УКРАЇНСЬКИЙ ДЕРЖАВНИЙ УНІВЕРСИТЕТ
ЗАЛІЗНИЧНОГО ТРАНСПОРТУ

ФАКУЛЬТЕТ УПРАВЛІННЯ ПРОЦЕСАМИ ПЕРЕВЕЗЕНЬ

Кафедра вищої математики

Н. Г. Панченко, М. Є. Резуненко

ВИЩА ТА ПРИКЛАДНА МАТЕМАТИКА

Конспект лекцій

Частина IV

Харків - 2021

Панченко Н. Г., Резуненко М. Є. Конспект лекцій. – Харків : УкрДУЗТ, 2021. – Ч. IV. – 64 с.

Конспект лекцій призначено для студентів освітнього рівня «бакалавр» економічного факультету.

Іл. 15, табл. 10, бібліогр.: 10 назв.

Конспект лекцій розглянуто та рекомендовано до друку на засіданні кафедри вищої математики 22 вересня 2020 р., протокол № 2.

Рецензент

доц. А. П. Рибалко (ХНЕУ)

ЗМІСТ

1 ВИПАДКОВІ ВЕЛИЧИНИ

1.1 Основні поняття про випадкові величини

Визначення. *Випадковою* називається величина, яка при кожному випробуванні набуває певного значення, заздалегідь невідомо якого, з множини можливих значень.

Відмінність випадкової величини від випадкової події полягає в такому: якщо випадкова подія може відбутися або не відбутися в результаті випробування, то випадкова величина в результаті випробування обов'язково набуває деякого значення.

Випадкові величини бувають дискретними і неперервними.

Визначення. *Дискретною випадковою величиною* (ДВВ) називають випадкову величину, всі можливі значення якої можна пронумерувати. Тобто множина значень дискретної величини або скінченна, або нескінченна злічення.

Визначення. *Неперервною випадковою величиною* (НВВ) називають випадкову величину, значення якої цілком заповнюють деякий числовий проміжок. Тобто множина значень неперервної величини нескінченна незлічення.

Випадкові величини позначаються великими латинськими літерами $X, Y, Z \ldots$, а їхні значення – відповідними малими літерами з індексами.

Приклад 1. Дискретні випадкові величини:

• X – кількість бракованих виробів у партії з трьох штук. ДВВ X може набувати значень

$$x_1 = 0,\ \ x_2 = 1,\ \ x_3 = 2,\ \ x_4 = 3.$$

• Y – кількість вагонів, що надійшли на вантажний двір протягом доби.

Приклад 2. Неперервні випадкові величини:

- час безвідмовної роботи приладу;
- час очікування пасажиром автобуса.

2 ДИСКРЕТНІ ВИПАДКОВІ ВЕЛИЧИНИ

2.1 Закон розподілу дискретної випадкової величини

Для повної характеристики випадкової величини необхідно враховувати як її можливі значення, так і ймовірності набуття цих значень, що досягається за допомогою закону розподілу.

Визначення. *Законом розподілу* випадкової величини називається відповідність між можливими значеннями випадкової величини і ймовірностями їхнього настання.

Закон розподілу може бути заданий:
- у вигляді таблиці;
- графічно;
- аналітично.

У випадку дискретної випадкової величини найзручнішою формою закону розподілу є ряд розподілу.

Визначення. *Рядом розподілу* дискретної випадкової величини називають перелік у порядку зростання всіх можливих її значень, для кожного з яких указана відповідна ймовірність його появи.

Зазвичай його оформлюють у вигляді таблиці.

X	x_1	x_2	$\ldots$	x_n	$\ldots$
P	p_1	p_2	$\ldots$	p_n	$\ldots$

(2.1)

де $p_i = P(X = x_i)$, $i = 1,2,...,n,...$;
$x_1 < x_2 < ... < x_n < ...$.

Оскільки події $\{X = x_i\}, i = 1,2,...,n,...$ утворюють повну групу, то

$$p_1 + p_2 + ... + p_n + ... = 1 \text{ (}\textit{умова нормування}\text{).} \qquad (2.2)$$

Ряд розподілу ДВВ можна подати *графічно*. Для цього в прямокутній системі координат будують точки з координатами $(x_i; p_i)$, $i = 1,2,...,n,...$ та послідовно з'єднують відрізками. Отриману ламану називають *багатокутником розподілу,* або *полігоном* (рисунок 2.1).

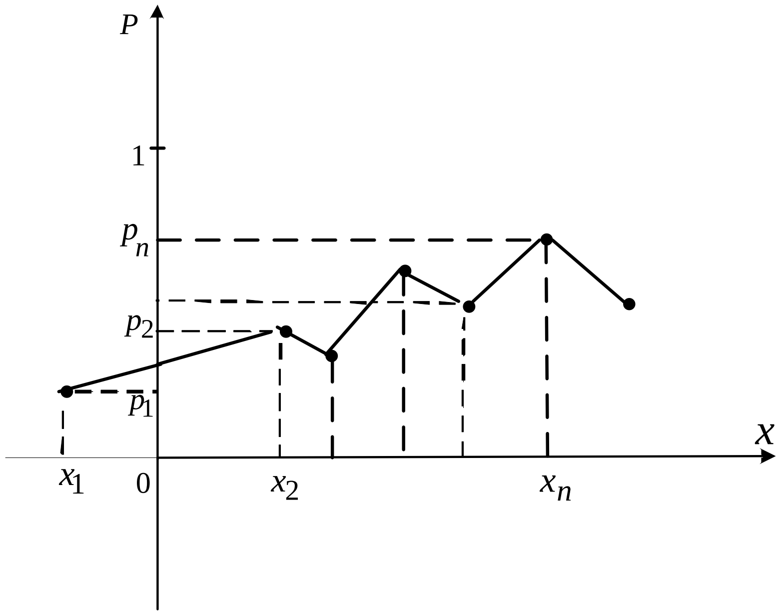

Рисунок 2.1 – Багатокутник розподілу (полігон) ДВВ

Ряд розподілу можна задати аналітично:

$$p_i = P(X = x_i) = \varphi(x_i),\ i = 1,2,...,n,..., \tag{2.3}$$

де $\varphi(x_i)$ – відома функція.

Зауваження:

- Аналітичний спосіб особливо важливий у випадку нескінченної множини значень $x_i,\ i = 1,2,...$.
- Зазвичай аналітично задаються спеціальні розподіли.

Визначення. *Модою* M_0 ДВВ X називається її найімовірніше значення. З геометричної точки зору, мода є абсцисою такої точки багатокутника розподілу (полігона), ордината якої максимальна.

Приклад 3. Задано ряд розподілу ДВВ X .

X	0	1	2	5	6	8
P	0,1	0,2	p_3	0,05	0,15	0,3

Необхідно:

1) знайти p_3;

2) побудувати багатокутник розподілу (полігон) ДВВ X ;

3) обчислити моду M_0.

Розв'язання:

1) за умовою нормування (2.2)

$$0{,}1+0{,}2+p_3+0{,}05+0{,}15+0{,}3=1,$$
$$p_3=0{,}2;$$

2) будуємо багатокутник розподілу ДВВ X (рисунок 2.2);

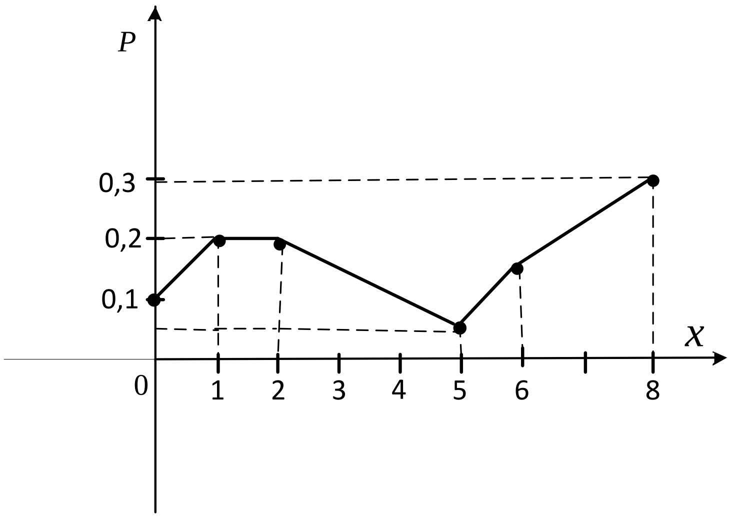

Рисунок 2.2 – Багатокутник розподілу ДВВ X прикладу 3

3) за визначенням мода ДВВ X $M_0=8$.

Відповідь: 1) $p_3=0{,}2$; 3) $M_0=8$.

Приклад 4. В регіоні три великих підприємства. Ймовірність невчасної сплати податку першим підприємством дорівнює 0,03, другим - 0,02, третім - 0,06. Скласти ряд розподілу ДВВ X – кількість підприємств, що вчасно сплатять податок.

Розв'язання. За умовою задачі ДВВ X може набувати чотирьох значень 0, 1, 2, 3. Розглянемо чотири події:

A_i={вчасно сплатять податок i підприємств}, $i=0,1,2,3$.

Обчислюючи ймовірності цих подій, отримаємо ймовірності того, що ДВВ X набуде кожного з розглянутих значень.

Позначимо через $p_i, i=1,2,3$ ймовірності вчасної сплати i-м підприємством. Тоді (дивись роботу [10], формули (4.9), (4.10)) ймовірність невчасної сплати $q_i=1-p_i$, $i=1,2,3$ відповідно. Таким чином, за умовою задачі отримаємо

$$p_1 = 0{,}97,\ q_1 = 0{,}03;$$
$$p_2 = 0{,}98,\ q_2 = 0{,}02;$$
$$p_3 = 0{,}94,\ q_3 = 0{,}06.$$

Тоді (дивись роботу [10], приклад 22)

$$P(X=0) = P(A_0) = q_1q_2q_3 = 0{,}03 \cdot 0{,}02 \cdot 0{,}06 = 0{,}000036;$$
$$P(X=1) = P(A_1) = p_1q_2q_3 + q_1p_2q_3 + q_1q_2q_3 =$$
$$= 0{,}97 \cdot 0{,}02 \cdot 0{,}06 + 0{,}03 \cdot 0{,}98 \cdot 0{,}06 + 0{,}03 \cdot 0{,}02 \cdot 0{,}94 = 0{,}003492;$$
$$P(X=2) = P(A_2) = p_1p_2q_3 + q_1p_2p_3 + p_1q_2p_3 =$$
$$= 0{,}97 \cdot 0{,}98 \cdot 0{,}06 + 0{,}03 \cdot 0{,}98 \cdot 0{,}94 + 0{,}97 \cdot 0{,}02 \cdot 0{,}94 = 0{,}102908;$$
$$P(X=3) = P(A_3) = p_1p_2p_3 = 0{,}97 \cdot 0{,}98 \cdot 0{,}94 = 0{,}893564.$$

Таким чином, ряд розподілу ДВВ X (або закон розподілу у табличній формі) має вигляд

X	0	1	2	3
P	0,000036	0,003492	0,102908	0,893564

Контроль (перевірка умови нормування (2.1)):

$$\sum_{i=1}^{4} p_i = 0{,}000036 + 0{,}003492 + 0{,}102908 + 0{,}893564 = 1.$$

Відповідь: ряд розподілу ДВВ X – кількість підприємств, що вчасно сплатять податок, має вигляд

X	0	1	2	3
P	0,000036	0,003492	0,102908	0,893564

Приклад 5. Підприємство дає замовлення незалежно трьом філіям. Імовірність того, що будь-яка філія не виконає свою роботу вчасно, дорівнює 0,05. Скласти ряд розподілу ДВВ X – кількість філій, які не вкладуться у термін виконання даного замовлення.

Розв'язання. За умовою ДВВ X має такі можливі значення:

$x_1 = 0$ – нема філій, які не вкладуться у термін;

$x_2 = 1$ – одна філія не вкладеться у термін;

$x_3 = 2$ – дві філії не вкладуться у термін;

$x_4 = 3$ – всі три філії не вкладуться у термін.

За формулою Бернуллі (дивись роботу [10], формулу (5.1)) $n = 3, p = 0{,}05, q = 0{,}95$ обчислюємо

$$P(X = 0) = P_3(0) = C_3^0 \cdot 0{,}05^0 \cdot 0{,}95^3 = 0{,}857375;$$
$$P(X = 1) = P_3(1) = C_3^1 \cdot 0{,}05^1 \cdot 0{,}95^2 = 0{,}135375;$$
$$P(X = 2) = P_3(2) = C_3^2 \cdot 0{,}05^2 \cdot 0{,}95^1 = 0{,}007125;$$
$$P(X = 3) = P_3(3) = C_3^3 \cdot 0{,}05^3 \cdot 0{,}95^0 = 0{,}000125.$$

Складаємо ряд розподілу ДВВ X (або закон розподілу у табличній формі).

X	0	1	2	3
P	0,857375	0,135375	0,007125	0,000125

Контроль (перевірка умови нормування (2.1)):

$$\sum_{i=1}^{4} p_i = 0{,}857375 + 0{,}135375 + 0{,}007125 + 0{,}000125 = 1.$$

Відповідь: ряд розподілу ДВВ X - кількість філій, які не вкладуться у термін виконання даного замовлення, має вигляд

X	0	1	2	3
P	0,857375	0,135375	0,007125	0,000125

Зауваження. Даний закон розподілу може бути заданий і аналітично, а саме

$$P(X = k) = C_3^k \cdot (0{,}05)^k \cdot (0{,}95)^{3-k}, \; k = 0,1,2,3.$$

Розподіли такого типу називаються *біноміальними* (дивись пп. 2.4.1).

2.2 Функція розподілу дискретної випадкової величини

Визначення. *Функцією розподілу ймовірностей* ДВВ X, або *інтегральною функцією,* називається ймовірність того, що в результаті випробування X набуде значення, меншого за аргумент x, тобто функція

$$F(x) = P(X < x), \qquad x \in R. \tag{2.4}$$

Зауваження. Дане визначення функції розподілу є спільним як для дискретних, так і для неперервних випадкових величин.

З визначення безпосередньо випливають властивості $F(x)$.

Властивості функції розподілу ймовірностей $F(x)$:

1) $0 \le F(x) \le 1$;

2) функція розподілу є неспадною, тобто якщо $x_1 < x_2$, то $F(x_1) \le F(x_2)$;

3) $\lim\limits_{x \to -\infty} F(x) = 0, \quad \lim\limits_{x \to +\infty} F(x) = 1$;

4) імовірність потрапляння випадкової величини X у півінтервал $[\alpha;\beta)$ дорівнює різниці значень функції розподілу в цих точках:

$$P(\alpha \le X < \beta) = F(\beta) - F(\alpha); \tag{2.5}$$

5) для ДВВ X функція розподілу має вигляд

$$F(x) = P(X < x) = \sum_{x_i < x} p_i, \quad x \in R, \tag{2.6}$$

тобто $F(x)$ дорівнює сумі ймовірностей значень X, що менші за x.

Зауваження. Функція розподілу ймовірностей ДВВ є *кусково-сталою*, а тому графік має *ступінчастий вигляд*.

Приклад 6. Побудувати функцію розподілу випадкової величини, заданої своїм багатокутником розподілу (рисунок 2.3).

Обчислити ймовірності $P(X < 2)$, $P(X \ge 4)$.

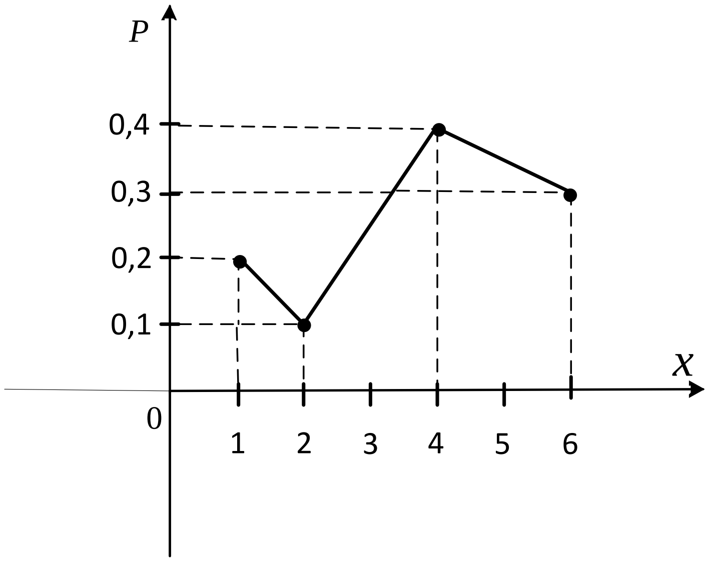

Рисунок 2.3 – Багатокутник розподілу прикладу 6

Розв'язання. За багатокутником розподілу (полігоном) (рисунок 2.3) складаємо ряд розподілу дискретної випадкової величини.

X	1	2	4	6
P	0,2	0,1	0,4	0,3

За формулою (2.6) складаємо функцію розподілу ДВВ X :

$$F(x)=\begin{cases}0, & x\le 1;\\ 0{,}2, & 1<x\le 2;\\ 0{,}3, & 2<x\le 4;\\ 0{,}7, & 4<x\le 6;\\ 1, & x>6,\end{cases}$$

і будуємо її графік (рисунок 2.4).

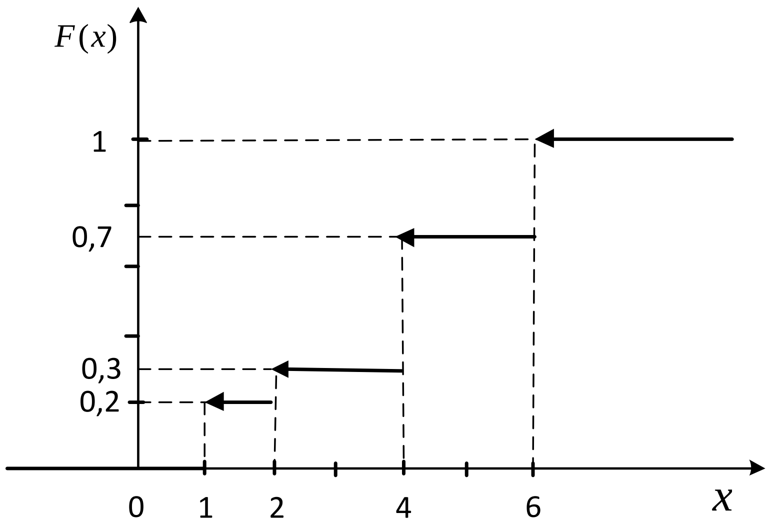

Рисунок 2.4 – Функція розподілу ДВВ X прикладу 6

За формулою (2.5) обчислюємо ймовірності $P(X<2),\ P(X\ge 4)$:

$$P(X<2)=F(2)-F(-\infty)=0{,}2-0=0{,}2\,,$$
$$P(X\ge 4)=P(4\le X<+\infty)=F(+\infty)-F(4)=1-0{,}3=0{,}7\,.$$

Відповідь: $P(X<2)=0{,}2$, $P(X\ge 4)=0{,}7$.

2.3 Числові характеристики дискретної випадкової величини

Закон розподілу повністю характеризує випадкову величину. Але для розв'язання багатьох практичних задач зручніше користуватися *числовими характеристиками* випадкової величини, що дають достатню інформацію про цю величину. Розглянемо найбільш важливі з них.

2.3.1 Математичне сподівання

Розглянемо ДВВ X, яку задано рядом розподілу

X	x_1	x_2	…	x_n
P	p_1	p_2	…	p_n

Визначення. *Математичним сподіванням* $M(X)$ *ДВВ* X називається сума добутків значень випадкової величини на їхні імовірності:

$$M(X) = x_1 \cdot p_1 + \ldots + x_n \cdot p_n = \sum_{i=1}^{n} x_i p_i . \tag{2.7}$$

Якщо ДВВ X набуває нескінченної кількості значень $x_1, x_2, ..., x_n, ...$ з відповідними ймовірностями $p_1, p_2, ..., p_n, ...$, то математичне сподівання обчислюється за формулою

$$M(X) = x_1 \cdot p_1 + \ldots + x_n \cdot p_n + ... = \sum_{i=1}^{+\infty} x_i p_i . \tag{2.8}$$

Імовірнісний зміст математичного сподівання полягає в тому, що $M(X)$ наближено дорівнює середньому арифметичному спостережуваних значень випадкової величини.

Іноді математичне сподівання $M(X)$ позначають через m_X .

Приклад 7. За умовою прикладу 4 обчислити середню кількість підприємств, що вчасно сплатять податок.

Розв'язання. Середню кількість підприємств, що вчасно сплатять податок, обчислюємо як математичне сподівання *ДВВ* X за формулою (2.7):

$$M(X) = 0 \cdot 0{,}000036 + 1 \cdot 0{,}003492 + 2 \cdot 0{,}102908 + 3 \cdot 0{,}893564 = 2{,}89 .$$

Відповідь: 2,89.

Властивості математичного сподівання $M(X)$:

1) $M(C) = C$, де $C = const$;

2) $M(CX) = CM(X)$;

3) $M(X + Y) = M(X) + M(Y)$;

3') $M(X_1 + \ldots + X_k) = M(X_1) + \ldots + M(X_k)$;

4) $M(XY) = M(X) \cdot M(Y)$ для незалежних величин X, Y ;

4') $M(X_1 \ldots X_k) = M(X_1) \cdot \ldots \cdot M(X_k)$ для попарно незалежних величин $X_1, \ldots, X_k$.

Приклад 8. Дискретні випадкові величини X та Y – незалежні і задані рядами розподілу

X	1	2	4
p	0,1	0,3	0,6

Y	0	5
p	0,2	0,8

Знайти $M(4-2X+Y+3XY)$.

Розв'язання. За формулою (2.7) обчислюємо математичні сподівання дискретних випадкових величин X та Y:

$$M(X)=\sum_{i=1}^{3}x_i\cdot p_i=1\cdot 0{,}1+2\cdot 0{,}3+4\cdot 0{,}6=3{,}1;$$
$$M(Y)=\sum_{j=1}^{2}y_j\cdot q_j=0\cdot 0{,}2+5\cdot 0{,}8=4.$$

Враховуючи властивості математичного сподівання дискретної випадкової величини, отримаємо

$$M(4-2X+Y+3XY)=M(4)-2M(X)+M(Y)+3M(X)M(Y)=$$
$$=4-2\cdot 3{,}1+4+3\cdot 3{,}1\cdot 4=39.$$

Відповідь: 39.

2.3.2 Дисперсія та середнє квадратичне відхилення

Для характеристики випадкової величини недостатньо знати тільки математичне сподівання цієї величини, оскільки одному і тому самому $M(X)$ може відповідати велика кількість випадкових величин з різними можливими значеннями. Тому потрібні ще додаткові числові характеристики, які б характеризували розсіювання випадкової величини навколо її середнього значення, тобто навколо математичного сподівання як центра розсіювання. Такими характеристиками є дисперсія та середнє квадратичне відхилення.

Визначення. *Дисперсією* $D(X)$ випадкової величини X називається математичне сподівання квадрата відхилення X від її математичного сподівання:

$$D(X)=M[X-M(X)]^2. \tag{2.9}$$

Імовірнісний зміст дисперсії полягає в тому, що $D(X)$ характеризує розсіювання значень випадкової величини навколо її математичного сподівання $M(X)$.

З визначення отримаємо формулу дисперсії для ДВВ X:

$$D(X)=\sum_{i=1}^{n}(x_i-M(X))^2\cdot p_i. \tag{2.10}$$

Якщо скористатися властивостями математичного сподівання, то отримаємо

$$D(X)=M[X-M(X)]^2=M\left(X^2-2X\cdot M(X)+M^2(X)\right)=$$
$$=M\left(X^2\right)-2M(X)\cdot M(X)+M^2(X)=M\left(X^2\right)-2M^2(X)+M^2(X)=$$
$$=M\left(X^2\right)-M^2(X).$$

Тобто дисперсія випадкової величини дорівнює різниці між математичним сподіванням квадрата цієї величини і квадратом її математичного сподівання

$$D(X)=M\left(X^2\right)-M^2(X) \tag{2.11}$$

та для дискретної випадкової величини обчислюється за формулою

$$D(X)=\sum_{i=1}^{n}x_i^2\cdot p_i-\left(\sum_{i=1}^{n}x_i\cdot p_i\right)^2. \tag{2.12}$$

Приклад 9. ДВВ X задано рядом розподілу

X	-2	0	1	3
P	0,2	0,1	0,4	0,3

Знайти $D(X)$.

Розв'язання. За формулою (2.7) обчислюємо математичне сподівання

$$M(X) = -2 \cdot 0{,}2 + 0 \cdot 0{,}1 + 1 \cdot 0{,}4 + 3 \cdot 0{,}3 = 0{,}9.$$

Обчислимо дисперсію двома способами. За визначенням (2.10)

$$D(X) = \sum_{i=1}^{n} \left(x_i - M(X)\right)^2 \cdot p_i = (-2-0{,}9)^2 \cdot 0{,}2 + (0-0{,}9)^2 \cdot 0{,}1 +$$

$$+ (1-0{,}9)^2 \cdot 0{,}4 + (3-0{,}9)^2 \cdot 0{,}3 = 3{,}09\,.$$

Щоб скористатись формулою (2.11), знайдемо спочатку розподіл ДВВ X^2

X^2	4	0	1	9
P	0,2	0,1	0,4	0,3

і математичне сподівання

$$M(X^2) = 4 \cdot 0{,}2 + 0 \cdot 0{,}1 + 1 \cdot 0{,}4 + 9 \cdot 0{,}3 = 3{,}9.$$

Тоді

$$D(X) = M\left(X^2\right) - M^2(X) = 3{,}9 - (0{,}9)^2 = 3{,}09.$$

Відповідь: $D(X) = 3{,}09$.

З визначення дисперсії і вже відомих властивостей математичного сподівання випливають властивості дисперсії.

Властивості дисперсії $D(X)$**:**

1) $D(X) \geq 0$;

2) $D(C) = 0$, де $C = const$;

3) $D(CX) = C^2 D(X)$;

4) $D(X+Y) = D(X-Y) = D(X) + D(Y)$ для незалежних X, Y;

4') $D(C+X) = D(X)$;

4'') $D(X_1 + \ldots + X_k) = D(X_1) + \ldots + D(X_k)$ для попарно незалежних $X_1, \ldots, X_k$.

Зауваження. Дисперсія має розмірність, що дорівнює квадрату розмірності випадкової величини. Тому розглядають ще одну характеристику розсіювання, розмірність якої збігається з розмірністю випадкової величини.

Визначення. *Середнім квадратичним відхиленням* називається величина

$$\sigma(X) = \sqrt{D(X)}, \tag{2.13}$$

що, як і дисперсія, характеризує розсіювання, але вимірюється в тих самих одиницях, що і сама випадкова величина.

Приклад 10. Задано дисперсії двох незалежних випадкових величин $D(X) = 3$, $D(Y) = 0{,}5$. Знайти:

1) $D(X + 4)$;
2) $D(4X + 2Y)$;
3) $D(X - 2Y)$;
4) $\sigma(X - Y)$.

Розв'язання. Скористаємось властивостями дисперсії:

1) $D(X + 4) = D(X) + D(4) = D(X) = 3$;

2) $D(4X + 2Y) = D(4X) + D(2Y) = 4^2 \cdot D(X) + 2^2 \cdot D(Y) =$
$= 16 \cdot 3 + 4 \cdot 0{,}5 = 50$;

3) $D(X - 2Y) = D(X) + 2^2 \cdot D(Y) = 3 + 4 \cdot 0{,}5 = 5$;

4) $\sigma(X - Y) = \sqrt{D(X - Y)} = \sqrt{D(X) + D(Y)} = \sqrt{3 + 0{,}5} \approx 1{,}871$.

Відповідь: 1) $D(X + 4) = 3$; 2) $D(4X + 2Y) = 50$; 3) $D(X - 2Y) = 5$; 4) $\sigma(X - Y) \approx 1{,}871$.

2.4. Деякі закони розподілу дискретних випадкових величин та їхні числові характеристики

2.4.1 Біноміальний закон розподілу

Нехай проводиться n незалежних випробувань, у кожному з яких імовірність появи події A однакова і дорівнює p.

Розглянемо дискретну випадкову величину X, яка визначає кількість появ події A (кількість успіхів) у цій серії випробувань.

Очевидно, що X може набувати значень $0,1,2,...,n$, імовірність яких обчислюють за формулою Бернуллі (дивись формулу (5.1) у роботі [10]):

$$P(X=k)=C_n^k \cdot p^k \cdot q^{n-k}, \quad k=0,1,2,...,n, \tag{2.14}$$

де $q=1-p$.

В такому випадку ДВВ X підкорюється *біноміальному* розподілу ймовірностей, або *розподілу Бернуллі*.

Функція розподілу ймовірностей має вигляд

$$F(x)=\begin{cases} 0, & x \le 0; \\ \sum\limits_{k=0}^{m-1} C_n^k \cdot p^k \cdot q^{n-k}, & m-1 \le x \le m,\ m=1,2,...,n; \\ 1, & x > n. \end{cases} \tag{2.15}$$

Числові характеристики біноміального розподілу

$$M(X)=np,\ D(X)=npq,\ \sigma(X)=\sqrt{npq}. \tag{2.16}$$

Приклад 11. Ймовірність банкрутства однієї компанії на кінець року дорівнює 0,3. Скільки в середньому компаній з 40 збанкротують на кінець року?

Розв'язання. За формулою (2.16) отримаємо

$$M(X)=40 \cdot 0{,}3=12.$$

Відповідь: 12 компаній.

2.4.2 Розподіл Пуассона

Нехай проводиться n незалежних випробувань, у кожному з яких імовірність появи події A однакова і дорівнює p. Якщо кількість випробувань n велика, а ймовірність p досить мала ($p \le 0{,}01$), то ймовірності появи ДВВ X - кількості появ події λ у цій серії випробувань, обчислюються за формулою

$$P(X = k) = \frac{\lambda^k}{k!} \cdot e^{-\lambda}, \; k = 0,1,2,..., \tag{2.17}$$

де

$$\lambda = np. \tag{2.18}$$

Цей закон розподілу називають *розподілом Пуассона,* або розподілом рідкісних подій.

Функція розподілу Пуассона має вигляд

$$F(x) = \begin{cases} 0, & x \le 0; \\ \sum\limits_{k=0}^{n-1} \frac{\lambda^k}{k!} \cdot e^{-\lambda}, & n-1 < x \le n, n = 1,2,3,.... \end{cases} \tag{2.19}$$

Числові характеристики закону Пуассона

$$M(X) = D(X) = \lambda. \tag{2.20}$$

Приклад 12. Підприємство відправило на склад 3000 виробів. Імовірність того, що при перевезенні виріб зазнає пошкодження, дорівнює 0,002. Знайти середню кількість пошкоджених виробів, що прибудуть на склад.

Розв'язання. За формулою (2.18) обчислюємо

$$\lambda = 3000 \cdot 0{,}002 = 6$$

і за формулою (2.20) знаходимо середню кількість пошкоджених виробів

$$M(X) = 6.$$

Відповідь: 6 виробів.

2.4.3 Геометричний розподіл

Нехай проводиться серія випробувань, у кожному з яких подія A (успіх) може з'явитись з імовірністю p і не з'явитись з імовірністю $q=1-p$. Випробування зупиняються при першому успіху.

Розглянемо ДВВ X - кількість випробувань, необхідних для появи події A. Множина можливих значень $x_1=1; x_2=2; x_3=3\ldots$ є необмеженою, а ймовірності відповідно дорівнюють $p_1=p;\ p_2=qp;\ p_3=q^2p\ldots$.

Геометричний розподіл описує кількість спроб, необхідних для отримання першого успіху в схемі незалежних випробувань Бернуллі. Геометричний закон розподілу дискретної випадкової величини X визначається формулою

$$P(X=k)=pq^{k-1},\ k=1,2,3\ldots, \tag{2.21}$$

де p − ймовірність успіху ($q=1-p$).

Геометричний розподіл називається також *розподілом Фарі*.

Функція геометричного розподілу має вигляд

$$F(x)=\begin{cases} 0, & x\le 1; \\ \sum\limits_{k=1}^{n-1} p\cdot q^{k-1}=p\cdot\dfrac{1-q^{n-1}}{1-q}, & n-1<x\le n. \end{cases} \tag{2.22}$$

Числові характеристики геометричного закону розподілу

$$M(X)=\frac{1}{p},\ D(X)=\frac{q}{p^2},\ \sigma(X)=\frac{\sqrt{q}}{p}. \tag{2.23}$$

Приклад 13. Гральний кубик підкидається до першої появи шістки. Обчислити числові характеристики ДВВ X – кількості здійснених підкидань.

Розв'язання. За умовою прикладу $p=\frac{1}{6}, q=\frac{5}{6}$, за формулами (2.23) знаходимо

$$M(X)=\frac{1}{1/6}=6,\quad D(X)=\frac{5/6}{1/36}=30,\ \sigma(X)=\sqrt{30}\approx 5{,}477.$$

Відповідь: $M(X)=6$, $D(X)=30$, $\sigma(X)\approx 5{,}477$.

3 НЕПЕРЕРВНІ ВИПАДКОВІ ВЕЛИЧИНИ

3.1 Інтегральна функція розподілу

Нагадаємо, що випадкову величину називають *неперервною*, якщо множина її можливих значень є проміжком (скінченним або нескінченним). Тому задати неперервну випадкову величину за допомогою таблиці неможливо. Для задавання неперервної випадкової величини (НВВ) застосовуються аналітичний та графічний способи.

Визначення. *Інтегральною функцією розподілу НВВ* X називається

$$F(x)=P(X<x),\quad x\in R, \tag{3.1}$$

де $F(x)$ є неперервною, диференційованою майже скрізь, за винятком можливо окремих ізольованих точок.

До властивостей інтегральної функції, що були перелічені вище (дивись пп. 2.2), додаються ще деякі властивості. Таким чином, отримаємо поповнений перелік.

Властивості інтегральної функції розподілу $F(x)$ для НВВ:

1) $0\le F(x)\le 1$;

2) якщо $x_1<x_2$, то $F(x_1)\le F(x_2)$;

3) $\lim\limits_{x\to-\infty} F(x)=0, \quad \lim\limits_{x\to+\infty} F(x)=1$;

4) імовірність того, що НВВ X набуває будь-якого конкретного значення x_0, дорівнює нулю

$$P(X=x_0)=0. \tag{3.2}$$

5) імовірності потрапляння *НВВ* X у проміжки $(\alpha;\beta), (\alpha;\beta], [\alpha;\beta)$ та $[\alpha;\beta]$ є рівними й обчислюються за формулою

$$P(\alpha \le X \le \beta)=P(\alpha < X \le \beta)=P(\alpha \le X < \beta)=P(\alpha < X < \beta)=F(\beta)-F(\alpha). \tag{3.3}$$

3.2 Щільність розподілу неперервної випадкової величини (диференціальна функція розподілу)

Оскільки інтегральна функція неперервної випадкової величини є диференційованою, то можна розглядати її похідну.

Визначення. *Щільністю розподілу ймовірностей НВВ* X називається похідна від інтегральної функції розподілу НВВ X

$$f(x)=F'(x), \; x\in R. \tag{3.4}$$

Визначення. Графік функції щільності розподілу $f(x)$ називається *кривою розподілу*.

Властивості функції щільності розподілу $f(x)$:

1) $f(x)\ge 0$;

2) $\int\limits_{-\infty}^{+\infty} f(x)dx=1$;

3) $P(\alpha < X < \beta)=\int\limits_{\alpha}^{\beta} f(x)dx.$ (3.5)

Зауваження:

- Щільність розподілу $f(x)$ існує лише для неперервних випадкових величин.

- Графік інтегральної функції розподілу $F(x)$ НВВ X завжди є неперервною кривою, в той час як графік щільності розподілу $f(x)$ може бути розривним.

Приклад 14. НВВ X задана своєю функцією розподілу

$$F(x)=\begin{cases}0, & x\le 0,\\ 2^x-1, & 0<x\le 1,\\ 1, & x>1.\end{cases}$$

Знайти щільність розподілу, побудувати графіки $f(x)$ і $F(x)$. Обчислити ймовірності $P(-0{,}5<X<0{,}5)$, $P(0{,}5\le X\le 1)$.

Розв'язання. За визначенням функції щільності розподілу ймовірностей (3.4) отримаємо

$$f(x)=F'(x)=\begin{cases}0, & x\le 0,\\ 2^x\ln 2, & 0<x\le 1,\\ 0, & x>1;\end{cases}=\begin{cases}2^x\ln 2, & x\in(0;1],\\ 0, & x\notin(0;1].\end{cases}$$

Графіки $f(x)$ та $F(x)$ зображено на рисунках 3.1, 3.2 відповідно.

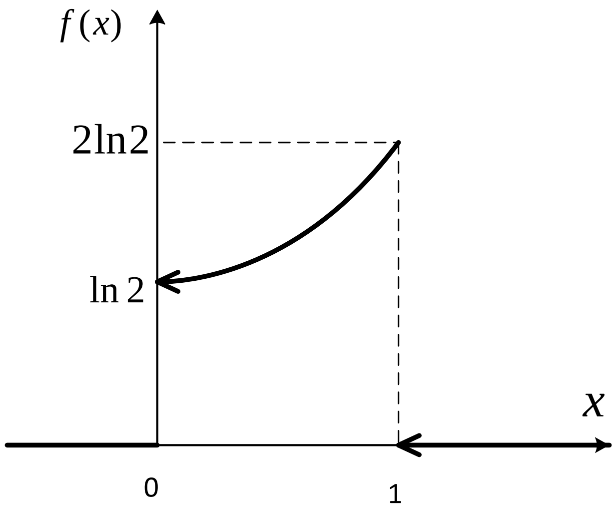

Рисунок 3.1 – Графік щільності розподілу $f(x)$

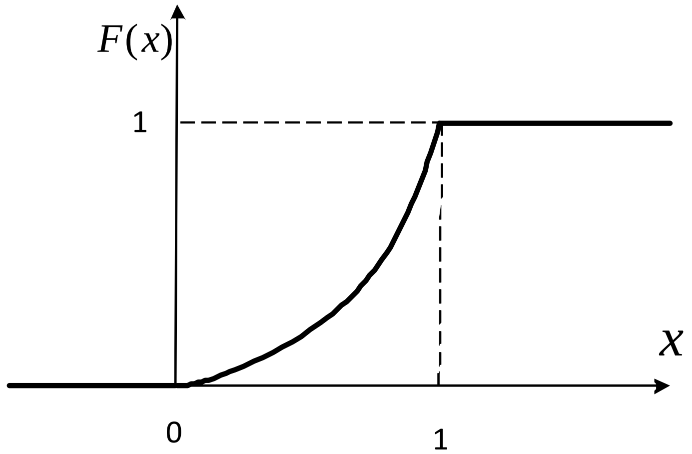

Рисунок 3.2 – Графік інтегральної функції розподілу $F(x)$

Для обчислення ймовірностей потрапляння випадкової величини X в інтервал скористуємось властивостями функції щільності розподілу (3.5):

$$P(-0{,}5 < X < 0{,}5) = \int_{-0{,}5}^{0{,}5} f(x)dx = \int_{-0{,}5}^{0} 0dx + \int_{0}^{0{,}5} 2^x \ln 2dx = \ln 2 \cdot \left.\frac{2^x}{\ln 2}\right|_0^{0{,}5} =$$

$$= 2^{0{,}5} - 2^0 = \sqrt{2} - 1;$$

$$P(0{,}5 \le X \le 1) = \int_{0{,}5}^{1} f(x)dx = \int_{0{,}5}^{1} 2^x \ln 2dx = \ln 2 \cdot \left.\frac{2^x}{\ln 2}\right|_{0{,}5}^{1} = 2 - 2^{0{,}5} = 2 - \sqrt{2}.$$

Зауважимо, що ймовірності $P(-0{,}5 < X < 0{,}5)$, $P(0{,}5 \le X \le 1)$ можна також знайти іншим способом, а саме за формулою (3.3)

$$P(-0{,}5 < X < 0{,}5) = F(0{,}5) - F(-0{,}5) = \left(2^{0{,}5} - 1\right) - 0 = \sqrt{2} - 1;$$
$$P(0{,}5 \le X \le 1) = F(1) - F(0{,}5) = \left(2^1 - 1\right) - \left(2^{0{,}5} - 1\right) = 2 - \sqrt{2}.$$

Відповідь: $P(-0{,}5 < X < 0{,}5) = \sqrt{2} - 1$; $P(0{,}5 \le X \le 1) = 2 - \sqrt{2}$.

3.3 Числові характеристики неперервної випадкової величини

Поняття математичного сподівання, дисперсії і середнього квадратичного відхилення мають такий самий зміст, як і для дискретних випадкових величин. Але у зв'язку з нескінченною множиною можливих значень НВВ формули обчислення числових характеристик є інтегральними.

Математичне сподівання $M(X)$ *НВВ* X визначається формулою

$$M(X) = \int_{-\infty}^{+\infty} x f(x)\,dx. \qquad (3.6)$$

Дисперсія $D(X)$ *НВВ* X обчислюється за формулою

$$D(X) = \int_{-\infty}^{+\infty} x^2 f(x)dx - \left(M(X)\right)^2. \qquad (3.7)$$

Незмінною залишається формула для *середнього квадратичного відхилення НВВ* X

$$\sigma(X) = \sqrt{D(X)}. \qquad (3.8)$$

Приклад 15. Задана щільність розподілу неперервної випадкової величини X

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

Необхідно:

1) побудувати графік $f(x)$;

2) обчислити числові характеристики НВВ X.

Розв'язання:

1) побудуємо графік функції щільності $f(x)$ (рисунок 3.3);

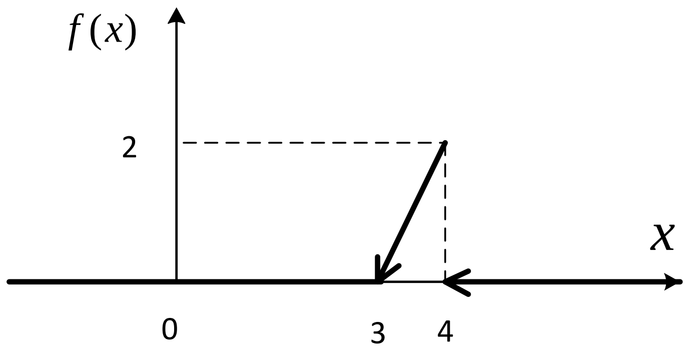

Рисунок 3.3 – Графік функції щільності $f(x)$

2) за формулами (3.6)-(3.8) обчислюємо числові характеристики:

$$M(X)=\int_3^4 x\,f(x)\,dx=\int_3^4 x\cdot(2x-6)\,dx=2\int_3^4\left(x^2-3x\right)dx=2\left(\frac{x^3}{3}-3\frac{x^2}{2}\right)\Bigg|_3^4=$$

$$=2\left[\left(\frac{4^3}{3}-3\cdot\frac{4^2}{2}\right)-\left(\frac{3^3}{3}-3\cdot\frac{3^2}{2}\right)\right]=\frac{11}{3};$$

$$D(X)=\int_3^4 x^2\,f(x)\,dx-\left(M(X)\right)^2=\int_3^4 x^2\cdot(2x-6)\,dx-\left(\frac{11}{3}\right)^2=$$

$$=2\int_3^4\left(x^3-3x^2\right)dx-\left(\frac{11}{3}\right)^2=2\left(\frac{x^4}{4}-3\frac{x^3}{3}\right)\Bigg|_3^4-\left(\frac{11}{3}\right)^2=$$

$$=2\left[\left(\frac{4^4}{4}-4^3\right)-\left(\frac{3^4}{4}-3^3\right)\right]-\left(\frac{11}{3}\right)^2=\frac{1}{18};$$

$$\sigma(X)=\sqrt{\frac{1}{18}}=\frac{1}{3\sqrt{2}}.$$

Відповідь: $M(X)=\dfrac{11}{3}$; $D(X)=\dfrac{1}{18}$; $\sigma(X)=\dfrac{1}{3\sqrt{2}}$.

Приклад 16. Задано щільність розподілу неперервної випадкової величини X

$$f(x)=\begin{cases}0, & x\le 3;\\ C_2x^2+C_1x+C_0, & 3<x\le 6;\\ 0, & x>6,\end{cases}$$

її математичне сподівання $M(X)=4$ і дисперсію $D(X)=\frac{3}{4}$.

Потрібно:

1) обчислити коефіцієнти C_2, C_1, C_0;

2) побудувати графік щільності розподілу ймовірностей $f(x)$.

Розв'язання:

1) за властивістю функції щільності розподілу (3.5) отримаємо

$$\int_3^6\left(C_2x^2+C_1x+C_0\right)dx=1;$$

$$\int_3^6\left(C_2x^2+C_1x+C_0\right)dx=\left(C_2\frac{x^3}{3}+C_1\frac{x^2}{2}+C_0x\right)\Bigg|_3^6=C_2\left(\frac{6^3}{3}-\frac{3^3}{3}\right)+$$

$$+C_1\left(\frac{6^2}{2}-\frac{3^2}{2}\right)+C_0(6-3)=63C_2+13{,}5C_1+3C_0=1.$$

За формулами (3.6), (3.7) обчислюємо числові характеристики НВВ X:

$$M(X)=\int_3^6 x\cdot\left(C_2x^2+C_1x+C_0\right)dx=\int_3^6\left(C_2x^3+C_1x^2+C_0x\right)dx=$$

$$=\left(C_2\frac{x^4}{4}+C_1\frac{x^3}{3}+C_0\frac{x^2}{2}\right)\Bigg|_3^6=C_2\left(\frac{6^4}{4}-\frac{3^4}{4}\right)+C_1\left(\frac{6^3}{3}-\frac{3^3}{3}\right)+C_0\left(\frac{6^2}{2}-\frac{3^2}{2}\right)=$$

$$=303{,}75C_2+63C_1+13{,}5C_0=4;$$

$$D(X)=\int_3^6 x^2\cdot\left(C_2x^2+C_1x+C_0\right)dx-4^2=\int_3^6\left(C_2x^4+C_1x^3+C_0x^2\right)dx-16=$$

$$=\left(C_2\frac{x^5}{5}+C_1\frac{x^4}{4}+C_0\frac{x^3}{3}\right)\Bigg|_3^6-16=C_2\left(\frac{6^5}{5}-\frac{3^5}{5}\right)+C_1\left(\frac{6^4}{4}-\frac{3^4}{4}\right)+$$

$$+C_0\left(\frac{6^3}{3}-\frac{3^3}{3}\right)-16=1506{,}6C_2+303{,}75C_1+63C_0-16=\frac{3}{4}.$$

За умовою прикладу складаємо систему рівнянь

$$\begin{cases}63C_2+13{,}5C_1+3C_0=1;\\303{,}75C_2+63C_1+13{,}5C_0=4;\\1506{,}6C_2+303{,}75C_1+63C_0-16=0{,}75;\end{cases}$$

отримаємо

$$C_2=\frac{5}{27},\ C_1=-\frac{17}{9},\ C_0=\frac{89}{18}.$$

Тоді щільність розподілу має вигляд

$$f(x)=\begin{cases}0, & x\le 3;\\ \frac{5}{27}x^2-\frac{17}{9}x+\frac{89}{18}, & 3<x\le 6;\\ 0, & x>6.\end{cases}$$

2) графік щільності розподілу ймовірностей $f(x)$ зображено на рисунку 3.4.

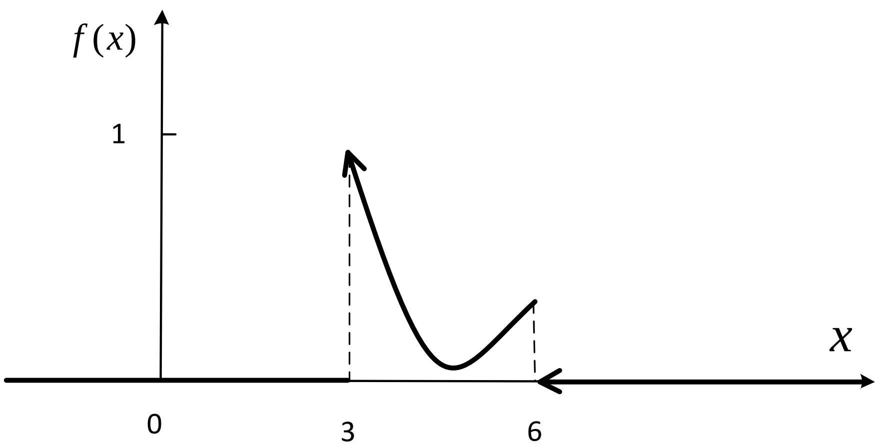

Рисунок 3.4 – Графік щільності розподілу $f(x)$

Зауваження. Аналогічно законам розподілу дискретних випадкових величин (дивись пп. 2.4) існують основні закони розподілу неперервних випадкових величин, до яких належать:

- рівномірний розподіл;
- показниковий (експоненціальний) розподіл;
- нормальний розподіл;
- логнормальний розподіл;
- розподіл Вейбула;
- гамма-розподіл;
- розподіл χ^2;
- розподіл Стьюдента тощо.

4 СИСТЕМИ ДВОХ ВИПАДКОВИХ ВЕЛИЧИН

4.1 Поняття багатовимірної випадкової величини

У попередніх розділах ми розглядали випадкові величини, можливі значення яких складаються з одного числа. Такі випадкові величини називаються *одновимірними.*

Але при вивченні масових випадкових явищ дуже часто необхідно декілька випадкових величин, що їх характеризують, розглядати сумісно, враховуючи зв'язки між ними.

Якщо випадкова величина визначається більш ніж одним числом, вона називається *багатовимірною*.

Визначення. Систему n випадкових величин $X_1; X_2; \ldots; X_n$, кожна з яких набуває того чи іншого значення у результаті одного й того самого випробування, позначають $(X_1; \ldots; X_n)$ і називають *n-вимірною випадковою величиною,* або *n-вимірним випадковим вектором*.

Кожна з випадкових величин $X_1, \ldots, X_n$ n-вимірної випадкової величини $(X_1; \ldots; X_n)$ називається її *складовою,* або *компонентою.*

Багатовимірна випадкова величина *дискретна,* якщо її компоненти є дискретними випадковими величинами. У *неперервної* багатовимірної випадкової величини компоненти неперервні.

Приклад 17. Підприємство виготовляє керамічну плитку. Якщо для кожного виробу контролюються його довжина X_1 і ширина X_2, то маємо двовимірну випадкову величину $(X_1;X_2)$. Якщо ж, крім зазначених параметрів, для кожного виробу контролюється також його товщина X_3, то маємо тривимірну випадкову величину $(X_1;X_2;X_3)$.

Як і для одновимірної випадкової величини, для випадкового вектора вводять поняття закону розподілу і числові характеристики.

Зауваження. В подальшому, для простоти, будемо розглядати тільки двовимірні випадкові величини $(X;Y)$. Отримані положення можна перенести за аналогією на випадкові вектори більшої розмірності.

4.2 Двовимірна дискретна випадкова величина. Закон розподілу

Визначення. *Двовимірною випадковою величиною* $(X;Y)$ називається величина, значення якої визначаються парою чисел $(x;y)$.

Двовимірна випадкова величина також називається *системою двох випадкових величин, або двовимірним випадковим вектором.*

Двовимірна випадкова величина $(X;Y)$ називається *дискретною*, якщо її компоненти X, Y є дискретними випадковими величинами.

Визначення. *Закон розподілу* двовимірної дискретної випадкової величини (ДДВВ) можна задати таблицею (*матрицею розподілу*), в якій перелічені всі можливі пари чисел $(x_i;y_j)$ та відповідні ймовірності їх появи $p_{ij}=p(X=x_i;Y=y_j)$ $i=\overline{1,n};\ j=\overline{1,m}$ (таблиця 4.1).

Таблиця 4.1 – Матриця розподілу

Y	X			
	x_1	x_2	$\ldots$	x_n
y_1	p_{11}	p_{21}	$\ldots$	p_{n1}
y_2	p_{12}	p_{22}	$\ldots$	p_{n2}
$\ldots$	$\ldots$	$\ldots$	$\ldots$	$\ldots$
y_m	p_{1m}	p_{2m}	$\ldots$	p_{nm}

Очевидно, ймовірності матриці розподілу задовольняють умову (*умову нормування*)

$$\sum_{i,j=1}^{n,m} p_{ij} = 1. \tag{4.1}$$

За законом сумісного розподілу системи двох випадкових величин $(X;Y)$ (таблиця 4.1) можна знайти закони розподілів її компонент X і Y.

Щоб знайти *закон розподілу компоненти* X ДДВВ $(X;Y)$, потрібно перелічити її значення x_i, $i=\overline{1,n}$, а ймовірності $p_i = p(X = x_i)$ цих значень знайти, сумуючи ймовірності в стовпчиках матриці розподілу:

$$p_i = p(X = x_i) = \sum_{j=1}^{m} p_{ij} . \tag{4.2}$$

Аналогічно, щоб знайти *закон розподілу компоненти* Y ДДВВ $(X;Y)$, потрібно перелічити значення y_j, $j=\overline{1,m}$, а їх імовірності $q_j = p(Y = y_j)$ знайти як суму в рядках матриці розподілу:

$$q_j = p(Y = y_j) = \sum_{i=1}^{n} p_{ij} . \tag{4.3}$$

Зауваження. Числові характеристики компонент X і Y знаходять як і для одновимірних дискретних випадкових величин, виходячи з законів розподілу X і Y (дивись пп. 2.3).

Приклад 18. Система двох випадкових величин $(X;Y)$ задана законом розподілу.

Y	X		
	-1	3	5
1	0,2	0,3	0,15
4	p_{12}	0,2	0,1

Знайти:

1) невідому ймовірність p_{12};

2) закони розподілу її компонент X і Y та їх числові характеристики $M(X)$, $D(X)$, $\sigma(X)$, $M(Y)$, $D(Y)$, $\sigma(Y)$.

Розв'язання:

1) за умовою нормування (4.1) обчислюємо невідому ймовірність

$$p_{12} = P(X = -1; Y = 4) = 1 - (0,2 + 0,3 + 0,15 + 0,2 + 0,1) = 0,05 .$$

Тоді закон сумісного розподілу ДДВВ $(X;Y)$ має вигляд

Y	X		
	-1	3	5
1	0,2	0,3	0,15
4	0,05	0,2	0,1

2) за формулами (4.2), (4.3) отримаємо закони розподілу компонент X та Y ДДВВ $(X;Y)$ відповідно

X	-1	3	5
P	0,25	0,5	0,25

Y	1	4
Q	0,65	0,35

Обчислюємо математичне сподівання, дисперсію та середнє квадратичне відхилення одновимірних випадкових величин X та Y за формулами (2.7), (2.11), (2.13):

$M(X) = (-1) \cdot 0{,}25 + 3 \cdot 0{,}5 + 5 \cdot 0{,}25 = 2{,}5$;
$D(X) = (-1)^2 \cdot 0{,}25 + 3^2 \cdot 0{,}5 + 5^2 \cdot 0{,}25 - (2{,}5)^2 = 4{,}75$;
$\sigma(X) = \sqrt{4{,}75} \approx 2{,}179$;
$M(Y) = 1 \cdot 0{,}65 + 4 \cdot 0{,}35 = 2{,}05$;
$D(Y) = 1^2 \cdot 0{,}65 + 4^2 \cdot 0{,}35 - (2{,}05)^2 = 2{,}0475$;
$\sigma(Y) = \sqrt{2{,}0475} \approx 1{,}431$.

Відповідь: $M(X) = 2{,}5$; $D(X) = 4{,}75$; $\sigma(X) \approx 2{,}179$;
$M(Y) = 2{,}05$; $D(Y) = 2{,}0475$; $\sigma(Y) \approx 1{,}431$.

4.3 Функція розподілу двовимірної випадкової величини та її властивості

Визначення. *Функцією розподілу (інтегральною функцією)* двовимірної випадкової величини $(X;Y)$ називається функція $F(x,y)$ двох змінних, яка для кожної пари $(x;y)$ значень своїх аргументів дорівнює імовірності того, що випадкові компоненти X і Y виявляться меншими за відповідні значення аргументів:

$$F(x,y) = P(X < x; Y < y). \tag{4.4}$$

З геометричної точки зору значення функції розподілу $F(x,y)$ дорівнює ймовірності потрапляння значення випадкової величини $(X;Y)$ в нескінченний (необмежений зліва і знизу) квадрант, права верхня вершина якого знаходиться в точці з координатами (x,y) (рисунок 4.1).

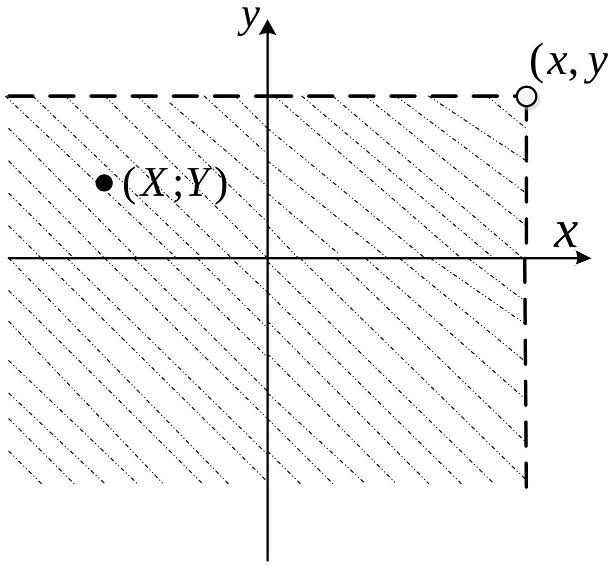

Рисунок 4.1 – Інтегральна функція розподілу двовимірної випадкової величини $(X;Y)$

З визначення випливають такі *властивості* інтегральної функції розподілу $F(x,y)$:

1) $0 \le F(x,y) \le 1$;

2) $F(x,y)$ неспадна функція за кожним аргументом, тобто:

- якщо $x_2 > x_1$, то $F(x_2,y) \ge F(x_1,y)$;
- якщо $y_2 > y_1$, то $F(x,y_2) \ge F(x,y_1)$;

3) справедливі граничні співвідношення

$$F(-\infty,-\infty)=0;\ F(-\infty,y)=0;\ F(x,-\infty)=0;\ F(+\infty,+\infty)=1;$$

4) при $y \to +\infty$ двовимірна інтегральна функція розподілу $F(x,y)$ визначає функцію розподілу $F_X(x)$ компоненти X

$$F_X(x) = F(x;+\infty).$$

Аналогічно, при $x \to +\infty$ двовимірна інтегральна функція розподілу $F(x,y)$ визначає функцію розподілу $F_Y(y)$ компоненти Y

$$F_Y(Y) = F(+\infty;y);$$

5) імовірності потрапляння випадкової точки $(X;Y)$:

- у півсмугу $\{x_1 < X < x_2; \quad Y < y\}$

$$P(x_1 \le X < x_2; y_1 \le Y < y_2) = F(x_2, y) - F(x_1, y);$$

- півсмугу $\{X < x; \quad y_1 < Y < y_2\}$

$$P(x_1 \le X < x_2; y_1 \le Y < y_2) = F(x, y_2) - F(x, y_1);$$

- прямокутник $\{x_1 < X < x_2; \quad y_1 < Y < y_2\}$

$$P(x_1 \le X < x_2; y_1 \le Y < y_2) = F(x_2, y_2) - F(x_1, y_2) - [F(x_2, y_1) - F(x_1, y_1)].$$

Зауваження. Наведене визначення (4.4) і властивості інтегральної функції розподілу $F(x,y)$ є спільними для дискретних і неперервних двовимірних випадкових величин.

4.4 Умовні закони розподілу дискретних двовимірних випадкових величин

Нехай $(X;Y)$ – дискретна двовимірна випадкова величина:
$x_i, i = \overline{1,n}$ – можливі значення компоненти X,
$y_j,\ j = \overline{1,m}$ – можливі значення компоненти Y.

Позначимо через $p(x_i \mid y_j)$ – імовірність того, що $X = x_i$ за умови, що $Y = y_j$.

Визначення. *Умовним розподілом X при $Y = y_j,\ j = \overline{1,m}$* називається розподіл, наведений в таблиці 4.2.

Таблиця 4.2 – Умовний розподіл X *при* $Y = y_j,\ j = \overline{1,m}$

X	x_1	x_2	...	x_n
$P(X \mid y_j)$	$p(x_1 \mid y_j)$	$p(x_2 \mid y_j)$	...	$p(x_n \mid y_j)$

Позначимо через $p(y_j \mid x_i)$ – імовірність того, що $Y = y_j$ за умови, що $X = x_i$.

Визначення. *Умовним розподілом* Y *при* $X = x_i, i = \overline{1,n}$ називається розподіл, наведений в таблиці 4.3.

Таблиця 4.3 – Умовний розподіл Y *при* $X = x_i, i = \overline{1,n}$

Y	y_1	y_2	...	y_m
$P(Y \mid x_i)$	$p(y_1 \mid x_i)$	$p(y_2 \mid x_i)$	...	$p(y_m \mid x_i)$

Умовні ймовірності (дивись пп. 4.7 у роботі [10]) обчислюють за формулами

$$p(x_i \mid y_j) = \frac{p(x_i; y_j)}{p(y_j)}, \tag{4.5}$$

$$p(y_j \mid x_i) = \frac{p(x_i; y_j)}{p(x_i)}, \tag{4.6}$$

де $p_{ij} = p(x_i; y_j) = p(X = x_i; Y = y_j)$ – імовірності появи пари $(x_i; y_j)$, $i = \overline{1,n}$, $j = \overline{1,m}$ (знаходиться з таблиці сумісного розподілу X і Y) (таблиця 4.2);

$p_i = p(x_i) = p(X = x_i)$ – імовірність появи значення x_i, $i = \overline{1,n}$ (знаходиться з безумовного розподілу компоненти X);

$p_j = p(y_j) = p(Y = y_j)$ – імовірності появи значення y_j, $j = \overline{1,m}$ (знаходиться з безумовного розподілу компоненти Y).

Зауваження. Як і для будь-якого розподілу, сума ймовірностей умовного розподілу завжди дорівнює одиниці.

Визначення. *Умовним математичним сподіванням* $M(X \mid Y = y_j)$ дискретної двовимірної випадкової величини називається

$$M(X \mid Y = y_j) = \sum_{i=1}^{n} x_i \cdot p(x_i \mid y_j). \tag{4.7}$$

Визначення. *Умовним математичним сподіванням* $M(Y \mid X = x_i)$ дискретної двовимірної випадкової величини (ДДВВ) називається

$$M(Y \mid X = x_i) = \sum_{j=1}^{m} y_j \cdot p(y_j \mid x_i). \tag{4.8}$$

Приклад 19. ДДВВ $(X;Y)$ задано законом розподілу.

Y	X		
	-1	3	5
1	0,2	0,3	0,15
4	0,05	0,2	0,1

Потрібно знайти:

1) умовний закон розподілу компоненти X при $Y=y_1=1$ та відповідне умовне математичне сподівання;

2) умовний закон розподілу компоненти Y при $X=x_2=3$ та відповідне умовне математичне сподівання.

Розв'язання: 1) з прикладу 18 маємо закон розподілу компоненти Y.

Y	1	4
Q	0,65	0,35

Обчислимо умовні ймовірності X при $Y=y_1=1$ за формулою (4.5):

$$p(x=-1\,|\,y=1)=\frac{p(X=-1;Y=1)}{p(Y=1)}=\frac{0{,}2}{0{,}65}=\frac{4}{13};$$

$$p(x=3\,|\,y=1)=\frac{p(X=3;Y=1)}{p(Y=1)}=\frac{0{,}3}{0{,}65}=\frac{6}{13};$$

$$p(x=5\,|\,y=1)=\frac{p(X=5;Y=1)}{p(Y=1)}=\frac{0{,}15}{0{,}65}=\frac{3}{13}.$$

Таким чином, за таблицею 4.2 умовний закон розподілу компоненти X при $Y=1$ має вигляд

X	-1	3	5
$P(X\,|\,Y=1)$	$\frac{4}{13}$	$\frac{6}{13}$	$\frac{3}{13}$

Контроль: $\frac{4}{13}+\frac{6}{13}+\frac{3}{13}\equiv 1$.

За отриманим умовним розподілом і формулою (4.7) обчислюємо умовне математичне сподівання компоненти X при $Y=1$:

$$M(X\,|\,Y=1)=(-1)\cdot\frac{4}{13}+3\cdot\frac{6}{13}+5\cdot\frac{3}{13}=\frac{29}{13};$$

2) з прикладу 18 маємо закон розподілу компоненти X.

X	-1	3	5
P	0,25	0,5	0,25

Обчислимо умовні ймовірності Y при $X=x_2=3$ за формулою (4.6):

$$p(y=1\,|\,x=3)=\frac{p(X=3;Y=1)}{p(X=3)}=\frac{0{,}3}{0{,}5}=0{,}6;$$
$$p(y=4\,|\,x=3)=\frac{p(X=3;Y=4)}{p(X=3)}=\frac{0{,}2}{0{,}5}=0{,}4.$$

Згідно з таблицею 4.3 умовний закон розподілу компоненти Y при $X=x_2=3$ має вигляд

Y	1	4
$P(Y\,|\,X=3)$	0,6	0,4

Контроль: $0{,}6+0{,}4\equiv1$.

За отриманим умовним розподілом знаходимо умовне математичне сподівання компоненти Y при $X=3$, використовуючи формулу (4.8):

$$M(Y\,|\,X=3)=1\cdot0{,}6+4\cdot0{,}4=2{,}2.$$

Відповідь: 1) $M(X\,|\,Y=1)=\dfrac{29}{13}$; 2) $M(Y\,|\,X=3)=2{,}2$.

4.5 Числові характеристики дискретної двовимірної випадкової величини. Кореляційний момент і коефіцієнт кореляції

Розглянемо найважливіші характеристики дискретної двовимірної випадкової величини.

Визначення. *Математичним сподіванням* дискретної двовимірної випадкової величини $(X;Y)$ називають невипадковий вектор $\left(m_x;m_y\right)$, координатами якого є математичні сподівання відповідних компонент випадкового вектора $(X;Y)$, обчислені за безумовними законами розподілу, тобто

$$\begin{aligned} m_x &= M(X) = \sum_{i=1}^{n} x_i \cdot p_i, \\ m_y &= M(Y) = \sum_{j=1}^{m} y_j \cdot q_j \end{aligned} . \qquad (4.9)$$

Зауваження. Точка з координатами $\left(m_x;m_y\right)$ визначає центр двовимірного розподілу.

Визначення. *Дисперсією* дискретної двовимірної випадкової величини $(X;Y)$ називають невипадковий вектор $\left(\sigma_x^2;\sigma_y^2\right)$, компонентами якого є дисперсії відповідних компонент випадкового вектора $(X;Y)$, обчислені за безумовними законами розподілу, тобто

$$\sigma_x^2 = D(X), \sigma_y^2 = D(Y). \qquad (4.10)$$

Крім числових характеристик компонент випадкового вектора $(X;Y)$, використовують величини, що характеризують зв'язок між його компонентами.

Визначення. *Кореляційним моментом* μ_{xy}, або *коваріацією* $\operatorname{cov}(X;Y)$, випадкових величин X і Y називається математичне сподівання добутку відхилень X і Y від своїх математичних сподівань:

$$\mu_{xy} = \mathrm{cov}(X;Y) = M\lfloor (X - m_x)\cdot(Y - m_y)\rfloor. \qquad (4.11)$$

Для ДДВВ $(X;Y)$ кореляційний момент визначається за формулою

$$\mu_{xy} = \sum_{i=1}^{n}\sum_{j=1}^{m}(x_i - m_x)\cdot\left(y_j - m_y\right)\cdot p_{ij}\,. \qquad (4.12)$$

На практиці для обчислення μ_{xy} можна також використовувати формулу

$$\mu_{xy} = M(XY) - M(X)M(Y)\,, \qquad (4.13)$$

де

$$M(XY) = \sum_{i=1}^{n}\sum_{j=1}^{m} x_i \cdot y_j \cdot p_{ij}\,. \qquad (4.14)$$

Зауваження. Кореляційний момент μ_{xy} характеризує ступінь розсіювання випадкових величин X і Y навколо їх математичного сподівання $\left(m_x;m_y\right)$, а також ступінь лінійної залежності між ними.

Для характеристики тільки ступеня лінійної залежності між випадковими величинами X і Y використовують коефіцієнт кореляції.

Визначення. *Коефіцієнтом кореляції* r_{xy} випадкових величин X і Y називається величина

$$r_{xy} = \frac{\mu_{xy}}{\sigma_x \cdot \sigma_y}, \qquad (4.15)$$

де σ_x, σ_y – середні квадратичні відхилення X і Y відповідно.

Зауваження:

- Значення коефіцієнта кореляції знаходиться в межах від -1 до +1.

• Якщо $r_{xy} \neq 0$, то величини X і Y називаються *корельованими.*

• Якщо X і Y незалежні між собою, то $r_{xy} = 0$, тобто X і Y - *некорельовані.*

• Якщо X і Y пов'язані *лінійною* функціональною залежністю $Y = aX + b$, то $r_{xy} = 1$ при $a > 0$ і $r_{xy} = -1$ при $a < 0$.

• Чим більше абсолютна величина коефіцієнта кореляції r_{xy}, тим ближче статистична залежність величин X і Y до лінійної функціональної.

Приклад 20. Розподіл дискретної двовимірної випадкової величини $(X;Y)$ задано таблицею.

Y	X		
	0	2	5
-3	0,15	0,05	0,1
1	0,3	0,25	0,15

Потрібно:

1) знайти кореляційний момент і коефіцієнт кореляції X і Y;

2) встановити, чи є компоненти ДДВВ $(X;Y)$ залежними (незалежними).

Розв'язання: 1) за законом розподілу ДДВВ $(X;Y)$ обчислюємо:

$$M(XY) = 0 \cdot (-3) \cdot 0{,}15 + 2 \cdot (-3) \cdot 0{,}05 + 5 \cdot (-3) \cdot 0{,}1 + 0 \cdot 1 \cdot 0{,}3 +$$
$$+ 2 \cdot 1 \cdot 0{,}25 + 5 \cdot 1 \cdot 0{,}15 \approx -0{,}55.$$

Складаємо закон розподілу компоненти X

X	0	2	5
P	0,45	0,3	0,25

і за формулами (2.7), (2.12), (2.13) обчислюємо числові характеристики X:

$$M(X)=0\cdot 0{,}45+2\cdot 0{,}3+5\cdot 0{,}25=1{,}85;$$
$$D(X)=0^2\cdot 0{,}45+2^2\cdot 0{,}3+5^2\cdot 0{,}25-(1{,}85)^2=4{,}0275;$$
$$\sigma(X)=\sqrt{4{,}0275}\approx 2{,}007.$$

Аналогічно складаємо закон розподілу компоненти Y

Y	-3	1
Q	0,3	0,7

та обчислюємо числові характеристики Y:

$$M(Y)=-3\cdot 0{,}3+1\cdot 0{,}7=-0{,}2;$$
$$D(Y)=(-3)^2\cdot 0{,}3+1^2\cdot 0{,}7-(-0{,}2)^2=3{,}36;$$
$$\sigma(Y)=\sqrt{3{,}36}\approx 1{,}833.$$

Кореляційний момент і коефіцієнт кореляції компонент X та Y знаходимо за формулами (4.13), (4.15):

$$\mu_{xy}=M(XY)-M(X)M(Y)=-0{,}55-1{,}85\cdot(-0{,}2)=-0{,}18;$$
$$r_{xy}=\frac{\mu_{xy}}{\sigma_x\cdot\sigma_y}=\frac{-0{,}18}{2{,}007\cdot 1{,}833}\approx -0{,}049;$$

2) оскільки кореляційний момент (а тому і коефіцієнт кореляції) не дорівнюють нулю, то компоненти X та Y ДДВВ $(X;Y)$ є корельованими і, як наслідок, залежними.

Відповідь: 1) $\mu_{xy}=-0{,}18$; $r_{xy}\approx -0{,}049$; 2) X і Y - залежні.

4.6 Лінійна регресія

Розглянемо ДДВВ $(X;Y)$, компоненти якої X та Y - залежні випадкові величини. На практиці виникає необхідність знаходження наближеного виразу компоненти Y через X (або навпаки) у вигляді лінійної залежності

$$Y=f(X)=aX+b,\ a,b\in R.$$

Визначення. *Середньоквадратичною регресією* Y *на* X називається така лінійна функція $Y = aX + b$, для якої $M[Y - f(X)]^2 \to \min$.

Лінійна середня квадратична регресія Y на X має вигляд

$$f(X) = m_y + r_{xy} \cdot \frac{\sigma_y}{\sigma_x} \cdot (X - m_x), \tag{4.16}$$

де m_x, m_y – математичні сподівання компонент X та Y відповідно;

σ_x, σ_y – середні квадратичні відхилення X та Y відповідно;

$r_{xy} = \dfrac{\mu_{xy}}{\sigma_x \sigma_y}$ – коефіцієнт кореляції X та Y.

З формулою (4.16) отримаємо рівняння прямої

$$y - m_y = r_{xy} \cdot \frac{\sigma_y}{\sigma_x} \cdot (x - m_x), \tag{4.17}$$

яка називається *прямою середньоквадратичної регресії* Y на X.

Аналогічно, *пряма середньоквадратичної регресії* X на Y має вигляд

$$x - m_x = r_{xy} \cdot \frac{\sigma_x}{\sigma_y} \cdot (y - m_y). \tag{4.18}$$

Приклад 21. ДДВВ $(X;Y)$ задана законом розподілу (приклад 20). Знайти рівняння прямої середньоквадратичної регресії Y на X.

Розв'язання. Враховуючи одержані результати в попередньому прикладі, за формулою (4.17) отримаємо

$$y - (-0{,}2) = -0{,}049 \cdot \frac{1{,}833}{2{,}007} \cdot (x - 1{,}85);$$

$$y = -0{,}045x - 0{,}117.$$

Відповідь: $y = -0{,}045x - 0{,}117$.

5 ЕЛЕМЕНТИ МАТЕМАТИЧНОЇ СТАТИСТИКИ

5.1 Предмет і завдання математичної статистики. Вибірковий метод і його основні поняття

Наука, яка займається розробленням методів збору й обробки дослідних даних з метою виявлення закономірностей масових випадкових явищ, називається *математичною статистикою*.

Важко уявити дослідження та прогнозування в економіці без використання методів, які спираються на математичну статистику.

Предметом математичної статистики є вивчення випадкових величин (або випадкових подій, процесів) за результатами спостережень, дослідів, повторних випробувань.

Завдання математичної статистики полягає у створенні методів збору та обробки статистичних даних для отримання практичних висновків.

Терміном *«сукупність»* у статистиці називається множина об’єктів, з яких здобувається вибірка.

Уся сукупність, що досліджується, називається *генеральною сукупністю.*

Генеральну сукупність можна вивчати шляхом суцільного вивчення всіх об’єктів або шляхом спостереження лише за частиною таких об’єктів. Частина об’єктів, які отримають з генеральної сукупності, називається *вибіркою,* або *вибірковою сукупністю.*

Загальна кількість об’єктів генеральної сукупності чи вибіркової сукупності називається їх об’ємом. Об’єм генеральної сукупності позначають через N, а об’єм вибіркової сукупності – через n. Очевидно, що $n \le N$.

Вибірка називається *випадковою*, якщо з генеральної сукупності будь-який елемент можна брати навмання і кожен з них може потрапити до неї з однаковою імовірністю.

Відбір об’єктів може бути повторним або безповторним.

Повторним називається відбір, коли відібраний об’єкт повертають у генеральну сукупність до відбору наступного об’єкта.

Безповторним називається відбір, коли відібраний об'єкт не повертають до генеральної сукупності.

Якщо випадкова вибірка достатньо повно характеризує генеральну сукупність, то така вибірка називається *репрезентативною.*

5.2 Варіаційний ряд розподілу та його числові характеристики

Припустимо, що з генеральної сукупності взято вибірку об'єктів для вивчення ознаки X. Ознака будь-якої сукупності приймає ряд значень, кількість яких може бути обмеженою або нескінченою. Позначимо можливі значення ознаки сукупності X через $x_1, x_2, ..., x_m$.

Визначення. Значення $x_1, x_2, ..., x_m$ називаються *варіантами* ознаки X.

Послідовність варіант, записаних у зростаючому порядку, називають *дискретним варіаційним рядом.*

Приклад 22. Нехай маємо вибірку з генеральної сукупності 23, 20, 7, 11, 5, 28. Скласти дискретний варіаційний ряд.

Розв'язання. Запишемо надану вибірку у вигляді зростаючої послідовності: 5, 7, 11, 20, 23, 28.

Відповідь: 5, 7, 11, 20, 23, 28.

Деякі значення варіант можуть повторюватись. Наприклад, значення x_1 - n_1 разів, значення x_2 - n_2 разів, … , x_m - n_m разів.

Визначення. *Частотою* $n_i, i = \overline{1, m}$ називають число появи варіанти $x_i, i = \overline{1, m}$.

Дискретний варіаційний ряд розподілу частот записують у вигляді таблиці, де вказують можливі значення $x_i, i = \overline{1, m}$ та відповідні їм частоти $n_i, i = \overline{1, m}$ (таблиця 5.1),

Таблиця 5.1 – Дискретний варіаційний ряд розподілу частот

x_i	x_1	x_2	...	x_m
n_i	n_1	n_2	...	n_m

причому

$$\sum_{i=1}^{m} n_i = n, \qquad (5.1)$$

де n – об’єм вибірки.

У дискретному варіаційному ряді розподілу замість частот можна вказувати відносні частоти.

Визначення. *Відносною частотою* $w_i, i=\overline{1,m}$ називають відношення частоти появи відповідної варіанти до загального об’єму вибірки

$$w_i = \frac{n_i}{n}. \qquad (5.2)$$

Очевидно, що

$$\sum_{i=1}^{m} w_i = 1. \qquad (5.3)$$

Тоді отримаємо дискретний варіаційний ряд розподілу відносних частот (таблиця 5.2).

Таблиця 5.2 – Дискретний варіаційний ряд розподілу відносних частот

x_i	x_1	x_2	...	x_m
w_i	w_1	w_2	...	w_m

При великому об’ємі вибірки користуватись дискретним варіаційним рядом розподілу незручно. У цьому випадку за вибіркою складають *інтервальний варіаційний ряд* розподілу.

Для його побудови весь інтервал розподілу поділяють на k однакових (іноді неоднакових) інтервалів довжиною h. Потім обчислюють частоту n_i' або відносну частоту w_i' потраплянь варіант у кожний інтервал.

Інтервальний варіаційний ряд розподілу частот можна подати у вигляді (таблиця 5.3).

Таблиця 5.3 – Інтервальний варіаційний ряд розподілу частот

Інтервал, $x_{i-1} \div x_i$	$x_0 \div x_1$	$x_1 \div x_2$	$x_2 \div x_3$	…	$x_{k-1} \div x_k$
Частота, n_i'	n_1'	n_2'	n_3'	…	n_k'

де n_i' – сума частот варіант, що потрапили до i-го інтервалу, причому

$$\sum_{i=1}^{k} n_i' = n. \qquad (5.4)$$

Аналогічно складається інтервальний варіаційний ряд розподілу відносних частот (таблиця 5.4).

Таблиця 5.4 – Інтервальний варіаційний ряд розподілу відносних частот

Інтервал, $x_{i-1} \div x_i$	$x_0 \div x_1$	$x_1 \div x_2$	$x_2 \div x_3$	...	$x_{k-1} \div x_k$
Відносна частота, w_i'	w_1'	w_2'	w_3'	...	w_k'

де w_i' – сума відносних частот варіант, що потрапили до i-го інтервалу, причому

$$\sum_{i=1}^{k} w_i' = 1. \qquad (5.5)$$

Приклад 23. За даними про обсяг соціальної підтримки населення за регіонами, у мільйонах грошових одиниць (таблиця 5.5), необхідно побудувати інтервальний варіаційний ряд розподілу частот для ознаки X – розміру наданої підтримки за регіонами, розділивши розмах варіації на 4 однакові інтервали.

Таблиця 5.5 – Дані прикладу 23

150	98	59	208	198	78	90	65	54	195	68	76	95
77	96	55	217	75	100	64	98	90	95	136	90	120

Розв'язання. За таблицею 5.5 обчислюємо розмах варіації $R = x_{\max} - x_{\min} = 217 - 54 = 163$ та довжину кожного інтервалу $h = \frac{R}{4} = \frac{163}{4} = 40{,}75$. Тоді інтервальний варіаційний ряд розподілу частот має вигляд

Інтервал, $x_{i-1} \div x_i$	$54 \div 94{,}75$	$94{,}75 \div 135{,}5$	$135{,}5 \div 176{,}25$	$176{,}25 \div 217$
Частота, n_i'	13	7	2	4

Контороль: $\sum_i n_i' = 13 + 8 + 1 + 4 = 26.$

Отже, найбільш розповсюдженим є розмір підтримки в межах від 54 до 94,75 млн грн.

Зауваження. Якщо варіанта потрапляє на межу інтервалу, то її включають лише один раз до лівого чи правого інтервалу, керуючись певним принципом.

5.3 Числові характеристики вибіркової сукупності

Розглянутий варіаційний ряд дає вичерпну характеристику статистичних даних. Проте іноді достатньо знати лише окремі ознаки даного варіаційного ряду. Числа, які є кількісним виразом таких ознак, називаються *числовими характеристиками вибірки*.

До основних числових характеристик вибірки належать *вибіркова середня, вибіркова дисперсія та вибіркове середнє квадратичне відхилення*. Ці числові характеристики ознаки X проведеної вибірки є аналогами відповідно математичного сподівання, дисперсії та середнього квадратичного відхилення випадкової величини X.

Визначення. *Вибірковою середньою* ознаки X вибірки називається число $\bar{x}_в$, що обчислюється за формулою

$$\bar{x}_в = \frac{1}{n} \cdot \sum_{i=1}^{m} x_i \cdot n_i , \tag{5.6}$$

де x_i, $n_i, i = \overline{1,m}$ – варіанти та відповідні їм частоти;

n – об’єм вибірки.

Зауваження. Коли частоти кожної варіанти, яка належить дискретному ряду розподілу, дорівнюють одиниці, вибіркова середня обчислюється за формулою

$$\bar{x}_в = \frac{1}{n} \cdot \sum_{i=1}^{n} x_i . \tag{5.7}$$

Вибіркове середнє $\overline{x}_в$ характеризує середнє значення ознаки X. Тому в статистиці часто замість $\overline{x}_в$ записують $\overline{x}$.

Визначення. *Вибірковою дисперсією* $D_в$ називається число, яке дорівнює середньому арифметичному квадратів відхилень варіант від вибіркової середньої:

$$D_в = \frac{1}{n} \cdot \sum_{i=1}^{m} \left(x_i - \overline{x}_в\right)^2 \cdot n_i. \tag{5.8}$$

Зручніше вибіркову дисперсію обчислювати за формулою

$$D_в = \frac{1}{n} \cdot \sum_{i=1}^{m} x_i^2 n_i - \left(\overline{x}_в\right)^2 = \overline{x^2} - \left(\overline{x}\right)^2, \tag{5.9}$$

де $\overline{x}$ – середнє значення ознаки X;

$\overline{x^2}$ – середнє значення ознаки X^2.

Визначення. *Вибірковим середнім квадратичним відхиленням* називається

$$\sigma_в = \sqrt{D_в}. \tag{5.10}$$

Зауваження. Якщо вибірка задана у вигляді інтервального варіаційного ряду розподілу частот (таблиця 5.3), то для обчислення $\overline{x}_в$, $D_в$, $\sigma_в$ треба побудувати дискретний варіаційний ряд розподілу частот вибірки (таблиця 5.6).

Таблиця 5.6 – Побудова дискретного варіаційного ряду розподілу частот вибірки за інтервальним рядом розподілу

z_i	z_1	z_2	…	z_m
n_i	n_1	n_2	…	n_m

де

$$z_i = \frac{x_i + x_{i+1}}{2}, \ i = 1,2,...,m \tag{5.11}$$

та за таблицею 5.6. знайти числові характеристики.

Вибірка також характеризується допоміжними числовими характеристиками, а саме модою, медіаною і коефіцієнтом варіації.

Визначення. *Модою* M_0 називається та варіанта варіаційного ряду, яка має найбільшу частоту.

Визначення. *Медіаною* M_e у статистиці називається варіанта, яка рівновіддалена від кінців варіаційного ряду. Медіана ділить ряд навпіл: по обидва боки від неї знаходиться однакова кількість одиниць вибірки.

Медіану M_e визначають за формулами

$$M_e = \frac{1}{2}\left(x_{\frac{n}{2}} + x_{\frac{n}{2}+1} \right), \text{ якщо } n - \text{парне};$$

$$M_e = x_{\frac{n+1}{2}}, \text{ якщо } n - \text{непарне}, \tag{5.12}$$

де індекси x_i дорівнюють нумерації варіант у варіаційному ряді.

Визначення. *Коефіцієнт варіації* обчислюється за формулою

$$\nu = \frac{\sigma_в}{\left|\overline{x}_в\right|} \cdot 100\% \tag{5.13}$$

і використовується для порівняння величин розсіювання по відношенню до вибіркової середньої двох варіаційних рядів.

Приклад 24. Задано вибірку 73, 70, 68, 73, 70, 74, 65, 73, 71, 66, 69, 78, 70, 67, 67, 67, 76, 71, 72, 68, яка характеризує квартальний прибуток підприємства, у грошових одиницях. Скласти дискретний варіаційний ряд розподілу частот. Обчислити числові характеристики вибірки.

Розв'язання. Складаємо дискретний варіаційний ряд розподілу частот за таблицею 5.1.

x_i	65	66	67	68	69	70	71	72	73	74	76	78
n_i	1	1	3	2	1	3	2	1	3	1	1	1

За формулою (5.1) об’єм вибірки дорівнює $n=20$.

За формулою (5.6) обчислюємо

$$\bar{x}_в=\frac{1}{20}\cdot(65\cdot1+66\cdot1+67\cdot3+68\cdot2+69\cdot1+70\cdot3+71\cdot2+72\cdot1+$$
$$+73\cdot3+74\cdot1+76\cdot1+78\cdot1)=70{,}4.$$

Знаходимо вибіркову дисперсію $D_в$ за формулою (5.9)

$$D_в=\frac{1}{20}\cdot(65^2\cdot1+66^2\cdot1+67^2\cdot3+68^2\cdot2+69^2\cdot1+70^2\cdot3+71^2\cdot2+$$
$$+72^2\cdot1+73^2\cdot3+74^2\cdot1+76^2\cdot1+78^2\cdot1)-(70{,}4)^2=11{,}14$$

та вибіркове середнє квадратичне відхилення за формулою (5.10)

$$\sigma_в=\sqrt{11{,}14}\approx3{,}338.$$

За визначенням мода $M_0=67$, $M_0=70$, $M_0=73$.

Для знаходження медіани M_e використовуємо формулу (5.12), враховуючи, що $n=20$,

$$M_e=\frac{1}{2}(x_{10}+x_{11})=\frac{1}{2}(70+70)=70.$$

За формулою (5.13) обчислюємо коефіцієнт варіації

$$v=\frac{3{,}338}{70{,}4}\cdot100\%\approx4{,}741\%.$$

Відповідь: $\bar{x}_в=70{,}4$; $D_в=11{,}14$; $\sigma_в\approx3{,}338$; $M_0=67=70=73$, $M_e=70$, $v\approx4{,}741\%$.

5.4 Емпірична функція розподілу

Статистичний ряд розподілу випадкової величини X, отриманий за емпіричними даними, також називають *емпіричним законом розподілу.*

Визначення. *Емпіричною функцією розподілу* (функцією розподілу вибірки) називається функція $F^*(x)$, яка визначає для кожного значення x відносну частоту події $X < x$, тобто

$$F^*(x) = \frac{n_x}{n}, \qquad (5.14)$$

де n_x – сумарна частота варіант, які менше від x;
n – об’єм вибірки.

Таким чином, емпірична функція розподілу визначається шляхом послідовного додавання відносних частот варіант, менших за x.

Графіком емпіричної функції розподілу $F^*(x)$ є *кумулята,* або графік накопичених відносних частот. Цей графік – ступенева фігура, яка має точки розриву при значеннях абсцис, що дорівнюють числовим значенням, яких набуває випадкова величина X.

Зауваження. Інтегральна функція розподілу $F(x)$ генеральної сукупності у математичній статистиці називається *теоретичною функцією* розподілу. Різниця між емпіричною і теоретичною функціями полягає в тому, що теоретична функція $F(x)$ визначає імовірність події $X < x$, а емпірична функція $F^*(x)$ – відносну частоту цієї події. Можна стверджувати, що при великих n значення функцій $F^*(x)$ і $F(x)$ мало відрізняються, а саме

$$\lim_{n \to +\infty} P\left\{\left|F(x) - F^*(x)\right| < \varepsilon\right\} = 1, \ \varepsilon > 0.$$

Таким чином, емпірична функція розподілу $F^*(x)$ може бути використана для оцінювання теоретичної функції розподілу $F(x)$ у генеральній сукупності.

Властивості емпіричної функції $F^*(x)$***:***

1) $0 \le F^*(x) \le 1$;

2) $F^*(x)$ – неспадна функція;

3) $F^*(x)=\begin{cases}0 & \text{при } x\le x_{\min},\\ 1 & \text{при } x> x_{\max},\end{cases}$ (5.15)

де $x_{\min}$ – найменша варіанта;

$x_{\max}$ – найбільша варіанта.

Приклад 25. Вибірка задана дискретним варіаційним рядом розподілу частот.

x_i	2	5	10	12
n_i	25	15	50	10

Знайти емпіричну функцію розподілу та побудувати її графік.

Розв'язання. За умовою $n=100,\ x_{\min}=2,\ x_{\max}=12.$

За другою властивістю емпіричної функції розподілу $F^*(x)=0$ при $x\le 2$. Значення $X<5$, а саме $x_1=2$ зустрічається 25 разів, отже, $F^*(x)=\dfrac{25}{100}=0{,}25$ при $2<x\le 5$. Значення $X<10$, а саме $x_1=2, x_2=5$ спостерігалось 25+15=40 разів, тому $F^*(x)=\dfrac{40}{100}=0{,}4$ при $5<x\le 10$. Значення $X<12$, тобто $x_1=2, x_2=5, x_3=10$ спостерігалось 25+15+50=90 разів, отже, $F^*(x)=\dfrac{90}{100}=0{,}9$ при $10<x\le 12$. Оскільки $x_{\max}=12$, то $F^*(x)=1$ при $x>12$.

Таким чином, емпірична функція розподілу має вигляд

$$F^*(x)=\begin{cases}0, & x\le 2;\\ 0{,}25, & 2<x\le 5:\\ 0{,}4, & 5<x\le 10;\\ 0{,}9, & 10<x\le 12;\\ 1, & x>12.\end{cases}$$

Графік цієї функції зображено на рисунку 5.1.

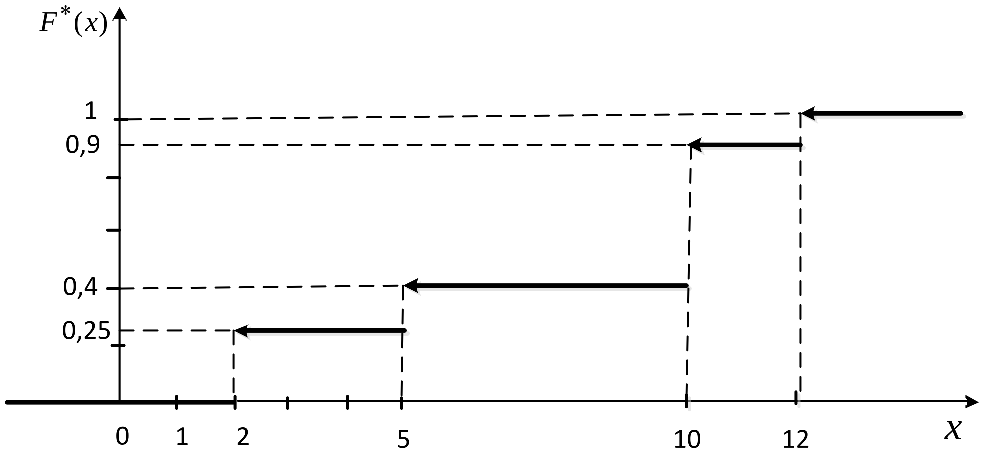

Рисунок 5.1 – Графік емпіричної функції розподілу $F^*(x)$ прикладу 25

Приклад 26. Знайти емпіричну функцію за даним інтервальним варіаційним рядом розподілу частот вибірки.

Інтервали	$0{,}5 \div 2{,}5$	$2{,}5 \div 4{,}5$	$4{,}5 \div 6{,}5$	$6{,}5 \div 8{,}5$
Частота, n_i	10	8	4	3
Накопичувана відносна частота, $\frac{n_i'}{n}$	$\frac{10}{25}=0{,}4$	$\frac{10+8}{25}=0{,}72$	$\frac{10+8+4}{25}=0{,}88$	$\frac{10+8+4+3}{25}=1$

Розв'язання. Записуємо у третій рядок даної таблиці накопичувану відносну частоту відповідно до кожного інтервалу та будуємо графік емпіричної функції $F^*(x)$ (рисунок 5.2).

Запишемо отриману функцію, складаючи рівняння прямих на кожному інтервалі [1]:

$$F^*(x)=\begin{cases} 0, & x \le 0{,}5; \\ 0{,}2(x-0{,}5), & 0{,}5 < x \le 2{,}5; \\ 0{,}16(x-2{,}5)+0{,}4, & 2{,}5 < x \le 4{,}5; \\ 0{,}08(x-4{,}5)+0{,}72, & 4{,}5 < x \le 6{,}5; \\ 0{,}06(x-6{,}5)+0{,}88, & 6{,}5 < x \le 8{,}5; \\ 1, & x > 8{,}5. \end{cases}$$

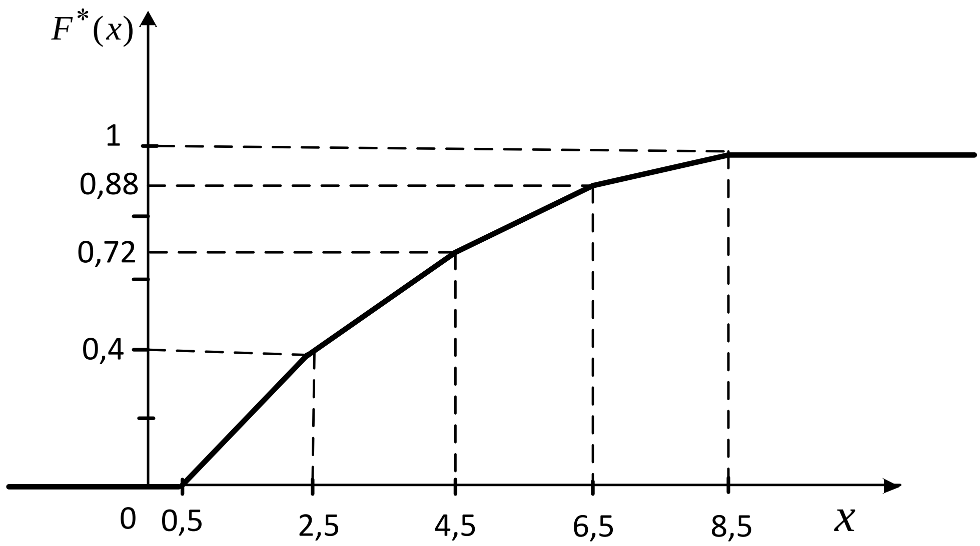

Рисунок 5.2 – Графік емпіричної функції $F^*(x)$ прикладу 26

Зауваження. Для дискретного статистичного розподілу (таблиця 5.1) графік емпіричної функції $F^*(x)$ – розривна східчаста лінія, а для інтервального розподілу (таблиця 5.3) – неперервна ламана лінія.

5.5 Полігон. Гістограма

Визначення. *Полігоном частот* називається ламана, відрізки якої послідовно з'єднують точки $M_1(x_1;n_1), M_2(x_2;n_2),...,M_m(x_m;n_m)$.

Полігоном відносних частот називається ламана, відрізки якої послідовно з'єднують точки $N_1(x_1;w_1), N_2(x_2;w_2),...,N_m(x_m;w_m)$.

Приклад 27. Побудувати полігон частот, полігон відносних частот за даним розподілом вибірки.

x_i	2,5	4,5	6,5	8,5
n_i	8	10	4	3

Розв'язання. Відкладаємо в системі координат точки $M_1(2{,}5;\,8), M_2(4{,}5;\,10)$, $M_3(6{,}5;\,4), M_4(8{,}5;\,3)$ та з'єднуємо їх відрізками прямих. Отримаємо полігон частот цієї вибірки (рисунок 5.3).

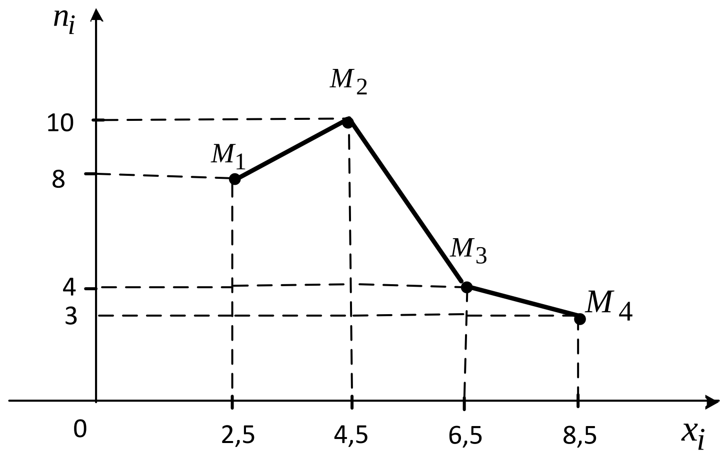

Рисунок 5.3 – Полігон частот

Для побудови полігона відносних частот складаємо дискретний варіаційний ряд розподілу відносних частот.

x_i	2,5	4,5	6,5	8,5
w_i	0,32	0,4	0,16	0,12

Відкладаємо в системі координат точки $N_1(2{,}5;\,0{,}32), N_2(4{,}5;\,0{,}4)$, $N_3(6{,}5;\,0{,}16), N_4(8{,}5;\,0{,}12)$ та з'єднуємо їх відрізками. Отримаємо полігон відносних частот (рисунок 5.4).

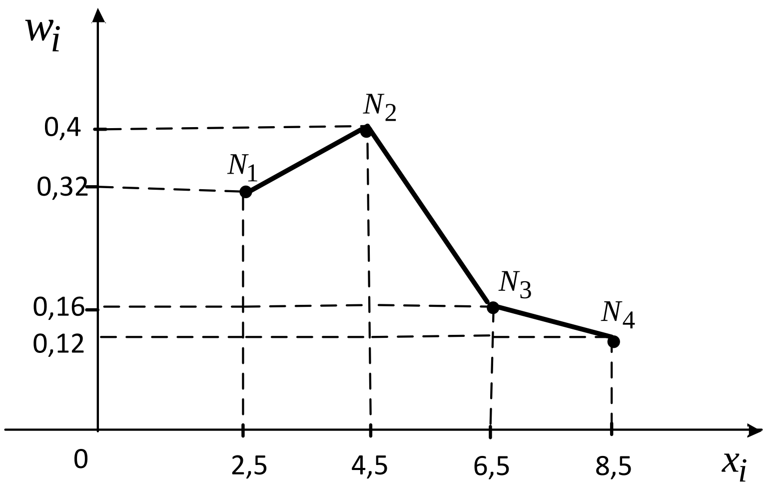

Рисунок 5.4 – Полігон відносних частот

У випадку інтервальних варіаційних рядів розподілу, отриманих для неперервної випадкової величини, для графічного зображення використовують гістограму. Розрізняють гістограму частот і гістограму відносних частот.

Визначення. *Гістограмою частот* називається східчаста фігура, яка складається з прямокутників, основами яких є часткові інтервали варіант довжиною h, а висоти дорівнюють щільності частоти $\frac{n_k}{h}$.

Для побудови гістограми частот на осі абсцис відкладають часткові інтервали довжиною h, а над ними проводять відрізки, паралельні осі абсцис на відстані $\frac{n_k}{h}$.

Площа k-го часткового прямокутника дорівнює $h \cdot \frac{n_k}{h} = n_k$ – сумі частот варіант k-го інтервалу. Площа гістограми дорівнює $S = \sum_{k=1}^{m} n_k = n$, тому площа гістограми дорівнює об'єму вибірки.

Визначення. *Гістограмою відносних частот* називається східчаста фігура, яка складається з прямокутників, основами яких є часткові інтервали довжиною h, а висоти дорівнюють щільності відносної частоти $\frac{w_k}{h}$.

Для побудови гістограми відносних частот на осі абсцис відкладають часткові інтервали довжиною h, а над ними проводять відрізки, паралельні осі абсцис на відстані $\frac{w_k}{h}$ від неї.

Площа k-го часткового прямокутника дорівнює $h \cdot \frac{w_k}{h} = w_k$ – відносній частоті k-го інтервалу. Площа всієї гістограми дорівнює $S = \sum_{k=1}^{m} w_k = 1$, тобто площа гістограми відносних частот дорівнює 1.

Зауваження. Вигляд гістограми відносних частот дає уявлення про щільність розподілу ймовірностей випадкової величини, яку досліджують. Гістограма відносних частот – це наближений графік щільності ймовірностей випадкової величини, тобто наближений графік диференціальної функції її розподілу.

Приклад 28. За даним розподілом вибірки побудувати гістограми частот і відносних частот.

Частковий інтервал, $x_{i-1} \div x_i$	$2 \div 4$	$4 \div 6$	$6 \div 8$	$8 \div 10$	$10 \div 12$	$12 \div 14$	$14 \div 16$
Частота, n_i	18	14	8	4	3	2	1

Розв'язання. Довжина кожного часткового інтервалу дорівнює $h = x_{i+1} - x_i = 2$. Для кожного інтервалу обчислимо значення щільності частот $\frac{n_i}{h}$ та щільності відносних частот $\frac{w_i}{h}$. Отримаємо таблицю.

Частковий інтервал, $x_{i-1} \div x_i$	$2 \div 4$	$4 \div 6$	$6 \div 8$	$8 \div 10$	$10 \div 12$	$12 \div 14$	$14 \div 16$
Щільність частоти інтервалу, $\frac{n_i}{h}$	9	7	4	2	1,5	1	0,5
Відносна частота, w_i	0,36	0,28	0,16	0,08	0,06	0,04	0,02
Щільність відносної частоти інтервалу, $\frac{w_i}{h}$	0,18	0,14	0,08	0,04	0,03	0,02	0,01

Побудуємо гістограму частот (рисунок 5.5) та гістограму відносних частот (рисунок 5.6).

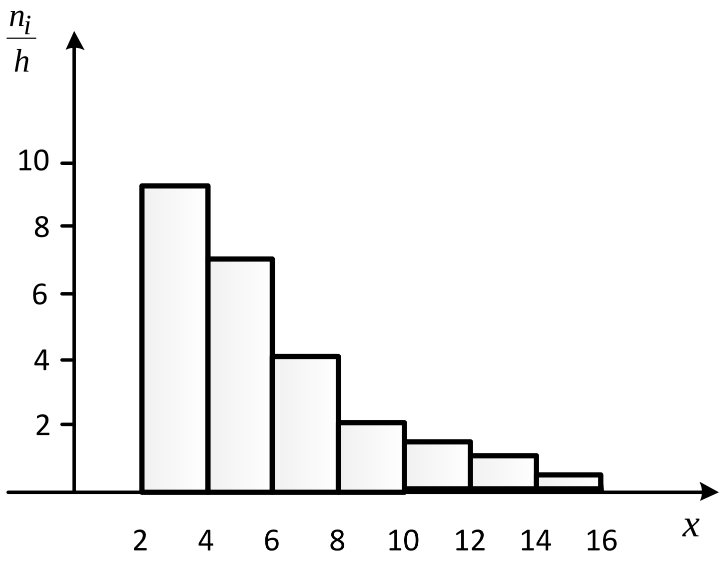

Рисунок 5.6 – Гістограма частот

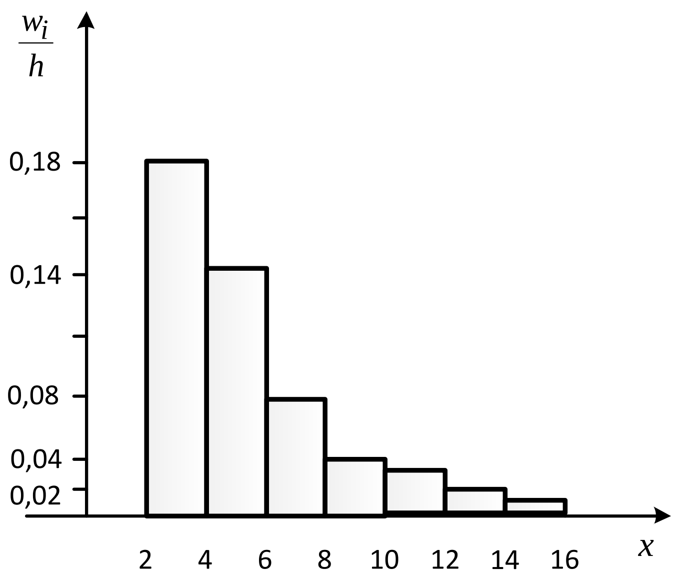

Рисунок 5.6 – Гістограма відносних частот

5.6 Коефіцієнт кореляції

Значне місце серед методів економічного аналізу посідає кореляційний аналіз.

Кореляційний аналіз – це метод дослідження взаємозалежності ознак у генеральній сукупності, які є випадковими величинами з нормальним характером розподілу.

Основними вимогами до застосування кореляційного аналізу є достатня кількість спостережень, сукупності факторних і результативних показників. Правильне застосування кореляційних методів дає змогу зрозуміти сутність процесів взаємозв’язків.

Для оцінювання тісноти взаємозв'язку в кореляційному аналізі використовується коефіцієнт кореляції, який дає кількісну оцінку статистичного взаємозв'язку між результатами вимірювань.

Коефіцієнт кореляції r_{yx} в статистиці – це показник кореляції (лінійної залежності) між двома змінними Y та X, який набуває значень від

«-1» до «+1» включно, та обчислюється за формулою

$$r_{yx} = \frac{\overline{y \cdot x} - \bar{y} \cdot \bar{x}}{\sigma_y \cdot \sigma_x}, \qquad (5.16)$$

де $\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$ – середнє значення величини X;

$\bar{y} = \frac{1}{n}\sum_{i=1}^{n} y_i$ – середнє значення величини Y;

$\overline{xy} = \frac{1}{n}\sum_{i=1}^{n} x_i y_i$, $\overline{x^2} = \frac{1}{n}\sum_{i=1}^{n} x_i^2$;

$\sigma_x^2 = \overline{x^2} - \left(\bar{x}\right)^2$ – вибіркова дисперсія величини X;

$\sigma_y^2 = \overline{y^2} - \left(\bar{y}\right)^2$ – вибіркова дисперсія величини Y;

σ_x, σ_y – середнє квадратичне відхилення X та Y відповідно.

Міцність зв’язку між двома ознаками Y та X виражається за допомогою значень $|r_{yx}|$ (таблиця 5.7).

Таблиця 5.7 – Кількісна та якісна міра кореляції

$\lvert r_{yx} \rvert = 0$	зв’язок відсутній
$0 < \lvert r_{yx} \rvert < 0{,}3$	зв’язок практично відсутній
$0{,}3 \le \lvert r_{yx} \rvert < 0{,}5$	зв’язок помірний
$0{,}5 \le \lvert r_{yx} \rvert < 0{,}7$	зв’язок середній
$0{,}7 \le \lvert r_{yx} \rvert < 0{,}9$	зв’язок сильний
$0{,}9 \le \lvert r_{yx} \rvert < 1$	зв’язок дуже сильний
$\lvert r_{yx} \rvert = 1$	випадкові величини пов’язані лінійною залежністю, цей зв’язок є функціональним

Приклад 29. Обчислити кореляційну залежність між величинами.

Номер спостереження	Обсяги торгів цінними паперами Y, млрд грн	Обсяги іноземних інвестицій X, млрд дол. США
1	39,2	3,9
2	68,5	4,4
3	108,6	5,3
4	203	6,7
5	321,3	8,4
6	403,8	16,4
7	492,8	21,2
8	754,3	29,5
9	883,4	35,7
10	1067,3	40,0

Розв’язання. Обчислюємо середні значення величин Y, X, XY, Y^2 та X^2:

$$\overline{y} \approx 434{,}22;\ \overline{x} \approx 17{,}15;\ \overline{xy} \approx 11863{,}98;\ \overline{y^2} \approx 305686{,}52;\ \overline{x^2} \approx 464{,}13;$$

вибіркові дисперсії величин Y та X:

$$\sigma_y^2 = 305686{,}52 - (434{,}22)^2 \approx 117139{,}51;$$
$$\sigma_x^2 = 464{,}13 - (17{,}15)^2 \approx 170{,}01.$$

Таким чином, за формулою (5.16) отримаємо коефіцієнт кореляції

$$r_{yx} = \frac{11863{,}98 - 434{,}22 \cdot 17{,}15}{\sqrt{117139{,}51} \cdot \sqrt{170{,}01}} \approx 0{,}99.$$

Відповідь: $r_{yx} \approx 0{,}99$, отже, зв’язок між обсягами торгів цінними паперами та обсягами іноземних інвестицій дуже сильний.

СПИСОК ЛІТЕРАТУРИ

1 Елементи теорії ймовірностей і математичної статистики в управлінні процесами перевезень : навч. посіб. / Т. В. Бутько, Р. В. Вовк, Н. Г. Панченко, А. П. Рибалко. Харків : УкрДАЗТ, 2011. 308 с.

2 Барковський В. В., Барковська Н. В. Математика для економістів: Вища математика. Київ : Національна академія управління, 1997. 397 с.

3 Валєєв К. Г., Джалладова І. А. Вища математика : навч. посіб. Київ : КНЕУ, 2001. Ч. 1. 546 с.

4 Васильченко І. П. Вища математика для економістів : підручник. Київ : Знання-Прес, 2002. 454 с.

5 Вища математика : підручник. Кн. 1. Основні розділи / за ред. Г. Л. Кулініча. Київ : Либідь, 2003. 400 с.

6 Вища математика : підручник. Кн. 2. Основні розділи / за ред. Г. Л. Кулініча. Київ : Либідь, 2003. 368 с.

7 Дубовик В. П., Юрик І. І. Вища математика : навч. посіб. Київ : А.С.К., 2001. 648 с.

8 Неміщ В. М., Процик А. І., Березька К. М. Вища математика (практикум) : навч. посіб. Тернопіль : Економічна думка, 2001. 266 с.

9 Коваленко Л. Б., Станішевський С. О. Збірник тестових завдань для менеджерів : навч. посіб. Харків : ХНАМГ, 2010. 423 с.

10 Панченко Н. Г., Резуненко М. Є. Вища та прикладна математика : конспект лекцій. Харків : УкрДУЗТ, 2020. Ч. 3. 50 с.

Н. Г. Панченко, М. Є. Резуненко

ВИЩА ТА ПРИКЛАДНА МАТЕМАТИКА

Конспект лекцій

Частина IV

Відповідальний за випуск Панченко Н. Г.

Редактор Третьякова К. А.

Підписано до друку 30.10.20 р.

Формат паперу 60х84 1/16. Папір писальний.

Умовн.-друк.арк. 3,5. Тираж 5. Замовлення №

Видавець та виготовлювач Український державний університет залізничного транспорту,
61050, Харків-50, майдан Фейєрбаха, 7.
Свідоцтво суб'єкта видавничої справи ДК № 6100 від 21.03.2018 р.