустити велику помилку буде виключена.

Означення. Незмщена оцінка $\theta_n^* = \theta_n^*(x_1, x_2, \dots, x_n)$ називається **ефективною**, якщо вона має найменшу дисперсію серед усіх незмщених оцінок параметра θ , обчислених за вибірками одного і того ж обсягу.

Означення. Точкова оцінка $\theta_n^* = \theta_n^*(x_1, x_2, \dots, x_n)$ параметра розподілу θ називається **слухною (або конзистентною, або змістовною)**, якщо θ_n^* збігається за ймовірністю до оцінюваного параметра при необмеженому зростанні обсягу вибірки, тобто виконується така рівність:

$$\forall \varepsilon > 0 \quad \lim_{n \rightarrow \infty} P\{|\theta_n^* - \theta| < \varepsilon\} = 1,$$

де $\varepsilon > 0$ як завгодно мале число.

Теорема (умова слухності незмщеної оцінки). Якщо дисперсія незмщеної оцінки при необмеженому зростанні обсягу вибірки прямує до нуля, то така оцінка є слухною, тобто

$$E(\theta_n^*) = \theta \quad \forall n, \quad \lim_{n \rightarrow \infty} D(\theta_n^*) = 0 \quad \Rightarrow \quad \forall \varepsilon > 0 \quad \lim_{n \rightarrow \infty} P\{|\theta_n^* - \theta| < \varepsilon\} = 1.$$

Доведення. Запишемо нерівність Чебишова для випадкової величини θ_n^* :

$$\forall \varepsilon > 0 \quad P\{|\theta_n^* - E(\theta_n^*)| \geq \varepsilon\} \leq \frac{D(\theta_n^*)}{\varepsilon^2} \xrightarrow{n \rightarrow \infty} 0.$$

Отже, $\forall \varepsilon > 0 \quad \lim_{n \rightarrow \infty} P\{|\theta_n^* - \theta| \geq \varepsilon\} = 0$, тобто $\forall \varepsilon > 0 \quad \lim_{n \rightarrow \infty} P\{|\theta_n^* - \theta| < \varepsilon\} = 1$, що й потрібно було довести. \square

У типових законах розподілу ймовірностей випадкової величини параметри розподілу виражаються через числові характеристики цієї величини (математичне сподівання, дисперсію, середнє квадратичне відхилення). Наприклад, якщо випадкова величина X нормально розподілена, то параметри розподілу $a = E(X)$, $\sigma = \sqrt{D(X)}$. Тому, перш за все нас цікавить задача про оцінки основних числових характеристик випадкової величини X .

Точкова оцінка математичного сподівання

Нехай x_1, x_2, \dots, x_n – вибірка, отримана в результаті n незалежних випробувань над випадковою величиною X – деякою ознакою генеральної сукупності, яка має математичне сподівання $E(X) = a$.

За точкову оцінку математичного сподівання $a = E(X)$ беруть вибіркове середнє:

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

Покажемо, що оцінка $a_n^* = \bar{x}$ є незміщеною для $E(X) = a$. Справді, нехай x_1, x_2, \dots, x_n – результати n незалежних спостережень над випадковою величиною X . За умовою $E(X) = a$ і оскільки x_1, x_2, \dots, x_n є незалежними випадковими величинами і мають однаковий закон розподілу, то $E(x_i) = a$ ($i = \overline{1, n}$). Використовуючи властивості математичного сподівання, маємо:

$$E(\bar{x}) = E\left(\frac{1}{n} \sum_{i=1}^n x_i\right) = \frac{1}{n} \sum_{i=1}^n E(x_i) = \frac{1}{n} \cdot na = a,$$

тобто $E(\bar{x}) = a$, що й потрібно було довести.

Припустимо додатково, що випадкова величина X має скінченну дисперсію $D(X) = \sigma^2$. Тоді можна стверджувати, що оцінка $a_n^* = \bar{x}$ є слушною. Для доведення обчислимо дисперсію вибіркового середнього \bar{x} . Використовуючи властивості дисперсії, знаходимо:

$$D(\bar{x}) = D\left(\frac{1}{n} \sum_{i=1}^n x_i\right) = \frac{1}{n^2} \sum_{i=1}^n D(x_i) = \frac{1}{n^2} \cdot n\sigma^2 = \frac{\sigma^2}{n}.$$

Оскільки $\lim_{n \rightarrow \infty} D(\bar{x}) = \lim_{n \rightarrow \infty} \frac{\sigma^2}{n} = 0$, то це означає, що оцінка \bar{x} є слушною для параметра a .

Прийmemo без доведення важливе для практики **твердження**: якщо випадкова величина X нормально розподілена з параметрами $E(X) = a$ і $D(X) = \sigma^2$, то оцінка \bar{x} має у класі всіх незміщених оцінок математичного сподівання a мінімальну дисперсію, яка дорівнює $\frac{\sigma^2}{n}$. Тому \bar{x} є ефективною оцінкою параметра a .

Точкова оцінка дисперсії. Підправлена дисперсія

Якщо випадкова вибірка складається з результатів n незалежних випробувань x_1, x_2, \dots, x_n над випадковою величиною X із математичним сподіванням $E(X) = a$ і дисперсією $D(X) = \sigma^2$, то за точкову оцінку дисперсії беруть вибірккову дисперсію

$$D_B = \frac{1}{n} \sum_{k=1}^n (x_k - \bar{x})^2,$$

яка є зміщеною оцінкою параметра $D(X) = \sigma^2$, або підправлену вибірккову дисперсію

$$s^2 = \frac{1}{n-1} \sum_{k=1}^n (x_k - \bar{x})^2,$$

яка є незміщеною оцінкою параметра $D(X) = \sigma^2$.

Той факт, що D_B є зміщеною оцінкою для $D(X)$, випливає з рівності

$$E(D_B) = \frac{n-1}{n} \sigma^2,$$

яку неважко встановити за допомогою безпосередніх обчислень.

Введемо позначення: $y_k = x_k - a$ ($k = \overline{1, n}$). Тоді

$$x_k - \bar{x} = x_k - a - (\bar{x} - a) = y_k - \frac{1}{n} \sum_{k=1}^n (x_k - a) = y_k - \frac{1}{n} \sum_{k=1}^n y_k = y_k - \bar{y};$$

$$E(x_k) = a, \quad E(y_k) = E(x_k - a) = E(x_k) - a = 0, \quad E(y_k^2) = E(x_k - a)^2 = D(x_k) = \sigma^2,$$

$$D(y_k) = D(x_k) = \sigma^2; \quad E(D_B) = E\left(\frac{1}{n} \sum_{k=1}^n (x_k - \bar{x})^2\right) = E\left(\frac{1}{n} \sum_{k=1}^n (y_k - \bar{y})^2\right) =$$

$$= E\left(\frac{1}{n} \sum_{k=1}^n y_k^2 - \bar{y}^2\right) = \frac{1}{n} \sum_{k=1}^n E(y_k^2) - E(\bar{y}^2) = \frac{n\sigma^2}{n} - E(\bar{y}^2) = \sigma^2 - E(\bar{y}^2);$$

$$E(\bar{y}^2) = E\left(\frac{1}{n} \sum_{k=1}^n y_k\right)^2 = \frac{1}{n^2} E\left(\sum_{k=1}^n y_k^2 + 2 \sum_{k < l} y_k y_l\right) = \frac{n\sigma^2}{n^2} + 2 \sum_{k < l} E(y_k y_l) = \frac{\sigma^2}{n},$$

оскільки $E(y_k y_l) = E(y_k)E(y_l) = 0$.

Отже,

$$E(D_B) = \sigma^2 - \frac{\sigma^2}{n} = \frac{n-1}{n} \sigma^2.$$

Далі, враховуючи співвідношення $s^2 = \frac{n}{n-1} D_B$, одержимо:

$$E(s^2) = \frac{n}{n-1} E(D_B) = \frac{n}{n-1} \cdot \frac{n-1}{n} \sigma^2 = \sigma^2 = D(X),$$

тобто підправлена дисперсія s^2 є незміщеною оцінкою для дисперсії $D(X) = \sigma^2$.

Дріб $\frac{n}{n-1}$ називають *поправкою Бесселя*. Для малих n поправка Бесселя значно відрізняється від одиниці. Для $n > 50$ практично немає різниці між D_B і s^2 .

Можна показати, що оцінки D_B і s^2 є слушними і не є ефективними.

У випадку, коли математичне сподівання a відоме і випадкова величина X нормально розподілена, то незміщеною, слушною та ефективною оцінкою дисперсії $D(X) = \sigma^2$ є оцінка

$$\frac{1}{n} \sum_{k=1}^n (x_k - a)^2.$$

Лекція 16

Вибіркове середнє нормально розподіленої ознаки генеральної сукупності

У цьому параграфі ми покажемо, що вибіркове середнє нормально розподіленої ознаки генеральної сукупності, знайдене за вибіркою обсягу n , є нормально розподіленою випадковою величиною.

Лема 1. Якщо X – нормально розподілена випадкова величина, то випадкова величина $Y = cX$, де $c = \text{const} > 0$, є також нормально розподіленою.

Доведення. Нехай щільність розподілу ймовірностей випадкової величини X задається функцією:

$$p_X(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-a)^2}{2\sigma^2}}, \quad -\infty < x < \infty.$$

Визначимо функцію розподілу і щільність розподілу ймовірностей випадкової величини Y :

$$F_Y(x) = P\{Y < x\} = P\{cX < x\} = P\left\{X < \frac{x}{c}\right\} = \int_{-\infty}^{x/c} p_X(u) du.$$

В одержаному інтегралі зробимо заміну: $u = \frac{y}{c}$, $du = \frac{dy}{c}$; $u = \frac{x}{c} \Rightarrow y = x$.

Тоді

$$F_Y(x) = \frac{1}{c} \int_{-\infty}^x p_X\left(\frac{y}{c}\right) dy;$$

$$p_Y(x) = \frac{dF_Y(x)}{dx} = \frac{1}{c} p_X\left(\frac{x}{c}\right) = \frac{1}{c} \cdot \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{\left(\frac{x}{c}-a\right)^2}{2\sigma^2}} = \frac{1}{c\sigma\sqrt{2\pi}} e^{-\frac{(x-ca)^2}{2(c\sigma)^2}},$$

а це означає, що Y – нормально розподілена випадкова величина з параметрами $E(Y) = ca$, $D(Y) = (c\sigma)^2$.

Лема 2. Якщо X_1, X_2 – незалежні нормально розподілені випадкові величини, то випадкова величина $Z = X_1 + X_2$ є також нормально розподіленою.

Доведення. Використаємо твердження, відоме в теорії ймовірностей як формула згортки: якщо випадкові величини X, Y – незалежні, то щільність розподілу ймовірностей випадкової величини $Z = X + Y$ визначається співвідношенням:

$$p(z) = \int_{-\infty}^{\infty} p_X(x) p_Y(z-x) dx,$$

де $p_X(x), p_Y(x)$ – щільності розподілу ймовірностей випадкових величин X і Y .

Нехай $E(X_1) = a_1$, $E(X_2) = a_2$, $D(X_1) = \sigma_1^2$, $D(X_2) = \sigma_2^2$. Тоді $E(Z) = a_1 + a_2$, $D(Z) = \sigma_1^2 + \sigma_2^2$, і можна показати, що

$$\begin{aligned}
 p(z) &= \int_{-\infty}^{\infty} p_{X_1}(x) p_{X_2}(z-x) dx = \frac{1}{\sigma_1 \sigma_1 \cdot 2\pi} \int_{-\infty}^{\infty} e^{-\frac{(x-a_1)^2}{2\sigma_1^2}} \cdot e^{-\frac{(z-x-a_2)^2}{2\sigma_2^2}} dx = \\
 &= \frac{1}{\sqrt{\sigma_1^2 + \sigma_2^2} \cdot \sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-\frac{(z-a_1-a_2)^2}{2(\sigma_1^2 + \sigma_2^2)}} dx,
 \end{aligned}$$

а це означає, що Z – нормально розподілена випадкова величина з параметрами $E(Z) = a_1 + a_2$, $D(Z) = \sigma_1^2 + \sigma_2^2$.

Теорема. Якщо X – нормально розподілена ознака генеральної сукупності, \bar{x} – вибіркове середнє, знайдене за вибіркою обсягу n з цієї генеральної сукупності, то \bar{x} – нормально розподілена випадкова величина.

Доведення. Нехай x_1, x_2, \dots, x_n – вибірка, отримана в результаті n незалежних випробувань над випадковою величиною X – нормально розподіленою ознакою генеральної сукупності, тоді випадкові величини x_1, x_2, \dots, x_n – незалежні. Спираючись на твердження леми 2, за допомогою методу математичної індукції неважко переконатися, що випадкова величина

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

є нормально розподіленою. Далі, використавши твердження леми 1, отримаємо, що вибіркове середнє

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

є нормально розподіленою випадковою величиною. \square

Поняття про інтервальні оцінки параметрів розподілу

Точкова оцінка θ^* параметра θ тим точніша, чим менша величина різниці $|\theta - \theta^*|$. Якщо би вдалося встановити, що $|\theta - \theta^*| < \delta$, то число $\delta > 0$ характеризувало б точність точкової оцінки θ^* параметра θ . Однак статистичні методи не дозволяють категорично стверджувати, що $|\theta - \theta^*| < \delta$, бо θ^* є випадкова величина. Можна лише казати про ймовірність γ , з якою ця нерівність виконується.

Означення. *Надійністю* точкової оцінки θ^* параметра розподілу θ називають імовірність γ , з якою виконується нерівність $|\theta - \theta^*| < \delta$, тобто

$$P\{|\theta - \theta^*| < \delta\} = \gamma. \quad (1)$$

На практиці надійність оцінки задається наперед, причому число γ вибирають близьким до одиниці: $\gamma = 0,95$; $\gamma = 0,99$; $\gamma = 0,999$. Наприклад, надійність оцінки $\gamma = 0,95$ означає, що за достатньо великої кількості вибірок 95% з них визначають такі інтервали довіри, в яких справді знаходиться невідомий параметр.

Співвідношення (1) перетворимо до рівносильного виразу:

$$P\{-\delta < \theta - \theta^* < \delta\} = \gamma \quad \text{або} \quad P\{\theta^* - \delta < \theta < \theta^* + \delta\} = \gamma.$$

Означення. Інтервал $(\theta^* - \delta, \theta^* + \delta)$, для якого виконується рівність (1), називається **інтервалом довіри (надійним інтервалом)**, а його межі $\theta^* - \delta$ і $\theta^* + \delta$ – **надійними межами** для параметра розподілу θ .

Інакше кажучи, інтервал довіри для параметра розподілу θ є інтервал $(\theta^* - \delta, \theta^* + \delta)$, який з імовірністю γ „накриває” точне значення цього параметра.

Зрозуміло, що завжди бажано, щоб для заданої близької до одиниці ймовірності γ довжина інтервалу довіри була якомога меншою. Однак практично завжди є така альтернатива: збільшення надійності γ призводить до збільшення довжини інтервалу довіри, і навпаки.

Загальний спосіб, за допомогою якого знаходять інтервал довіри, полягає в тому, що розв’язують рівняння (1) і визначають з нього число δ . А для цього потрібно обчислити ймовірність $P\{\theta^* - \delta < \theta < \theta^* + \delta\}$. Останнє обчислення можна зробити, якщо відомий закон розподілу точкової оцінки $\theta^*(x_1, x_2, \dots, x_n)$ або пов’язаної з нею іншої випадкової величини, бо тоді можна використати відомі формули з теорії ймовірностей:

$$P\{\alpha \leq \theta^* < \beta\} = F(\beta) - F(\alpha), \quad \text{або} \quad P\{\alpha \leq \theta^* < \beta\} = \int_{\alpha}^{\beta} p(x) dx,$$

де $F(x)$ – функція розподілу і $p(x)$ – щільність розподілу випадкової величини θ^* .

Дві теореми про надійні межі для математичного сподівання

Теорема 1. Нехай X – нормально розподілена ознака генеральної сукупності, для якої $E(X) = a$, $D(X) = \sigma^2$, \bar{x} – вибіркве середнє, обчислене за вибіркою обсягу n з цієї генеральної сукупності. Тоді

$$\forall t > 0 \quad P\left\{|\bar{x} - a| < \frac{\sigma}{\sqrt{n}} t\right\} = 2\Phi(t), \quad (2)$$

$$\text{де } \Phi(x) = \frac{1}{\sqrt{2\pi}} \int_0^x e^{-\frac{t^2}{2}} dt.$$

Доведення. З того, що X – нормально розподілена випадкова величина, випливає, що вибіркве середнє \bar{x} – нормально розподілена випадкова величина, для якої $E(\bar{x}) = a$, $D(\bar{x}) = \frac{\sigma^2}{n} = \bar{\sigma}^2$, $\bar{\sigma} = \frac{\sigma}{\sqrt{n}}$. Для нормально розподіленої випадкової величини \bar{x} з параметрами розподілу a і $\bar{\sigma}$ ймовірність того, що модуль відхилення її значення від параметра a не перевищує числа ε , як відомо, обчислюється за формулою:

$$P\{|\bar{x} - a| < \varepsilon\} = 2\Phi\left(\frac{\varepsilon}{\bar{\sigma}}\right).$$

Покладемо у цій формулі $\varepsilon = \frac{\sigma}{\sqrt{n}}t$. Тоді $\frac{\varepsilon}{\bar{\sigma}} = \frac{\varepsilon}{\sigma/\sqrt{n}} = \frac{\sigma/\sqrt{n}}{\sigma/\sqrt{n}}t = t$ і ми одержимо співвідношення (2).

Теорема 2. Нехай X – довільно розподілена ознака генеральної сукупності, для якої $E(X) = a$, $D(X) = \sigma^2$, \bar{x} – вибіркоче середнє, обчислене за вибіркою обсягу n з цієї генеральної сукупності. Тоді

$$\forall t > 0 \quad \lim_{n \rightarrow \infty} P\left\{|\bar{x} - a| < \frac{\sigma}{\sqrt{n}}t\right\} = 2\Phi(t). \quad (3)$$

Доведення. Елементи вибірки обсягу n – значення випадкової величини X x_1, x_2, \dots, x_n – це незалежні однаково розподілені випадкові величини, для яких $E(x_k) = a$, $D(x_k) = \sigma^2$ ($k = \overline{1, n}$). Застосуємо до цих випадкових величин теорему Ліндеберга-Леві, згідно з якою

$$\lim_{n \rightarrow \infty} P\left\{\frac{\sum_{k=1}^n x_k - na}{\sigma\sqrt{n}} < t\right\} = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^t e^{-\frac{x^2}{2}} dx.$$

Отже, можемо записати:

$$\lim_{n \rightarrow \infty} P\left\{-t < \frac{\sum_{k=1}^n x_k - na}{\sigma\sqrt{n}} < t\right\} = \frac{1}{\sqrt{2\pi}} \int_{-t}^t e^{-\frac{x^2}{2}} dx = 2\Phi(t).$$

Після домноження кожної з частин нерівності

$$-t < \frac{\sum_{k=1}^n x_k - na}{\sigma\sqrt{n}} < t$$

на $\frac{\sigma}{\sqrt{n}}$, одержимо співвідношення (3). \square

Інтервальні оцінки для математичного сподівання

Нехай x_1, x_2, \dots, x_n – результати n незалежних спостережень за випадковою величиною X , на підставі яких необхідно знайти інтервал довіри для невідомого параметра $a = E(X)$.

Оскільки для математичного сподівання точковою оцінкою є вибіркоче середнє \bar{x} , то для знаходження інтервалу довіри $\bar{x} - \delta < a < \bar{x} + \delta$ потрібно розв'язати рівняння:

$$P\{|\bar{x} - a| < \delta\} = \gamma \quad \Leftrightarrow \quad P\{\bar{x} - \delta < a < \bar{x} + \delta\} = \gamma. \quad (4)$$

Якщо середнє квадратичне відхилення σ випадкової величини X відоме, то розв'язок рівняння (4) можна знайти, використовуючи рівності (2) або (3).

Так, якщо σ відоме, X – нормально розподілена випадкова величина або обсяг вибірки значний ($n > 30$), то ми можемо записати, що

$$P\left\{|\bar{x} - a| < \frac{\sigma}{\sqrt{n}}t\right\} = 2\Phi(t) = \gamma.$$

Тоді, якщо $t = t_\gamma$ – розв'язок рівняння $2\Phi(t) = \gamma$, то з надійністю γ інтервал

$$\bar{x} - \frac{\sigma}{\sqrt{n}}t_\gamma < a < \bar{x} + \frac{\sigma}{\sqrt{n}}t_\gamma$$

є інтервалом довіри для математичного сподівання a .

Якщо середнє квадратичне відхилення σ невідоме, але обсяг вибірки значний ($n > 30$), то інтервал довіри можна записати у вигляді

$$\bar{x} - \frac{s}{\sqrt{n}}t_\gamma < a < \bar{x} + \frac{s}{\sqrt{n}}t_\gamma, \quad (5)$$

де s – підправлене середнє квадратичне відхилення, знайдене за вибіркою обсягу n .

Якщо середнє квадратичне відхилення σ невідоме, обсяг вибірки незначний ($n < 30$), але X – нормально розподілена випадкова величина, то інтервал довіри також записують у вигляді (5), де значення $t_\gamma = t(\gamma, n)$ шукають за таблицями як розв'язок рівняння

$$P\{|T| < t_\gamma\} = \int_{-t_\gamma}^{t_\gamma} s(x, n) dx = 2 \int_0^{t_\gamma} s(x, n) dx = \gamma,$$

де $T = \frac{\bar{x} - a}{s/\sqrt{n}}$ – випадкова величина, розподілена за законом Ст'юдента з $k = n - 1$ ступенями вільності, який характеризується щільністю розподілу

$$s(x, n) = B_n \left(1 + \frac{x^2}{n-1}\right)^{-\frac{n}{2}}, \quad x \in (-\infty, +\infty),$$

де B_n – деяка нормуюча константа. Розподіл Ст'юдента залежить лише від одного параметра n і при $n \rightarrow \infty$ наближається до нормального закону розподілу.

Приклад 1. Випадкова величина X розподілена нормально з відомим середнім квадратичним відхиленням $\sigma = 3$. Знайти інтервал довіри з надійністю $\gamma = 0,95$ для оцінки невідомого математичного сподівання a , якщо вибіркоче середнє $\bar{x} = 20,02$ знайдене за даними вибірки обсягу $n = 36$.

Розв'язання. З рівняння $2\Phi(t) = 0,95 \Leftrightarrow \Phi(t) = 0,475$ за допомогою таблиць функції Лапласа знаходимо $t = t_\gamma = 1,96$. Межі інтервалу довіри шукаємо за формулами:

$$\bar{x} - \frac{\sigma}{\sqrt{n}}t_\gamma = 20,02 - \frac{3}{\sqrt{36}} \cdot 1,96 = 19,04; \quad \bar{x} + \frac{\sigma}{\sqrt{n}}t_\gamma = 20,02 + \frac{3}{\sqrt{36}} \cdot 1,96 = 21,00.$$

Отже, $a \in (19,04; 21,00)$ з надійністю $\gamma = 0,95$.

Приклад 2. Ознака X генеральної сукупності розподілена нормально. За вибіркою обсягу $n=16$ знайдено вибіркове середнє $\bar{x}=20,2$ і підправлене середнє квадратичне відхилення $s=0,8$. Оцінити невідоме математичне сподівання a за допомогою інтервалу довіри з надійністю $\gamma = 0,95$.

Розв'язання. Оскільки $n=16 < 30$ і середнє квадратичне відхилення σ невідоме, то для знаходження меж інтервалу довіри використаємо формулу (5), де значення $t_\gamma = t(\gamma, n)$ шукаємо за допомогою таблиць. Тоді

$$t_\gamma = t(0,95; 16) = 2,13; \quad \bar{x} - \frac{s}{\sqrt{n}}t_\gamma = 19,774; \quad \bar{x} + \frac{s}{\sqrt{n}}t_\gamma = 20,626.$$

Отже, $a \in (19,774; 20,626)$ з надійністю $\gamma = 0,95$.

Оцінка істинного значення вимірюваної величини

Нехай здійснюються n незалежних вимірювань деякої фізичної величини з однаковою точністю приладу (одним приладом), і до того ж істинне значення цієї величини невідоме. Результати вимірювань x_1, x_2, \dots, x_n – це незалежні однаково розподілені випадкові величини, оскільки вони мають одне і те саме математичне сподівання – істинне значення вимірюваної величини, і однакові дисперсії, бо вимірювання здійснюються з однаковою точністю. На підставі центральної граничної теореми можемо також стверджувати, що ці випадкові величини розподілені нормально. Отже, істинне значення вимірюваної величини можна оцінити за середнім арифметичним окремих вимірювань за допомогою інтервалів довіри.

Приклад. За даними 9-ти незалежних вимірювань фізичної величини, здійснених за допомогою одного приладу, знайдено середнє арифметичне результатів окремих вимірювань $\bar{x} = 42,319$ і підправлене середнє квадратичне відхилення $s = 5,0$. Оцінити істинне значення вимірюваної величини з надійністю $\gamma = 0,95$.

Розв'язання. Оскільки $n=9 < 30$ і середнє квадратичне відхилення σ невідоме, то, як і в попередньому прикладі, межі інтервалу довіри шукаємо за формулою (5), а значення $t_\gamma = t(\gamma, n)$ – за допомогою таблиць:

$$t_\gamma = t(0,95; 9) = 2,31; \quad \bar{x} - \frac{s}{\sqrt{n}}t_\gamma = 38,469; \quad \bar{x} + \frac{s}{\sqrt{n}}t_\gamma = 46,169.$$

Отже, з надійністю $\gamma = 0,95$ істинне значення вимірюваної величини покривається інтервалом $(38,469; 46,169)$.

Інтервали довіри для оцінки середнього квадратичного відхилення нормально розподіленої випадкової величини

Нехай ознака X генеральної сукупності розподілена нормально. Знайдемо надійні межі для середнього квадратичного відхилення σ з заданою надійністю γ . Оскільки точковою оцінкою для параметра σ є підправлене середнє квадратичне відхилення s , то для цього потрібно розв'язати рівняння:

$$P\{|\sigma - s| < \delta\} = \gamma \Leftrightarrow P\{s - \delta < \sigma < s + \delta\} = \gamma.$$

Перетворимо подвійну нерівність $s - \delta < \sigma < s + \delta$:

$$s\left(1 - \frac{\delta}{s}\right) < \sigma < s\left(1 + \frac{\delta}{s}\right) \Leftrightarrow s(1 - q) < \sigma < s(1 + q), \quad (6)$$

де $q = \frac{\delta}{s}$.

Залишається знайти q . З цією метою розглянемо випадкову величину $\chi = \sqrt{n-1}s/\sigma$, де n – обсяг вибірки. Відомо, що випадкова величина $(n-1)s^2/\sigma^2$ розподілена за законом χ^2 з $n-1$ ступенями вільності, тому квадратний корінь з неї позначають через χ . Щільність розподілу χ має вигляд

$$R(x, n) = A_n x^{n-2} e^{-x^2/2},$$

де A_n – деяка стала. Цей розподіл не залежить від оцінюваного параметра σ , а залежить лише від обсягу вибірки n . Перетворимо нерівність (6) до вигляду $\chi_1 < \chi < \chi_2$. Ймовірність цієї нерівності дорівнює заданій імовірності γ , тобто

$$P\{\chi_1 < \chi < \chi_2\} = \int_{\chi_1}^{\chi_2} R(x, n) dx = \gamma.$$

Припускаючи, що $q < 1$, перепишемо нерівність (6) так:

$$\frac{1}{s(1+q)} < \frac{1}{\sigma} < \frac{1}{s(1-q)} \Leftrightarrow \frac{\sqrt{n-1}}{1+q} < \frac{s\sqrt{n-1}}{\sigma} < \frac{\sqrt{n-1}}{1-q}$$

або

$$\frac{\sqrt{n-1}}{1+q} < \chi < \frac{\sqrt{n-1}}{1-q}.$$

Отже,

$$\begin{aligned} P\{s - \delta < \sigma < s + \delta\} &= P\{s(1 - q) < \sigma < s(1 + q)\} = \\ &= P\left\{\frac{\sqrt{n-1}}{1+q} < \chi < \frac{\sqrt{n-1}}{1-q}\right\} = \int_{\chi_1}^{\chi_2} R(x, n) dx = \gamma, \end{aligned}$$

де $\chi_1 = \sqrt{n-1}/(1+q)$, $\chi_2 = \sqrt{n-1}/(1-q)$. З цього рівняння можна за заданими n і γ (за допомогою таблиці $q = q(\gamma, n)$) знайти q .

Обчисливши за вибіркою s і знайшовши за таблицею q , отримаємо шуканий інтервал довіри (6), який покриває параметр σ з заданою надійністю γ .

Якщо $q > 1$, то нерівність (6) набуде вигляду

$$0 < \sigma < s(1 + q).$$

У цьому випадку q також шукають за таблицею значень $q = q(\gamma, n)$.

Оцінка точності вимірювань

В теорії помилок прийнято точність вимірювань (точність приладу) характеризувати за допомогою середнього квадратичного відхилення σ випадкових помилок вимірювань. Для оцінки σ використовують підправлене середнє квадратичне відхилення s .

Приклад. За результатами 15-ти вимірювань, здійснених одним приладом, обчислили підправлене середнє квадратичне відхилення $s = 0,12$. Оцінити точність приладу з надійністю $\gamma = 0,99$.

Розв'язання. Задача зводиться до відшукування інтервалу довіри

$$s(1 - q) < \sigma < s(1 + q),$$

який покриває σ з заданою надійністю $\gamma = 0,99$. За таблицею $q = q(\gamma, n)$ за $\gamma = 0,99$ і $n = 15$ знаходимо $q = 0,73$. Шуканий інтервал довіри

$$0,12(1 - 0,73) < \sigma < 0,12(1 + 0,73) \text{ або } 0,03 < \sigma < 0,21.$$

Лекція 17

Вибірковий коефіцієнт кореляції

Кореляційний аналіз досліджує наявність і характер зв'язків між випадковими величинами X і Y – ознаками генеральної сукупності.

З теорії ймовірностей відомо, що ступінь зв'язку між випадковими величинами X і Y визначається такими числовими характеристиками їхнього сумісного розподілу як *коваріація* $\text{cov}(X, Y)$ і *коефіцієнт кореляції* $\rho(X, Y)$, які обчислюються за формулами:

$$\text{cov}(X, Y) = E(XY) - E(X)E(Y);$$

$$\rho(X, Y) = \frac{\text{cov}(X, Y)}{\sqrt{D(X) \cdot D(Y)}} = \frac{\text{cov}(X, Y)}{\sigma(X) \cdot \sigma(Y)}.$$

Основна задача *кореляційного аналізу* полягає у виявленні залежності між випадковими величинами X і Y і може бути розв'язана шляхом побудови статистичних оцінок коефіцієнта кореляції.

Точкову оцінку для коефіцієнта кореляції обчислюють за формулою:

$$\begin{aligned} r = r(X, Y) &= \frac{\frac{1}{n} \sum_{i=1}^n x_i y_i - \bar{x} \cdot \bar{y}}{\sqrt{\left(\frac{1}{n} \sum_{i=1}^n x_i^2 - \bar{x}^2 \right) \left(\frac{1}{n} \sum_{i=1}^n y_i^2 - \bar{y}^2 \right)}} = \\ &= \frac{\sum_{i=1}^n x_i y_i - n\bar{x} \cdot \bar{y}}{\sqrt{\left(\sum_{i=1}^n x_i^2 - n\bar{x}^2 \right) \left(\sum_{i=1}^n y_i^2 - n\bar{y}^2 \right)}}. \end{aligned} \quad (1)$$

Означення. Точкова оцінка r коефіцієнта кореляції між випадковими величинами X і Y , яка обчислюється за формулою (1), називається **вибірковим коефіцієнтом кореляції**.

Вибірковий коефіцієнт кореляції характеризує зв'язок між випадковими величинами X і Y (ознаками генеральної сукупності): а) якщо $r > 0$, то зв'язок між X і Y є додатний і вони зменшуються або збільшуються одночасно; б) якщо $r < 0$, то зв'язок між X і Y є від'ємний і зі збільшенням однієї з них друга зменшується або навпаки; в) якщо $r = 0$, то випадкові величини X і Y є некорельовані і це означає лише відсутність лінійного зв'язку між ними.

Вибірковий коефіцієнт кореляції задовольняє нерівність $|r| \leq 1$.

Вибіркове рівняння прямої регресії Y на X

Регресійний аналіз встановлює аналітичну форму залежності між випадковими величинами X і Y – ознаками генеральної сукупності.

Якщо $r(X, Y) \neq 0$, то X і Y – корельовані випадкові величини і з наближенням величини $|r(X, Y)|$ до одиниці залежність між цими випадковими

величинами наближається до лінійної залежності вигляду $Y = \alpha X + \beta$.

Нагадаємо, що рівняння прямої регресії Y на X має вигляд:

$$y = \alpha x + \beta, \quad (2)$$

де

$$\alpha = \rho(X, Y) \frac{\sigma(Y)}{\sigma(X)} = \frac{\text{cov}(X, Y)}{\sigma(X)\sigma(Y)} \frac{\sigma(Y)}{\sigma(X)} = \frac{\text{cov}(X, Y)}{D(X)}, \quad \beta = E(Y) - \alpha E(X). \quad (3)$$

Означення. Рівняння (2) називається **вибірковим рівнянням прямої регресії Y на X** , якщо коефіцієнти у ньому вибрано у вигляді точкових оцінок α і β , визначених співвідношеннями (3).

Отже, у вибірковому рівнянні прямої регресії Y на X коефіцієнти α і β визначаються у вигляді:

$$\alpha = \frac{\text{cov}(X, Y)}{D(X)} = \frac{\frac{1}{n} \sum_{i=1}^n x_i y_i - \bar{x} \cdot \bar{y}}{\frac{1}{n} \sum_{i=1}^n x_i^2 - \bar{x}^2} = \frac{\sum_{i=1}^n x_i y_i - n\bar{x} \cdot \bar{y}}{\sum_{i=1}^n x_i^2 - n\bar{x}^2};$$

$$\beta = \bar{y} - \alpha \bar{x} = \frac{1}{n} \sum_{i=1}^n y_i - \alpha \frac{1}{n} \sum_{i=1}^n x_i.$$

Приклад. Зв'язок між кількісними ознаками X і Y генеральної сукупності задається таблицею:

№ п/п	1	2	3	4	5	6	7	8	9	10
X	2	3	4	5	6	7	8	9	10	11
Y	1	5	7	11	13	17	19	23	25	29

Записати рівняння прямої регресії Y на X .

Розв'язання. Обчислюємо:

$$\bar{x} = 6,5; \quad \bar{y} = 15; \quad \sum_{i=1}^{10} x_i y_i = 1225; \quad \sum_{i=1}^{10} x_i^2 = 505;$$

$$\alpha = \frac{\sum_{i=1}^{10} x_i y_i - 10 \cdot \bar{x} \cdot \bar{y}}{\sum_{i=1}^{10} x_i^2 - 10 \cdot \bar{x}^2} = \frac{1225 - 10 \cdot 6,5 \cdot 15}{505 - 10 \cdot (6,5)^2} = 3,03; \quad \beta = \bar{y} - \alpha \bar{x} = 15 - 3,03 \cdot 6,5 = -4,70.$$

Отже, вибіркове рівняння регресії:

$$y = 3,03x - 4,70.$$

Щоб переконатися в тому, що наше припущення про лінійність зв'язку між X і Y було правильним, обчислимо вибірковий коефіцієнт кореляції за формулою:

$$r(X, Y) = \frac{\sum_{i=1}^{10} x_i y_i - n\bar{x} \cdot \bar{y}}{\sqrt{\left(\sum_{i=1}^{10} x_i^2 - 10 \cdot \bar{x}^2 \right) \left(\sum_{i=1}^{10} y_i^2 - 10 \cdot \bar{y}^2 \right)}} =$$

$$= \frac{1225 - 10 \cdot 6,5 \cdot 15}{\sqrt{(505 - 10 \cdot (6,5)^2)(3010 - 10 \cdot (15)^2)}} = 0,9984.$$

Оскільки вибірковий коефіцієнт кореляції $r(X, Y)$ є досить близьким до одиниці, то припущення про лінійність зв'язку між X і Y – правильне. Крім цього, $r > 0$, тому зв'язок між X і Y є додатний і ці випадкові величини збільшуються одночасно.

Лекція 18

Статистична перевірка статистичних гіпотез

Означення статистичної гіпотези і задача про її статистичну перевірку

Дані вибіркового спостереження часто становлять основу для прийняття одного з кількох альтернативних рішень (продукція може бути бракованою або якісною, точність обробки виробу в межах норми або нижча від норми і т. д.). Із загальнометодологічного погляду тут йдеться про висунення деякої гіпотези, яку відхиляють або приймають після проведення деякого експерименту. Якщо цей експеримент має статистичний (стохастичний) характер, кажуть, що гіпотеза є *статистичною*.

Означення. *Статистичною* називають гіпотезу про властивості генеральної сукупності, що перевіряється на основі вибірки.

Статистичними гіпотезами можуть бути, наприклад, такі твердження: розподіл імовірностей випадкової величини є нормальний; розподіл імовірностей випадкової величини є пуассонівський; у нормальному розподілі випадкової величини параметри $a = 20$ і $\sigma = 1,5$; у показниковому розподілі випадкової величини параметр $\alpha = 5$; випадкові величини X і Y незалежні і т. п.

У математичній статистиці виділяють два основні *типи статистичних гіпотез*:

- *гіпотези про закон розподілу ймовірностей випадкової величини (ознаки генеральної сукупності);*
- *гіпотези про значення параметрів розподілу випадкової величини (ознаки генеральної сукупності).*

Статистичні гіпотези першого типу називають непараметричними, а другого типу – параметричними.

Означення. *Основною (нульовою)* називають висунуту гіпотезу і позначають H_0 .

Альтернативною (конкуруючою) називають гіпотезу, яка повністю або частково логічно заперечує нульову гіпотезу, і позначають H_1 .

Наприклад, математичне сподівання нормально розподіленої випадкової величини $a = 10$ – основна гіпотеза; математичне сподівання нормально розподіленої випадкової величини $a \neq 10$ – альтернативна гіпотеза. Записують це так: $H_0 : a = 10$; $H_1 : a \neq 10$.

Задача про статистичну перевірку статистичних гіпотез формулюється так. Розглядають деяку гіпотезу про те, що розподіл імовірностей деякої випадкової величини має той чи інший вигляд, або параметри розподілу мають ті чи інші значення.

Задача полягає у тому, щоб на основі вивчення статистичних даних (вбірки) підтвердити справедливість висунутої гіпотези чи спростувати її. При цьому вказується також імовірність того, що прийняте рішення є правильним

або помилковим. Проблема зменшення ймовірності того, що прийняте рішення є помилковим, є також однією із задач математичної статистики.

У результаті статистичної перевірки гіпотези може бути прийняте одне з двох правильних рішень: 1) гіпотеза приймається і вона істинна; 2) гіпотеза відхиляється і вона неістинна.

Поряд із тим у результаті статистичної перевірки статистичної гіпотези можуть бути допущені помилки (прийняті неправильні рішення) двох типів: 1) гіпотеза відхиляється, але вона істинна (помилка першого роду); 2) гіпотеза приймається, але вона неістинна (помилка другого роду).

Виявляється, що помилка першого роду має вагоміші наслідки, ніж помилка другого роду.

Щоб застрахувати себе від помилки першого роду або принаймні звести до мінімуму ризик її допущення, вводиться спеціальне число α , яке виражає ймовірність відхилення правильної гіпотези.

Означення. *Ймовірність допущення помилки першого роду називають рівнем значущості і позначають через α .*

Число α задають наперед і найчастіше його вибирають рівним 0,1; 0,05; 0,01. Якщо $\alpha = 0,05$, то це означає, що ймовірність допустити помилку першого роду є мала, а саме – ми ризикуємо її допустити у 5-ти випадках зі 100.

Критерій статистичної перевірки гіпотези

Для перевірки нульової гіпотези вводять певну числову характеристику, яку обчислюють на основі вибірки і на підставі якої вирішують: прийняти основну гіпотезу чи альтернативну. Зрозуміло, що вибрана числова характеристика для різних вибірок матиме, загалом кажучи, різні значення, і тому вона є випадковою величиною.

Означення. *Статистичним критерієм гіпотези (або просто критерієм гіпотези) називається випадкова величина K , за допомогою якої проводиться перевірка гіпотези.*

Випадкову величину K вибирають такою, щоб її закон розподілу ймовірностей був відомий.

Означення. *Значення випадкової величини K , обчислене на основі даних певної вибірки, називають емпіричним значенням критерію гіпотези і позначають $K_{емп}$.*

Виявляється, що за одних значень $K_{емп}$ гіпотеза H_0 приймається, а за інших його значень – відхиляється.

Означення. *Сукупність значень критерію K , за яких нульова гіпотеза H_0 відхиляється, називається критичною областю, а сукупність значень критерію K , за яких нульову гіпотезу H_0 приймають, називається областю прийняття гіпотези.*

Звідси маємо таке **правило перевірки статистичних гіпотез**: якщо емпіричне значення критерію $K_{емп}$ належить критичній області, то нульову

гіпотезу H_0 відхиляють; якщо емпіричне значення критерію $K_{\text{емп}}$ належить області прийняття гіпотези, то нульову гіпотезу H_0 приймають.

Якщо випадкова величина K є одновимірною, то критична область, як правило, є множиною точок певних інтервалів на прямій, які відокремлені від області прийняття гіпотези так званими критичними точками $k_{\text{кр}}$. Тобто для знаходження критичної області достатньо визначити критичні точки.

Залежно від конкуруючої гіпотези розглядають три види критичних областей:

- *правостороння критична область* – це та область на числовій прямій, що визначається нерівністю $K > k_{\text{кр}}$;
- *лівостороння критична область* – це та область на числовій прямій, що визначається нерівністю $K < k_{\text{кр}}$;
- *двостороння критична область* – це та область на числовій прямій, що визначається сумою інтервалів $K < -k_{\text{кр}}$ і $K > k_{\text{кр}}$.

Для знаходження критичної області задаються рівнем значущості α і шукають критичні точки $k_{\text{кр}}$ із таких співвідношень:

а) для правосторонньої критичної області:

$$P\{K > k_{\text{кр}}\} = \alpha \quad (k_{\text{кр}} > 0);$$

б) для лівосторонньої критичної області:

$$P\{K < k_{\text{кр}}\} = \alpha \quad (k_{\text{кр}} < 0);$$

в) для двосторонньої симетричної критичної області:

$$P\{K > k_{\text{кр}}\} = \alpha / 2 \quad (k_{\text{кр}} > 0).$$

Зрозуміло, що для певної гіпотези можна побудувати багато різних критеріїв її перевірки, і за кожним критерієм можемо одержувати різні результати щодо прийняття нульової гіпотези H_0 на основі тієї самої вибірки. Тому для визначення кращого критерію вводиться характеристика, яка називається *потужністю* критерію.

Означення. *Потужністю критерію називають імовірність потрапляння критерію у критичну область за умови, що конкуруюча гіпотеза є істинною.*

Тобто потужність критерію визначається як імовірність не допустити помилку другого роду при вибраному критерії.

Перевірка гіпотези про закон розподілу. Критерій згоди Пірсона

Критерієм згоди називають статистичний критерій перевірки гіпотези про закон розподілу ймовірностей випадкової величини (ознаки генеральної сукупності). Є кілька критеріїв згоди: критерій Колмогорова, критерій Смірнова, критерій Пірсона та ін. Розглянемо критерій згоди Пірсона (критерій χ^2), який ґрунтується на порівнянні емпіричних і теоретичних частот.

Нехай висунуто гіпотезу H_0 : випадкова величина X розподілена за законом A .

Здійснивши вибірку обсягу n , знаходять і записують у вигляді таблиці інтервальный статистичний розподіл частот:

$[z_{i-1}, z_i)$	$[z_0, z_1)$	$[z_1, z_2)$...	$[z_{m-1}, z_m)$
n_i	n_1	n_2	...	n_m

$$\sum_{i=1}^m n_i = n.$$

Оскільки перевіряється гіпотеза про те, що розподіл ознаки X генеральної сукупності описується певною (конкретною) функцією розподілу $F(x)$, або, що те ж саме, щільністю розподілу $p(x)$, то для кожного інтервалу $[z_{i-1}, z_i)$ можна визначити теоретичні ймовірності p_i попадання значень випадкової величини X у цей інтервал, а отже, і теоретичні частоти $n'_i = np_i$.

Для обчислення ймовірностей p_i використовують формули:

$$p_i = P\{z_{i-1} \leq X < z_i\} = F(z_i) - F(z_{i-1}) = \int_{z_{i-1}}^{z_i} p(x) dx, \quad i = \overline{1, m}. \quad (1)$$

Зазначимо, що для обчислення ймовірностей p_1 і p_m у формулах (1) покладають, відповідно, $z_0 = -\infty$ і $z_m = +\infty$. Тоді $\sum_{i=1}^m p_i = 1$.

Отримані результати обчислень зручно записати у формі таблиці:

$[z_{i-1}, z_i)$	$[z_0, z_1)$	$[z_1, z_2)$...	$[z_{m-1}, z_m)$
n_i	n_1	n_2	...	n_m
p_i	p_1	p_2	...	p_m
$n'_i = np_i$	n'_1	n'_2	...	n'_m

Згідно з критерієм Пірсона для перевірки гіпотези H_0 вводиться випадкова величина (статистика) K :

$$K = \chi^2 = \sum_{i=1}^m \frac{(n_i - n'_i)^2}{n'_i},$$

де m – кількість груп у статистичному розподілі вибірки; n_i – емпірична частота ознаки X в i -тій групі; $n'_i = np_i$ – теоретична частота; p_i – імовірність того, що значення X належить i -тій групі.

Відомо, що при $n \rightarrow \infty$ закон розподілу статистики K прямує до закону розподілу χ^2 з $k = m - r - 1$ ступенями вільності, де m – кількість груп у статистичному розподілі вибірки; r – кількість параметрів гіпотетичного розподілу A (наприклад, $r = 2$ для нормального розподілу, $r = 1$ для розподілу Пуассона, $r = 0$ для рівномірного розподілу).

Для критерію χ^2 будують правосторонню критичну область за правилом:

$$P\{\chi^2 > \chi_{кр}^2\} = \alpha. \quad (2)$$

За заданим рівнем значущості α і кількістю ступенів вільності k із таблиці критичних точок розподілу χ^2 (в якій дано розв'язки рівняння (2)) знаходять критичну точку $k_{кр} = \chi_{кр}^2(\alpha, k)$.

На підставі даних вибірки, записаних у таблиці, обчислюють емпіричне значення критерію Пірсона:

$$K_{емп} = \sum_{i=1}^m \frac{(n_i - n'_i)^2}{n'_i}.$$

Порівнюємо значення $K_{емп}$ і $k_{кр}$: якщо $K_{емп} \geq k_{кр}$, то гіпотезу H_0 відхиляють; якщо ж $K_{емп} < k_{кр}$, то гіпотезу H_0 приймають.

Застосування критерію χ^2 вимагає дотримання таких умов: 1) експериментальні дані мають бути незалежними, тобто вибірка має бути випадковою; 2) обсяг вибірки має бути достатньо великим (практично не меншим ніж 50 одиниць), а частота кожної групи – не меншою за 5. Якщо остання умова не виконується, то проводиться попереднє об'єднання нечисленних груп.

Критерій згоди Пірсона дає відповідь на питання, чи розбіжність між емпіричними і теоретичними частотами зумовлена випадковістю, чи вона є значущою. Як і будь-який інший критерій він *не доводить справедливості* гіпотези H_0 , а лише дозволяє встановити на прийнятному рівні значущості *узгодженість чи неузгодженість* гіпотези H_0 з даними спостережень.

Приклад 1. У таблиці наведено покази 500 навмання вибраних механічних годинників, виставлених у вітринах годинникових магазинів. Для рівня значущості $\alpha = 0,05$ перевірити гіпотезу H_0 про те, що покази годинників рівномірно розподілені на проміжку $[0; 12)$.

i	0	1	2	3	4	5	6	7	8	9	10	11
n_i	41	34	54	39	49	45	41	33	37	41	47	39

$$\sum_{i=0}^{11} n_i = 500.$$

Тут i – номер проміжка від i -ої години до $(i+1)$ -ої, $i = \overline{0,11}$; n_i – кількість годинників, покази яких потрапляють в i -ий проміжок.

Розв'язання. Для рівномірного розподілу щільність розподілу ймовірностей визначається функцією:

$$p(x) = \begin{cases} 0, & x \notin [0; 12); \\ 1/12, & x \in [0; 12). \end{cases}$$

Для $i = \overline{0,11}$ обчислюємо:

$$p_i = P\{X \in [i, i+1)\} = \int_i^{i+1} \frac{dx}{12} = \frac{1}{12}; \quad n'_i = np_i = 500 \cdot \frac{1}{12} = 41,7.$$

$$K_{\text{емп}} = \sum_{i=0}^{11} \frac{(n_i - n'_i)^2}{n'_i} = \frac{1}{41,7} \left((41 - 41,7)^2 + \dots + (39 - 41,7)^2 \right) = 9,99.$$

Кількість ступенів вільності $k = m - r - 1 = 12 - 0 - 1 = 11$. За таблицею критичних точок розподілу χ^2 знаходимо: $k_{\text{кр}} = \chi^2_{\text{кр}}(0,05; 11) = 19,7$. Отже, $K_{\text{емп}} < k_{\text{кр}}$, тому гіпотезу H_0 приймаємо.

Приклад 2. У таблиці наведено емпіричні частоти n_i та теоретичні частоти n'_i , обчислені, виходячи з гіпотези H_0 про нормальний розподіл генеральної сукупності. Для рівня значущості $\alpha = 0,01$ перевірити гіпотезу H_0 про нормальний розподіл генеральної сукупності.

n_i	8	16	40	72	36	18	10
n'_i	6	18	36	76	39	18	7

Розв'язання. Обчислимо емпіричне значення критерію Пірсона ($m = 7$):

$$K_{\text{емп}} = \sum_{i=1}^7 \frac{(n_i - n'_i)^2}{n'_i} = \frac{(8-6)^2}{6} + \frac{(16-18)^2}{18} + \dots + \frac{(10-7)^2}{7} = 3,06.$$

За таблицею критичних точок розподілу χ^2 для рівня значущості $\alpha = 0,01$ і кількості ступенів вільності $k = m - r - 1 = 7 - 2 - 1 = 4$ ($r = 2$ для нормального розподілу) знаходимо: $k_{\text{кр}} = \chi^2_{\text{кр}}(0,01; 4) = 13,3$. Отже, $K_{\text{емп}} < k_{\text{кр}}$, тому гіпотезу H_0 приймаємо.

Перевірка гіпотези про порівняння середнього значення (математичного сподівання) ознаки генеральної сукупності зі стандартом

У критеріях для перевірки гіпотези $H_0: a = a_0$ про те, що значення математичного сподівання $a = E(X)$ досліджуваної ознаки X генеральної сукупності збігається зі стандартом a_0 , використовують статистику \bar{x} – вибіркове середнє. Залежно від інформації щодо генеральної сукупності, якою володіємо, розрізняємо такі моделі.

Модель А. *Гіпотеза про значення математичного сподівання нормального закону розподілу за відомої дисперсії.*

Нехай випадкова величина X нормально розподілена з невідомим математичним сподіванням $a = E(X)$, але відомою дисперсією $\sigma^2 = D(X)$. Потрібно на основі вибірки перевірити нульову гіпотезу $H_0: a = a_0$ про рівність математичного сподівання a певному числу a_0 . При цьому припускаємо, що відомі такі величини: дані вибірки обсягу n ; середнє квадратичне відхилення $\sigma = \sigma(X)$; гіпотетичне значення математичного сподівання a_0 ; рівень значущості α .

Із вивченого нами матеріалу випливає, що вибіркове середнє \bar{x} для вибірки з нормального розподілу з параметрами (a, σ^2) має нормальний розподіл з параметрами $(a, \sigma^2/n)$, тому за умови істинності гіпотези H_0 (коли $E(\bar{x}) = a_0$) випадкова величина

$$U = \frac{(\bar{x} - a_0)\sqrt{n}}{\sigma},$$

яку беруть за критерій перевірки гіпотези H_0 , також розподілена нормально з параметрами $(0, 1)$.

Справді,

$$E(U) = \frac{\sqrt{n}}{\sigma} E(\bar{x} - a_0) = \frac{\sqrt{n}}{\sigma} (E(\bar{x}) - a_0) = 0;$$

$$D(U) = \frac{n}{\sigma^2} D(\bar{x} - a_0) = \frac{n}{\sigma^2} D(\bar{x}) = \frac{n}{\sigma^2} \cdot \frac{\sigma^2}{n} = 1.$$

Отже, щільність розподілу випадкової величини U має вигляд:

$$p(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}.$$

Тому

$$P\{U \in (0, z)\} = \frac{1}{\sqrt{2\pi}} \int_0^z e^{-\frac{x^2}{2}} dx = \Phi(z).$$

Якщо конкуруюча гіпотеза має вигляд: $H_1: a \neq a_0$, то розглядають двосторонню симетричну критичну область, для якої критичну точку шукають із співвідношення:

$$P\{U > u_{кр}\} = \alpha/2.$$

Оскільки $P\{0 < U < \infty\} = \Phi(\infty) = 0,5$, то

$$P\{0 < U < \infty\} = P\{0 < U < u_{кр}\} + P\{U > u_{кр}\} = 0,5,$$

тобто

$$\Phi(u_{кр}) + \frac{\alpha}{2} = \frac{1}{2} \quad \text{або} \quad \Phi(u_{кр}) = \frac{1-\alpha}{2}.$$

Правило 1. Якщо нульова гіпотеза $H_0: a = a_0$, а конкуруюча гіпотеза $H_1: a \neq a_0$, то перевірку гіпотези H_0 проводимо за такою схемою:

- обчислюємо емпіричне значення критерію за формулою:

$$U_{емп} = \frac{(\bar{x} - a_0)\sqrt{n}}{\sigma}; \quad (3)$$

- знаходимо за таблицею значень функції Лапласа критичне значення $u_{кр}$, використовуючи рівняння:

$$\Phi(u_{кр}) = \frac{1-\alpha}{2}; \quad (4)$$

- робимо висновок про висунуту гіпотезу: якщо $|U_{емп}| < u_{кр}$, то гіпотезу H_0 приймаємо; якщо $|U_{емп}| \geq u_{кр}$, то відхиляємо гіпотезу H_0 на користь альтернативи H_1 .

Якщо конкуруюча гіпотеза має вигляд: $H_1: a > a_0$, то розглядають правосторонню критичну область, для якої критичну точку шукають із співвідношення:

$$P\{U > u_{кр}\} = \alpha.$$

Тоді

$$\Phi(u_{кр}) + \alpha = \frac{1}{2}, \quad \text{тобто} \quad \Phi(u_{кр}) = \frac{1-2\alpha}{2}.$$

Якщо конкуруюча гіпотеза має вигляд: $H_1: a < a_0$, то розглядають лівосторонню критичну область, для якої $P\{U < -u_{кр}\} = \alpha$.

Правило 2. Якщо нульова гіпотеза $H_0: a = a_0$, а конкуруюча гіпотеза $H_1: a > a_0$ або $H_1: a < a_0$, то перевірку гіпотези H_0 також проводимо за схемою правила 1 з такими змінами:

- замість рівняння (4) для знаходження критичного значення $u_{кр}$ використовуємо рівняння:

$$\Phi(u_{кр}) = \frac{1-2\alpha}{2}; \quad (5)$$

- робимо висновок стосовно висунутої гіпотези H_0 :

1) якщо $U_{емп} < u_{кр}$, то нема підстав відхилити гіпотезу H_0 ; якщо $U_{емп} \geq u_{кр}$, то відхиляємо гіпотезу H_0 на користь альтернативи $H_1: a > a_0$;

- 2) якщо $U_{емп} > -u_{кр}$, то нема підстав відхилити гіпотезу H_0 ; якщо $U_{емп} \leq -u_{кр}$, то гіпотезу H_0 відхиляємо і приймаємо гіпотезу $H_1: a < a_0$.

Приклад. Із нормально розподіленої генеральної сукупності з відомим середнім квадратичним відхиленням $\sigma = 0,6$ одержано вибірку обсягом $n = 36$ і за нею знайдено вибіркоче середнє $\bar{x} = 21,6$. Потрібно для рівня значущості $\alpha = 0,05$ перевірити нульову гіпотезу $H_0: a = a_0 = 21$ за наявності різних альтернативних гіпотез:

- а) $H_1: a \neq a_0$;
 б) $H_1: a > a_0$;
 в) для $\bar{x} = 20,85$ розглянути альтернативну гіпотезу $H_1: a < a_0$.

Розв'язання. Обчислимо емпіричне значення критерію за формулою (3):

$$U_{емп} = \frac{(21,6 - 21)\sqrt{36}}{0,6} = 6.$$

Тепер розглянемо наведені в задачі випадки:

а) для альтернативної гіпотези $H_1: a \neq a_0$ знаходимо $u_{кр}$ за таблицею значень функції Лапласа, використовуючи формулу (4):

$$\Phi(u_{кр}) = \frac{1 - \alpha}{2} = \frac{1 - 0,05}{2} = 0,475 \Rightarrow u_{кр} = 1,96.$$

Оскільки $|U_{емп}| = 6 > 1,96 = u_{кр}$, то відхиляємо гіпотезу H_0 і приймаємо гіпотезу $H_1: a \neq a_0$;

б) для альтернативної гіпотези $H_1: a > a_0$ знаходимо $u_{кр}$ за таблицею значень функції Лапласа, використовуючи формулу (5):

$$\Phi(u_{кр}) = \frac{1 - 2\alpha}{2} = \frac{1 - 2 \cdot 0,05}{2} = 0,45 \Rightarrow u_{кр} = 1,645.$$

Оскільки $U_{емп} = 6 > 1,645 = u_{кр}$, то відхиляємо гіпотезу H_0 і приймаємо гіпотезу $H_1: a > a_0$.

в) якщо $\bar{x} = 20,85$, то $U_{емп} = \frac{(20,85 - 21)\sqrt{36}}{0,6} = -1,5$. Як і в пункті б)

$$\Phi(u_{кр}) = \frac{1 - 2\alpha}{2} = \frac{1 - 2 \cdot 0,05}{2} = 0,45 \Rightarrow u_{кр} = 1,645.$$

Оскільки $U_{емп} = -1,5 > -u_{кр} = -1,645$, то гіпотезу H_0 приймаємо.

Модель Б. Гіпотеза про значення математичного сподівання нормального закону розподілу за невідомої дисперсії.

Нехай випадкова величина X нормально розподілена з невідомими математичним сподіванням $a = E(X)$ і дисперсією $\sigma^2 = D(X)$. Потрібно на основі вибірки перевірити нульову гіпотезу $H_0: a = a_0$ про рівність математичного сподівання a певному числу a_0 . При цьому припускаємо, що

відомі такі величини: дані вибірки обсягу n ; гіпотетичне значення математичного сподівання a_0 ; рівень значущості α .

Оскільки середнє квадратичне відхилення $\sigma = \sigma(X)$ невідоме, то для перевірки гіпотези H_0 тут ми вже не зможемо скористатися статистикою U через те, що для неї неможливо буде обчислити емпіричне значення $U_{емп}$. У даному випадку використовуємо статистику

$$T = \frac{(\bar{x} - a_0)\sqrt{n}}{s},$$

де \bar{x} – вибіркове середнє, а s – підправлене середнє квадратичне відхилення. Можна показати, що за умови істинності гіпотези H_0 випадкова величина T має розподіл Ст'юдента з $k = n - 1$ ступенями вільності.

Подальша побудова критичної області для дво- та односторонніх перевірок гіпотези здійснюється аналогічно як у випадку моделі А, з тією лише різницею, що критичні точки (тут замість $u_{кр}$ вони будуть позначатися через $t_{кр}$) визначаються за таблицею критичних точок розподілу Ст'юдента, а не значень функції Лапласа. За того самого рівня значущості α значення $t_{кр}$ буде більшим, аніж $u_{кр}$.

Правило 1. Якщо нульова гіпотеза $H_0: a = a_0$, а конкуруюча гіпотеза $H_1: a \neq a_0$, то перевірку гіпотези H_0 проводимо за такою схемою:

- обчислюємо емпіричне значення критерію:

$$T_{емп} = \frac{(\bar{x} - a_0)\sqrt{n}}{s}; \quad (6)$$

- знаходимо за таблицею критичних точок розподілу Ст'юдента за заданим рівнем значущості α (для двосторонньої критичної області) і кількістю ступенів вільності $k = n - 1$ критичну точку $t_{кр} = t_{кр}(\alpha, k)$;
- робимо висновок про висунуту гіпотезу: якщо $|T_{емп}| < t_{кр}$, то гіпотезу H_0 приймаємо; якщо $|T_{емп}| \geq t_{кр}$, то відхиляємо гіпотезу H_0 на користь альтернативи H_1 .

Правило 2. Якщо нульова гіпотеза $H_0: a = a_0$, а конкуруюча гіпотеза $H_1: a > a_0$ або $H_1: a < a_0$, то перевірку гіпотези H_0 також проводимо за схемою правила 1 з такими змінами:

- за таблицею критичних точок розподілу Ст'юдента за заданим рівнем значущості α (для односторонньої критичної області) і кількістю ступенів вільності $k = n - 1$ знаходимо критичну точку $t_{кр} = t_{кр}(\alpha, k)$;
- робимо висновок стосовно висунутої гіпотези H_0 :
1) якщо $T_{емп} < t_{кр}$, то нема підстав відхилити гіпотезу H_0 ; якщо $T_{емп} \geq t_{кр}$, то відхиляємо гіпотезу H_0 на користь альтернативи $H_1: a > a_0$;

2) якщо $T_{емп} > -t_{кр}$, то нема підстав відхилити гіпотезу H_0 ; якщо $T_{емп} \leq -t_{кр}$, то гіпотезу H_0 відхиляємо і приймаємо гіпотезу $H_1: a < a_0$.

Приклад. За вибіркою обсягу $n=20$ з нормально розподіленої генеральної сукупності знайдено вибіркоче середнє $\bar{x}=16$ і підправлене середнє квадратичне відхилення $s=4,5$. Потрібно для рівня значущості $\alpha=0,05$ перевірити нульову гіпотезу $H_0: a = a_0 = 15$ за наявності конкуруючої гіпотези: а) $H_1: a \neq a_0$; б) $H_1: a > a_0$.

Розв'язання. Обчислимо емпіричне значення критерію за формулою (6):

$$T_{емп} = \frac{(16-15)\sqrt{20}}{4,5} = 0,99.$$

а) Знаходимо за таблицею критичних точок розподілу Ст'юдента за заданим рівнем значущості $\alpha=0,05$ (для двосторонньої критичної області) і кількістю ступенів вільності $k=20-1=19$ критичну точку $t_{кр} = t_{кр}(0,05; 19) = 2,09$. Оскільки $|T_{емп}| = 0,99 < 2,09 = t_{кр}$, то гіпотезу H_0 приймаємо.

б) За таблицею критичних точок розподілу Ст'юдента за заданим рівнем значущості $\alpha=0,05$ (для односторонньої критичної області) і кількістю ступенів вільності $k=20-1=19$ знаходимо критичну точку $t_{кр} = t_{кр}(0,05; 19) = 1,73$. Оскільки $T_{емп} = 0,99 < 1,73 = t_{кр}$, то гіпотезу H_0 приймаємо.

Перевірка гіпотези про порівняння дисперсії нормальної генеральної сукупності зі стандартом

Нехай відомо, що випадкова величина X нормально розподілена, але її дисперсія $\sigma^2 = D(X)$ невідома. Потрібно на основі вибірки перевірити нульову гіпотезу $H_0: \sigma^2 = \sigma_0^2$ про рівність дисперсії σ^2 певному числу σ_0^2 . При цьому припускаємо, що відомі такі величини: дані вибірки обсягу n ; підправлена дисперсія s^2 ; гіпотетичне значення дисперсії σ_0^2 ; рівень значущості α .

На практиці дану гіпотезу перевіряють, якщо треба перевірити точність приладів, інструментів, станків, стійкість технологічних процесів. Наприклад, якщо допустима характеристика розсіювання контрольованого розміру деталей, які виготовляє станок-автомат, рівна σ_0^2 , а знайдена за вибіркою виявиться значно більшою, то станок необхідно підналагодити.

За критерій перевірки гіпотези H_0 беруть випадкову величину

$$\chi^2 = \frac{(n-1)s^2}{\sigma_0^2},$$

яка розподілена за законом χ^2 з $k=n-1$ ступенями вільності. Вигляд критичної області залежить від конкуруючої гіпотези H_1 .

Якщо конкуруюча гіпотеза має вигляд: $H_1 : \sigma^2 > \sigma_0^2$, то розглядають правосторонню критичну область, для якої критичну точку шукають із співвідношення:

$$P\{\chi^2 > \chi_{кр}^2(\alpha, k)\} = \alpha.$$

Для перевірки гіпотези H_0 обчислюють емпіричне значення критерію

$$\chi_{емп}^2 = \frac{(n-1)s^2}{\sigma_0^2} \quad (7)$$

і за таблицею критичних точок розподілу χ^2 за заданим рівнем значущості α і кількістю ступенів вільності $k = n - 1$ знаходять критичну точку $\chi_{кр}^2(\alpha, k)$.

Якщо $\chi_{емп}^2 < \chi_{кр}^2(\alpha, k)$, гіпотезу H_0 приймають; якщо $\chi_{емп}^2 > \chi_{кр}^2(\alpha, k)$, то гіпотезу H_0 відхиляють і приймають гіпотезу $H_1 : \sigma^2 > \sigma_0^2$.

Якщо конкуруюча гіпотеза має вигляд: $H_1 : \sigma^2 \neq \sigma_0^2$, то розглядають двосторонню критичну область, для якої критичну точку шукають із співвідношення:

$$P\{\chi^2 < \chi_{лів.кр}^2\} + P\{\chi^2 > \chi_{пр.кр}^2\} = \alpha,$$

тобто

$$P\{\chi^2 < \chi_{лів.кр}^2(\alpha/2, k)\} = P\{\chi^2 > \chi_{пр.кр}^2(\alpha/2, k)\} = \frac{\alpha}{2}.$$

В таблиці критичних точок розподілу χ^2 вказані лише „праві” критичні точки, але, враховуючи, що

$$P\{\chi^2 > \chi_{лів.кр}^2\} = 1 - P\{\chi^2 < \chi_{лів.кр}^2\} = 1 - \frac{\alpha}{2},$$

„ліву” критичну точку можна шукати як „праву”, використовуючи одержану рівність. Отже, обчисливши емпіричне значення критерію $\chi_{емп}^2$ за формулою (7), за заданими α і $k = n - 1$ за допомогою таблиці критичних точок розподілу χ^2 знаходять „ліву” критичну точку $\chi_{лів.кр}^2 = \chi_{кр}^2(1 - \alpha/2, k)$ і „праву” критичну точку $\chi_{пр.кр}^2 = \chi_{кр}^2(\alpha/2, k)$. Якщо $\chi_{лів.кр}^2 < \chi_{емп}^2 < \chi_{пр.кр}^2$, гіпотезу H_0 приймають; якщо $\chi_{емп}^2 < \chi_{лів.кр}^2$ або $\chi_{емп}^2 > \chi_{пр.кр}^2$ то гіпотезу H_0 відхиляють і приймають гіпотезу $H_1 : \sigma^2 \neq \sigma_0^2$.

Якщо конкуруюча гіпотеза має вигляд: $H_1 : \sigma^2 < \sigma_0^2$, то розглядають лівосторонню критичну область, для якої критичну точку шукають із співвідношення:

$$P\{\chi^2 < \chi_{кр}^2(\alpha, k)\} = \alpha.$$

Тоді

$$P\{\chi^2 > \chi_{кр}^2(\alpha, k)\} = 1 - \alpha.$$

Тому за допомогою таблиці критичних точок розподілу χ^2 за заданими α і $k = n - 1$ знаходять $\chi_{кр}^2 = \chi_{кр}^2(1 - \alpha, k)$. Якщо $\chi_{емп}^2 > \chi_{кр}^2$, гіпотезу H_0 приймають; якщо $\chi_{емп}^2 < \chi_{кр}^2$, то гіпотезу H_0 відхиляють і приймають гіпотезу $H_1: \sigma^2 < \sigma_0^2$.

Приклади. 1) За вибіркою обсягу $n = 13$ з нормально розподіленої генеральної сукупності знайдено підправлену дисперсію $s^2 = 14,6$. Потрібно для рівня значущості $\alpha = 0,01$ перевірити нульову гіпотезу $H_0: \sigma^2 = \sigma_0^2 = 12$ за наявності конкуруючої гіпотези $H_1: \sigma^2 > \sigma_0^2$.

2) За вибіркою обсягу $n = 13$ з нормальної генеральної сукупності знайдено підправлену дисперсію $s^2 = 10,3$. Потрібно для рівня значущості $\alpha = 0,02$ перевірити нульову гіпотезу $H_0: \sigma^2 = \sigma_0^2 = 12$ за наявності конкуруючої гіпотези $H_1: \sigma^2 \neq \sigma_0^2$.

3) За даними прикладу 2 для рівня значущості $\alpha = 0,05$ перевірити нульову гіпотезу $H_0: \sigma^2 = \sigma_0^2 = 12$ за наявності конкуруючої гіпотези $H_1: \sigma^2 < \sigma_0^2$.

Розв'язання. 1) Обчислимо емпіричне значення критерію за формулою (7):

$$\chi_{емп}^2 = \frac{(13 - 1) \cdot 14,6}{12} = 14,6.$$

За допомогою таблиці критичних точок розподілу χ^2 знаходимо:

$$\chi_{кр}^2 = \chi_{кр}^2(\alpha, k) = \chi_{кр}^2(0,01; 12) = 26,2.$$

Оскільки $\chi_{емп}^2 < \chi_{кр}^2$, то гіпотезу H_0 приймаємо.

2) За формулою (7) знаходимо:

$$\chi_{емп}^2 = \frac{(13 - 1) \cdot 10,3}{12} = 10,3.$$

За допомогою таблиці критичних точок розподілу χ^2 визначаємо „ліву” критичну точку

$$\chi_{лів.кр}^2 = \chi_{кр}^2(1 - \alpha/2, k) = \chi_{кр}^2(0,99; 12) = 3,57$$

і „праву” критичну точку

$$\chi_{пр.кр}^2 = \chi_{кр}^2(\alpha/2, k) = \chi_{кр}^2(0,01; 12) = 26,2.$$

Оскільки $3,57 < 10,3 < 26,2$, тобто $\chi_{лів.кр}^2 < \chi_{емп}^2 < \chi_{пр.кр}^2$, то гіпотезу H_0 приймаємо.

3) Як і раніше, $\chi_{емп}^2 = \frac{(13 - 1) \cdot 10,3}{12} = 10,3$. Для гіпотези $H_1: \sigma^2 < \sigma_0^2$

$$\chi_{кр}^2 = \chi_{кр}^2(1 - \alpha, k) = \chi_{кр}^2(0,95; 12) = 5,23.$$

Оскільки $\chi_{емп}^2 > \chi_{кр}^2$, то гіпотезу H_0 приймаємо.

ЛІТЕРАТУРА

1. Єлейко Я.І., Тріщ Б.М. *Теорія ймовірностей*. — Львів: ЛНУ ім. Івана Франка, 2001.
2. Єлейко Я.І., Тріщ Б.М. *Методичні вказівки до вивчення курсу „Теорія ймовірностей і математична статистика”*. Основи вибіркового методу. — Львів: ЛНУ ім. Івана Франка, 2001.
3. Бобик О.І., Берегова Г.І., Копитко Б.І. *Теорія ймовірностей і математична статистика*. Львів: ЛБІ НБУ, 2003.
4. Гмурман В.Е. *Теория вероятностей и математическая статистика*. — М.: Высш. шк., 1978.
5. Гмурман В.Е. *Руководство к решению задач по теории вероятностей и математической статистике*. — М.: Высшая школа, 2001.
6. Шефтель З.Г. *Теорія ймовірностей*. — К.: Вища школа, 1977.
7. Черняк О.І., Обушна О.М., Ставицький А.В. *Теорія ймовірностей та математична статистика. Збірник задач*. — К.: Знання, 2002.
8. Гнеденко Б.В. *Курс теории вероятностей*. — М.: Физматгиз, 1961.
9. Жлуктенко В.І., Наконечний С.І. *Теорія ймовірностей з елементами математичної статистики*. — К.: УМК ВО, 1991.
10. Жлуктенко В.І., Наконечний С.І. *Теорія ймовірностей і математична статистика: Ч.1*. — К.: КНЕУ, 2000.
11. Кулініч Г.Л., Максименко Л.О., Плахотник В.В., Призва Г.Й. *Вища математика: основні означення, приклади і задачі: У 2-х кн.* — К.: Либідь, 1994. — Кн.1.