МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ
ДЕРЖАВНИЙ ВИЩИЙ НАВЧАЛЬНИЙ ЗАКЛАД
«НАЦІОНАЛЬНИЙ ГІРНИЧИЙ УНІВЕРСИТЕТ»

ЕЛЕМЕНТИ ТЕОРІЇ ЙМОВІРНОСТЕЙ ТА МАТЕМАТИЧНОЇ СТАТИСТИКИ В ГІРНИЦТВІ

Навчальний посібник

Дніпропетровськ
НГУ
2015

УДК 519.21:519.22 (075.8)
ББК 22.17:33.1 (я 73)
Е 50

Рекомендовано вченою радою НГУ як навчальний посібник для студентів технічних спеціальностей різних форм навчання (протокол № 9 від 13.11.2014).

Рецензенти:

Назаренко В.А. – д-р техн. наук, проф. (Державний вищий навчальний заклад «Національний гірничий університет», проф. кафедри маркшейдерії);

Сторчай В.Ф. – канд. фіз.-мат. наук, доц. (Державний вищий навчальний заклад «Національний гірничий університет», проф. кафедри вищої математики)

Е 50 Елементи теорії ймовірностей та математичної статистики в гірництві: Навч. посібник. / О.О. Сдвижкова, О.В. Бугрим, Д.В. Бабець, О.С. Іванов; М-во освіти і науки України, Нац. гірн. ун-т. – Д. : НГУ, 2015. – 103 с.

ISBN 978-966-350-540-4

Посібник містить короткі відомості з теорії ймовірностей та математичної статистики, що необхідні для побудови статистичних моделей. Також в книзі наведено приклади безпосереднього використання для розв’язання практичних завдань. Основною метою посібника є ілюстрація застосування статистичних методів для вирішення завдань гірництва, зокрема геомеханіки.

УДК 519.21:519.22 (075.8)
ББК 22.17:33.1 (я 73)

ISBN 978-966-350-540-4

ЗМІСТ

Вступ 5
Статистичні методи в проблемах гірництва 6
Тема 1. Основні поняття, що використовуються в теорії ймовірностей 8
1.1. Випадкові події. Імовірність події. Частота події 8
1.2. Види випадкових подій 9
1.3. Основні теореми теорії ймовірностей. Алгебра подій 10
1.3.1. Теорема додавання ймовірностей 10
1.3.2. Теорема множення ймовірностей 11
1.3.3. Формула повної ймовірності 12
1.3.4 Імовірність гіпотез. Формула Баєса 13
Тема 2. Випадкові величини і закони їх розподілу 16
2.1. Поняття випадкової величини 16
2.2. Інтегральна функція розподілу 17
2.3. Диференціальна функція розподілу випадкової величини 20
Тема 3. Побудова статистичного розподілу кількісної ознаки (елементи математичної статистики) 24
3.1. Сутність вибіркового методу 24
3.2. Побудова інтервального ряду. Гістограма частот 25
Тема 4. Числові характеристики випадкової величини. Моменти розподілу. Визначення їх за даними досліду 29
4.1. Основні характеристики розташування і варіації випадкової величини 29
4.2. Моменти розподілу 31
4.3. Визначення числових характеристик кількісної ознаки за дослідними даними. Параметри статистичного розподілу (елементи математичної статистики) 34
Тема 5. Деякі теоретичні закони розподілу 39
5.1. Показниковий закон розподілу 39
5.2. Нормальний закон розподілу 42
5.3. Логарифмічно нормальний розподіл 47
5.4. Гамма-розподіл 51
5.5. Вибір теоретичного розподілу за експериментальними даними. 52
5.6. Деякі теоретичні розподіли, що застосовуються в математичній статистиці. Розподіл «Хі-квадрат» 54
5.7. Розподіл Стьюдента 55

Тема 6. Інтервальне оцінювання невідомих параметрів розподілу 59
6.1. Характеристики похибок випробування 59
6.2. Довірчий інтервал для математичного сподівання нормальної випадкової величини X при відомому середньоквадратичному відхиленні 60
6.3. Довірчий інтервал для математичного очікування нормальної величини X при невідомому квадратичному відхиленні σ 62
Тема 7. Перевірка гіпотези про закон розподілу генеральної сукупності. Критерій Пірсона 65
Тема 8. Використання методів теорії ймовірностей та математичної статистики для вирішення деяких задач геомеханіки 71
8.1. Визначення коефіцієнта структурного послаблення 71
8.1.1. Гіпотеза про нормальний розподіл величини R – міцності структурних елементів масиву 66
8.1.2. Гіпотеза про логарифмічно нормальний розподіл величини R – міцності структурних елементів масиву 74
8.1.3. Врахування макродефектів при визначенні коефіцієнта структурного послаблення 76
8.2. Приклад виконання розрахункового завдання 78
Тема 9. Кореляційна залежність. Метод найменших квадратів 84
9.1. Стохастичний зв'язок. Регресія 84
9.2. Лінійна кореляція. Метод найменших квадратів 86
9.3. Обробка кореляційної таблиці 90
ДОДАТОК А. Значення нормованої інтегральної функції нормального розподілу 94
ДОДАТОК Б. Таблиця значень функції Лапласа: Ф(t) 96
ДОДАТОК В. Критичні точки розподілу Стьюдента 97
ДОДАТОК Г. Критичні точки розподілу χ^2 98
ДОДАТОК Д. Варіанти для виконання індивідуального завдання №1 99
ДОДАТОК Е. Варіанти для виконання індивідуального завдання №2 101
Бібліографічний список 102

ВСТУП

Метою вивчення розділів «Елементи теорії ймовірностей» та «Математична статистика в гірництві» є підготовка фахівців до вирішення різних гірничотехнічних завдань на основі обробки статистичної інформації і прогнозу поведінки стохастичних систем шляхом побудови ймовірних моделей того чи іншого технологічного процесу або природного явища.

Даний посібник містить короткі відомості з теорії ймовірностей та математичної статистики, що необхідні для побудови статистичних моделей, які безпосередньо використовуються для вирішення практичних завдань.

Основною метою вивчення даних розділів є ілюстрація застосування статистичних методів для вирішення завдань гірництва, зокрема геомеханіки.

CRDF GLOBAL

Автори висловлюють вдячність Фонду цивільних досліджень та розвитку США (CRDF Global) за підтримку при розробці даного навчального посібника в рамках спільного Українсько-Американського проекту «Науково-освітній центр «Стійкість геотехнічних систем: процеси, явища, ризики».

СТАТИСТИЧНІ МЕТОДИ В ПРОБЛЕМАХ ГІРНИЦТВА

Під час дослідження різних фізичних і технічних процесів часто трапляються явища особливого роду, які прийнято називати випадковими. *Випадкове явище* при багаторазовому повторенні того самого досліду проходить щоразу по-іншому [1].

Наприклад, при багаторазовому визначенні характеристик міцності зразків однієї і тієї самої гірської породи результати окремих випробувань будуть трохи різними. Ці розходження обумовлені впливом багатьох факторів, а саме незначними відмінностями геометричної форми зразків, різницями в структурі гірської породи в окремих зразках і т.п. [2]. Як би точно і докладно не були дотримані умови експериментів, неможливо досягти повного збігу результатів.

Подібна картина спостерігається при вимірах зсувів контуру виробки в перетинах, що перебувають в однакових умовах, за той самий проміжок часу; при вимірюванні продуктивності робіт в очисних або підготовчих вибоях, проведених у тих самих гірничо-геологічних умовах із застосуванням тих самих механізмів.

У ряді практичних задач випадковими елементами зневажають, розглядаючи замість реального об'єкта його спрощену схему – модель. При цьому з множини факторів, що впливають, виділяють найбільш вагомі, а впливом другорядних факторів нехтують. Потім застосовується той чи інший математичний апарат (наприклад, складаються й інтегруються диференціальні рівняння, що описують дане явище). Таким чином, виявляється основна закономірність, що властива даному явищу. Чим більше факторів буде враховано при математичному моделюванні, тим детальніше досліджується явище, а також точніше стає науковий прогноз.

Описана схема дає вірогідний результат при розв’язуванні задач, де підсумок досліду залежить від невеликої кількості основних факторів, які залишаються постійними від досліду до досліду. Але ця схема непридатна при розв’язуванні таких задач, де результат досліду залежить не тільки від основних факторів, але й значною мірою від кількості другорядних, пов’язаних між собою, випадкових факторів [2]. Саме такі питання виникають у гірництві. Наприклад, навантаження на кріплення гірської виробки, зсуви її контуру формуються під впливом цілого ряду факторів, серед яких безліч випадкових, причому таких, які неможливо врахувати в рішеннях, заснованих на фундаментальних законах. Наприклад, структурна неоднорідність породного середовища (наявність тріщин, флуктуацій, включень), похибки в технології зведення кріплення, недосконаліcть тампонажу закріпного простору тощо. Крім того, ті фізичні константи, що фігурують у розрахунках – модуль пружності порід, коефіцієнт Пуассона, об'ємна вага, межі міцності на розтягання і стиск – тісно пов'язані з випадковою природою явища, тобто самі є *випадковими* величинами.

Для розв’язання задач, що містять випадкові величини, тобто елементи невизначеності та багатопричинності, використовують *ймовірністно - статистичні методи.* Вони спираються на *теорію ймовірностей* і *математичну статистику*.

Практика показує, якщо випадкові явища мають масовий характер, у них можна знайти цілком визначені закономірності. Вивчення цих закономірностей становить *предмет теорії ймовірностей*.

Відзначимо, що саме масовість випадкових явищ забезпечує прояв у них специфічних, так званих «статистичних» закономірностей. Завданням *математичної статистики* є систематизація статистичних даних для встановлення закономірностей перебігу випадкових масових явищ. *Статистичними даними* називаються відомості про кількість об'єктів, що володіють тими чи іншими ознаками.

Статистичні методи, що використовуються в різних галузях знань, у тому числі й у гірничому виробництві, мають загальні риси. Для глибокого вивчення цих методів і використання їх у завданнях гірництва варто усвідомити основні поняття теорії ймовірностей і математичної статистики, викладені в численних навчальних посібниках. Нижче наведено основні відомості, необхідні для вирішення ряду специфічних завдань геомеханіки і гірничого виробництва. Більш повний і докладний виклад понять, якими оперує теорія ймовірностей і математична статистика, містять підручники [1 – 3].

1. ОСНОВНІ ПОНЯТТЯ, ЩО ВИКОРИСТОВУЮТСЯ В ТЕОРІЇ ЙМОВІРНОСТЕЙ

1.1. **Випадкові події. Імовірність події. Частота події**

Випадкова подія – факт, який у результаті досліду може відбутися, а може і не відбутися.

Приклади подій:

A – «руйнування зразка породи розміром 7х7х7 см під дією навантаження Р=500 Н»;

B – «руйнування зразка породи розміром 7х7х7 см під дією навантаження Р=1000 Н»;

C – «безвідмовна робота механізму протягом 100 годин»;

D – «безвідмовна робота механізму протягом 50 годин»;

E – «безвідмовна робота механізму протягом 25 годин».

Безумовно, що подія *B* більш можлива, ніж подія *A*, а подія *E* більш можлива, ніж події *C* и *D*. Можливість появи події характеризується числом, що називають *імовірністю* події.

Класична формула визначення ймовірності події *A* має вигляд:

$$P(A) = \frac{m}{n} \quad , \tag{1.1}$$

де n – загальна кількість рівноможливих випадків (результатів) досліду ; m – кількість тих випадків, що сприяють появі події.

Нехай, наприклад, у скриньці знаходяться 100 деталей, з яких 10 деталей є бракованими. Якщо не дивлячись витягти одну деталь, то ймовірність події *A* – «витягнена деталь є бракованою» –

$$P = 10/100=1/10.$$

Із класичного визначення випливає, що ймовірність випадкової події коливається в межах від *0* до *1*. За одиницю прийнята ймовірність ***вірогідної*** події, тобто події, що неодмінно відбудеться в результаті досліду. Наприклад, у досліді з киданням гральної кості вірогідною є подія: «число, що випало, не *більше* 6».

Нулю дорівнює ймовірність ***неможливої*** події, тобто такої, що не може відбутися в даному досліді. Наприклад, при киданні гральної кості подія «*число, що випало*, *дорівнює* 16» є неможливою в досліді.

За класичною формулою (1.1) ймовірність події можна визначити за умовами досліду, не здійснюючи його. Ясно, наприклад, і без кидання гральної кості, що ймовірність події «*число, що випало*, *дорівнює 5*» становить *1/6*. Тому ймовірність, що отримана за класичною формулою (1.1), називають характеристикою «*апріорі*». Слабкою стороною класичного визначення ймовірності є те, що далеко не завжди результати досліду можна подати у вигляді сукупності елементарних подій. Тому формула (1.1) має обмежене застосування.

На практиці аналогом імовірності є ***відносна частота*** події – величина, що визначається після проведення серії дослідів («*апостеріорі*») за формулою:

$$w(A) = \frac{m}{n}, \tag{1.2}$$

тут n – загальна кількість дослідів; m – кількість дослідів, у результаті яких відбулася подія A.

Приклад 1.1

У вибої відібрано 35 проб руди. З них 7 проб мали вміст металу не більш 1%. Частота події A – «вміст металу в пробі не більш 1 %» –

$$w(A) = \frac{7}{35} = \frac{1}{5}.$$

Якби ми повторили дослід, тобто відібрали на цій самій ділянці інші 35 проб, то результат міг би бути іншим. Наприклад, тільки 5 проб містили металу не більш 1 %. Тоді б частота події A складала 1/7. Однак, тривалі спостереження показали, якщо в однакових умовах здійснюють досліди, у кожнім із яких кількість випробувань досить велика, то відносна частота змінюється в кожному досліді мало (тим менше, чим більше зроблено випробувань), коливаючись біля деякого постійного числа. Це число і є ймовірністю події – теоретичною величиною, до якої прагне емпірична величина – відносна частота – при нескінченному збільшенні кількості випробувань.

1.2. Види випадкових подій

Події називаються ***несумісними****, якщо вони не можуть відбутися разом в одному досліді.*

Приклади несумісних подій: «випадіння герба» і «випадіння цифри» при одному підкиданні монети; «зразок зруйнувався» і «зразок не зруйнувався» у тому самому досліді.

Кілька подій утворюють **повну групу**, якщо в результаті досліду з'явиться хоча б одна з них. Іншими словами, поява хоча б однієї з подій повної групи є достовірною подією. Наприклад, у результаті досліду – кидання гральної кості – обов’язково відбудеться одна з подій: A_1 – «випало число 1»; A_2 – «випало число 2»; A_3 – «випало число 3»;.... A_6 – «випало число 6». Ці несумісні події утворюють повну групу. В результаті досліду – стрілок зробив постріл по цілі – обов'язково відбудеться одна з подій: «попадання», «промах». Ці дві несумісні події утворюють повну групу.

Події називають рівноможливими, якщо є підстави вважати, що жодне з них не є більш можливим, ніж інше.

Наприклад, події «випадіння герба» і «випадіння цифри» при одному підкиданні монети є *рівноможливими*. Дійсно, передбачається, що монета виготовлена з однорідного матеріалу, має правильну циліндричну форму і наявність карбування не впливає на випадання тієї чи іншої сторони монети.

1.3. Основні теореми теорії ймовірностей. Алгебра подій
1.3.1.Теорема додавання ймовірностей

Сумою А+В двох подій А і В називають подію, яка полягає у появі події А або події В, або їх обох.

Сумою декількох подій називають подію, яка полягає у появі хоча б однієї з цих подій. Наприклад, подія А + В + С полягає в появі однієї з наступних подій: А, В, С, А і В, А і С, В і С, А і В і С.

Теорема. *Якщо події A, B, C,......N несумісні, то ймовірність суми цих подій дорівнює сумі ймовірностей подій:*

$$P(A+B+C+...+N)=P(A)+P(B)+P(C)+......+P(N). \quad (1.3)$$

Із цієї теореми випливає дуже важливий **висновок 1.**

Якщо несумісні події A_1, A_2, A_3,...... A_n вичерпують собою всі можливі випадки досліду (тобто утворюють ***повну групу*** подій), то сума ймовірностей цих подій дорівнює одиниці. Тобто:

$$P(A_1)+(A_2)+(A_3)......+(A_n) = 1. \quad (1.4)$$

Наприклад, при киданні гральної кості події: A_1 – «випало число 1»; A_2 – «випало число 2»; A_3 – «випало число 3»;.... A_6 – «випало число 6» є несумісними, вичерпують собою всі можливі випадки і, таким чином, утворюють повну групу подій. Ймовірність кожної з цих подій відповідно до класичної формули (1.1) дорівнює 1/6. Сума ймовірностей цих подій:

$$P(A_1)+P(A_2)+P(A_3)+.......+P(A_6)=1/6+1/6+1/6+...+1/6=6/6=1.$$

Перед тим, як сформулювати другий висновок з теореми додавання дамо таке визначення. ***Протилежними*** *називаються дві єдино можливі події, що утворюють повну групу.* Наприклад, у досліді – кидання монети – події A – «випадіння герба» і A^*– «випадіння цифри» є протилежними; у досліді – навантаження зразка – події B – «зразок зруйнувався» і B^* – «зразок не зруйнувався» також є протилежними.

Висновок 2. Сума протилежних подій дорівнює одиниці. Тобто:

$$P(A)+P(A^*)=1.$$

По суті, висновок 2 – окремий випадок висновку 1. Таким чином, якщо відома ймовірність події A, завжди можна визначити ймовірність протилежної події *A**. Цей факт широко використовується для розв'язування задач. Якщо, наприклад, протилежна подія *A** розпадається на меншу кількість елементарних випадків, ніж подія *A*, доцільно підрахувати ймовірність події *A* як різницю:

$$P(A)= 1- P(A^*).$$

1.3.2. Теорема множення ймовірностей

Добутком двох подій A і B називають подію AB, яке полягає в спільній появі (поєднанні) цих подій. Наприклад, якщо A – «деталь придатна», B – «деталь пофарбована», то AB – «деталь придатна і пофарбована». Добутком кількох подій називають подію, яка полягає в спільній появі всіх цих подій.

Перш ніж сформулювати теорему, наведемо таке визначення.

Події називають ***незалежними****, якщо поява однієї з них не змінює ймовірностей появи інших подій.*

Подія *A* називається *залежною* від події *B,* якщо ймовірність події *A* змінюється залежно від того, відбулася подія *B* чи ні. ***Умовною*** називають імовірність події *A*, обчислену з урахуванням того, що подія *B* відбулася. Таку ймовірність позначають $P_B(A)$. ***Умовну*** ймовірність події *B* обчислюють з урахуванням того, що подія A відбулася, і позначають $P_A(B)$.

Теорема. *Ймовірність спільної появи* ***незалежних*** *подій A, B, C,……N дорівнює добутку ймовірностей цих подій:*

$$P(A,B,C,\ldots N)=P(A)P(B)P(C)\ldots P(N)\ldots \qquad (1.5)$$

Імовірність спільної появи кількох ***залежних*** *подій дорівнює добутку ймовірності одного з них на умовні ймовірності всіх інших, причому, ймовірність кожної наступної події обчислюється з урахуванням того, що всі попередні події вже відбулися.* Зокрема для трьох залежних подій *A, B, C* ймовірність спільної появи

$$P(A, B, C)=P(A)P_A(B)P_{AB}(C). \qquad (1.6)$$

Приклад 1.2

Нехай ділянка електричного кола містить два послідовно з'єднаних елемента. Нехай за даними випробувань перший елемент працює безвідмовно з ймовірністю $p_1=0,9$, а другий – з імовірністю $p_2=0,8$. Події *A* – «перший елемент спрацював безвідмовно» і *B* – «другий елемент спрацював безвідмовно» є незалежними. Подія «коло спрацювало безвідмовно» полягає в спільній появі незалежних подій *A* і *B*. Імовірність безвідмовної роботи кола

$$P(A,B) = p_{1*}\, p_2=0,9*0,8=0,72.$$

Цей приклад наочно ілюструє відомий факт: чим більше послідовно з'єднаних елементів містить електричне коло, тим менше ймовірність його безвідмовної роботи, навіть якщо кожний з елементів має високу ймовірність безвідмовної роботи.

Приклад 1.3

Технічний пристрій складається з 7-ми елементів. Кожен елемент, незалежно від інших, може мати несправність, ймовірність якої дорівнює 0,05. Якщо хоча б один елемент несправний, то в процесі роботи пристрою відбудеться аварія. Знайти ймовірність цієї події.

Нехай подія B – «елемент несправний». Ймовірність $P(B)=0,05$. Протилежна подія B^* – «елемент справний». Ймовірність протилежної події відповідно висновку 2

$$P(B^*)=1-P(B)=1-0,005=0,95.$$

Позначимо через A подію «аварія: хоча б один елемент несправний». Протилежна подія A^* – «усі 7 елементів справні». За теоремою множення ймовірність $P(A^*)=0,95^7=0,695$. Отже, шукана ймовірність події A

$$P(A)=1-P(A^*)=1-0,695=0,305.$$

Таким чином, *у 30,5 % випадків можлива аварія*. І це при ймовірності несправності кожного елемента, яка дорівнює усього лише 0,05. Усе тому, що пристрій складається з 7 елементів, і відмовлення одного з них зовсім не малоймовірна подія.

Приклад 1.4

Нехай партія з 100 деталей має 10 бракованих. Випадково відбирають одну деталь, потім другу. Визначимо ймовірність того, що обидві вони браковані.

Імовірність події A – «перша з відібраних деталей – бракована» має ймовірність $P(A)= 10/100=1/10$.

Імовірність події B – «друга з відібраних деталей – бракована» залежить від того, чи відбулася подія A. З урахуванням того, що вона дійсно відбулася, умовна ймовірність події B

$$P_A(B)=9/99=1/11.$$

Імовірністю того, що обидві деталі браковані, є ймовірність спільної появи двох залежних подій A і B:

$$P(AB)=P(A)P_A(B)=1/10*1/11=1/110.$$

1.3.3. Формула повної ймовірності

Висновком з теорем додавання і множення ймовірностей є формула повної ймовірності.

Нехай подія A може відбутися тільки разом з однією з подій H_1, H_2, H_3,…H_n, попарно несумісних, що утворюють повну групу. Події H_j, $j=\overline{1,n}$ називаються гіпотезами для події A. У досліді подія A може відбутися тільки у вигляді AH_1, або AH_2, або AH_n. Отже, ймовірність цієї події за теоремою додавання

$$P(A)=P(AH_1)+P(AH_2)+\ldots+P(AH_n)...$$

Відповідно до теореми множення ймовірностей для залежних подій одержимо формулу повної ймовірності:

$$P(A)=P(H_1)P_{H1}(A)+P(H_2)P_{H2}(A)+\ldots P(H_n)P_{Hn}(A)=\sum_{j=1}^{n}P(H_j)P_{H_j}(A)\,. \qquad (1.7)$$

Приклад 1.5

На дворі копальні С (рис. 1.1) групуються вагонетки для забору породи з прохідницьких вибоїв, позначених точками. Вагонетки в рівних кількостях вирушають до вибоїв по магістралях h_1, h_2, h_3.

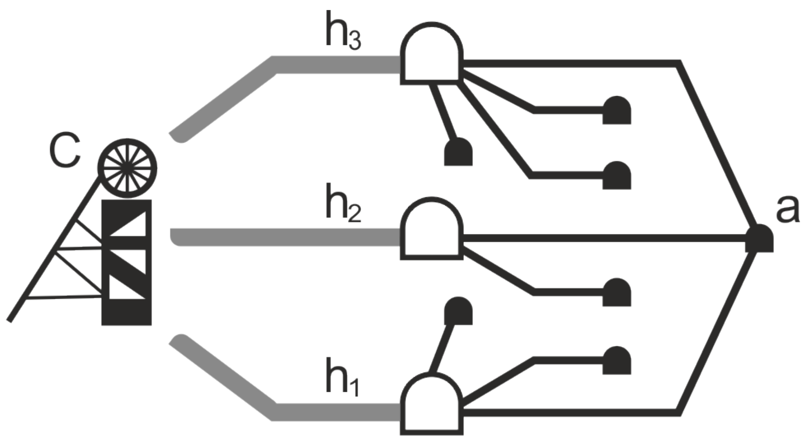


Рис. 1.1. Схема до прикладу 1.5

Вагонетка з фіксованим номером може потрапити на кожну з цих магістралей. Яка ймовірність того, що саме ця вагонетка потрапить до прохідницького вибою a.

<u>Розв’язок.</u> Висунемо гіпотези:

H_1 – *«вагонетка потрапила на магістраль h_1»;* $P(H_1)=1/3$,

H_2 – *«вагонетка потрапила на магістраль h_2»;* $P(H_2)=1/3$,

H_3 – *«вагонетка потрапила на магістраль h_3»;* $P(H_3)=1/3$.

За формулою повної ймовірності (1.7) отримаємо

$$P(A)=P(H_1)P_{H1}(A)+P(H_2)P_{H2}(A)+P(H_3)P_{H3}(A)=$$
$$=1/3*1/3+1/3*1/2+1/3*1/4=13/36.$$

1.3.4. Імовірність гіпотез. Формула Баєса

Нехай подія А може відбутися тільки разом з однією з подій H_j, $j=1,n$. Нехай також подія A відбулася. Яка ймовірність того, що появі події А сприяла саме гіпотеза H_j? Отримаємо

$$P(AH_j)= P(A)\,P_A(H_j) = P(H_j)P_{Hj}(A).$$

Звідси
$$P_A(H_j)= P(H_j)P_{Hj}(A)\,/P(A).$$

З урахуванням формули повної ймовірності (1.7) дістанемо формулу Баєса:

$$P_A(H_j) = \frac{P(H_j)P_{H_j}(A)}{\sum_{j=1}^{n} P(H_j)P_{H_j}(A)} \quad . \tag{1.8}$$

Ця формула визначає «питому вагу» ймовірності появи події A разом з гіпотезою H_j серед усіх можливих реалізацій події A.

Приклад 1.6

При обстеженні деякої автоматизованої системи розглядаються дві причини, що можуть вивести її із строю. Імовірності цих причин відповідно дорівнюють *0,6* і *0,4*. Для уточнення обстеження призначаються додаткові іспити, результатом яких є позитивний «+» чи негативний «-» сигнали датчика. У випадку реалізації першої причини ймовірність сигналу «+» дорівнює *0,9*, сигналу «-» – *0,1*. У випадку реалізації другої причини сигнал «-» і сигнал «+» мають імовірність 0,5. Іспит зробили двічі й обидва рази був отриманий сигнал «-». Потрібно знайти ймовірність реалізації кожної з причин можливого виходу системи з ладу після проведення іспитів.

Розв’язок. Позначимо

H_1 – *«причина №1 »;* $P(H_1)=0{,}6$,

H_2 – *«причина №2 »;* $P(H_2)=0{,}4$.

Нехай подія A – «обидва сигнали негативні». У випадку причини №1 ймовірність негативного сигналу $P_{H1}(A)=0{,}1*0{,}1=0{,}01$, а у випадку причини №2 $P_{H2}(A) =0{,}5*0{,}5=0{,}25$. За формулою (1.8) маємо

$$P_A(H_1) = \frac{P(H_1)P_{H_1}(A)}{P(H1)P_{H_1}(A) + P(H_2)P_{H_2}(A)} = \frac{0{,}6*0{,}01}{0{,}6*0{,}01+0{,}4*0{,}25} \cong 0{,}06;$$

$$P_A(H_2) = \frac{P(H_2)P_{H2}(A)}{P(H1)P_{H_1}(A) + P(H_2)P_{H_2}(A)} = \frac{0{,}4*0{,}25}{0{,}6*0{,}01+0{,}4*0{,}25} \cong 0{,}94.$$

Отримані результати дають вагомі підстави вважати, що система може вийти з ладу через причину №2.

Контрольні питання

1. Що називається випадковим явищем?
2. Що складає предмет теорії ймовірностей?
3. У чому полягає завдання математичної статистики?
4. Наведіть класичну формулу ймовірності випадкової події.
5. Що є емпіричним аналогом ймовірності?
6. Які події називаються несумісними?
7. Сформулюйте теорему додавання ймовірностей для несумісних подій.
8. Які події утворюють повну групу? Чому дорівнює сума ймовірностей подій, що утворюють повну групу?
9. Які події називають незалежними?
10. Сформулюйте теорему множення ймовірностей для залежних і незалежних подій?

2. ВИПАДКОВІ ВЕЛИЧИНИ І ЗАКОНИ ЇХ РОЗПОДІЛУ

2.1. Поняття випадкової величини

Випадковою називають величину, яка у результаті досліду набуває будь-якого значення. Заздалегідь воно невідоме і обумовлене випадковими причинами.

Якщо значення випадкової величини можна перелічити, вона називається ***дискретною***. Приклад – кількість відбійних молотків, що виходять із ладу протягом зміни. Тут випадкова величина може приймати тільки цілі значення.

Якщо випадкова величина в ході досліду приймає будь-яке значення з деякого інтервалу, вона називається ***неперервною.*** До неперервних випадкових величин відносяться: значення вмісту металу в руді за результатами випробування, час безвідмовної роботи механізму, продуктивність праці робітника, похибки вимірювань і т.і. Майже всі величини, що характеризують властивості гірських порід (межа міцності, модуль деформацій, питома вага), є неперервними випадковими величинами.

Надалі випадкові величини будемо позначати великими літерами (*X, У,* і т.д.), а їх можливі значення – відповідними малими (x_1, x_2 , x_n; $y_1, y_2, \dots y_n$).

Розглянемо дискретну випадкову величину X з її можливими значеннями x_1, x_2 , ... x_n . Кожного з цих значень величина X набуває з деякою ймовірністю, а саме: значення $X=x_1$ – з ймовірністю p_1; значення $X=x_2$ – з імовірністю $p_2, \dots$ значення $X=x_n$ – з імовірністю p_n.

Події «$X=x_2$», «$X=x_2$»,…,«$X=x_n$» несумісні і утворюють повну групу. Отже, сума ймовірностей цих подій

$$\sum_{i=1}^{n} p_i = 1. \qquad (2.1)$$

Ця сумарна ймовірність певним чином розподілена між окремими значеннями випадкової величини X. Випадкова величина буде цілком описана з імовірнісної точки зору, якщо ми будемо точно знати, як саме розподілена сумарна ймовірність, тобто яка ймовірність кожної з подій «$X=x_2$», «$X=x_2$»,…,«$X=x_n$».

Визначення. *Відповідність між значеннями випадкової величини і ймовірностями, з якими ці значення з'являються в результаті досліду, називається* ***законом розподілу****.*

Про випадкову величину говорять, що вона *підпорядкована* даному закону. Найпростішою формою завдання закону розподілу є таблиця, що містить значення випадкової величини і відповідні їм імовірності (табл. 2.1). Така таблиця називається *рядом розподілу.*

Таблиця 2.1

X	x_1	x_2	……	x_n
p	p_1	p_2	……	p_n

Приклад 2.1

На руднику протягом 100 робочих змін фіксувалася кількість відбійних молотків, що виходять з ладу з різних причин. Результати спостережень наведено в табл. 2.2, що являє собою ряд розподілу для випадкової величини X – кількості молотків, що вийшли з ладу протягом робочого часу.

Таблиця 2.2

Кількість молотків, що вийшли з ладу протягом робочого часу	Кількість спостережень m	Відносні частоти w
0	16	0,16
1	56	0,56
2	24	0,24
3	3	0,03
4	1	0,01
	$\sum_{i=1}^{n} m_i = 100$	$\sum_{i=1}^{n} w_i = 1$

Щоб додати ряду розподілу наочність, будують так звану полігональну криву: по осі абсцис відкладають значення випадкової величини, по осі ординат – імовірності цих значень (у даному випадку відносні частоти) (рис. 2.1).

З наведеного прикладу видно, що в умовах даної шахти найбільш імовірним є вихід з ладу одного відбійного молотка протягом робочого часу.

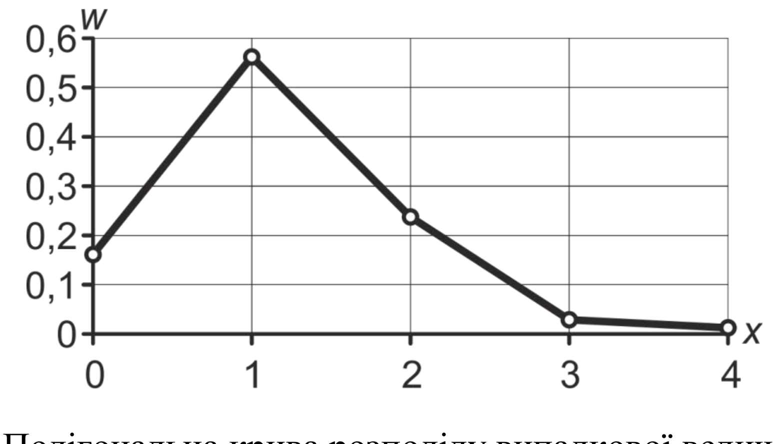


Рис. 2.1. Полігональна крива розподілу випадкової величини X – кількості відбійних молотків, що вийшли з ладу протягом зміни

2.2. Інтегральна функція розподілу

Ряд розподілу вичерпно характеризує дискретну випадкову величину. Однак ця характеристика не є універсальною. Для неперервної випадкової величини побудувати такий ряд неможливо, оскільки не можна перелічити всі її значення. Крім того, як ми переконаємося надалі, ймовірність того, що неперервна випадкова величина набуває якогось окремого значення x дорівнює нулю. Для неперервної випадкової величини має сенс говорити тільки про ймовірність того, що вона приймає значення з деякого, нехай навіть дуже малого інтервалу.

Тому для кількісної характеристики розподілу неперервної випадкової величини розглядають не ймовірність події $X=x$, а ймовірність події $X<x$. Зрозуміло, що ймовірність цієї події залежить від значення x, тобто є функцією x, що називається ***функцією розподілу*** випадкової величини і позначається $F(x)$:

$$F(x) = P(X<x). \tag{2.2}$$

Функція розподілу є універсальною формою закону розподілу. Вона також придатна для опису неперервних і дискретних величин.

Властивості функції розподілу

1. Із визначення інтегральної функції $F(x)$ як імовірності виходить, що її значення належать відрізку [0,1]:

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

2. Інтегральна функція $F(x)$ – неспадна функція свого аргументу, тобто

$$F(x_2) \ge F(x_1), \text{ якщо } x_2 \ge x_1.$$

Дійсно, нехай $x_2 > x_1$. Подія $X < x_2$ розпадається на дві несумісні події: «$x_1 \le X < x_2$» або «$X < x_1$». За теоремою додавання

$$P(X < x_2)=P(x_1 \le X < x_2)+P(X < x_1).$$

Звідси

$$P(x_1 \le X < x_2)=P(X < x_2) - P(X < x_1)=F(x_2) - F(x_1).$$

Отже, виходить, що $F(x_2) \ge F(x_1)$, оскільки ймовірність у лівій частині рівності більше або дорівнює нулю.

З останньої формули випливає важливий висновок: *імовірність того, що випадкова величина отримає значення з інтервалу* (x_1, x_2), *дорівнює приросту функції на цьому інтервалі*:

$$P(x_1 \le X < x_2) = F(x_2) - F(x_1). \tag{2.3}$$

3. Якщо можливі значення випадкової величини належать інтервалу (a, b), то:

1) $F(x)=0$ при $x \le a$ (ймовірність події «$X \le a$», а, отже, і значення функції в точці $x=a$ дорівнює нулю, оскільки ця подія неможлива);

2) $F(x)=1$ при $x \ge b$ (ймовірність події «$X \ge b$», а, отже, і значення функції в точці $x=b$ дорівнює одиниці, оскільки ця подія достовірна);

Із формули (2.3) випливає, що ймовірність того, що неперервна випадкова величина X прийме одне визначене значення $X=x_1$ дорівнює нулю.

Цей факт цілком відповідає вимогам практичних задач. Наприклад, важливо визначити ймовірністю того, що розміри деталей не виходять за дозволені межі, але не ставлять питання про ймовірність їх збігу з номінальним значенням.

Відмітимо, що було б неправильно вважати: якщо ймовірність $P(X=x_1)$ дорівнює нулю, то подія $X=x_1$ неможлива (якщо, звичайно, не обмежуватися класичним визначенням імовірності). Дійсно, у результаті досліду випадкова величина обов'язково отримає одне з можливих значень. Зокрема, це значення може виявитися таким, що дорівнює x_1.

Для дискретної випадкової величини графік інтегральної функції має східчастий вигляд. Наприклад, на рис. 2.2 зображена інтегральна функція, побудована для ряду розподілу, заданого табл. 2.2.

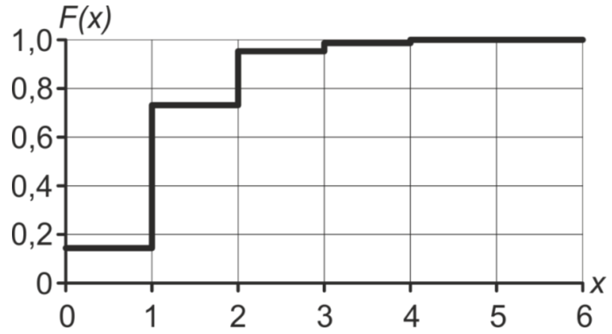


Рис. 2.2. Графік інтегральної функції розподілу для випадкової величини Х – числа відбійних молотків, що вийшли з ладу протягом робочого часу

Коли величина Х набуває дискретних значень 0, 1 ,2, 3, 4, інтегральна функція змінюється стрибкоподібно, причому величини стрибків дорівнюють ймовірностям цих значень (відповідно 0,16; 0,56; 0,24; 0,03; 0,01). Сума всіх стрибків функції F(x) дорівнює одиниці.

Зі збільшенням можливих значень Х кількість стрибків збільшується, а стрибки зменшуються. Східчаста лінія при цьому наближається до плавної кривої. Для неперервної випадкової величини графік інтегральної функції – неперервна лінія. Її графік наведений на рис. 2.3.

Зрозуміло, що у випадку, коли можливі значення неперервної випадкової величини заповнюють всю числову вісь, справедливі граничні співвідношення, що випливають із визначення інтегральної функції: $F(-\infty)=0$; $F(+\infty)=1$.

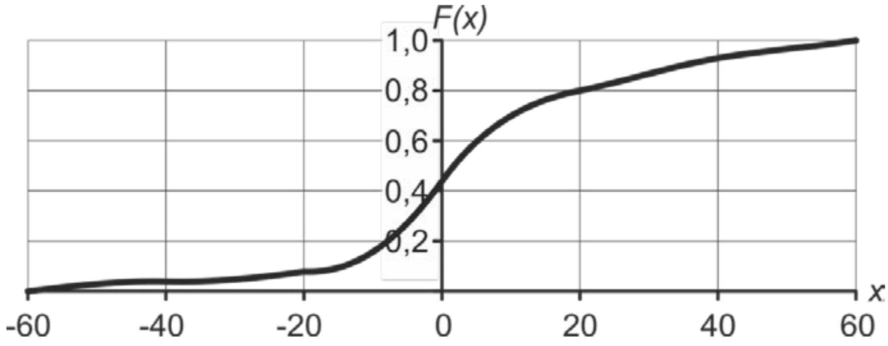


Рис. 2.3. Графік інтегральної функції для неперервної випадкової величини

Приклад 2.2

Випадкова величина X задана функцією розподілу (рис. 2.4):

$$F(x) = \begin{cases} 0 & при \quad x \le 2 \\ 0{,}5x - 1 & при \quad 2 < x \le 4 \\ 1 & при \quad x < 4 \end{cases}$$

Знайти ймовірність того, що в результаті випробувань випадкова величина прийме значення: а) з інтервалу *(2,5; 3,0)*; б) не менше *3*.

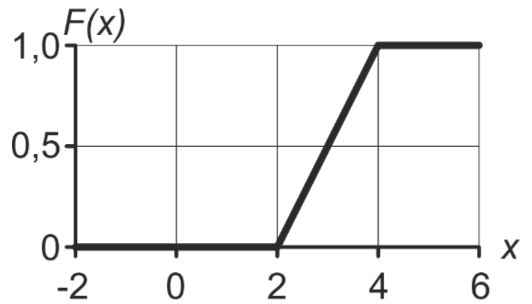


Рис. 2.4. Графік заданої інтегральної функції розподілу (приклад 2.2)

Розв'язок

а) $P(2{,}5<X < 3{,}0) = F(3{,}0)-F(2{,}5)=0{,}5-0{,}25=0{,}25$.

б) Події $X \ge 3$ і $X < 3$ є протилежними.

Використовуючи функцію розподілу, знайдемо ймовірність того, що $X < 3$:
$P(x < 3) = F(3) =0{,}5$, тоді $P(x \ge 3) = 1- p(x < 3) =1-0{,}5=0{,}5$.

2.3. Диференціальна функція розподілу випадкової величини

Розглянемо неперервну випадкову величину X з функцією розподілу *F(X)*. Обчислимо ймовірність потрапляння цієї випадкової величини на ділянку від x до $x + \Delta x$:

$$P(x < X < x + \Delta x) = F(x + \Delta x) - F(x). \qquad (2.4)$$

Розглянемо відношення цієї ймовірності до довжини ділянки, тобто середню ймовірність, що приходиться на одиницю довжини цієї ділянки, і будемо наближати Δx до нуля, тобто перейдемо до границі при $\Delta x \to 0$:

$$\lim_{\Delta x \to 0} \frac{F(x+\Delta x) - F(x)}{\Delta x}.$$

Цей вираз є не чим іншим як похідною функції розподілу $F'(x)$.

Її називають ***диференціальною функцією розподілу*** та позначають $f(x)$. Таким чином,

$$f(x) = F'(x) = \lim_{\Delta x \to 0} \frac{F(x+\Delta x) - F(x)}{\Delta x}. \tag{2.5}$$

Функція *f(x)* показує, як щільно розподіляються значення випадкової величини в інтервалі від x до $x+\ \Delta x$. Звідси її інша назва – ***щільність розподілу*** випадкової величини.

Щільність розподілу, так само як і інтегральна функція розподілу, є однією з форм закону розподілу. Ми побачимо надалі, що вона більш наочна, хоча і не така універсальна. Вона існує тільки для неперервної випадкової величини.

Із (2.4 – 2.5) виходить, що ймовірність потрапляння значення випадкової величини на проміжок $(x, \quad x+\Delta x)$ (з точністю до нескінченно малих)

$$P(x < X < x+\Delta x) = f(x)dx. \tag{2.6}$$

Величину $f(x)dx$ називають елементом імовірності. Геометрично можна зобразити площею елементарного прямокутника з висотою, що дорівнює $f(x)$, і основою dx (рис. 2.5). Очевидно, що ймовірність влучення випадкової величини в довільний інтервал від x_1 до x_2 дорівнює сумі елементів імовірності на цій ділянці, тобто інтегралу:

$$P(x_1 < X < x_2) = \int_{x_1}^{x_2} f(x)dx. \tag{2.7}$$

Геометрично ймовірність влучення випадкової величини на ділянку (x_1, x_2) дорівнює площі криволінійної трапеції, обмеженої кривою $f(x)$ і прямими $x=x_1$, $x=x_2$.

Функція розподілу *F(x)* для щільності розподілу *f(x)* є первісною. Тому її називають **інтегральною** функцію розподілу. За визначенням

$$F(x) = P(X < x) = P(-\infty < X < x) \ ,$$

тоді за формулою (2.7) отримаємо

$$F(x) = \int_{-\infty}^{x} f(x)dx \tag{2.8}$$

Геометрично *F(x)* є не чим іншим, як площею під кривою щільності розподілу, що лежить ліворуч від точки *x*.

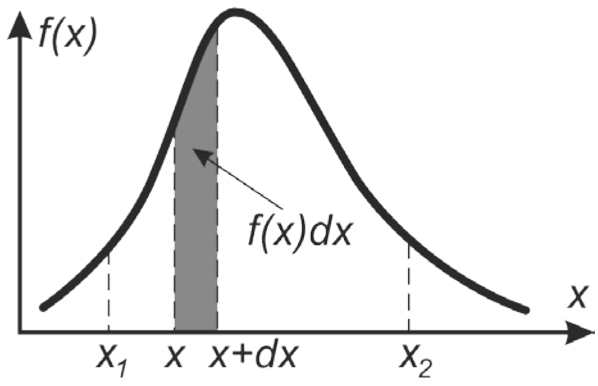


Рис. 2.5. Графік диференціальної функції (щільності розподілу)

Диференціальна функція *f(x)* (щільність розподілу) та інтегральна функція *F(x)* – різні форми закону розподілу випадкової величини X. Знаючи хоча б одну з цих функцій, ми можемо відповісти на найбільш важливе практичне запитання: із якою ймовірністю досліджувана випадкова величина набуває значення з того чи іншого інтервалу. З урахуванням (2.7) і формули Ньютона – Лейбніца ця ймовірність

$$P(x_1 < X < x_2) = \int_{x_1}^{x_2} f(x)dx = F(x_2) - F(x_1). \tag{2.10}$$

Далі буде показано, що саме ця формула є основою розв’язання більшості статистичних задач.

При цьому істотне значення мають властивості щільності розподілу:

1. $f(x) \geq 0$. Геометрично це означає, що крива розподілу завжди лежить не нижче осі абсцис.

2. $$\int_{-\infty}^{+\infty} f(x)dx = 1. \tag{2.11}$$

Дійсно, цей інтеграл є ймовірністю того, що випадкова величина *X* набуває значення з інтервалу $(-\infty, \ +\infty)$. Така подія є достовірною. Імовірність її дорівнює одиниці. Геометрично це означає, що вся площа під кривою розподілу дорівнює одиниці.

Приклад 2.3

Визначимо щільність розподілу для випадкової величини, заданої в попередньому прикладі 2.2 інтегральною функцією:

$$F(x) = \begin{cases} 0 & при \quad x \leq 2 \\ 0{,}5x - 1 & при \quad 2 < x \leq 4 \\ 1 & при \quad x > 4 \end{cases}.$$

Диференціюючи *F(x)* одержимо:

$$f(x) = F'(x) = \begin{cases} 0 & при \quad x \le 2 \\ 0{,}5 & при \quad 2 < x \le 4 \\ 0 & при \quad x > 4 \end{cases}.$$

Графік отриманої диференціальної функції зображений на рис. 2.6. На всьому інтервалі зміни випадкової величини X щільність розподілу є постійною. Такий вид розподілу називається *законом рівномірної щільності* (рис. 2.6).

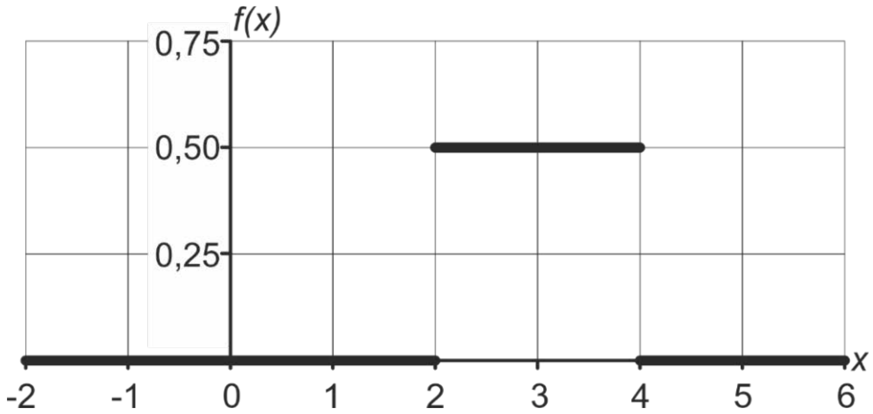


Рис. 2.6. Графік диференціальної функції (закон рівномірної щільності)

Покажемо, тепер, як може бути отримана щільність розподілу випадкової величини практичним шляхом.

Контрольні питання

1. Яку величину називають випадковою?
2. Які випадкові величини називають дискретними і які безперервними? Наведіть приклади.
3. Що називають законом розподілу випадкової величини?
4. Яким чином може бути заданий закон розподілу?
5. Що називають інтегральною функцією розподілу? Назвіть властивості цієї функції. Який вигляд має її графік ?
6. Що таке диференціальна функція розподілу? Чому її називають щільністю розподілу?
7. Напишіть, як зв'язана інтегральна функція розподілу з диференціальною функцією.
8. Як визначити ймовірність улучення випадкової величини в заданий інтервал? Що являє собою ця ймовірність на графіках інтегральної і диференціальної функцій?

3. ПОБУДОВА СТАТИСТИЧНОГО РОЗПОДІЛУ КІЛЬКІСНОЇ ОЗНАКИ (Елементи математичної статистики)

3.1. Сутність вибіркового методу

Вище вказувалося, що прогноз поведінки гірських порід навколо підземної споруди можна здійснювати на підставі аналізу деяких кількісних ознак. Зокрема такими є фізико-механічні властивості масиву. Внаслідок неоднорідності породного масиву більшість з кількісних ознак є випадковими величинами. Наприклад, у результаті випробування зразків гірської породи на стиск ми одержуємо різні значення міцності навіть для зразків однієї літологічної різниці, що обумовлено випадковими причинами. Таким чином, дана кількісна ознака – міцність породи на стиск – є випадковою величиною, що може приймати значення з певного інтервалу. Дослідника, що виконує геомеханічні розрахунки, цікавить, які значення досліджуваної ознаки є найбільш імовірними, як оцінити наявний розкид цих значень, як він вплине на остаточні результати тощо. Для відповіді на ці питання дослідник має підібрати для отриманих даних теоретичний закон розподілу у вигляді інтегральної чи диференціальної функцій розподілу. Першим кроком на шляху підбору теоретичного розподілу є статистична обробка даних, тобто їх ***статистичний аналіз.***

При статистичному аналізі процесів гірничого виробництва використовують ***вибіркові дані***. З обстежуваної ***генеральної сукупності*** однорідних об'єктів, кількість яких N, відбирають випадково певну кількість об’єктів (n одиниць) **,** тобто формують ***вибірку***.

Нехай вивчається деяка кількісна ознака X, що властива генеральній сукупності однорідних об'єктів. Наприклад, досліджується міцність деякої гірської породи на одноосьовий стиск. Звичайно ця характеристика встановлюється шляхом випробування в лабораторних умовах зразків породи об’ємом v, виготовлених із проби, що у свою чергу відібрана в досліджуваному масиві гірської породи. Таким чином, застосовується вибірковий метод: з генеральної чисельності зразків $N = V/n$ (V – увесь обсяг досліджуваного масиву гірської породи) відбирається тільки n зразків. Величина n називається ***обсягом вибірки***.

Для того, щоб за даними вибірки можливо було достатньо впевнено судити про важливі ознаки генеральної сукупності, необхідно, аби об'єкти вибірки правильно її представляли. Коротко ця вимога формулюється так: вибірка повинна бути *репрезентативною*. Вибірка є репрезентативною, якщо її створювати випадково, тобто так, щоб всі об'єкти генеральної сукупності мали рівні шанси потрапити у вибірку.

Нехай у результаті вимірів отримані значення (варіанти) кількісної ознаки X: x_1 x_2 x_3 ... x_n. Якщо ці значення записані в тій послідовності, у якій вони спостерігалися в дійсності, вони називаються ***незгрупованими*** даними.

Якщо таких даних кілька десятків чи сотень наш розум не в змозі охопити зміст явища, що спостерігається. Для виявлення характерних рис досліджуваної ознаки корисно згрупувати дані.

Нехай у результаті дослідів кількісної ознаки X значення x_1 з'явилося m_1 раз, x_2 – з'явилося m_2 разів, значення x_r – з'явилося m_r раз. Числа m_i $(i = 1,r)$, що показують скільки разів спостерігається кожне із значень ознаки, називаються ***частотами***.

Зрозуміло, що сума частот повинна дорівнювати обсягу вибірки

$$\sum_{i=1}^{r} m_i = n. \tag{3.1}$$

Величини $\omega_i = \frac{m_i}{n}$ називають ***відносними частотами***. Очевидно, що

$$\sum_{i=1}^{r} \omega_i = 1. \tag{3.2}$$

Якщо розташувати варіанти в порядку, що зростає чи спадає, і вказати для кожної варіанти її частоту, одержимо ***розподіл ознаки*** у вигляді впорядкованого ***варіаційного ряду*** (табл. 3.1).

Таблиця 3.1

Загальний вигляд варіаційного ряду.

Значення кількісної ознаки X	x_1	x_2	x_3	…	x_n	
Частоти m	m_1	m_2	m_3	…	m_n	$\Sigma m_i=n$
Відносні частоти w	w_1	w_2	w_3	…	w_n	$\Sigma w_i=1$

Як бачимо, варіаційний ряд аналогічний ряду розподілу дискретної випадкової величини, що представлений в таблиці 2.1. Раніше наведена табл. 2.2, яка характеризує статистичний розподіл кількості відбійних молотків, котрі вийшли з ладу за певний термін, являє собою приклад варіаційного ряду.

3.2. Побудова інтервального ряду. Гістограма частот

У табл. 3.2 як приклад наведено фрагмент журналу досліду з визначення межі міцності зразків алевроліту на одноосьовий стиск у вигляді незгрупованих даних. Якщо для цих даних побудувати варіаційний ряд, він вийде дуже громіздким (табл. 3.3).

Таблиця 3.2

№ зразка	Міцність зразка на стиск
1 2 3	52,8 48,4 60,0
…	...
31	56,3

32 33	49,7 56,4
і т.д...	...

Таблиця 3.3

Варіанти X_i, мПа	Частота, m_i	Варіанти X_i, мПа	Частота, m_i	Варіанти X_i, мПа	Частота, m_i
48,4	1	51,2	1	53,4	3
48,9	1	51,4	1	54,3	4
49,2	1	51,8	1	54,5	1
49,3	2	52,2	2	64,6	1
49,7	1	52,4	1	55,4	1
50,3	2	52,7	2	56,4	3
50,8	2	52,8	4	57,2	1
51,0	2	53,0	1	60,0	1

Для кількісної ознаки, що являє собою *неперервну випадкову величину*, варто побудувати так званий ***інтервальний ряд***. Для цього:

1) серед незгрупованих даних вибирають найбільше x_{max} і найменше x_{min} значення;

2) відрізок (x_{max}, x_{min}) поділяють на інтервали (розряди) довжиною l:

$$l = \frac{x_{\max} - x_{\min}}{k}, \tag{3.3}$$

де k – кількість інтервалів. Це число можна вибрати довільно, а можна скористатися формулою

$$k = 1 + 3{,}2\lg n\,; \tag{3.4}$$

3) визначають частоту влучення значень величини X у кожний з інтервалів;

4) обчислюють середину кожного інтервалу, і надалі розглядають її як варіанту.

Таблиця 3.4

Інтервали міцності зразка на стиск, u_{i-1} - u_i , мПа	Значення середини інтервалу , u_i	Частота, m_i	Відносна частота ,w_i
48-50	49	6	0,15
50,1-52	51	9	0,225
52,1-54	53	13	0,325
54,1-56	55	7	0,175
56,1-58	57	4	0,1
58,1-60	59	1	0,025
		$\sum_{i=1}^{r} m_i = 40$	$\sum_{i=1}^{r} w_i = 1{,}0$

Таким чином, в інтервальному ряді частоти відносяться не до окремих значень ознаки, а до середин інтервалів. Наприклад, дані, що наведені в табл. 3.3, розбиті на 6 інтервалів довжиною *l=2* (табл. 3.4).

Наочним графічним зображенням інтервального ряду є ***гістограма частот.***

На осі абсцис відкладають відрізки, які дорівнюють довжинам інтервалів. На кожному з відрізків будують прямокутник, висота якого складає:

$$h_i = \frac{w_i}{l_i}, \quad i = \overline{1,r} \;, \tag{3.5}$$

де r – кількість інтервалів. Тоді площа кожного *i-го* прямокутника дорівнює відносній частоті w_i, а площа всієї гістограми дорівнює одиниці (рис. 3.1).

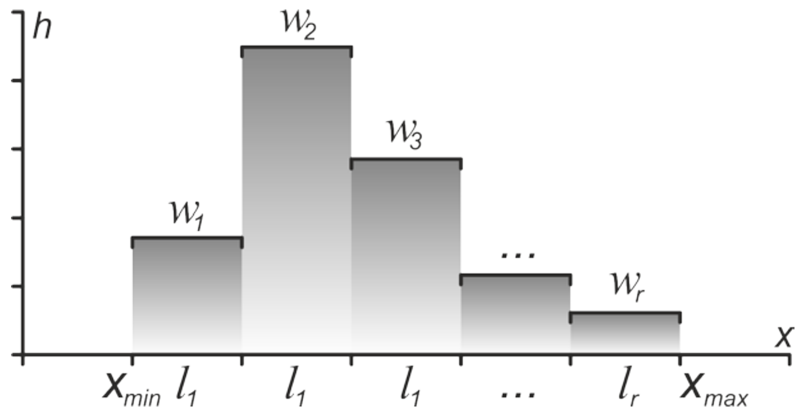


Рис. 3.1. Гістограма відносних частот

Наприклад, на рис. 3.2 побудована гістограма відносних частот для величини X – міцності на стиск алевроліту (за даними табл. 3.4).

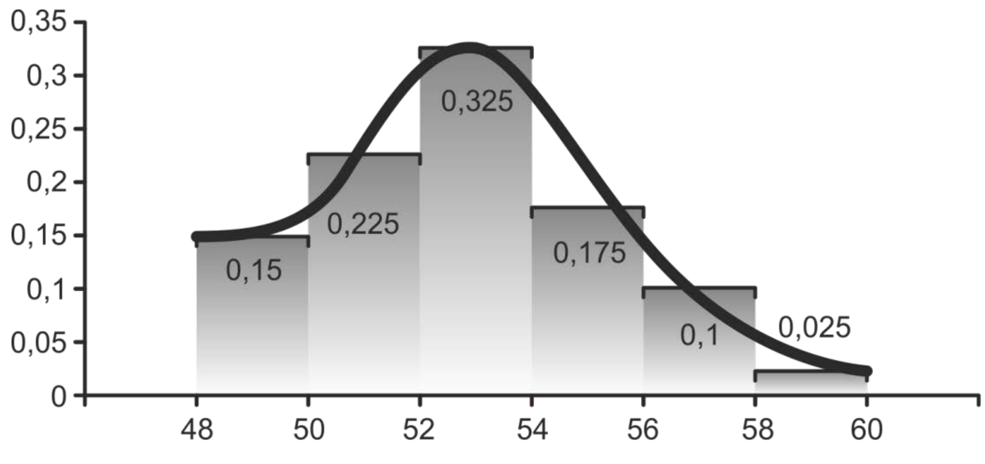


Рис. 3.2. Гістограма відносних частот межі міцності на стиск алевроліту (за даними таблиці 3.4)

Проведемо приблизно через середини сторін прямокутників лінію *f**. Очевидно, що при збільшенні числа дослідів можна вибирати усе більш дрібні інтервали. При цьому лінія *f** буде все більше наближатися до деякої теоретичної кривої, що обмежує площу, яка дорівнює одиниці. Неважко переконатися, що ця теоретична крива являє собою графік щільності розподілу досліджуваної кількісної ознаки X, тобто графік функції *f(x)*.

Подальший аналіз полягає в підборі аналітичного виразу для функції *f(x)*. При цьому дослідник аналізує вигляд гістограми та лінії ***f****, що у першому наближенні вказує на вигляд графіка щільності розподілу *f(x)*. Далі буде показано, що для вибору функції *f(x)* має значення фізична сутність досліджуваної величини. Але на першому етапі підбору аналітичного виразу для *f(x)* головну роль відіграють так звані числові характеристики випадкової величини, що будуть розглянуті далі.

Контрольні питання

1. У чому є сутність вибіркового методу**?** Поясніть це на прикладі вивчення властивостей гірських порід.
2. Якій основній вимозі повинна задовольняти вибірка з генеральної сукупності**?**
3. Як групують статистичні дані**?** Який вигляд має варіаційний ряд**?**
4. Що являє собою інтервальний ряд**?**
5. Що є графічним зображенням інтервального ряду**?** Які висновки може зробити дослідник після побудови гістограми частот. Які подальші кроки він може почати для вивчення розподілу кількісної ознаки**?**

4. ЧИСЛОВІ ХАРАКТЕРИСТИКИ ВИПАДКОВОЇ ВЕЛИЧИНИ. МОМЕНТИ РОЗПОДІЛУ. ВИЗНАЧЕННЯ ЇХ ЗА ДАНИМИ ДОСЛІДУ

4.1. Основні характеристики розташування і варіації випадкової величини

Числовими характеристиками випадкової величини називають числа, що визначають істотні риси розподілу випадкової величини, наприклад, деяке середнє значення, навколо якого групуються можливі значення випадкової величини; яке-небудь число, що характеризує ступінь варіації можливих значень щодо середнього. У теорії ймовірностей застосовується велика кількість числових характеристик. Розглянемо лише ті з них, що використовуються найбільш часто для характеристики кількісних ознак.

1. Математичне сподівання випадкової величини (іноді його називають просто середнім значенням).

Нехай X – *дискретна* випадкова величина, що приймає значення $x_1, x_2 \ldots x_n$ з ймовірностями $p_1, p_2 \ldots p_n$. Математичним сподіванням $M(X)$ (чи середнім значенням x_{cp}) називають величину, що дорівнює сумі добутків її можливих значень на ймовірності, з якими ці значення з'являються:

$$M(X)= x_{cp} = x_1 p_1 + x_2 p_2 + \ldots x_n p_{n\ldots} = \sum_{i=1}^{n} x_i p_i . \qquad (4.1)$$

Приклад 4.1. У шухляді є 100 заготівок вагою 1кг, 20 заготівок вагою 2 кг, 5 заготівок – вагою 3 кг. Визначити середню вагу заготівки.

У даному прикладі середня вага заготівки – це і є математичне сподівання (середнє значення) випадкової величини - ваги заготівки. Визначимо середнє значення за формулою (4.1):

$$x_{cp} = M(x) = 1\frac{100}{125} + 2\frac{20}{125} + 3\frac{5}{125} = \frac{155}{125} \text{ (кг).}$$

Якщо Х – *неперервна* випадкова величина з щільністю розподілу $f(x)$, її математичне сподівання виражається вже не сумою, а інтегралом

$$M(X) = \int_{-\infty}^{+\infty} x f(x) dx . \qquad (4.2)$$

Із визначення математичного сподівання випливають його властивості:

а) математичне сподівання постійної величини Х дорівнює цій величині:

$$M(C)=C; \quad C=const;$$

б) математичне сподівання добутку незалежних випадкових величин дорівнює добутку їх математичних сподівань. Наприклад, для трьох незалежних величин $X,У,Z$:

$$M(X,У,Z)= M(X)M(У)M(Z);$$

звідси виходить, що постійний множник можна виносити за знак математичного сподівання:

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

в) математичне сподівання алгебраїчної суми випадкових величин дорівнює сумі математичних сподівань цих величин:

$$M(X+У+ Z)= M(X)+M(У)+M(Z).$$

Приклад 4.2 Розглянемо дві випадкові величини *X* и *У*, що мають такі ряди розподілу:

X_i	-2	-1	0	1	2
P_i	0,05	0,2	0,5	0,2	0,05

Y_i	-20	-10	0	10	20
P_i	0,05	0,2	0,5	0,2	0,05

Неважко переконатися, що $M(X)=0$ і $M(У)=0$.

Як бачимо, випадкові величини *X* и *У* мають однакові математичні сподівання, але відхилення значень цих величин від математичних сподівань істотно відрізняються. *Відхиленням* називають різницю між значенням випадкової величини і її математичним сподіванням: $|x-M(X)|$

У теорії ймовірностей розкид значень випадкової величини, тобто ступінь варіації ознаки, характеризують дисперсією.

2. ***Дисперсією*** (розсіюванням) називають математичне сподівання квадрата відхилень випадкової величини:

$$D(X)=M\left(|x-M(X)|^2\right). \qquad (4.3)$$

Формула для визначення дисперсії має вигляд:

а) для дискретної випадкової величини:

$$D(X)=\sum_{i=1}^{n}\left(x_i-M(X)\right)^2 p_i; \qquad (4.4)$$

б) для неперервної випадкової величини з щільністю розподілу *f(x)*

$$D(X)=\int_{-\infty}^{+\infty}\left(x-M(X)\right)^2 f(x)dx. \qquad (4.5)$$

На практиці обчислення дисперсії за формулами (4.4) і (4.5) трохи громіздкі. З урахуванням властивостей математичного сподівання формулу (4.4) можна перетворити до вигляду

$$D(X) = M\left(X^2\right) - \left[M(X)\right]^2 . \qquad (4.6)$$

(Доказ наведено у [3.с.81]).

Тоді для дискретної випадкової величини

$$D(X) = \sum_{i=1}^{n} x_i^{\,2} p_i - \left[M(X)\right]^2 . \qquad (4.7)$$

Для неперервної випадкової величини

$$D(X) = \int_{-\infty}^{+\infty} x^2 f(x)dx - \left[M(X)^2\right]. \qquad (4.8)$$

Для попереднього приклада дисперсії величин *X* и *У* відповідно дорівнюють

D(X)=0,8;   D(Y)=80.

Оскільки при піднесенні в квадрат змінюється розмірність випадкової величини, для характеристики розкиду використовують величину, яка обчислюється:

$$\sigma(X) = \sqrt{D(X)} \qquad (4.9)$$

і називається ***середнім квадратичним відхиленням*** або ***стандартом***.

Для попереднього прикладу зазначені характеристики відповідно: $\sigma(X) = 0{,}89$, $\sigma(Y) = 8{,}9$ **.**

4.2. Моменти розподілу

Розглянемо дискретну випадкову величину X, що задана розподілом

X	1	2	5	100
P	0,6	0,2	0,19	0,01

Математичне сподівання цієї величини

*M(X)= 1*0,6+2*0,2+5*0,19+100*0,01=2,95.*

Складемо тепер ряд із квадратів значень величини X и знайдемо математичне сподівання для величини X^2:

X^2	1	4	25	10000
P	0,6	0,2	0,19	0,01

$$M(X^2)= 1*0,6+4*0,2+25*0,19+10000*0,01=106,15.$$

$M(X^2)$ істотно більше *M(X)* з тієї причини, що серед можливих значень величини X є одне, що набагато переважує всі інші, хоча і має дуже малу ймовірність появи. Для того, щоб урахувати вплив таких значень у теорії ймовірностей уведено моменти розподілу, початкові і центральні, котрі, як ми побачимо далі, узагальнюють поняття числових характеристик випадкової величини.

Початковим моментом *к-го* порядку називають математичне сподівання величини $X^к$, тобто

$$\nu_k = M(X^k).$$

Так, для дискретної величини формула має вигляд:

$$\nu_k = \sum_{i=1}^{n} x^k{}_i\, p_i, \tag{4.10}$$

а для неперервної величини:

$$\nu_k = \int_{-\infty}^{+\infty} x^k f(x)dx.$$

Зокрема, початковий момент першого порядку є не чим іншим, як математичним сподіванням:

$$\nu_1 = M(X), \tag{4.11}$$

початковий момент другого порядку є математичним сподіванням квадрата випадкової величини:

$$\nu_2 = M(X^2) \quad \text{і т.д.}$$

Центральним моментом *к-го* порядку називають математичне сподівання величини *$(X-M(X))^к$:*

$$\mu_k = M\left((X - M(X))^k\right). \tag{4.12}$$

Зокрема, з (4.12) випливає, що центральний момент першого порядку (тобто математичне сподівання відхилень) дорівнює нулю. Центральний момент другого порядку являє собою дисперсію:

$$\mu_2 = M\left((X - M(X))^2\right) = D(X). \tag{4.13}$$

З формули (4.6) випливає, що дисперсія легко може бути виражена через початкові моменти першого і другого порядків:

$$D(X) = \mu_2 = \nu_2 - \nu_1^{\,2} \quad . \qquad (4.14)$$

Третій і четвертий центральні моменти також можна виразити через початкові моменти:

$$\mu_3 = \nu_3 - 3\nu_2\nu_1 + 2\nu_1^3 \quad , \qquad (4.15)$$

$$\mu_4 = \nu_4 - 4\nu_3\nu_1 + 6\nu_2\nu_1^2 - 3\nu_1^4 .$$

Моменти вищих порядків використовуються для опису закону розподілу. З ними зв'язані:

1) **асиметрія**

$$A = \frac{\mu_3}{\sigma^3} = \frac{\mu_3}{\mu_2^{\,3/2}} \; , \qquad (4.16)$$

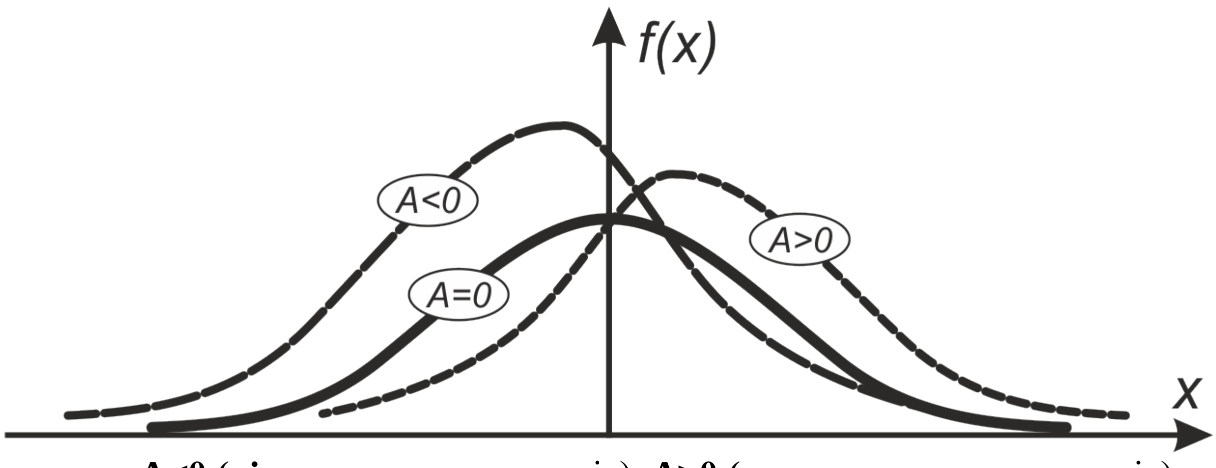
що характеризує несиметричність кривої щільності розподілу випадкової величини X відносно її математичного сподівання (рис. 4.1);

2) **ексцес**

$$E = \frac{\mu_4}{\sigma^4} - 3 = \frac{\mu_4}{\mu_2^{\,2}} - 3 , \qquad (4.17)$$

що характеризує гостровершинність графіка диференціальної функції розподілу (рис. 4.1).

а)



A<0 (лівостороння асиметрія), **A>0 (правостороння** асиметрія) б)

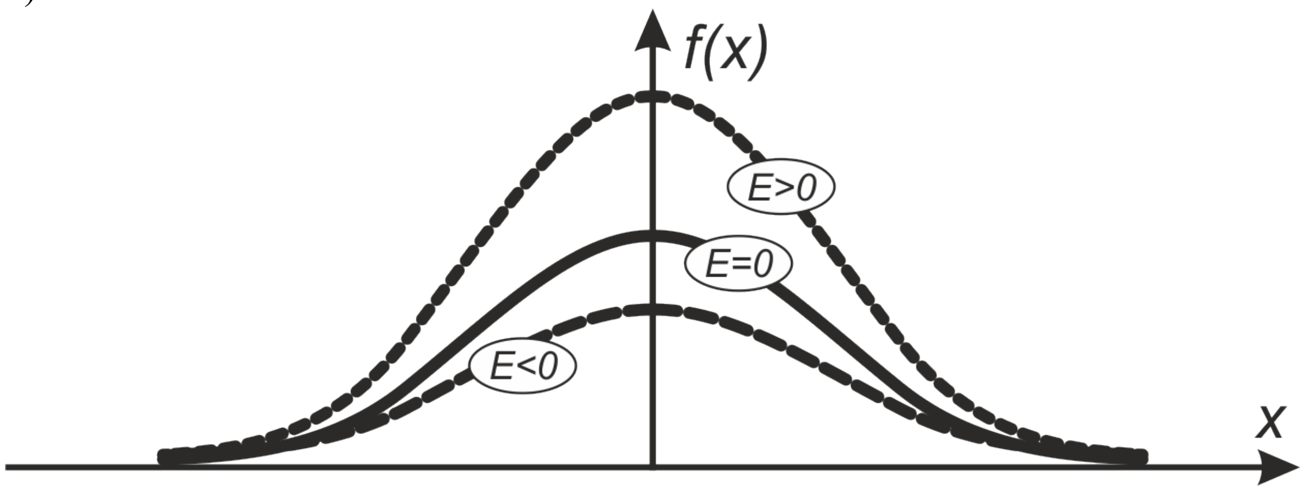


Рис. 4.1. До поняття асиметрії й ексцесу випадкової величини

4.3. Визначення числових характеристик кількісної ознаки за дослідними даними. Параметри статистичного розподілу. (Елементи математичної статистики)

Усі моменти розподілу випадкової величини, розглянуті в попередньому розділі, мають свої статистичні аналоги. Вони називаються ***моментами статистичного розподілу*** і визначаються за дослідними даними.

Якщо дані згруповані у варіаційний ряд, подібний до наведеного у табл. 3.1, статистичні моменти визначаються за тими самими формулами, що і їх теоретичні аналоги для дискретної випадкової величини, але із заміною ймовірності $\boldsymbol{p_i}$ на відносну частоту $\boldsymbol{w_i}$ (табл. 4.1). Статистичні параметри будемо позначати *.

Наприклад, статистичний аналог математичного сподівання, тобто статистичний момент першого порядку, називається ***вибірковою середньою***, позначається $\overline{x}^*$ і визначається формулою:

$$\overline{x}^* = \nu_1^* = \sum_{i=1}^{r} x_i\, w_i \ ,$$

r – кількість значень випадкової ознаки, що спостерігаються.

Статистичним аналогом дисперсії є ***вибіркова дисперсія***, що визначається як статистичний центральний момент другого порядку:

$$D^* = \mu_2 = \sum_{i=1}^{r} x_i^2 w_i - (\overline{x}^*)^2 \ .$$

Тут варто зробити **зауваження**.

1. Статистичні величини ν_k^*, μ_k^*, у тому числі середня вибіркова $\bar{x}^*$ і дисперсія D^*, є для теоретичних величин μ_k, ν_k (у тому числі $\bar{x}$, $D(x)$) так званими точковими оцінками, тобто такими, що виражаються одним числом.

2. У повному курсі математичної статистики доводиться, що середня вибіркова дає *незміщену* оцінку дійсної величини математичного сподівання. На відміну від неї вибіркова дисперсія дає *зміщену* оцінку дійсної дисперсії досліджуваної кількісної ознаки. *Незміщену* оцінку дає величина, що обчислюється:

$$S^2 = \frac{n}{n-1} D^* \quad .$$

Її називають ***виправленою дисперсією.*** Оцінкою середнього квадратичного відхилення (стандарту) є корінь з виправленої дисперсії.

Статистичні оцінки асиметрії й ексцесу пов'язані зі статистичними центральними моментами третього і четвертого порядків.

Якщо дослідні дані сформовані у вигляді інтервального ряду в ролі варіант $\boldsymbol{x_i}$ виступають середини інтервалів ($\boldsymbol{u_i}$).

Моменти статистичного розподілу можуть бути визначені і за незгрупованими даними (табл. 4.1).

Таблиця 4.1
Статистичні моменти розподілу

Назва статистичного моменту	**Незгруповані дані**	**Згруповані дані**
1	2	3
1.Початковий момент k-го порядку *Окремий випадок*: початковий момент першого порядку – середня вибіркова $\bar{x}^* = \nu_1$	$\nu_k^* = \frac{1}{n}\sum_{i=1}^{n} x_i^k$ *n – обсяг вибірки* $\bar{x}^* = \nu_1^* = \frac{1}{n}\sum_{i=1}^{n} x_i$	$\nu_k^* = \sum_{i=1}^{r} x_i^k \cdot w_i$ *r – кількість значень, що спостереджуються* $\bar{x}^* = \nu_1^* = \sum_{i=1}^{r} x_i \ w_i$

2.Центральний момент к-го порядку *Окремі випадки:* 2.1) центральний момент другого порядку - вибіркова дисперсія $D^* = \mu_2^* = \nu_2^* - (\nu_1^*)^2$ 2.2) Центральний момент третього порядку $\mu_3^* = \nu_3^* - 3\nu_2^*\nu_1^* + 2\nu_1^{*3}$ 2.3) центральний момент четвертого порядку $\mu_4^* = \nu_4^* - 4\nu_3^*\nu_1^* + 6\nu_2^*\nu_1^{*2} - 3\nu_1^{*4}$	$\mu_k^* = \frac{1}{n}\sum_{i=1}^{n}(x_i - \overline{x}^*)^k$ $D^* = \mu_2 = \frac{1}{n}\sum_{i=1}^{n}x_i^2 - (\overline{x}^*)^2$ $\mu_3^* = \frac{1}{n}\sum_{i=1}^{n}(x_i - \overline{x}^*)^3$ $\mu_4^* = \frac{1}{n}\sum_{i=1}^{n}(x_i - \overline{x}^*)^4$	$\mu_k^* = \sum_{i=1}^{r}(x_i - \overline{x}^*)^k \cdot w_i$ $D^* = \mu_2 = \sum_{i=1}^{r}x_i^2 w_i - (\overline{x}^*)^2$ $\mu_3^* = \sum_{i=1}^{r}(x_i - \overline{x}^*)^3 \cdot w_i$ $\mu_4^* = \sum_{i=1}^{r}(x_i - \overline{x}^*)^4 \cdot w_i$
3. Виправлена дисперсія	$S^2 = \frac{n}{n-1}D^*$	
4. Середнє квадратичне відхилення (стандарт)	$S = \sqrt{S^2}$	
5. Відносна варіація, коефіцієнт варіації	$\eta^* = \frac{S}{\overline{x}^*}$; $V^* = \eta^* * 100\%$.	
6. Статистична оцінка асиметрії	$A^* = \frac{\mu_3^*}{S^3} = \frac{\mu_3^*}{\mu_2^{*\,3/2}}$	
7. Статистична оцінка ексцесу	$E^* = \frac{\mu_4^*}{S^4} - 3 = \frac{\mu_4^*}{\mu_2^{*\,2}} - 3$	

Приклад 4.2

Визначити моменти статистичного розподілу для даних випробувань межі міцності на стиск зразків алевроліту, що наведені у табл. 3.3 і згруповані в інтервальний ряд (табл.3.4).

1. Визначимо початковий момент першого порядку, тобто вибіркову середню:

$$\overline{x}^* = \nu_1^* = 49*0{,}15 + 51*0{,}225 + 53*0{,}325 + 55*0{,}175 + 57*0{,}1 + 59*0{,}025 = 52{,}85\,\text{МПа}.$$

2. Визначимо характеристики варіації значень кількісної ознаки:

- статистичний початковий момент другого порядку:

$$\nu_2^* = 49^2 * 0{,}15 + 51^2 * 0{,}225 + 53^2 * 0{,}325 + 55^2 * 0{,}175 + 57^2 * 0{,}1 + 59^2 * 0{,}025 = 2799{,}6;$$

- статистичний центральний момент другого порядку, тобто вибіркову дисперсію:

$$D^* = \mu_2^* = \nu_2^* - (\nu_1^*)^2 = 2799{,}6 - 52{,}85^2 = 6{,}48$$

- виправлена дисперсія

$$S^2 = \frac{40}{40-1} * 6{,}48 = 6{,}64 \ (МПа)^2;$$

- середнє квадратичне відхилення (стандарт):

$$S = \sqrt{S^2} = 2{,}58 \ МПа;$$

- відносна варіація, коефіцієнт варіації:

$$\eta^* = \frac{2{,}58}{52{,}85} = 0{,}05;$$

$$V^* = 0{,}05 * 100 = 5\% .$$

3. Визначимо величини, що характеризують асиметрію та ексцес:

- статистичні початкові моменти третього і четвертого порядків:

$$\nu_3^* = 49^3 * 0{,}15 + 51^3 * 0{,}225 + 53^3 * 0{,}325 + 55^3 * 0{,}175 + 57^3 * 0{,}1 + 59^3 * 0{,}025 = 148648{,}3,$$

$$\nu_4^* = 49^4 * 0{,}15 + 51^4 * 0{,}225 + 53^4 * 0{,}325 + 55^4 * 0{,}175 + 57^4 * 0{,}1 + 59^4 * 0{,}025 = 7911190;$$

статистичні центральні моменти третього і четвертого порядків:

$$\mu_3^* = \nu_3^* - 3\nu_2^*\nu_1^* + 2\nu_1^{*3} = 148648{,}3 - 4 * 2799{,}6 * 52{,}85 + 2 * 52{,}85^3 = 4{,}72;$$

$$\mu_4^* = \nu_4^* - 4\nu_3^*\nu_1^* + 6\nu_2^*\nu_1^{*2} - 3\nu_1^{*4} =$$

$$= 7911190 - 4 * 148648{,}3 * 52{,}85 + 6 * 2799{,}6 * 52{,}85^2 - 3 * 52{,}85^4 = 104{,}76.$$

- статистична оцінка асиметрії:

$$A^* = \frac{4{,}72}{2{,}58^3} = 0{,}28 .$$

- статистична оцінка ексцесу:

$$E^* = \frac{104{,}76}{2{,}58^4} - 3 = -0{,}63 .$$

Аналізуючи статистичні моменти досліджуваної ознаки – міцності гірської породи на одноосьовий стиск – можна зробити такі висновки. Середнє значення ознаки складає 52,85 МПа. Значення ознаки коливаються навколо свого середнього з невеликою варіацією, що складає 5 %. Розподіл ознаки має невелику правобічну асиметрію. Його можна вважати майже симетричним. Значення ексцесу також близько до нуля. За виглядом гістограми можна припустити, що розподіл даної ознаки описується деякою функцією *f(x)*, графік якої має максимум у точці x=52,85, крива майже симетрична щодо вертикальної прямої, проведеної через цю точку.

Для прогнозу появи значень ознаки з тією чи іншою ймовірністю потрібно підібрати для функції *f(x)* аналітичний вираз. Існує ряд функцій, що задовільно описують розподіл достатньо вивчених випадкових величин. Вони називаються теоретичними законами розподілу. Нижче будуть наведені деякі з них.

Контрольні питання

1. Що характеризує математичне сподівання випадкової величини?
2. Наведіть формули для визначення математичного сподівання дискретної і безперервної випадкових величин.
3. Що характеризує дисперсія випадкової величини?
4. Наведіть формули для визначення дисперсії дискретної і безперервної випадкових величин.
5. Що таке початковий момент розподілу *к-го* порядку? Яку числову характеристику дає початковий момент 1-го порядку?
6. Що таке центральний момент *к-го* порядку? Яку числову характеристику дає центральний момент 2-го порядку?
7. Що характеризують асиметрія й ексцес випадкової величини? Наведіть формули для визначення цих характеристик.
8. Як визначити числові характеристики випадкової величини за даними досліду? Наведіть формули для середньої вибіркової і вибіркової дисперсії для згрупованих і незгрупованих даних.
9. Що таке виправлена дисперсія? Наведіть формулу для її визначення.
10. Що таке коефіцієнт варіації випадкової величини?

5. ДЕЯКІ ТЕОРЕТИЧНІ ЗАКОНИ РОЗПОДІЛУ

5.1. Показниковий закон розподілу

Така величина, як час безвідмовної роботи деякого механізму описується щільністю розподілу (диференціальною функцією), графік якої наведено на рис. 5.1.

Видно, що ця крива являє собою експоненту, що спадає на проміжку $(0, \ +\infty)$. Тобто для зазначеного інтервалу $f(x) = C\, e^{-\lambda x}$. Для того, щоб дана функція дійсно описувала закон розподілу ймовірностей випадкової величини, вона повинна задовольняти умові (2.11), тобто

$$C \int_0^{+\infty} e^{-\lambda x} dx = 1 .$$

За цією умовою знайдемо значення постійної С.

$$C \int_0^{+\infty} e^{-\lambda x} dx = -\frac{1}{\lambda} C \lim_{b \to \infty}\left(e^{-\lambda b} - e^0\right) = \frac{C}{\lambda} \text{ ; звідси } \frac{C}{\lambda} = 1 \Rightarrow C = \lambda$$

Таким чином, щільність розподілу показникового закону (його також називають експоненціальним) має вигляд:

$$f(x) = \lambda\ e^{-\lambda x} \ . \tag{5.1}$$

Величину λ називають ***параметром розподілу***. Для опису розподілу конкретної кількісної ознаки параметр розподілу має бути визначений на основі статистичних даних. Для цього використовують ***метод моментів***. Він полягає в прирівнюванні емпіричних і теоретичних моментів одного порядку:

$$\nu_k^* = \nu_k; \quad \mu_k^* = \mu_k \tag{5.2}$$

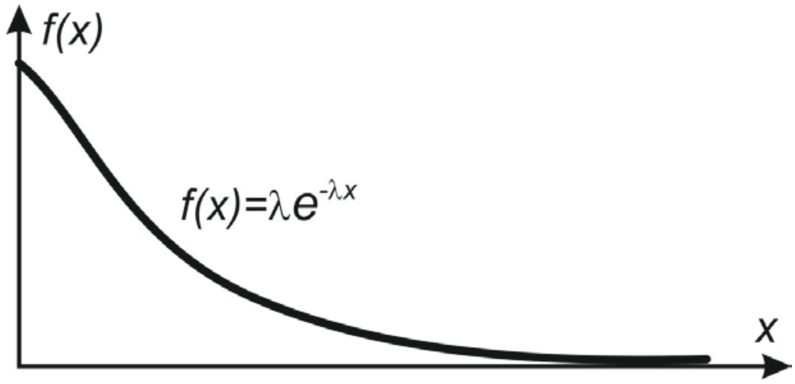


Рис. 5.1. Показниковий закон розподілу

Ми побачимо нижче, що теоретичні моменти для того чи іншого розподілу часто пов'язані з параметром даного розподілу. Тому, насамперед, варто визначити теоретичні моменти для даного розподілу. Знайдемо математичне сподівання показникового розподілу, тобто початковий момент першого порядку. За формулою (4.2) дістанемо:

$$M(x)=\int_{0}^{+\infty} x\cdot f(x)dx=\lambda\int_{0}^{+\infty} xe^{-\lambda x}dx=$$

$$\left\{\begin{matrix} x=u & du=dx \\ e^{-\lambda x}=dV & V=-\dfrac{1}{\lambda}e^{-\lambda x}\end{matrix}\right\}=\lambda\lim_{b\to\infty}\left(-\frac{1}{\lambda}xe^{-\lambda x}+\frac{1}{\lambda}\int_{0}^{b} e^{-\lambda x}dx\right)=$$

$$=\lambda\lim_{b\to\infty}\left(-\frac{1}{\lambda}be^{-\lambda b}-\frac{1}{\lambda^2}e^{-\lambda b}+\frac{1}{\lambda^2}\right)=\frac{1}{\lambda}.$$

Таким чином для показникового розподілу:

$$M(x)=\frac{1}{\lambda}. \tag{5.3}$$

Аналогічно, за формулою (4.8) виконуючи два рази інтегрування за частинами і переходячи до границі функції, знайдемо дисперсію показникового розподілу, тобто центральний момент другого порядку

$$D(x)=\lambda\int_{0}^{+\infty} x^2e^{-\lambda x}dx-\left(\frac{1}{\lambda}\right)^2=\frac{1}{\lambda^2}.$$

Середнє квадратичне відхилення відповідно до формули (4.9) дорівнює:

$$\sigma(x)=\sqrt{\frac{1}{\lambda^2}}=\frac{1}{\lambda}. \tag{5.4}$$

Як бачимо, параметр розподілу λ тісно пов'язаний з числовими характеристиками випадкової величини. Крім того, із (5.3) і (5.4) виходить, що для показникового закону характерна рівність середнього квадратичного відхилення і математичного сподівання. Цей факт допоможе надалі підбирати закон розподілу для емпіричних даних.

Застосуємо тепер метод моментів для визначення λ. Дорівняємо теоретичний і емпіричний моменти першого порядку, тобто математичне сподівання і середню вибіркову:

$$M(x) = \overline{x}^*.$$

Оскільки для показового закону $M(x) = \frac{1}{\lambda}$, отримаємо, що $\overline{x}^* = \frac{1}{\lambda} \quad \Rightarrow \quad \lambda = \frac{1}{\overline{x}^*}$.

Таким чином, для знаходження параметра показникового розподілу досить мати середню вибіркову за емпіричними даними.

Приклад 5.1

Нехай відомі результати перевірки деякого приладу на іспитовому стенді. Випробувано 100 зразків, вивчався час безвідмовної роботи приладу. Він склав: для 80 приладів менш 10 годин; для 10 приладів 10 – 20 годин; для 5 приладів 20 – 30 годин, для 5 приладів 30 – 40 годин. Визначити ймовірність того, що час роботи приладу складе не менш 20 годин.

Інтервальний ряд для цих даних подано у табл. 5.1.

Таблиця 5.1

Час безвідмовної роботи	0 – 10	10 – 20	20 – 30	30 – 40
Середини інтервалів	5	15	25	35
Кількість приладів (частота ознаки)	80	10	5	5
Відносна частота	0,8	0,1	0,05	0,05

Спираючись на вигляд гістограми (рис. 5.2) у грубому наближенні можна висловити припущення про те, що дана кількісна ознака – час безвідмовної роботи приладу – розподілена за показниковим законом.

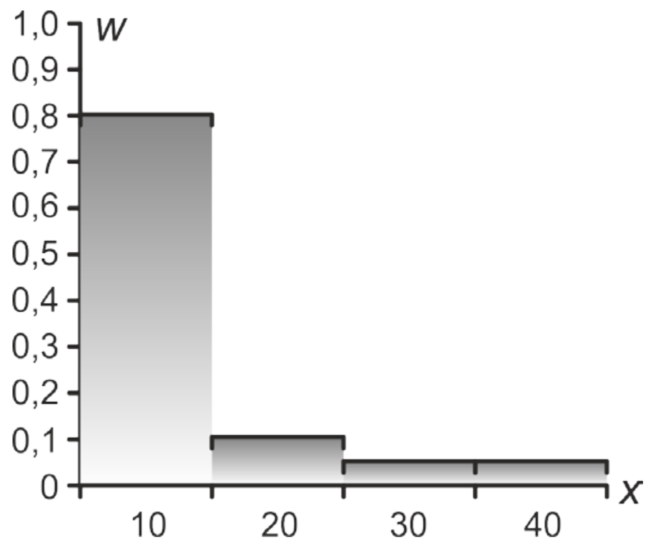


Рис. 5.2. Гістограма відносних частот для даних табл. 5.1

Визначимо емпіричні моменти розподілу.

Середня вибіркова:

$$\overline{X}^* = 5\cdot 0{,}8 + 15\cdot 0{,}1 + 25\cdot 0{,}05 + 35\cdot 0{,}05 = 4 + 1{,}5 + 3 = 8{,}5\ .$$

Вибіркова дисперсія:

$$D^* = 25\cdot 0{,}8 + 225\cdot 0{,}1 + 625\cdot 0{,}05 + 1225\cdot 0{,}05 - (8{,}5)^2 =$$
$$= 20 + 22{,}5 + 92{,}5 - 72{,}25 = 62{,}75.$$

Виправлена дисперсія і середнє квадратичне відхилення:

$$S^2 = \frac{n}{n-1}\cdot 62{,}75 = \frac{100}{99}\cdot 6{,}75 = 63{,}38;$$
$$\sigma^* = \sqrt{S^2} \quad = 7{,}9.$$

Бачимо, що математичне сподівання і середнє квадратичне відхилення є близькими за значеннями. Це також свідчить на користь показникового закону з диференціальною функцією $f(x) = \lambda\ e^{-\lambda x}$.

Визначимо за допомогою методу моментів параметр розподілу λ :

$$\lambda = \frac{1}{\overline{x}^*} = \frac{1}{8{,}5} = 0{,}117 \approx 0{,}12 \quad .$$

Тепер, маючи теоретичний закон розподілу даної кількісної ознаки можна прогнозувати ймовірність появи того чи іншого значення цієї випадкової величини. Відповімо на запитання задачі. Імовірність того, що час безвідмовної роботи приладу буде не менш 20 годин, визначимо за формулою (2.8)

$$P\,(20 \prec t \prec +\infty) = 0{,}12\int_{20}^{+\infty} e^{-0{,}12t}dt = 0{,}12\cdot\left(\frac{1}{-0{,}12}\right)\lim_{b\to+\infty}(e^{-b} - e^{-2{,}4}) = 0{,}091\ .$$

Додамо, що показниковий (експоненціальний) закон відіграє велику роль у теорії надійності механізмів і машин.

5.2. Нормальний закон розподілу

Звернемося тепер до гістограми, що зображена на рис. 3.2. Крива, що проведена через середини сторін багатокутників, у першому наближенні показує, який вигляд має диференціальна функція розподілу *f(x)*. У даному випадку крива має форму, майже симетричну щодо вертикальної прямої $x = \overline{x}^*$.

Характерно те, що велика частина значень кількісної ознаки групуються поблизу точки $x = \bar{x}*$, що є точкою максимуму кривої.

Велика кількість випадкових величин має подібну щільність розподілу ймовірностей. Її називають ***нормальним законом*** або законом ***Гауса*** (рис. 5.3). Вираз для диференціальної функції розподілу закону Гауса має вигляд:

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-a)^2}{2\sigma^2}} \tag{5.5}$$

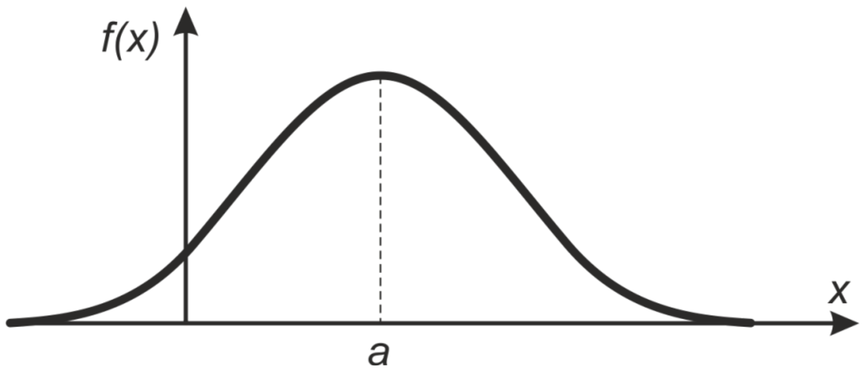


Рис. 5.3. Графік диференціальної функції нормального розподілу

Перш за все виникає питання, чи є функція (5.5) щільністю розподілу. Перевіримо рівність (2.11). Для цього зробимо заміну $\frac{x-a}{\sigma} = t \Rightarrow dt = \frac{1}{\sigma}dx$ та візьмемо до уваги інтеграл Ейлера – Пуассона

$$\int_{-\infty}^{+\infty} e^{\frac{-t^2}{2}} dt = \sqrt{2\pi} .$$

Тоді отримаємо:

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

Таким чином, рівність (2.11) виконується, площа під кривою (5.5) дорівнює одиниці.

Максимальна ордината кривої, що дорівнює $\frac{1}{\sigma\sqrt{2\pi}}$, відповідає точці $x=a$. При віддаленні від точки a щільність розподілу зменшується і при $x \to \pm\infty$ крива асимптотично наближається до осі абсцис.

З'ясуємо значення параметрів a і σ, що входять до виразу (5.5). Для цього знайдемо числові характеристики нормальної величини.

Відповідно до формули (4.2)

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

Використовуючи попередню заміну та інтеграл Ейлера – Пуассона отримаємо $M(x) = a$. Аналогічно можна показати, що

$$D(x) = \sigma^2, \quad \sigma(x) = \sigma .$$

Отже, параметр a являє собою математичне сподівання і характеризує центр розподілу; параметр σ характеризує розсіювання випадкової величини щодо центра розподілу і обумовлює крутість кривої: чим більше σ, тим більше розсіювання, тим менше крутість кривої.

Можна показати, що для нормальної випадкової величини (тобто для величини, що розподілена за нормальним законом) центральний момент 3-го порядку дорівнює нулю ($\mu_3 = 0$), а центральний момент четвертого порядку

$$\mu_4 = 3\sigma^4 .$$

Це означає, що і асиметрія і ексцес для нормального закону дорівнюють нулю ($A = \frac{\mu_3}{\sigma^3} = 0$; $E = \frac{\mu_4}{\sigma^4} - 3 = 0$).

Зауваження. Закон розподілу, що виражається диференціальною функцією

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-a)^2}{2\sigma^2}} ,$$

називається ***загальним нормальним*** розподілом. Для нього $M(x) = a ; \sigma(x) = \sigma$. Часто при розв'язуванні задач переходять до так званої *нормованої* величини $\frac{x-a}{\sigma} = t$. Для неї щільність розподілу має вигляд :

$$f^0(t) = \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}} \qquad (5.6)$$

і називається ***нормованим нормальним*** розподілом.

Математичне сподівання і середнє квадратичне відхилення для нормованої величини відповідно $M(t)=0$; $\sigma(t)=1$.

Інтегральна функція нормального розподілу має вигляд:
а) для загального розподілу

$$F(x)=\frac{1}{\sigma\sqrt{2\pi}}\int_{-\infty}^{x} e^{-\frac{(x-a)^2}{2\sigma^2}}dx\ , \tag{5.7}$$

б) для нормованого розподілу

$$F_0(t)=\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{t} e^{-\frac{t^2}{2}}dt\ . \tag{5.8}$$

Інтеграли (5.7), (5.8) не виражаються в елементарних функціях. Але для інтеграла (5.8) існують таблиці, де наведені значення функції $F_0(t)$ для різних значень аргументу *t* (додаток А). А у додатку Б наведено значення функції:

$$\Phi(t)=\frac{1}{\sqrt{2\pi}}\int_{0}^{t} e^{-\frac{t^2}{2}}dt\ , \tag{5.9}$$

яка називається функцією Лапласа.

Імовірність улучення нормальної випадкової величини в заданий інтервал

Відповідно до формули (2.8) ймовірність влучення випадкової величини в інтервал (x_1, x_2):

$$P(x_1<x<x_2)=\int_{x_1}^{x_2} f(x)dx=\frac{1}{\sigma\sqrt{2\pi}}\int_{x_1}^{x_2} e^{-\frac{(x-a)^2}{2\sigma^2}}dx\ .$$

Зробимо заміну: $t=\dfrac{x-a}{\sigma}$; $\quad dx=\sigma dt$; $\quad t_1=\dfrac{x-a}{\sigma}$; $\quad t_2=\dfrac{x-a}{\sigma}$.

Тоді отримаємо:

$$P(x_1<x<x_2)==\frac{1}{\sqrt{2\pi}}\int_{t_1}^{t_2} e^{-\frac{t^2}{2}}dt=F_0(t_2)-F_0(t_1)=F_0\left(\frac{x_2-a}{\sigma}\right)-F_0\left(\frac{x_1-a}{\sigma}\right).$$

Тут $F_0(t)$ – інтегральна функція (5.8) . Можна показати, що при такій самій заміні може бути використана і функція Лапласа (5.9). Таким чином, для нормального закону розподілу ймовірність потрапляння випадкової величини в заданий інтервал

$$P(x_1 < x < x_2) = F_0\left(\frac{x_2 - a}{\sigma}\right) - F_0\left(\frac{x_1 - a}{\sigma}\right) = \Phi\left(\frac{x_2 - a}{\sigma}\right) - \Phi\left(\frac{x_1 - a}{\sigma}\right). \quad (5.10)$$

Імовірність заданого відхилення. Правило «3-х сигм»

Знайдемо ймовірність того, що відхилення нормальної випадкової величини X від її математичного сподівання a не перевищує величини δ, тобто знайдемо ймовірність події $P(|x-a|<\delta)$. За формулою (5.10) дістанемо:

$$\begin{aligned} P(|x-a|<\delta) &= P(a-\delta < x < a+\delta) = \Phi\left(\frac{a+\delta-a}{\sigma}\right) - \Phi\left(\frac{a-\delta-a}{\sigma}\right) = \\ &= \Phi\left(\frac{\delta}{\sigma}\right) - \Phi\left(-\frac{\delta}{\sigma}\right) = 2\Phi\left(\frac{\delta}{\sigma}\right). \end{aligned} \quad (5.11)$$

Тут використаний той факт, що функція (5.9) непарна. Проаналізуємо такі випадки.

1. Нехай $\delta=\sigma$. Тоді

$$P(|x-a|<\sigma) = 2\Phi\left(\frac{\sigma}{\sigma}\right) = 2\Phi(1) = 0{,}68.$$

Тут значення функції $\Phi(1) = 0{,}34$ визначене за таблицею із додатку Б.

2. Нехай тепер δ=2σ. Тоді

$$P(|x-a|<2\sigma) = 2\Phi\left(\frac{2\sigma}{\sigma}\right) = 2\Phi(2) = 0{,}95$$

Тут $\Phi(2) = 0{,}478$.

3. Нехай δ=3σ. Тоді

$$P(|x-a|<3\sigma) = 2\Phi\left(\frac{3\sigma}{\sigma}\right) = 2\Phi(3) = 0{,}997.$$

Тут $\Phi(2) = 0{,}4986$.

Бачимо, що з імовірністю 0,997, тобто з *практичною вірогідністю*, значення нормальної величини потрапляють в інтервал

$$a - 3\sigma < x < a + 3\sigma. \qquad (5.12)$$

Цей факт називають «правилом 3-х сигм». За його допомогою можна судити про те, чи є дана випадкова величина розподіленою за нормальним законом чи ні. Таким чином, однією з підстав для вибору гіпотези про нормальний закон розподілу є виконання умов:

1) 99,7 % значення випадкової величини належать інтервалу

$$a - 3\sigma < x < a + 3\sigma;$$

2) асиметрія й ексцес близькі до нуля ($A \cong 0$, $E \cong 0$).

Нормальний закон розподілу відіграє особливу роль у теорії ймовірностей і статистики. У теорії ймовірностей доводиться так звана центральна гранична теорема, відповідно до якої сума великої кількості незалежних випадкових величин, підпорядкованих будь-якому розподілу, приблизно підпорядковується нормальному закону. Ця закономірність тим точніше, чим більша кількість випадкових величин складає цю суму.

Наприклад, помилка вимірів – це випадкова величина, що є сумою елементарних помилок, кожна з яких викликана дією певної причини, яка не залежить від інших. Якщо всі ці помилки відіграють у сумі рівномірно малу роль, то сама сума розподілена за нормальним законом.

Оскільки умови, що визначають нормальний розподіл, зустрічаються часто, даний закон розподілу ймовірностей здобув широке застосування. Зокрема, узагальнюючи результати іспитів гірських порід, дослідники часто дотримуються гіпотези про нормальний розподіл, параметри якого мають ясний фізичний зміст і легко визначаються за характеристиками вибірки.

Приклад 5.2

У розділах 3.1 та 4.3 як приклад вивчався статистичний розподіл такої кількісної ознаки, як міцність на одноосьовий стиск гірської породи. Крива ***f****, що вирівнює гістограму на рис. 3.4, має приблизно такий самий вигляд, як і нормальна щільність розподілу (рис. 5.3). На користь нормального розподілу свідчать і близькі до нуля значення асиметрії й ексцесу, а також виконання правила «3-х сигм». Дійсно, максимальне і мінімальне значення випадкової величини (x_{min}= *48,4*; x_{max}= *60,0)* укладаються в інтервал $(52{,}8 - 2{,}58*3; \quad 52{,}8 + 2{,}58*3)$. Тому в першому наближенні можна висунути гіпотезу про те, що значення досліджуваної кількісної ознаки – міцності на стиск деякої гірської породи – підпорядковуються нормальному закону розподілу.

Визначимо параметри цього розподілу на основі дослідних даних. Використовуємо метод моментів. Дорівняємо початкові статистичний і теоретичний моменти розподілу:

$$M(x) = \overline{x}^*.$$

Оскільки для нормального закону $M(x)=a$, отримаємо, що

$\bar{x}^*=a$, тобто $a = 52,8.$

Далі, щоб знайти ще один невідомий параметр (σ), дорівняємо центральні статистичний і теоретичний моменти другого порядку:

$$D(x)=S^2 .$$

Оскільки для нормального закону $D(x)=\sigma^2$ дістанемо, що

$$S^2=\sigma^2 \quad \rightarrow \quad \sigma = S = 2,58 .$$

Таким чином, для даної кількісної ознаки щільність розподілу ймовірностей описується функцією:

$$f(x)=\frac{1}{2,58\sqrt{2\pi}}\, e^{-\frac{(x-52,8)^2}{2*2,58^2}} .$$

Визначимо, наприклад, імовірність того, що міцність на стиск буде не менше 50 МПа. Використовуємо формулу (5.10):

$$P(x \ge 50)=1-P(x<50)=1-F_0\left(\frac{50-52,8}{2,58}\right)=1-F_0(-1,08)=1-0,1401=0,8599 .$$

Зауваження. У прикладі 5.2 теза про теоретичний закон розподілу ознаки була прийнята в першому наближенні. У відповідальних розрахунках потрібна перевірка статистичної гіпотези за так званими критеріями згоди. Цьому питанню присвячений розділ 7.

Іоганн Карл Фрідріх Гаус (1777 – 1855) — німецький математик, механік, фізик, астроном і геодезист. Вважається одним з найвеличніших математиків всіх часів, «королем математиків».

5.3. Логарифмічно нормальний розподіл

Якщо кількісна ознака має значно асиметричний розподіл, гіпотеза про нормальний закон є неправильною. Однак нормальному закону можуть підпорядковуватися логарифми значень досліджуваної ознаки. У цьому випадку розподіл ознаки називають ***логарифмічно нормальним*** (скорочено ***логнормальним*)**. Як приклад на рис. 5.4 наведено розподіли потужності пластів вугілля і логарифмів значень потужності.

Логарифмічно нормальний розподіл описує випадкову величину *X*, логарифм якої має нормальний розподіл з параметрами *a* і *σ*. Щільність розподілу величини $Z = \ln(X)$ має вигляд (5.3):

$$f(z) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(Z-a)^2}{2\sigma^2}}.$$

Тоді випадкова величина $X=e^Z$ розподілена за логарифмічно нормальним законом:

$$f(x) = \frac{1}{x\sigma\sqrt{2\pi}} e^{-\frac{(\ln x-a)^2}{2\sigma^2}}. \qquad (5.13)$$

а)

б)

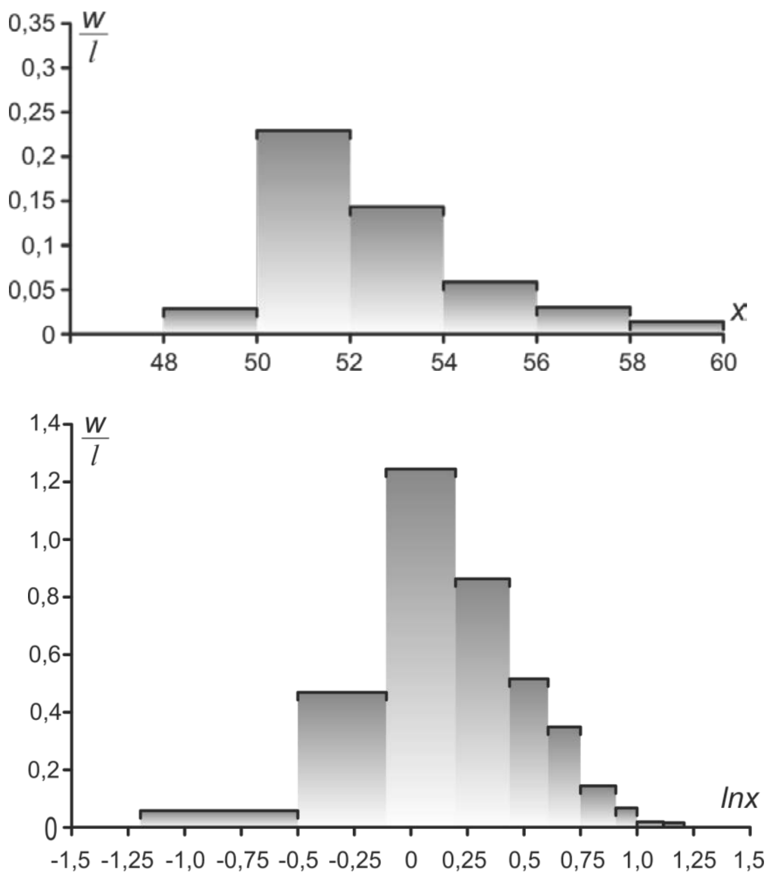


Рис. 5.4. Гістограма відносних частот:
а) потужностей пласта; б) логарифмів потужностей пласта

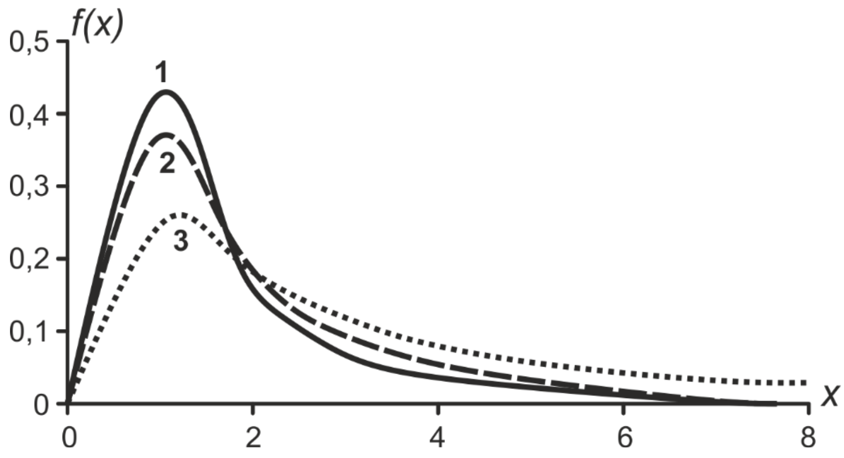
Важливо відзначити, що параметри a і σ для нормально розподіленої величини $Z = \ln(X)$ відповідно дорівнюють математичному сподіванню і середньому квадратичному відхиленню, але для логнормальної величини *X* це просто параметри форми і масштабу. Однак вони теж пов'язані з числовими характеристиками випадкової величини. Математичне сподівання і дисперсія для логнормальної величини виражаються через параметри розподілу таким чином:

$$M(X) = e^{a+\frac{\sigma^2}{2}}; \qquad (5.14)$$

$$D(X) = e^{2a+\sigma^2}(e^{\sigma^2} - 1). \qquad (5.15)$$

Форма кривої щільності логнормального розподілу залежить від співвідношення параметрів a і σ (рис. 5.5). Для неї характерна істотна правобічна асиметрія.

а)



б)

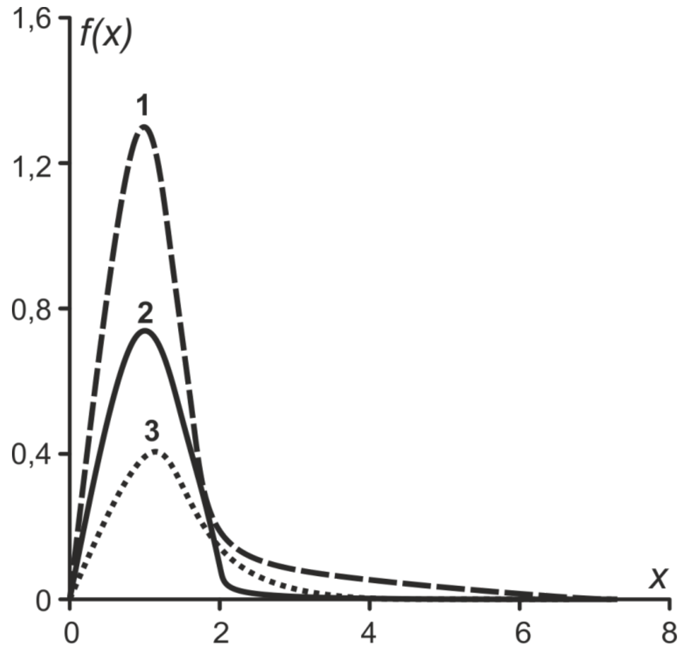


Рис. 5.5. Логарифмічно нормальний розподіл з різними значеннями параметрів: а) 1 – a = 0, σ^2=1; 2 – a = 0,3, σ^2=1; 3 – a = 1, σ^2=1; б) 1– a = 0, σ^2=0,1; 2 – a = 0, σ^2=0,3; 3 – a = 0, σ^2=1

Покажемо, як визначити параметри логнормального розподілу μ і σ за методом моментів. Використовуємо основні співвідношення:

$$M(x) = \overline{x}*,$$
$$D(x) = S^2.$$

З урахуванням формул (5.14) та (5.15) утворюємо систему щодо двох невідомих μ і σ:

$$\begin{cases} \overline{x}^* = \exp\left(\sigma^2/2 + \mu\right) \\ S^2 = \exp\left(\sigma^2 + 2\mu\right)\left[\exp\left(\sigma^2\right) - 1\ \right] \end{cases}.$$

Логарифмуючи обидві частини цих рівностей, дістанемо:

$$\begin{cases} \ln \overline{x}^* = \dfrac{\sigma^2}{2} + \mu \\ \ln S^2 = \sigma^2 + 2\mu + \ln\left(\exp\left(\sigma^2\right) - 1\right) \end{cases}.$$

Виключимо μ з рівнянь:

$$\ln \frac{S^*}{\overline{x}^{*2}} = \ln\left(\exp\left(\sigma^2\right) - 1\right).$$

Враховуючи, що $\dfrac{\sqrt{S^2}}{\overline{x}^2} = \eta^2$, отримаємо:

$$\sigma = \sqrt{\ln\left(\eta^2 + 1\right)}$$
$$\mu = \ln \overline{x}^* - \frac{1}{2}\ln\ \left(\eta^2 + 1\right). \tag{5.16}$$

Формули (5.16) є остаточними для визначення параметрів логнормального розподілу за вибірковими даними.

Логарифмічно нормальний розподіл ймовірностей є досить широко розповсюдженою статистичною моделлю опису явищ і процесів у науках про Землю. Цьому розподілу підпорядковується вміст елементів і мінералів у вивержених гірських породах, розміри часток осадових порід, розміри часток при подрібненні твердих тіл зосередженою силою, величини граничних руйнівних напружень для деяких типів порід тощо.

Значний обсяг випробувань зразків осадових, магнетичних і метаморфічних гірських порід на розтягання і стискання показали, що найбільш прийнятною статистичною моделлю для границі міцності є логарифмічно нормальний розподіл.

Логарифмічно нормальний розподіл можна вивести як статистичну модель випадкової величини, значення якої утворюються у результаті множення великої кількості елементарних помилок, аналогічно тому, як нормальний розподіл є моделлю їх суми.

Властивості логарифмічно нормального закону виходять з властивостей відповідного нормального розподілу. Крім того, цей розподіл має найважливішу особливість: розподіл добутку *n* незалежних позитивних випадкових величин з логарифмічно нормальними розподілами знову підпорядковується цьому розподілу. Для нього має місце аналог центральної граничної теореми: розподіл добутку незалежних позитивних випадкових величин при деяких спільних умовах прагне до логарифмічно нормального закону при необмеженому зростанні числа співмножників.

5.4. Гамма-розподіл

Диференціальна функція гамма-розподілу має вигляд:

$$f(x) = \begin{cases} \dfrac{\lambda^{\eta}}{\Gamma(\eta)} x^{\eta-1} \cdot e^{-\lambda x} \; при \quad x \geq 0; \quad \lambda > 0 \quad \eta > 0 \\ 0 \qquad в \quad інших \quad випадках. \end{cases} .$$

Тут $\Gamma(\eta)$ – гамма-функція:

$$\Gamma(\eta) = \int_0^{\infty} x^{\eta-1}\ e^{-x} dx .$$

Залежно від параметрів η і λ крива $f(x)$ приймає різні форми (рис. 5.6).

Математичне сподівання і дисперсія гамма-розподілу зв'язані з параметрами розподілу співвідношеннями:

$$M(X) = \frac{\eta}{\lambda}; \qquad D(X) = \frac{\eta}{\lambda^2} .$$

При $\eta=1$ гамма-функція приймає значення:

$$\Gamma(1) = \int_0^{\infty} e^{-x} dx = -e^{-x}\Big|_0^{+\infty} = -(\frac{1}{e^{\infty}} - e^0) = 1$$

і тоді гамма-розподіл має вигляд:

$$f(x) = \begin{cases} \lambda e^{-\lambda x} \ при \quad x \geq 0; \quad \lambda > 0 \\ 0 \qquad в \quad інших \quad випадках \end{cases}.$$

Це ні що інше, як щільність показникового (експоненціального) розподілу. Таким чином, показниковий розподіл є частковим випадком гамма-розподілу.

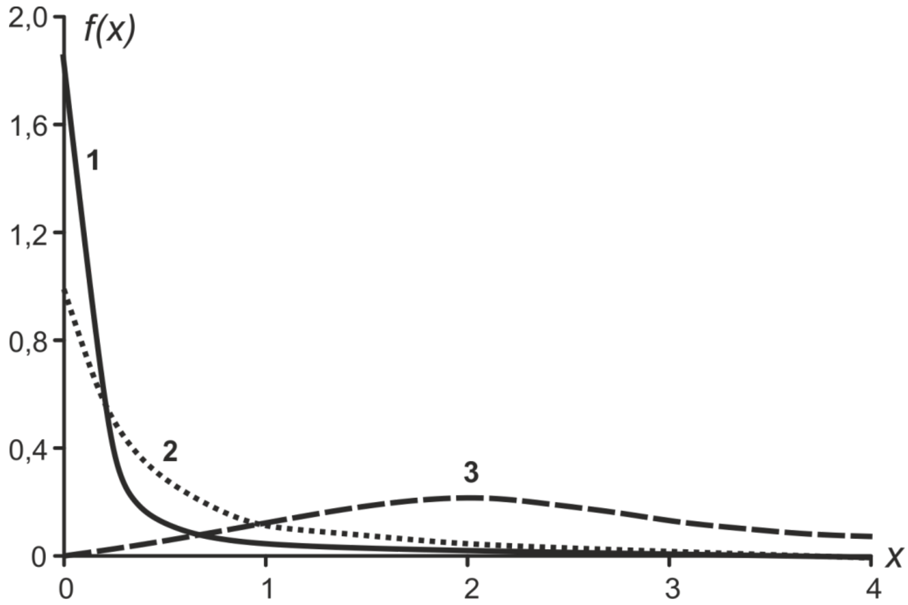


Рис. 5.6. Гамма-розподіл з параметрами:
1 *-η =0,5; λ=1;* ***2*** *-η =1; λ=1;* ***3*** *-η =3; λ=1*

Гамма-розподіл також як і показниковий має широке застосування в задачах надійності, у теорії масового обслуговування. Крім того, експериментально встановлено, що цьому розподілу підпорядковуються такі величини як вміст корисного компонента в пробах гірських порід.

Докладніше про теоретичні розподіли, які є основою для побудови статистичних моделей у техніці, можна ознайомитися в монографії [4].

5.5. Вибір теоретичного розподілу за експериментальними даними

Для вибору теоретичного розподілу за експериментальними даними автори [4] рекомендують використовувати графік К. Пірсона (рис. 5.7). На цьому графіку по осі абсцис відкладають нормований показник асиметрії:

$$\beta_1 = \left(\frac{\mu_3}{\sigma^3} \right)^2 \quad , \tag{5.17}$$

по осі ординат – показник ексцесу:

$$\beta_2 = \frac{\mu_4}{\sigma^4} \quad . \tag{5.18}$$

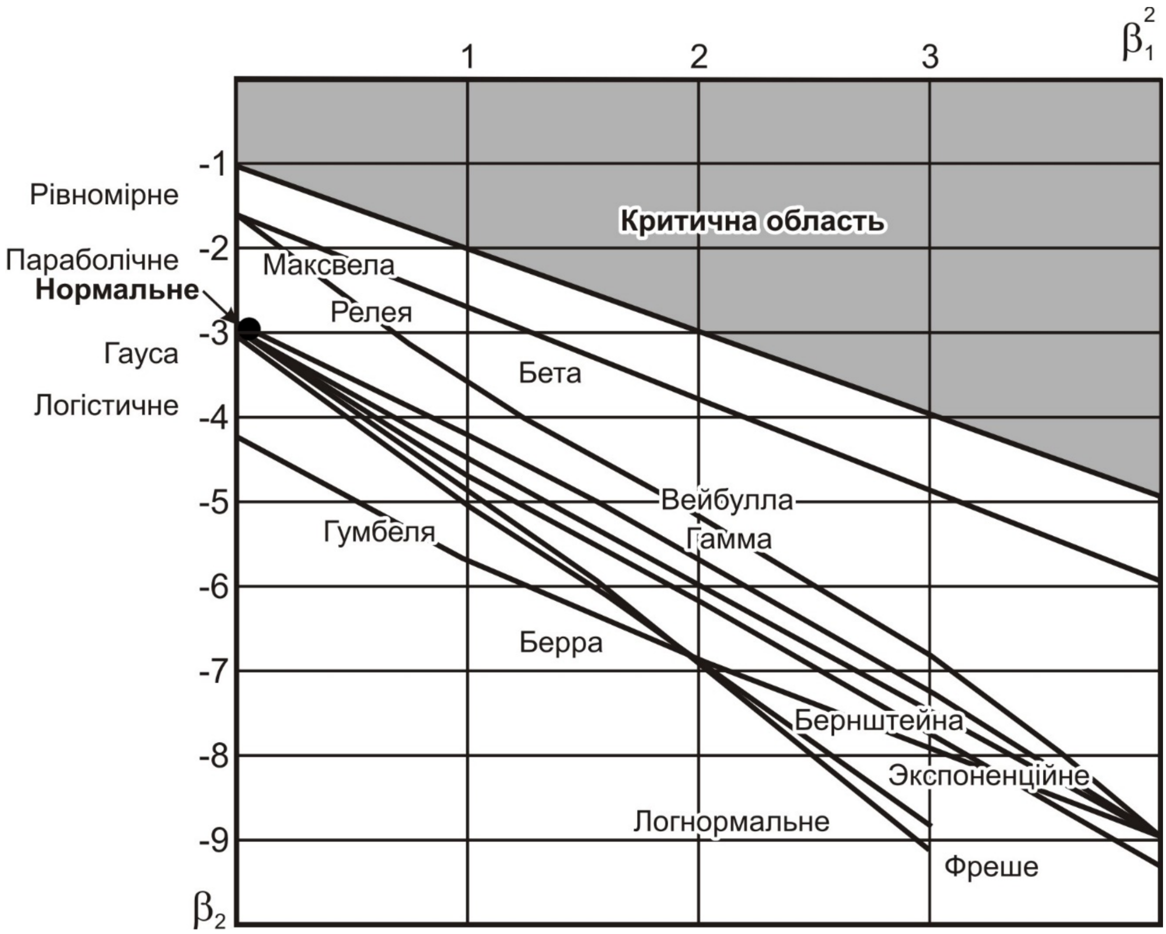


Рис. 5.7. Графік Пірсона для вибору теоретичних розподілів

Неважко переконатися, що для нормального розподілу $\beta_1 = 0$, а $\beta_2 = 3$. Таким чином, на графіку Пірсона нормальному розподілу відповідає точка з координатами (0, 3). Для показового розподілу $\beta_1 = 4$ і $\beta_2 = 9$, і йому також відповідає на графіку точка з координатами (4, 9). Для інших розподілів, таких як логарифмічно нормальний та гамма-розподіл, показники асиметрії й ексцесу змінюються залежно від параметрів розподілу, тому на графіку їм відповідають вже не точки, а лінії (рис. 5.7). Деякі розподіли займають цілу ділянку. Для того, щоб підібрати розподіл для статистичних даних емпіричні точки з координатами *(β_1^*; β_2^*),* наносять на графік Пірсона. Якщо ці точки групуються поблизу лінії або точки, що відповідає тому чи іншому розподілу, висувають гіпотезу про те, що статистичні дані підпорядковуються саме цьому розподілу. Далі прийняту гіпотезу варто підтвердити за одним з критеріїв згоди, що наведені у підручниках [1 – 4] та будуть розглянуті нижче в розділі 7.

Приклад 5.3

Підберемо за графіком Пірсона закон розподілу кількісної ознаки, що розглянута в розділі 4.3 (приклад 4.2.). Для наведених статистичних даних нормований показник асиметрії

$$\beta_1 = \left(\frac{\mu_3}{\sigma^3}\right)^2 = \left(\frac{4,72}{2,58^3}\right) = 0,28^2 = 0,0784 \quad .$$

Нормований показник ексцесу

$$\beta_2 = \frac{\mu_4}{\sigma^4} = \frac{104,76}{2,58^4} = 2,37 .$$

Таким чином, даному статистичному розподілу на графіку Пірсона відповідає точка (0,078; 2,37), що розташована поблизу точки (0; 3), яка відповідає нормальному закону розподілу. Тоді, додатково до виконання правила «3-х сигм» (приклад 5.2), маємо ще одну підставу висунути гіпотезу про підпорядкування розглянутих в прикладі 4.2 статистичних даних нормальному розподілу.

5.6. Деякі теоретичні розподіли, що застосовуються в математичній статистиці. Розподіл «Хі-квадрат»

Цей розподіл («χ» – грецька буква, вимовляється «хі») тісно пов'язаний з нормальним. Вивчається варіанта:

$$\chi^2 = \sum_{\mathrm{i}=1}^{n} y_i^2 , \tag{5.19}$$

де y_i^2 – незалежні випадкові величини, розподілені за нормальним законом з параметрами $a = 0,\quad \sigma = 1$. При цьому величина χ^2 має свій закон розподілу (так званий «χ -квадрат» або «χ^2», вимовляється «хі-квадрат»). Можна довести, що цей закон має щільність розподілу

$$f(x,v)=\begin{cases}\frac{1}{2^{\frac{v}{2}}\Gamma(v/2)}\cdot x^{\frac{v}{2}-1}\cdot e^{-\frac{x}{2}}, & x>0\\ 0 & x\le 0\end{cases} \quad . \tag{5.20}$$

При цьому

$$v=n-r \quad , \tag{5.21}$$

де r – число зв'язків, накладених на величину y_i^2. Якщо $y_i^2=\frac{x_i-x}{\sigma}$ і приймається, що $\bar{x}=\overline{x^*}=\frac{1}{n}\sum_{i=1}^{n}x_i$, то на y_i^2 накладається один зв'язок і r = 1; коли приймається, крім цього, що $\sigma=\sigma^*$, то r = 2 тощо; v – називається **числом степенів свободи**.

З формули (5.20) видно, що гамма-розподіл при $\beta=\frac{1}{2}, \alpha=\frac{v}{2}$ збігається із розподілом χ^2. Із збільшенням v розподіл χ^2 повільно наближається до нормального (рис. 5.8).

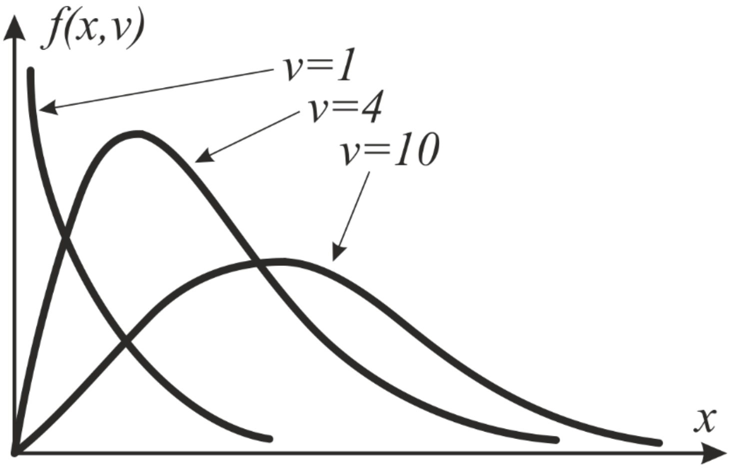


Рис. 5.8. Щільність розподілу закону χ^2 при різних значеннях параметру v

Можна показати, що

$$M(\chi^2)=v \tag{5.22}$$
$$D(\chi^2)=2v \tag{5.23}$$

Вперше χ^2-розподіл було розглянуто Р. Хельмертом (1876 р.) і К.Пірсоном (1900 р.).

Карл (Чарльз) Пірсон (1857 – 1936 рр.) – видатний англійський математик, біолог, філософ, член Лондонського королівського товариства, професор Лондонського університету. Його основні праці пов'язані з математичною статистикою, теорією кореляції, критеріями згоди.

5.7. Розподіл Стьюдента

Нехай Y – нормальна випадкова величина з $a = 0$, $\sigma = 1$, Z – незалежна від Y випадкова величина, розподілена за законом χ^2 з v степенями свободи. Можна довести, що випадкова величина

$$T = \frac{Y}{\sqrt{Z/v}} \quad . \tag{5.24}$$

розподілена за **законом Стьюдента** із щільністю ймовірності

$$t(x; v) = \frac{\Gamma\left(\frac{v-1}{2}\right)}{\sqrt{\pi v}\,\Gamma\left(\frac{v}{2}\right)} \cdot \left(1 + \frac{x^2}{v}\right)^{-\frac{v+1}{2}}. \tag{5.25}$$

Цей розподіл, як і χ^2, містить один параметр v – число степенів свободи, що обчислюється за формулою (5.21).

Зі зростанням v розподіл Стьюдента швидко наближується до нормального (рис. 5.9).

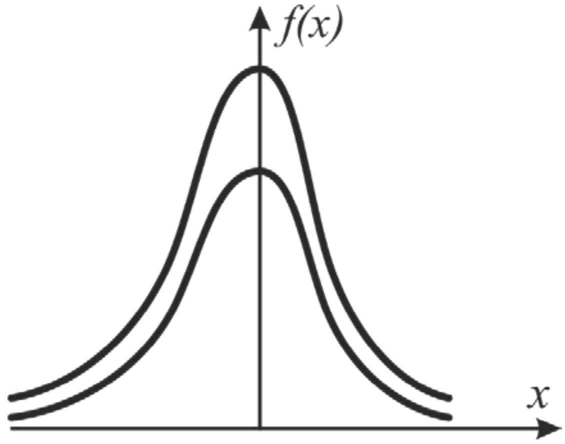


Рис. 5.9. Щільність розподілу закону Стьюдента

Стьюдент (англійською – «студент») – псевдонім **Вільяма Сілі Госсета** (1876 – 1937 рр.), англійського статистика, що вивчав довірчі інтервали, критерії згоди.

Контрольні питання

1. Наведіть графік і аналітичний вираз для щільності розподілу показникового закону. Яка випадкова величина описується показниковим законом розподілу**?**
2. У чому полягає метод моментів оцінки параметрів розподілу**?**
3. Наведіть графік і аналітичний вираз для щільності розподілу нормального закону. Які випадкові величини описуються нормальним законом розподілу**?**
4. Як зв'язані параметри розподілу нормального закону з числовими характеристиками випадкової величини**?**

5. Що таке нормована нормальна величина**?**
6. Чим відрізняється загальний нормальний розподіл від нормованого нормального розподілу**?**
7. Як знайти ймовірність улучення нормальної випадкової величини в заданий інтервал**?**
8. Як визначити ймовірність заданого відхилення**?**
9. У чому полягає правило «трьох сигм»**?**
10. Чому дорівнює асиметрія й ексцес нормального розподілу**?**
11. Назвіть інші закони розподілу, що використовуються для опису кількісних ознак.
12. Як підібрати теоретичний закон розподілу для емпіричних даних**?**

6. ІНТЕРВАЛЬНЕ ОЦІНЮВАННЯ НЕВІДОМИХ ПАРАМЕТРІВ РОЗПОДІЛУ

В розділі 4.3 отримані точкові оцінки невідомих параметрів розподілу за вибірковими даними, що є результатом досліду. Вони мають суттєвий недолік: невідомо з якою точністю визначають параметр, що оцінюється. Якщо при великої кількості спостережень точність звичайно буває достатньою для практичних висновків, то для вибірок невеликого об'єму питання про точність оцінювання є суттєвим. Цей факт говорить про важливість іншого підходу, що пов'язаний з інтервальним оцінюванням невідомих параметрів розподілу.

6.1. Характеристики похибок випробування

Нехай θ – шуканий параметр розподілу, а θ^* – його точкова оцінка, знайдена за даними вибірки. Точність оцінки характеризують за допомогою величини ε, такою, що $\left|\theta-\theta^*\right|<\varepsilon$. Параметр θ невідомий, тому при заданому ε можна поставити питання лише про ймовірність події $\left|\theta-\theta^*\right|<\varepsilon$. Її називають ***довірчою ймовірністю*** або ***надійністю*** та позначають γ, тобто

$$\gamma = P\left(\left|\theta-\theta^*\right|<\varepsilon\right).$$

Нерівність $\left|\theta-\theta^*\right|<\varepsilon$ можна записати у вигляді:

$$\theta^*-\varepsilon<\theta<\theta^*+\varepsilon .$$

Тоді довірча ймовірність (надійність) має вигляд:

$$\gamma = P\left(\theta^*-\varepsilon<\theta<\theta^*+\varepsilon\right). \qquad (6.1)$$

Звичайно надійність γ задають заздалегідь. Залежно від змісту задачі γ покладають таким, що дорівнює одному із значень: *γ=0,95; 0,98; 0,99; 0,999.*

Так у випадку, якщо *γ=0,99*, то рівність (6.1) означає, що з ймовірністю 0,99 параметр θ попадає в інтервал $I_\gamma=\left(\theta^*-\varepsilon,\theta^*+\varepsilon\right)$. Цей інтервал називають ***довірчим.*** Іншими словами, γ – імовірність того, що інтервал I_γ покриває невідомий параметр θ.

Поняття ***довірчого інтервалу*** дозволяє охарактеризувати похибку випробування з допомогою двох величин – довжини інтервалу 2ε і надійності γ. Якщо для випадкової величини $\left|\theta-\theta^*\right|$ відома щільність розподілу $f(x)$, то

$$\gamma = P\left(\left|\theta - \theta^*\right| < \varepsilon\right) = \int_{-\varepsilon}^{\varepsilon} f(x)dx. \tag{6.2}$$

Як відомо, для низки найбільш вживаних законів розподілу значення інтегралів типу (6.2) обчислені та зведені у спеціальні таблиці довідкової літератури.

У випадку нормального закону розподілу отримаємо надійність за формулою:

$$\gamma = \int_{-\varepsilon}^{\varepsilon} f(x)dx = \Phi\left(\frac{\varepsilon - a}{\sigma_1}\right) - \Phi\left(\frac{-\varepsilon - a}{\sigma_1}\right), \tag{6.3}$$

де $a = M(\theta - \theta^*)$; $\sigma_1 = \sqrt{D(\theta - \theta^*)}$.

Для незміщеної оцінки $M(\theta^*) = \theta$ та $a = 0$. У такому випадку рівність (6.3) має вигляд:

$$\gamma = 2\Phi\left(\frac{\varepsilon}{\sigma_1}\right). \tag{6.4}$$

6.2. Довірчий інтервал для математичного сподівання нормальної випадкової величини *X* при відомому середньоквадратичному відхиленні

У тих випадках, коли вимірювання виконуються за допомогою добре відомого методу, ми заздалегідь знаємо похибки виміру. Це означає, що заздалегідь відома величина середньоквадратичного відхилення σ. Нехай для шуканого параметру $\bar{x}$ отримана оцінка : $\bar{x}^* = \frac{1}{n}\sum_{i=1}^{n} x_i$.

Тоді

$$\sigma_1^2 = D(\bar{x} - \bar{x}^*) = D(\bar{x}) - D(\bar{x}^*) = D\left(\frac{1}{n}\sum_{i=1}^{n} x_i\right) = \frac{1}{n^2} nD(x) = \frac{\sigma^2}{n} \quad \text{або} \quad \sigma_1 = \frac{\sigma}{\sqrt{n}}. \tag{6.5}$$

При виведенні формули (6.5) були враховані відомі властивості дисперсії:

$$\begin{gathered} D(x \pm y) = D(x) + D(y); \\ D(c) = 0, \quad \text{де} \quad c = const; \\ D(cx) = c^2 D(x); \\ D\left(\sum_{i=1}^{n} x_i\right) = nD(x_i) = nD(x). \end{gathered} \tag{6.6}$$

Підставимо (6.5) у (6.4) та отримаємо:

$$\gamma = 2\Phi\left(\frac{\varepsilon\sqrt{n}}{\sigma}\right) = 2\Phi(t), \qquad (6.7)$$

де $t = \dfrac{\varepsilon\sqrt{n}}{\sigma}$.

Таким чином, для того щоб визначити довірчий інтервал для математичного очікування нормальної випадкової величини при відомому σ необхідно:

1. для γ за таблицею функції Лапласа $\Phi(t)$ (Додаток Б) знайти таке значення аргументу t, щоб виконувалась умова: $\gamma=2\Phi(t)$, тобто $\Phi(t)= \gamma/2$ (число t іноді називають квантилем);
2. за допомогою формули (6.7) для знайденого значення t обчислити ε:

$$\varepsilon = \frac{\sigma t}{\sqrt{n}}; \qquad (6.8)$$

3. записати нерівності:

$$x^* - \varepsilon < \bar{x} < \bar{x}^* + \varepsilon$$

$$\text{або} \qquad (6.9)$$

$$x^* - \frac{\sigma t}{\sqrt{n}} < \bar{x} < \bar{x}^* + \frac{\sigma t}{\sqrt{n}}.$$

Отже, з надійністю γ теоретичний параметр, що оцінюється, $\bar{x}$ потрапляє в інтервал $\left(\bar{x}^* - \dfrac{\sigma t}{\sqrt{n}}, \bar{x}^* + \dfrac{\sigma t}{\sqrt{n}}\right)$.

З формули (6.8) визначаємо необхідний обсяг вибірки n, при якому значення $\bar{x}$ буде визначено із заданою похибкою ε і заданою надійністю γ:

$$n = \frac{\sigma^2 t^2}{\varepsilon^2}.$$

Приклад 6.1

Повернемось до даних випробувань межі міцності на стиск зразків алевроліту, що представлені вибіркою обсягом $n = 40$ та згруповані в інтервальний ряд (табл. 3.4). В прикладі 4.2 розраховані параметри статистичного розподілу цієї кількісної ознаки: середня вибіркова $\bar{x}^* = 52{,}85$ МПа та середнє квадратичне відхилення (стандарт) $s = 2{,}58$ МПа. Припустимо, що таке відхилення є стандартним для випробувань в даних лабораторних умовах, тобто будемо вважати, що величина $\sigma = 2{,}58$ є похибкою методу.

За оцінками асиметрії та ексцесу приймаємо гіпотезу про нормальний закон розподілу (приклад 5.3). Визначимо з надійністю 0,98 довірчий інтервал для теоретичного середнього (математичного сподівання) досліджуваної ознаки:

1. За умовою $\gamma = 2\Phi(t) = 0{,}98$, тому Ф(t)=0,49. Для значення функції Лапласа 0,49 знаходимо (додаток Б) *t=2,32*.

2. За формулою (6.8) визначаємо: $\varepsilon = \frac{\sigma \cdot t}{\sqrt{n}} = \frac{2{,}58 \cdot 2{,}32}{\sqrt{40}} = 0{,}946.$

3. Отже, довірчий інтервал для математичного сподівання $\overline{x}$ має вигляд:

$$52{,}85 - 0{,}946 < \overline{x} < 52{,}85 + 0{,}946$$

Тобто з імовірністю 0,98 шуканий теоретичний параметр $\overline{x}$ (міцність на стиск алевроліту) знаходиться в інтервалі $51{,}90 < \overline{x} < 53{,}80$.

Надійність $\gamma = 0{,}98$ означає, що тільки в середньому в двох випадках з 100 похибка при визначенні $\overline{x}$ перевищує за модулем $\varepsilon = 0{,}946$ МПа.

Приклад 6.2

Визначимо тепер, яким повинен бути обсяг вибірки *n* (скільки зразків необхідно випробувати), щоб гранична помилка $\varepsilon = \pm 0{,}946$ була гарантована з надійністю *γ=0,997*.

У даному випадку відомі *γ=0,997*, $\overline{x}^* = 52{,}85$ МПа, $\sigma = 2{,}58$ МПа. Закон розподілу кількісної ознаки – нормальний. Оскільки за умовою $2\Phi(t) = \gamma = 0{,}997$, то Ф(t)=0. По таблиці знаходимо відповідне t=3. За формулою (6.8) визначимо необхідний обсяг вибірки:

$$n = \frac{t^2\sigma^2}{\varepsilon^2} = \frac{3^2 \cdot 2{,}58^2}{0{,}946^2} = 66{,}94 = 67 \text{ зразків}$$

Таким чином, для підвищення надійності (від 0,98 до 0,997) покриття довірчим інтервалом шуканого теоретичного параметру – міцності на стиск алевроліту – необхідно збільшити кількість випробувань з 40 до 67 зразків.

6.3. Довірчий інтервал для математичного очікування нормальної величини *X* при невідомому квадратичному відхиленні σ

Як правило, похибка методу вимірювання заздалегідь невідома і її треба визначати в процесі вимірювань. При порівняно невеликому обсязі вимірювань ми маємо можливість визначати лише величину s – виправлену дисперсію, яку не можна розцінювати як стандартну похибку методу σ.

Нехай X – нормальна випадкова величина. Тоді випадкова величина

$$Y = \frac{\overline{x} - \overline{x}^*}{\sigma / \sqrt{n}} \qquad (6.10)$$

також розподілена за нормальним законом з параметрами $M(Y) = 0$ та $\sigma(Y) = 1$.

Дійсно $M(\overline{x} - \overline{x}^*) = M(\overline{x}) - M(\overline{x}^*) = \overline{x} - M\left(\frac{1}{n}\sum_{i=1}^{n} x_i\right) = \overline{x} - \frac{1}{n} \cdot n\overline{x} = 0.$

$$D(Y)=D\left(\frac{\bar{x}-\bar{x}^*}{\sigma/\sqrt{n}}\right)=\frac{n}{\sigma^2}\cdot\left(D(\bar{x})-D(\bar{x}^*)\right)=\frac{n}{\sigma^2}D\left(\frac{1}{n}\sum_{i=1}^{n}x_i\right)=\frac{n}{\sigma^2}\cdot\frac{1}{n^2}\cdot nD(x)=1.$$

При доведенні цих фактів були використані відомі властивості математичного очікування та дисперсії (6.6):

1. $M(c)=c,\quad \textit{якщо}\quad c=const;$
2. $M(cx)=cM(x),\quad \textit{якщо}\quad c=const;$
3. $M(x+y)=M(x)+M(y).$

Розглянемо випадкову величину:

$$Z=\frac{(n-1)S^2}{\sigma^2}=\left\{\begin{array}{c} S^2=\dfrac{n}{n-1}D(x) \\ (n-1)S^2=nD(x)\end{array}\right\}=\frac{nD(x)}{\sigma^2}=\frac{n}{\sigma^2}\sum_{i=1}^{n}\frac{(x_i-\bar{x}^*)^2}{n}=\sum_{i=1}^{n}\left(\frac{x_i-\bar{x}^*}{\sigma}\right)^2. \qquad (6.11)$$

Такі величини підпорядковуються закону χ^2 (див. розділ 5.6) з числом ступенів свободи, що обчислюється: $\nu=n-1$. Розглянемо також величину $t=\dfrac{Y}{\sqrt{Z/(n-1)}}$, яка з урахуванням співвідношень (6.10) та (6.11) записується у вигляді:

$$t=\frac{Y}{\sqrt{Z/(n-1)}}=\frac{\sqrt{n}(\bar{x}-\bar{x}^*)}{\sigma\sqrt{S^2/\sigma^2}}=\frac{\sqrt{n}}{S}(\bar{x}-\bar{x}^*)\,. \qquad (6.12)$$

Така величина змінюється за законом Стьюдента, що наведений в розділі 5.6. Враховуючи (6.12), виконаємо тотожні перетворення:

$$\left(\left|\bar{x}-\bar{x}^*\right|<\varepsilon\right)\Rightarrow\left(\frac{\sqrt{n}}{S}\left|\bar{x}-\bar{x}^*\right|<\frac{\varepsilon\sqrt{n}}{S}\right)\Rightarrow\left(t<t_\gamma\right), \qquad (6.14)$$

де покладено:

$$t_\gamma=\frac{\varepsilon\sqrt{n}}{S}. \qquad (6.13)$$

Отже, точність визначення $\bar{x}$ у випадку нормальної випадкової величини при невідомому σ можна оцінити довірчою ймовірністю (надійністю) $\gamma=p(|t|<t_\gamma),$ яка може бути визначена:

$$\gamma=p(|t|<t_\gamma)=\int_{-t_\gamma}^{t_\gamma}f(x,\nu)dx, \qquad (6.14)$$

де $f(x,v)$ – щільність розподілу Стьюдента (5.25). Значення інтегралу (5.25) можна одержати за таблицею розподілів Стьюдента у додатку В.

Таким чином, якщо відомі γ та n, за таблицею з додатку В знаходимо t_γ. Після цього за формулою (6.13) знаходимо $\varepsilon = \frac{St_\gamma}{\sqrt{n}}$. Тим самим визначаємо довірчий інтервал:

$$x^* - \frac{St_\gamma}{\sqrt{n}} < \bar{x} < \bar{x}^* + \frac{St_\gamma}{\sqrt{n}}. \qquad (6.15)$$

Зауважимо, що інтервал (6.15) істотно відрізняється від інтервалу (6.9) лише для малих вибірок. Прийнято при $n > 30$ користуватись вже не виразом (6.15), а виразом (6.5).

Приклад 6.3

Припустимо, що випробування міцності алевроліту (приклад 6.1) виконувались тільки на 10-ти зразках. Тобто обсяг вибірки $n = 10$. Припустимо, що отримані такі самі характеристики розподілу: середня вибіркова $\bar{x}^* = 52{,}85$кг/см^2 та середнє квадратичне відхилення (стандарт) $s = 2{,}58$ МПа. В цьому випадку у зв'язку з невеликим обсягом вибірки не можна вважати величину $s = 2{,}58$ відомою похибкою методу. Тому для знаходження довірчого інтервалу, що покриває з надійністю істинне значення міцності даної гірської породи, використовуємо формулу (6.15).

Розв'язок. Будемо визначати довірчий інтервал з надійністю *γ=0,99*. За допомогою таблиці (додаток В) для $n = 10$ і $\gamma = 0{,}99$ знаходимо $t_\gamma = 3{,}25$. За формулою (6.15) визначаємо:

$$\varepsilon = \frac{t_\gamma S}{\sqrt{n}} = \frac{3{,}25 \cdot 2{,}58}{\sqrt{10}} = 2{,}651.$$

Таким чином, теоретичне значення міцності (істинне значення) $\bar{x}$ з ймовірністю 0,99 належить інтервалу.

$$52{,}85 - 2{,}65 < \bar{x} < 52{,}85 + 2{,}65;$$
$$50{,}20 < \bar{x} < 55{,}50.$$

Бачимо, що цей інтервал значно більше ніж той, що ми отримали в прикладі 6.1. Цей результат обумовлений малим обсягом вибірки. Для більш точної оцінки треба збільшувати кількість випробувань. Так, при обсязі $n = 40$ і $\gamma = 0{,}99$ отримаємо $t_\gamma = 2{,}708$.

Тоді $\varepsilon = \dfrac{t_\gamma S}{\sqrt{n}} = \dfrac{2{,}708 \cdot 2{,}58}{\sqrt{40}} = 1{,}104$.

Цей результат близький до того, що отриманий за формулою 6.8 у прикладі 6.1.

Контрольні питання

1. Що називають довірчою ймовірністю?
2. Що таке довірчий інтервал?
3. Як визначається довірчий інтервал для математичного сподівання нормальної випадкової величини?
4. Як визначається довірчий інтервал для математичного очікування нормальної величини при заздалегідь невідомому квадратичному відхиленні?

7. ПЕРЕВІРКА ГІПОТЕЗИ ПРО ЗАКОН РОЗПОДІЛУ ГЕНЕРАЛЬНОЇ СУКУПНОСТІ. КРИТЕРІЙ ПІРСОНА

На основі обробки статистичних даних, графічного зображення статистичного розподілу у вигляді гістограми частот, аналізу точкових оцінок моментів розподілу, а також виходячи з фізичного змісту досліджуваної кількісної ознаки можна висунути гіпотезу H (зробити припущення) про те, що випадкова величина X, що досліджується, підпорядковується деякому закону розподілу. Розподіл, що має випадкова величина за висунутою гіпотезою, називають теоретичним (гіпотетичним).

Для остаточного висновку про закон розподілу необхідно перевірити, наскільки припущення про закон розподілу узгоджується з статистичними даними, тобто експериментом. І навіть, якщо зроблене припущення про закон розподілу є правильним, експериментальний (статистичний) закон розподілу буде в якийсь мірі відрізнятся від теоретичного. Тому потрібно вирішити задачу: чи є розходження між статистичним $f^*(x)$ та теоретичним $f(x)$ законами розподілу наслідком обмеженої кількості спостережень, тобто випадковим, чи це розходження є суттєвим і пов'язане з тим, що дійсний розподіл випадкової величини X відрізняється від гіпотетичного. Тому необхідні критерії, які дозволяють вирішити, чи узгоджуються значення випадкової величини X: $x_1, x_2, ..., x_n$ з гіпотезою щодо її щільності розподілу. Такі критерії називають критеріями згоди.

Обмежимося описом застосування критерію χ^2, що був запропонований К. Пірсоном. Відповідно цьому критерію порівнюються емпіричні частоти та теоретичні ймовірності. Розглянемо величину:

$$\chi^2 = \sum_{i=1}^{l} \frac{(m_i - np_i)^2}{np_i}, \tag{7.1}$$

де l – кількість розрядів даного інтервального ряду; n – обсяг вибірки; m_i – інтервальні частоти (емпіричні дані); p_i – імовірності потрапляння випадкової величини у i-й проміжок, що обчислюються відповідно гіпотетичною функцією $f(x)$:

$$p(x_{i-1} \le x \le x_i) = \int_{x_{i-1}}^{x_i} f(x)dx. \tag{7.2}$$

Величина (7.1) є випадковою, тому що у різних випробуваннях вона приймає різні, заздалегідь невідомі значення. Зрозуміло, що чим менше різняться емпіричні та теоретичні частоти, тим менше величина (7.1).

Якщо закон розподілу критерію згоди відомий, можна вказати таке критичне значення критерію $\chi^2 = \chi^2_{кр.пр.}$ або $\chi^2 = \chi^2_{кр.лв.}$ (рис. 7.1) при якому відповідна ймовірність події $\chi^2 > \chi^2_{кр.пр.}$ або $\chi^2 < \chi^2_{кр.лв.}$ буде близькою до нуля.

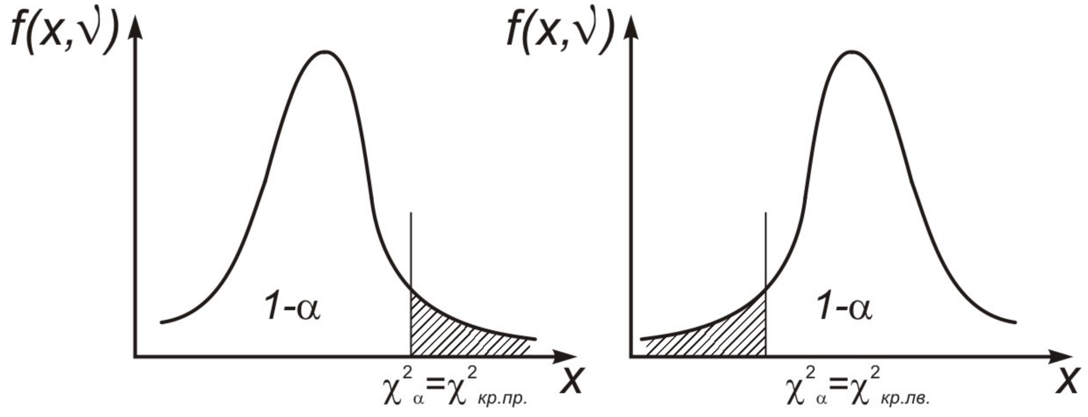


Рис. 7.1. Критичні значення критерію χ^2

Геометрично ці ймовірності являють собою площі під кривою розподілу, так звані ***правосторонню*** та ***лівосторонню*** критичні області. Ймовірність потрапляння в ці області в одиничному випробуванні дуже мала, близька до нуля. Цю ймовірність відповідно змісту задачі приймають такою, що дорівнює 0,01; 0,025; 0,05 та називають ***рівнем значущості*** і позначають α.

Зрозуміло, що оцінювання гіпотези за критерієм полягає в порівнянні величини χ^2, що визначена відповідно до формули (7.1) з критичним значенням $\chi^2_{кр.пр.}$ або $\chi^2_{кр.лв.}$, що знайдені залежно від обраного рівня значущості α. Якщо $\chi^2 > \chi^2_{кр.пр.}$ або $\chi^2 < \chi^2_{кр.лв.}$, то значення досліджуваної величини X потрапляє в критичну область (відповідно праву або ліву) і обрану гіпотезу про вид функції $f(x)$ слід відкинути. Якщо, навпаки, $\chi^2 < \chi^2_{кр.пр.}$ або $\chi^2 > \chi^2_{кр.лв.}$, то величина χ^2 знаходиться поза критичною областю і гіпотеза може бути прийнята.

Таким чином, для перевірки гіпотези необхідно визначити закон розподілу випадкової величини χ^2. Доведено, що при $n \to \infty$ закон розподілу випадкової величини (7.1) незалежно від закону розподілу генеральної сукупності X, прямує до розподілу χ^2 з ν ступенями свободи (див. розділ 5.6) Тому критерій (7.1) і називають χ^2 («хі – квадрат»).

Кількість ступенів свободи знаходять за формулою $\nu = l - r$, де r – кількість зв'язків, що накладені на частоти m_i. Наприклад, якщо $\sum_{i=1}^{l} m_i = n$ та в ролі теоретичних параметрів $\bar{x}$ і σ, які входять до теоретичної функції $f(x)$, приймаються їх точкові оцінки $\bar{x}^*$ і S, то приймають $r = 3$. Критичні значення величини χ^2 наведені в таблиці додатку Г залежно від значень α та ν.

Розрахунки виконують в такій послідовності: .

1) висувають гіпотезу про закон розподілу $f(x)$, за вибіркою визначаються параметри розподілу (наприклад, $a = \overline{x}^*$ та $\sigma = S$, якщо висунута гіпотеза про нормальний закон розподілу), а також кількість ступенів свободи $\nu = l - r$ ($r = 3$ для нормального закону розподілу).
2) вибирають рівень значущості α. За таблицею для даних α та ν знаходять критичне значення критерію $\chi^2_{кр.пр.}$
3) за формулою (7.2) визначають ймовірності p_i потрапляння випадкової величини X до інтервалів (x_{i-1}, x_i).
4) за формулою (7.1) обчислюють емпіричне значення χ^2.
5) порівнюємо значення χ^2 з критичним:

   а) у випадку правосторонньої критичної області гіпотеза, що висувалася, відкидається, якщо $\chi^2 > \chi^2_{кр.пр.}$;

   б) у випадку лівосторонньої критичної області гіпотеза, що висувалася, відкидається, якщо $\chi^2 < \chi^2_{кр.лв.}$;

   в) у випадку двосторонньої критичної області гіпотеза, що висувалася, буде **правильною,** якщо виконується умова:

$$p\left(\chi^2 < \chi^2_{кр.лв.}\right) + p\left(\chi^2 > \chi^2_{кр.пр.}\right) = \alpha. \qquad (7.3)$$

Очевидно, що критичні точки можуть бути вибрані нескінченою множиною способів. Якщо розподіл критерію симетричний відносно нуля і є підстави обирати симетричні відносно нуля точки $\chi^2_{кр.лв.} < 0$ і $\chi^2_{кр.пр.} > 0$, то приймаємо

$$p\left(\chi^2 < \chi^2_{кр.лв.}\right) = p\left(\chi^2 > \chi^2_{кр.пр.}\right).$$

Враховуючи (7.3), одержимо:

$$p\left(\chi^2 < \chi^2_{кр.лв.}\right) = p\left(\chi^2 > \chi^2_{кр.пр.}\right) = \frac{\alpha}{2}.$$

У додатку Г вказані лише «праві» критичні точки. Знайти «ліву» точку можна враховуючи, що події $\chi^2 > \chi^2_{кр.пр.}$ і $\chi^2 < \chi^2_{кр.лв.}$ протилежні і сума їх ймовірностей дорівнює 1:

$$p\left(\chi^2 < \chi^2_{кр.лв.}\right) = 1 - p\left(\chi^2 > \chi^2_{кр.пр.}\right) = 1 - \frac{\alpha}{2} \qquad (7.4)$$

Приклад 7.1

Перевіримо гіпотезу про нормальний закон розподілу для даних, що представлені вибіркою обсягом $n = 40$ та згруповані в інтервальний ряд (табл. 3.4). В прикладі 4.2 розраховані параметри статистичного розподілу цієї кількісної ознаки: середня вибіркова $\overline{x}^* = 52{,}85$ та середнє квадратичне відхилення (стандарт) $s = 2{,}58$. В прикладі 5.2 висунута гіпотеза про підпорядкування цих даних нормальному закону розподілу і визначені параметри цього розподілу $a = \overline{x}^* = 52{,}8$; $\sigma = s = 2{,}58$. Знаходимо число ступенів свободи: $\nu = l - r = 5 - 3 = 2$. Розглянемо двосторонню критичну область за таблицею (додаток Г) для $\nu = 2$ і $\alpha/2 = 0{,}025$, визначимо $\chi^2_{кр.пр.} = 7{,}4$, а для $\left(1 - \dfrac{\alpha}{2}\right) = 0{,}975$ і того самого ν знаходимо $\chi^2_{кр.лев.} = 0{,}051$.

Знайдемо тепер ймовірності p_i потрапляння значення випадкової величини в інтервали (x_{i+1}, x_i). Скористуємось функцією Лапласа відповідно до формули 5.10:

$$P_i = \Phi\left(\frac{x_i - a}{\sigma}\right) - \Phi\left(\frac{x_{i-1} - a}{\sigma}\right).$$

Отримаємо:

$$P_1 = \Phi\left(\frac{50 - 52{,}8}{2{,}58}\right) - \Phi\left(\frac{48 - 52{,}8}{2{,}58}\right) = \Phi(-1{,}08) - \Phi(-1{,}86) = -0{,}3599 + 0{,}4686 = 0{,}11;$$

$$P_2 = \Phi\left(\frac{52 - 52{,}8}{2{,}58}\right) - \Phi\left(\frac{50 - 52{,}8}{2{,}58}\right) = \Phi(-0{,}31) - \Phi(-1{,}08) = -0{,}1217 + 0{,}3599 = 0{,}24;$$

$$P_3 = \Phi\left(\frac{54 - 52{,}8}{2{,}58}\right) - \Phi\left(\frac{52 - 52{,}8}{2{,}58}\right) = \Phi(0{,}46) - \Phi(-0{,}27) = 0{,}1772 + 0{,}1217 = 0{,}3;$$

$$P_4 = \Phi\left(\frac{56 - 52{,}8}{2{,}58}\right) - \Phi\left(\frac{54 - 52{,}8}{2{,}58}\right) = \Phi(1{,}24) - \Phi(0{,}5) = 0{,}3925 - 0{,}1772 = 0{,}22;$$

$$P_5 = \Phi\left(\frac{58 - 52{,}8}{2{,}58}\right) - \Phi\left(\frac{56 - 52{,}8}{2{,}58}\right) = \Phi(2{,}01) - \Phi(1{,}28) = 0{,}4772 - 0{,}3925 = 0{,}09;$$

$$P_6 = \Phi\left(\frac{60 - 52{,}8}{2{,}58}\right) - \Phi\left(\frac{58 - 52{,}8}{2{,}58}\right) = \Phi(2{,}79) - \Phi(2{,}05) = 0{,}4974 - 0{,}4772 = 0{,}02.$$

Таблица 7.1

Інтервали міцності зразка на стиск, u_{i-1} - u_i, мПа	Значення середини інтервалу, u_i	Частота, m_i	Відносна частота ,w_i	Теоретичні ймовірності
48 – 50	49	6	0,150	0,11
50 – 52	51	9	0,225	0,24
52 – 54	53	13	0,325	0,30
54 – 56	55	7	0,175	0,22
56 – 58	57	4	0,100	0,09
58 – 60	59	1	0,025	0,02

За формулою (1) визначимо емпіричне значення критерію:

$$(\chi)^2 = \sum_{i=1}^{l} \frac{(m_i - np_i)^2}{np_i} = \sum_{i=1}^{6} \frac{(m_i - 40p_i)^2}{40p_i} =$$

$$= \frac{(6-4{,}4)^2}{4{,}4} + \frac{(9-9{,}6)^2}{9{,}6} + \frac{(13-12)^2}{12} + \frac{(7-8{,}8)^2}{8{,}8} + \frac{(4-3{,}6)^2}{3{,}6} + \frac{(1-0{,}8)^2}{0{,}8} =$$

$$= 0{,}58 + 0{,}037 + 0{,}08 + 0{,}37 + 0{,}04 + 0{,}05 = 1{,}157.$$

В даному випадку $\chi^2_{кр.лев.} < \chi^2 < \chi^2_{кр.пр.} (0{,}051 < 1{,}157 < 7.4)$.

Тому можна стверджувати, що гіпотеза про розподіл випадкової величини, що спостерігається X за нормальним законом із параметрами $a=52{,}8$ і $\sigma = 2{,}58$ не протирічить статичним даним.

8. ВИКОРИСТАННЯ МЕТОДІВ ТЕОРІЇ ЙМОВІРНОСТЕЙ ТА МАТЕМАТИЧНОЇ СТАТИСТИКИ ДЛЯ ВИРІШЕННЯ ДЕЯКИХ ЗАДАЧ ГЕОМЕХАНІКИ

8.1. Визначення коефіцієнта структурного послаблення

Однією з найважливіших геомеханічних характеристик є міцність порід на одноосьове стиснення, що визначається в лабораторних умовах за відомими методиками. Відмінність міцності на стиск породного масиву $R_м$ від середньої міцності лабораторних зразків порід $\overline{R}$ оцінюється коефіцієнтом структурного послаблення:

$$Кс = \frac{R_М}{\overline{R}} \ . \tag{8.1}$$

Відповідно до статистичної теорії міцності масив гірських порід можна уявити як деяку систему (агрегат), що складається з необмеженої кількості структурних елементів (рис.8.1). Зразки, які досліджуються в лабораторних умовах, теж можна вважати такими елементами.

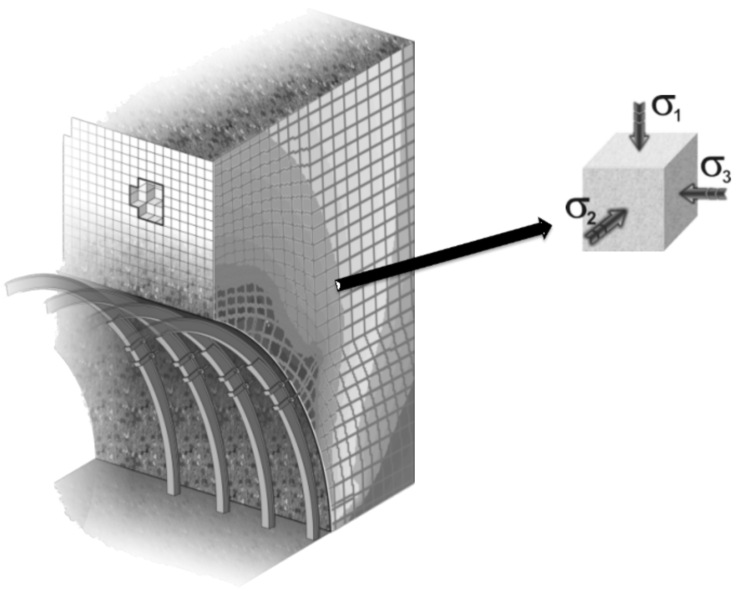


Рис. 8.1. Уявлення про породний масив як сукупність структурних елементів

Міцність R кожного з структурних елементів цієї системи є випадкова величина, що підпорядкована тому чи іншому закону розподілу. Нехай $F(R)$ – інтегральна функція цього розподілу.

Міцність усієї системи (тобто міцність породного масиву) повинна оцінюватися величиною $R_м$ (рис. 8.2) такою, що міцність його структурних елементів була б *не нижче* значення $R_м$. Імовірність такої події визначається формулою:

$$P(R \geq Rм) = 1 - F(Rм)\,. \tag{8.2}$$

Ця ймовірність залежить від виду розподілу, тобто від виду функції *F(R)*. Відносно цієї функції можна висунути різні гіпотези.

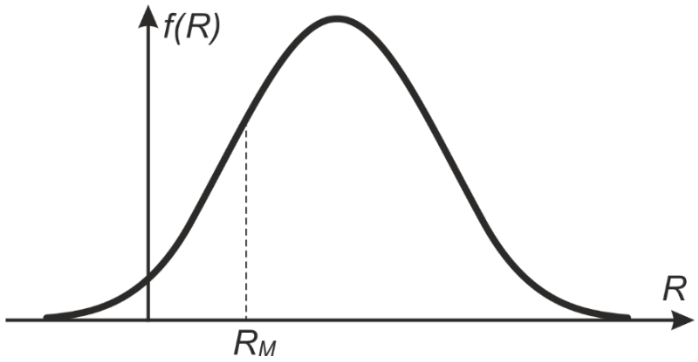


Рис. 8.2. Гіпотетичний розподіл міцності структурних елементів масиву

8.1.1. Гіпотеза про нормальний розподіл величини R – міцності структурних елементів масиву

Нехай міцність структурних елементів підпорядковується нормальному закону розподілу, відповідно якому диференціальна функція ймовірностей має вигляд:

$$f(R) = \frac{1}{\sigma\sqrt{2\pi}} \cdot e^{-\frac{(R-a)^2}{2\sigma^2}} ,$$

де a – математичне сподівання величини R; σ – її середнє квадратичне відхилення. Тоді формула (8.2) приймає такий вигляд:

$$p(R \geq R_M) = 1 - F_0(\frac{R_M - a}{\sigma}) .$$

Розв'яжемо це рівняння щодо величини $R_м$:

$$F_0(\frac{R_M - a}{\sigma}) = 1 - p \Rightarrow \frac{R_M - a}{\sigma} = t ,$$

де $t = \arg F_0(1-p)$ – аргумент функції $F_0(t) = \frac{1}{\sqrt{2\pi}} \int\limits_{-\infty}^{t} e^{-\frac{t^2}{2}} dt$ при її значенні , яке дорівнює *1 – p*. Далі отримаємо:

$$R_M = \sigma \cdot \arg F_0(1-p) + a . \quad (8.3)$$

За формулою (8.1), використовуючи метод моментів ($M(R)=\overline{R}$), знаходимо коефіцієнт структурного послаблення:

$$Kc=\frac{R_M}{M(R)}.$$

Для нормального розподілу $M(R)=a$, тоді, розділивши обидві частини виразу (8.3) на величину *a*, дістанемо:

$$Kc=\frac{\sigma}{a}\cdot\arg F_0(1-p)+1\ .$$

Тут $\frac{\sigma}{a}=\eta$ – відносна варіація міцності структурних елементів. Остаточно дістанемо:

$$Kc=\eta\arg F_0(1-p)+1 \qquad (8.4)$$

Отже, ми здобули коефіцієнт структурного послаблення як величину, що залежить, по-перше, від відносної варіації η, яка по суті характеризує ступінь неоднорідності середовища; по-друге – від ймовірності *P*, що характеризує собою рівень значимості об'єкта.

Аналізуючи графік залежності (8.4) (рис. 8.3) помітимо, що при $\eta>0,4$, коефіцієнт структурного послаблення може приймати від’ємні значення, що природно суперечить фізичній суті даної величини. Очевидно, що це недолік імовірнісної моделі. Нормальний закон задовільно описує тільки такі величини, для яких варіація значень відносно середнього не перевищує 33 % (це випливає з правила «3-х сигм»).

Приклад 8.1

Визначити розрахункове значення міцності на одноосьовий стиск алевроліту, якщо за даними іспитів середнє значення міцності лабораторних зразків $\overline{R}c$ складає 40 МПа, варіація значень складає 30 %.

З рівняння (8.1) знаходимо розрахункову міцність породного масиву:

$$R_{расч}=R_M=\overline{R}c\cdot Kc\ .$$

Будемо визначати коефіцієнт послаблення з імовірністю *p=0,95*. Визначимо аргументу t нормованої нормальної функції $F_0(t)$ при її значенні, що дорівнює *1–0,95=0,05*. Із таблиці додатка А визначаємо, що значенню інтегральної функції $F_0(t)=\frac{1}{2\sqrt{\pi}}\int_{-\infty}^{t}e^{-\frac{t^2}{2}}dt=0{,}05$ відповідає значення аргументу

$t = -1{,}64$. Тобто $\arg F_0(0{,}05) = -1{,}64$. Тоді коефіцієнт структурного послаблення

$$K_c = 0{,}3 \cdot (-1{,}64) + 1 = 0{,}508 \ .$$

Таким чином, розрахункове значення міцності

$$R_{расч} = 0{,}508 \cdot 40 = 21 \text{ МПа.}$$

Зауваження

Якщо користатися таблицею, де наведені значення функції Лапласа (5.8), то для заданої ймовірності *p=0,95* потрібно знайти аргумент функції $\Phi(t) = \frac{1}{\sqrt{2\pi}} \int\limits_0^t e^{-\frac{t^2}{2}} dt$ при її значенні, що дорівнює $1-(0{,}95+0{,}5)$. При цьому варто врахувати, що функція *Ф(t)* непарна, і від'ємному значенню функції відповідає від'ємне значення аргументу.

Вище відзначалося, що для величин, варіація яких перевершує 33%, гіпотезу про нормальний розподіл слід відкинути. Однією з можливих статистичних моделей для величин із варіацією $\eta > 0{,}33$ є логарифмічно нормальний закон розподілу. Здобудемо вираз для коефіцієнта структурного послаблення виходячи з цієї гіпотези.

8.1.2. Гіпотеза про логарифмічно нормальний розподіл величини *R* – міцності структурних елементів масиву

Нехай *R* розподілена *логнормально*, тоді величина $Z = \ln R$ має нормальний розподіл з параметрами a і σ. Нехай $Z_м$ – оцінює логарифм граничного значення міцності породного масиву. Знайдемо ймовірність того, що всі значення величини *Z* є не меншими за величину $Z_м$:

$$P(Z > Zм) = 1 - F_0\left(\frac{Zм - a}{\sigma}\right) \ .$$

Розв'яжемо це рівняння відносно $Z_м$:

$$Z_м = a + \sigma \arg F(1-p) \ .$$

Повертаючись до випадкової величини *R*, із огляду на те, що $Z = \ln R \Rightarrow R = e^Z$, дістанемо:

$$R_м = e^{a + \sigma \cdot \arg(1-p)} \ . \tag{8.5}$$

Щоб перейти до коефіцієнта структурного послаблення, розділимо обидві частини на величину математичного сподівання *M(R)*. Для логнормального закону математичне сподівання і дисперсія виражаються через параметри розподілу за формулами (5.13) і (5.14).

Із (8.5) маємо:

$$Kc = \frac{e^{a+\sigma \cdot \arg F(1-P)}}{e^{\frac{\sigma^2}{2}+a}} \quad . \tag{8.6}$$

Врахуємо, що для логнормального закону: $\eta^2 = \frac{D(R)}{M^2(R)} = e^{\sigma^2} + 1$. Тоді остаточно величина коефіцієнта структурного послаблення

$$Kc = \frac{e^{\arg F(1-P)\sqrt{\ln(\eta^2+1)}\cdot}}{\sqrt{\eta^2+1}} . \tag{8.7}$$

Графіки залежності (8.7) наведено на рис. 8.3. Як бачимо, у даному випадку при необмеженому зростанні варіації значень міцності структурних елементів коефіцієнт структурного послаблення асимптотично прямує до нуля. Появу від'ємних значень виключено завдяки властивостям логнормальної функції розподілу.

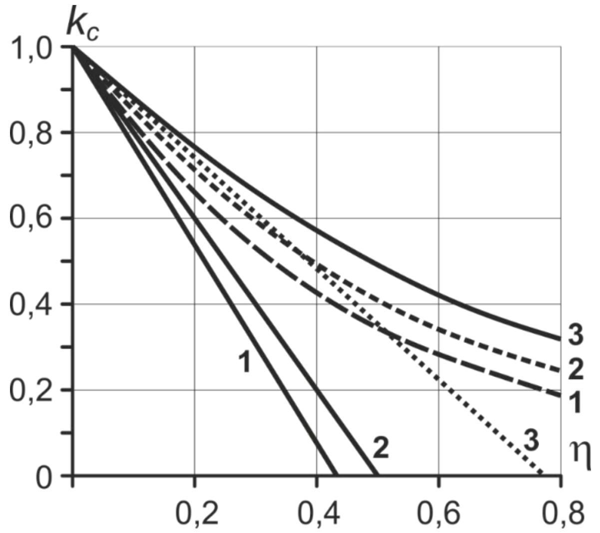


Рис. 8.3. Графіки залежності коефіцієнта структурного послаблення від величини відносної варіації значень міцності структурних елементів масиву:
1 – p=0,99; 2 – p=0,95; 3 – p=0,9;
пунктир – нормальний закон розподілу;
неперервна – логнормальний закон розподілу

Приклад 8.2

Визначимо розрахункове значення міцності на одноосьовий стиск алевроліту, якщо за даними лабораторного випробування Rc=40 МПа, а розкид даних складає 50 %, тобто *η=0,5*. Такий великий розкид значень міцності свідчить про істотну неоднорідність середовища. Застосовувати гіпотезу про нормальний розподіл міцності тут не можна, тому будемо виходити з гіпотези про логнормальний розподіл міцності. Коефіцієнт структурного послаблення визначимо за формулою (8.7) з ймовірністю р=0,95. Отримаємо:

$$\eta^2+1=1{,}25\ ;\ \sqrt{\eta^2+1}=1{,}12\ ;\ \ln(\eta^2+1)=0{,}22\ ;$$

$$\sqrt{\ln(\eta^2+1)}=0{,}47\ ;\qquad \arg F(1-0{,}95)=-1{,}64\ ;$$

$$Kc=\frac{e^{-1{,}64\cdot 0{,}472}}{1{,}118}=0{,}412\ .$$

Розрахункове значення міцності

$$R_M=R\bar{c}\cdot Kc=40\cdot 0{,}412=16{,}1\ \text{МПа}\ .$$

Як бачимо, значна варіація значень міцності структурних елементів масиву, яка врахована за допомогою гіпотези про логнормальний розподіл, уплинула в остаточному підсумку на значення міцності породного масиву. Воно вийшло на 23 % менше ніж у попередньому прикладі.

Треба, однак, відзначити, що, як правило, для результатів лабораторного випробування зразків такий великий розкид даних не характерний. Звичайно варіація міцності зразків однієї літологічної групи не перевершує 30-35 %, а статистичний розподіл, що побудований за даними вибірки, зазвичай близький до нормального. На цій підставі дослідники і приймають гіпотезу про нормальний розподіл. На рис. 8.4 наведені статистичні розподіли межі міцності на стиск пісковиків, що побудовані за результатами лабораторних випробувань в лабораторії геомеханіки НГУ.

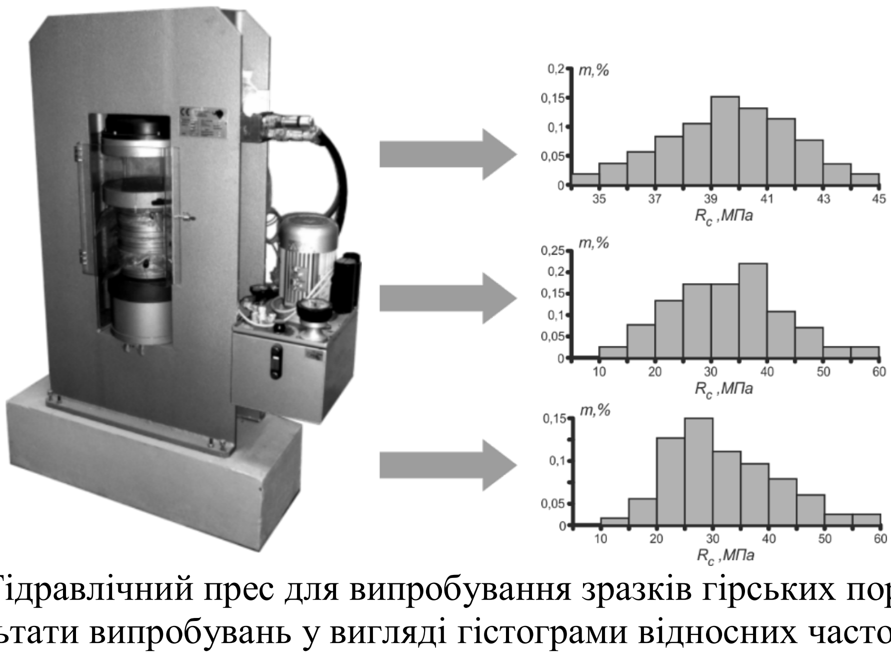


Рис. 8.4. Гідравлічний прес для випробування зразків гірських порід та результати випробувань у вигляді гістограми відносних частот

Отримані дані не суперечать нормальному закону розподілу. Але виникає питання, чи дійсно вибірка, що отримана як результат лабораторного випробування, відображує властивості генеральної сукупності. Чи виконується основна вимога до вибірки – її репрезентативність, тобто рівність шансів для всіх елементів генеральної сукупності потрапити у вибірку? Розглянемо це питання докладніше.

8.1.3. Врахування макродефектів при визначенні коефіцієнта структурного послаблення

Відомо, що в гірській породі присутня природна тріщинуватість. Відомо також, що наявність тріщин значно зменшує міцність гірської породи. Дійсно, структурний елемент, що пересічений тріщиною, має значно меншу несучу здатність ніж інші елементи. Але такий елемент не має шансів потрапити у вибірку, бо він зруйнується при виготовленні зразків із проби ще до початку дослідів. Таким чином, структурні елементи, які містять макродефекти не беруть участь у формуванні статистичного розподілу. Однак як реально існуючі вони повинні бути включені в статистику випробування (рис. 8.5). У монографії [6] запропонований спосіб урахування таких елементів. В основу способу покладено, що у вибірці обсягом n міститься n_T порушених елементів, міцність яких близька до нуля. Величина n_T залежить від відстані між тріщинами. Наявність нульових елементів змінює статистичний розподіл і всі моменти, що характеризують цей розподіл: середню вибіркову, дисперсію, відносну варіацію, показники асиметрії й ексцесу. У зв'язку з цим розподіл відхиляється від симетричного, гіпотеза про нормальний розподіл стає неможливою. У [6] встановлено залежність між статистичними характеристиками звичайного варіаційного ряду (ν_k, μ_k) і «виправленого», тобто отриманого шляхом умовного додавання у вибірку елементів з нульовою міцністю (ν'_k, μ'_k). Для початкових моментів ця залежність має вигляд:

$$m'_k = K_{TP} m_k \ , \tag{8.8}$$

де $K_{TP} = \dfrac{l_T}{l_T + l_0}$, l_T – відстань між тріщинами в масиві, l_0 – розмір стандартного випробовуваного зразка. «Виправлені» центральні моменти з урахуванням їх вираження через початкові моменти за формулами (4.14) і (4.15) мають вигляд:

$$\begin{aligned}
\mu'_2 &= K_{TP} m_2 - K_{TP}^2 m_1^2; \\
\mu'_3 &= K_{TP} m_3 - 3K_{TP}^2 m_2 m_1 + 2K_{TP}^3 m_1^3; \\
\mu'_4 &= K_{TP} m_4 - 4K_{TP}^2 m_3 m_1 + 6K_{TP}^3 m_2 m_1^2 - 3K_{TP}^4 m_1^4.
\end{aligned} \tag{8.9}$$

Вираз для відносної варіації «виправленого» розподілу має вигляд:

$$\eta' = \sqrt{(\eta^2+1)\frac{l_T+l_0}{l_T}-1} \tag{8.10}$$

Таким чином, при обробці статистичних даних при відомій відстані між тріщинами моменти розподілу варто перерахувати за формулами (8.8) і (8.9); визначити відповідно “виправлені “ значення показників асиметрії й ексцесу за формулами (5.16) і (5.17); за графіком Пірсона (рис. 5.7) вибрати найбільш близький до точки (β_1,β_2) розподіл; визначити коефіцієнт структурного послаблення породи виходячи з цієї гіпотези (наприклад за гіпотезою про логнормальний розподіл).

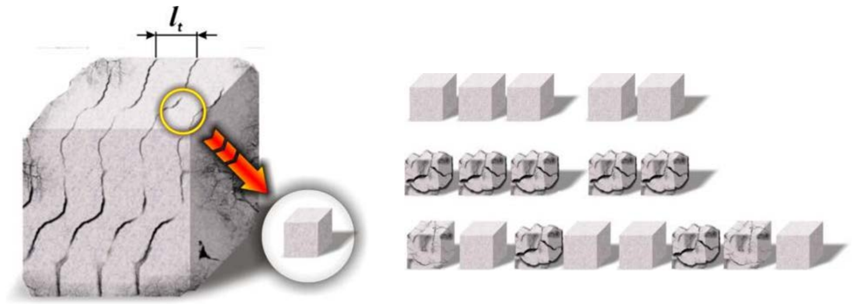


Рис. 8.5. Уявне включення в вибірку дефектних елементів

8.2. Приклад виконання розрахункового завдання

Надано вибірку значень кількісної ознаки X – міцності гірської породи на стиск за результатами лабораторних випробувань:

20,15	26,2	20,76	22,57	29,23	27,41
29,83	34,67	37,09	43,14	20,15	29,83
34,67	24,99	28,62	25,59	37,69	26,81
28,62	14,7	21,97	35,23	29,23	29,83
27,41	35,28	28,62	23,78	28,62	29,23

<u>Завдання</u>

1. Побудувати статистичний розподіл кількісної ознаки з використанням вибіркових даних. Для цього:

а) згрупувати дані у вигляді інтервального ряду;

б) побудувати гістограму частот або відносних частот.

2. Визначити моменти статистичного розподілу: середню вибіркову, виправлену дисперсію, середньоквадратичне відхилення, коефіцієнт варіації, асиметрію та ексцес.

3. Висунути гіпотезу про теоретичний закон розподілу випадкової величини – границі міцності на одноосьовий стиск.

4. З використанням методу моментів підібрати параметри теоретичного розподілу.

5. Перевірити гіпотезу про теоретичний закон розподілу за критерієм Пірсона.

6. Визначити коефіцієнт структурного послаблення гірської породи та розрахункове значення міцності породного середовища.

<u>Розв'язання</u>

1. Знайдемо максимальне та мінімальне значення елементів вибірки:

$x_{\min} = 14{,}7$; $x_{\max} = 43{,}14$.

2. Довільно виберемо кількість інтервалів:

$k = 5$.

3. Довжина кожного інтервалу:

$$l = \frac{x_{\max} - x_{\min}}{k} = \frac{43{,}14 - 14{,}7}{5} = 5{,}69.$$

4. Формуємо інтервали та визначаємо частоту потрапляння значень в кожний з інтервалів:

1. -$14{,}7 - 20{,}39$: $m_1 = 3$;
2. - $20{,}39 - 26{,}08$: $m_2 = 6$;
3. -$26{,}08 - 31{,}77$: $m_3 = 14$;
4. -$31{,}77 - 37{,}46$: $m_4 = 5$;
5. -$37{,}46 - 43{,}14$: $m_5 = 2$.

5. Визначаємо відносні частоти: $w_i = \dfrac{m_i}{n}$;

$$w_1 = \frac{3}{30} = 0{,}1;\ w_2 = \frac{6}{30} = 0{,}2;\ w_3 = \frac{14}{30} = 0{,}47;\ w_4 = \frac{5}{30} = 0{,}17;\ w_5 = \frac{2}{30} = 0{,}06;$$

6. Визначаємо середини інтервалів:

$$u_1 = \frac{x_{\min} + x_1}{2} = \frac{14{,}7 + 20{,}39}{2} = 17{,}55;$$

$u_2 = u_1 + l = 23{,}24$;

$u_3 = u_2 + l = 28{,}93$;

$u_4 = u_3 + l = 34{,}62$;

$u_5 = u_4 + l = 40{,}31$.

7. Визначимо для кожного інтервалу висоту прямокутника h_i: $h_i = \dfrac{w_i}{l}$;

$$h_1 = \frac{0.1}{5{,}69} = 0{,}02;\ h_2 = \frac{0{,}2}{5{,}69} = 0.04;\ h_3 = \frac{0.47}{5{,}69} = 0.094;\ h_4 = \frac{0.17}{5{,}69} = 0.03;\ h_5 = \frac{0.06}{5{,}69} = 0{,}012.$$

9. Формуємо інтервальний ряд:

Інтервал	14,7–20,39	20,40–26,09	26,09–31,77	31,78–37,46	37,47–43,14
Частоти	3	6	14	5	2
Відносна частота	0,1	0,2	0,47	0,17	0,06
Середини інтервалів	17,55	23,24	28,93	34,62	40,31
Висоти	0,02	0,04	0,094	0,03	0,012

9. Графічним зображенням інтервального ряду є гістограма частот (рис. 8.6):

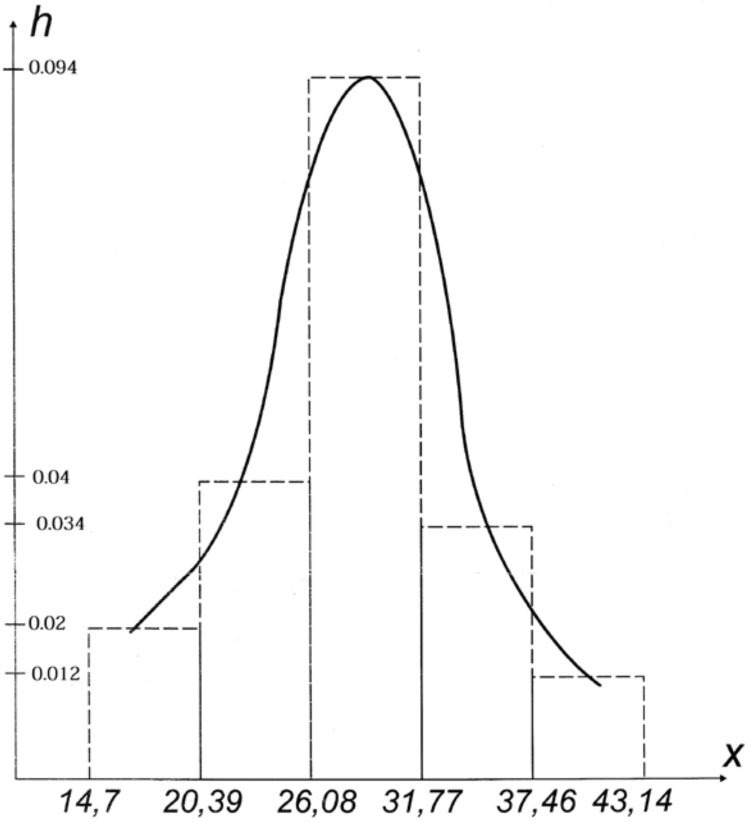


Рис. 8.6. Гістограма відносних частот.

10. Визначимо початковий момент першого порядку, тобто вибіркову середню:

$$\overline{x}^* = \nu_1^* = 17{,}55*0{,}1+23{,}24*0{,}2+28{,}93*0{,}47+34{,}62*0{,}17+40{,}31*0{,}06=28{,}31;$$

11. Визначимо характеристики варіації значень кількісної ознаки:

- статистичний початковий момент другого порядку:

$$\nu_2^* = 17{,}55^2*0{,}1+23{,}24^2*0{,}2+28{,}93^2*0{,}47+34{,}62^2*0{,}17+40{,}31^2*0{,}06=833{,}53;$$

- статистичний центральний момент другого порядку, тобто вибіркову дисперсію

$$D^* = \mu_2^* = \nu_2^* - (\nu_1^*)^2 = 833{,}53 - 28{,}31^2 = 32{,}53 \;;$$

- виправлену дисперсію:

$$S^2 = \frac{30}{30-1} * 32{,}53 = 33{,}65 \;;$$

- середнє квадратичне відхилення (стандарт):

$$S = \sqrt{S^2} = 5{,}8;$$

- відносну варіацію, коефіцієнт варіації:

$$\eta^* = \frac{5{,}8}{28{,}31} = 0{,}2; \quad V^* = 0{,}2*100 = 20\%.$$

12. Визначимо величини, що характеризують асиметрію та ексцес:

- статистичні початкові моменти третього і четвертого порядків:

$$\nu_3^* = 17{,}55^3*0{,}1+23{,}24^3*0{,}2+28{,}93^3*0{,}47+34{,}62^3*0{,}17+40{,}31^3*0{,}06=25411{,}6;$$

$$\nu_4^* = 17{,}55^4*0{,}1+23{,}24^4*0{,}2+28{,}93^4*0{,}47+34{,}62^4*0{,}17+40{,}31^4*0{,}06=799675{,}43;$$

- статистичні центральні моменти третього і четвертого порядків:

$$\mu_3^* = \nu_3^* - 3\nu_2^*\nu_1^* + 2\nu_1^{*3} = 25411{,}6 - 3*833{,}42*28{,}31 + 2*28{,}31^3 = 35{,}04;$$

$$\begin{aligned}\mu_4^* &= \nu_4^* - 4\nu_3^*\nu_1^* + 6\nu_2^*\nu_1^{*2} - 3\nu_1^{*4} = \\ &= 799675{,}43 - 4*25411{,}6*28{,}31 + 6*833{,}42*28{,}31^2 - 3*28{,}31^4 = 2767{,}47\end{aligned};$$

- статистичну оцінку асиметрії

$$A^* = \frac{35{,}04}{5{,}8^3} = 0{,}18;$$

- статистичну оцінку ексцесу

$$E^* = \frac{2767{,}47}{5{,}8^4} - 3 = -0{,}55.$$

13. Висуваємо гіпотезу про теоретичний розподіл випадкової величини. Базуючись на тому, що коефіцієнти асиметрії та ексцесу близькі до нуля, а гістограма частот має симетричний вигляд, а також виходячи з фізичного змісту випадкової величини вважатимемо, що випадкова величина розподілена за нормальним законом.

14. Перевіримо цю гіпотезу за **правилом «3-х сигм»**. Визначимо інтервал:

$$\left[\overline{x^*} - 3S; \quad \overline{x^*} + 3S\right] \Rightarrow \left[28{,}31 - 3\cdot 5{,}8; \quad 28{,}31 + 3\cdot 5{,}8\right] \Rightarrow \left[10{,}91; \quad 45{,}71\right].$$

Значення $x_{\min} = 14{,}7$ та $x_{\max} = 43{,}14$ потрапляють до цього інтервалу.

15. Використовуючи **метод моментів** знаходимо параметри теоретичного розподілу:

$$M(x) = \overline{x^*} \Rightarrow \ a = \overline{x^*} = 28{,}31;$$

$$D(x) = S^2 \Rightarrow \ \sigma(x) = S = 5{,}8 \Rightarrow \ \sigma = S = 5{,}8 \ .$$

Диференціальна функція нормального закону розподілу (закон Гауса) має вигляд:

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} \cdot e^{-\frac{(x-a)^2}{2\sigma^2}} \ .$$

Для досліджуваної кількісної ознаки

a = 28,31 – середня вибіркова;

σ = 5,8 – стандарт.

Таким чином, одержимо:

$$f(x) = \frac{1}{5{,}8\sqrt{2\pi}} \cdot e^{-\frac{(x-28{,}31)^2}{2\cdot 33{,}65}}.$$

16. Перевіримо гіпотезу про нормальний закон розподілу за **критерієм Пірсона**:

- Знайдемо число ступенів свободи: $\nu = l - r = 5 - 3 = 2$.
- Розглянемо двобічну критичну область та за таблицею у додатку А для $\nu = 2$, $\alpha/2 = 0{,}025$ визначимо $\chi^2_{кр.пр} = 7{,}4$, а для $(1-\alpha/2) = 0{,}975$ та того самого ν знаходимо $\chi^2_{кр.лв} = 0{,}051$.
- Імовірність P_i потрапляння в розряди (x_{i-1}, x_i) знайдемо як для нормального розподілу (з використанням функції Лапласа) за формулою:

$$P_i = \Phi\left(\frac{x_i - a}{\sigma}\right) - \Phi\left(\frac{x_{i-1} - a}{\sigma}\right).$$

Для кожного інтервалу маємо:

$$P_1=\Phi\left(\frac{20,39-28,31}{5,8}\right)-\Phi\left(\frac{14,7-28,31}{5,8}\right)=\Phi(-1,36)-\Phi(-2,35)=0,08;$$

$$P_2=\Phi\left(\frac{26,09-28,31}{5,8}\right)-\Phi\left(\frac{20,39-28,31}{5,8}\right)=\Phi(-0,4)-\Phi(-1,36)=0,26;$$

$$P_3=\Phi\left(\frac{31,77-28,31}{5,8}\right)-\Phi\left(\frac{26,09-28,31}{5,8}\right)=\Phi(0,6)-\Phi(-0,4)=0,37;$$

$$P_4=\Phi\left(\frac{37,46-28,31}{5,8}\right)-\Phi\left(\frac{31,77-28,31}{5,8}\right)=\Phi(1,58)-\Phi(0,6)=0,22;$$

$$P_5=\Phi\left(\frac{43,14-28,31}{5,8}\right)-\Phi\left(\frac{37,46-28,31}{5,8}\right)=\Phi(2,56)-\Phi(1,58)=0,05;$$

За формулою (7.1) визначимо емпіричне значення критерію:

$$\chi^2=\sum_{i=1}^{l}\frac{(m_i-np_i)^2}{np_i}=\sum_{i=1}^{5}\frac{(m_i-30p_i)^2}{30p_i}=\frac{(3-2,4)^2}{2,4}+\frac{(6-7,8)^2}{7,8}+\frac{(14-11,1)^2}{11,1}+\frac{(5-6.6)^2}{6.6}+\frac{(2-1,5)^2}{1.5}=1,89$$

Отримано співвідношення $\chi^2_{кр.лв}<\chi^2<\chi^2_{кр.пр}\Rightarrow$ (0,051<1,89<7,4). Тому можна стверджувати, що гіпотеза про нормальний розподіл випадкової величини X з параметрами $a=28,31$ та $\sigma=5,8$ не суперечить статистичним даним.

Визначимо **коефіцієнт структурного послаблення**:

$Kc=\eta^{*}argF_o(1\text{-}p)+1$ $\qquad$ $argF_o(1\text{-}P) = -1.64$

$Kc=0.2\cdot(-1.64)+1=0.672$

$$R_C=\bar{x}^*\cdot K_C$$

$$R_C=28,31\cdot 0,672=19,02$$

9. КОРЕЛЯЦІЙНА ЗАЛЕЖНІСТЬ. МЕТОД НАЙМЕНШИХ КВАДРАТІВ

9.1. Стохастичний зв'язок. Регресія

Випадкові величини X і Y називають *стохастично залежними*, якщо зміна однієї з них приводить до зміни розподілу іншої. Зокрема, може змінюватись той чи інший параметр розподілу. Якщо із зміною однієї випадкової величини зміщується центр розподілу іншої, тобто її середнє значення, то стохастичний зв'язок між величинами називають *кореляційним.*

Наведемо приклад випадкової величини Y, яка не пов'язана з величиною X функціонально, але зв'язана кореляційно. Нехай Y – видобуток вугілля, X – кількість комбайнів. З однакових за площею ділянок при рівних кількостях техніки і обслуговуючого персоналу, однаковій організації праці видобувають різну кількість вугілля в місяць, тобто Y не є функцією від X. Це пояснюється впливом випадкових чинників (варіацією фізико-механічних властивостей порід, зміною потужності пласта та ін.). Разом з тим, як показує досвід, середній видобуток є функцією від кількості комбайнів, що використовуються, тобто Y пов'язаний з X кореляційною залежністю.

Як приклад розглянемо кореляційну табл. 9.1.

Таблиця 9.1

Y \ X	30	**35**	**40**	**45**	**50**	**55**	n_Y	$h_1 = 5$
18	**4**	**6**	**-**	**-**	**-**	**-**	**10**	
28	**-**	**8**	**10**	**-**	**-**	**-**	**18**	
38	**-**	**-**	**4**	**35**	**5**	**-**	**44**	
48	**-**	**-**	**4**	**12**	**6**	**-**	**22**	
58	**-**	**-**	**-**	**1**	**3**	**2**	**6**	
n_X	**4**	**14**	**18**	**48**	**14**	**2**	$n = 100$	
$h_2 = 10$								

Числа n_{XY} в клітинках таблиці визначають число появ тієї або іншої пари значень X і Y; n_X і n_Y загальне число появ того або іншого значення величин X і Y; n – обсяг вибірки.

Кожен рядок таблиці з числами n_{XY} разом з першим рядком є умовним розподілом величини X при даному значенні Y. Перший рядок разом з останнім – безумовний розподіл X. Аналогічний зміст мають стовпці.

Для умовних розподілів можна знаходити звичайні характеристики розподілу.

Умовні середні визначаються за формулами:

$$\overline{y}_X = \frac{\sum\limits_{K=1}^{m} n_{XY} y_K}{n_X}; \quad \overline{x}_y = \frac{\sum\limits_{i=1}^{l} n_{XY} y_i}{n_y}, \tag{9.1}$$

де m – число рядків; l – число стовбців (табл. 9.1).

Використовуючи дані табл. 9.1, одержимо:

$X = 30$ $\overline{y}_X = \frac{4 \cdot 18 + 0 \cdot 28 + ...}{4} = 18;$

$X = 35$ $\overline{y}_X = \frac{6 \cdot 18 + 8 \cdot 28}{14} = 237;$

$X = 40$ $\overline{y}_X = \frac{10 \cdot 28 + 4 \cdot 38 + 4 \cdot 48}{18} = 34{,}67;$

$X = 45$ $\overline{y}_X = \frac{35 \cdot 38 + 14 \cdot 48 + 1 \cdot 58}{48} = 41;$

$X = 50$ $\overline{y}_X = \frac{51 \cdot 38 + 6 \cdot 48 + 3 \cdot 58}{14} = 46{,}5;$

$X = 55$ $\overline{y}_X = \frac{2 \cdot 58}{2} = 58.$

Зведемо ці дані в табл. 9.2.

Таблиця 9.2

X	30	35	40	45	50	55
$\overline{y}_X$	18	23,7	34,67	41	46,5	58
n_X	4	14	18	48	14	2

Як бачимо, кожному значенню X відповідає єдине значення $\overline{y}_X$. Тому залежність $\overline{y}_X$ от X є функціональною:

$$\overline{y}_X = f(x) \tag{9.2}$$

Рівняння вигляду (9.2) визначає кореляційну залежність Y от X. Воно зветься рівнянням регресії Y по X. Ломана, що побудована за даними табл. 9.2, є емпіричною лінією регресії. Теоретичною лінією регресії є графік функції $f(x)$.

Рівняння регресії X по Y задається рівністю

$$\overline{x_y} = \varphi(y). \qquad (9.3)$$

Якщо обидві функції $f(x)$ и $\varphi(y)$ в (9.2) та (9.3) лінійні, то кореляцію між X та Y називають лінійною. Інакше говорять про нелінійну кореляцію.

В теорії кореляції розглядаються дві основні задачі:

- встановлення функцій регресії $f(x)$ та $\varphi(y)$;
- оцінювання тісноти кореляційного зв'язку випадкових величин.

9.2. Лінійна кореляція. Метод найменших квадратів

Якщо теоретичні передумови відсутні, то вибір функцій (9.2) і (9.3) можливо здійснити таким чином: передбачити, наприклад, що $f(x)$ – лінійна функція, що задовольняє рівнянню

$$\overline{y}_X = \rho \cdot x + b, \qquad (9.4)$$

і параметри ρ і b знайти за вибірковими даними ***методом найменших квадратів***. Пояснимо суть цього методу.

Нехай в результаті випробувань отримано n пар чисел $(x_1, y_1), (x_2, y_2), \ldots\ldots, (x_n, y_n)$. Будемо вимагати, щоб пряма (9.4) була найбільш близька до всієї сукупності даних точок. Оскільки точки (x_i, y_i), як правило, не лежать на прямій (9.4), то підстановка координат (x_i, y_i) в рівняння (9.4) приводить до нев'язкі (відхилення):

$$\varepsilon_i = y_i - \rho \cdot x_i + b \rho = tg\varphi,$$

яка у загальному випадку є відмінною від нуля (рис. 9.1).

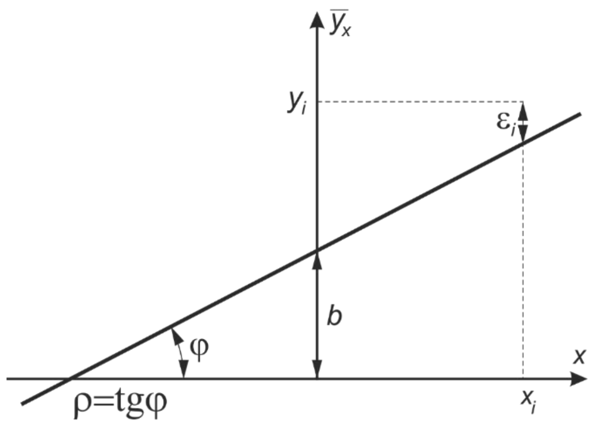


Рис. 9.1. До опису методу найменших квадратів

Величина $F=\sum_{i=1}^{n}\varepsilon_i^2$ характеризує сумарну похибку наближення даної сукупності точок $(x_i.y_i)$ до прямої (9.4). Із зміною ρ і b змінюється величина F, тобто F залежить від ρ і b:

$$F=F(\rho,b).$$

Підберемо параметри ρ і b так, щоб сума квадратів відхилень точок $(x_i.y_i)$ від прямої (9.4) була мінімальною, тобто щоб мінімальною була функція:

$$F=\sum_{i=1}^{n}\varepsilon_i^2=\sum_{i=1}^{n}(y_i-\rho x_i-b)^2. \tag{9.5}$$

У цьому і полягає суть методу найменших квадратів.

Для опису мінімуму функції (9.5) прирівняємо нулю відповідні частинні похідні:

$$\left.\begin{aligned}&\frac{dP}{d\rho}=2\sum_{i=1}^{n}x_i(y_i-\rho x_i-b)=0;\\&\frac{dP}{db}=2\sum_{i=1}^{n}(y_i-\rho x_i-b)=0;\end{aligned}\right\}$$

або

$$\left.\begin{aligned}&\rho\sum_{i=1}^{n}x_i^2+b\sum_{i=1}^{n}x_i=\sum_{i=1}^{n}x_iy_i;\\&\rho\sum_{i=1}^{n}x_i+b\cdot n=\sum_{i=1}^{n}y_i.\end{aligned}\right\}\quad . \tag{9.6}$$

Оскільки пара значень (x_i,y_i) в табл. 9.1 трапляється n_{XY} раз, то в кожну суму в рівності (9.6) необхідно внести множник n_{XY}. Крім того, на підставі відомих формул:

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

запишемо

$$\sum_{i=1}^{n} n_{xy} x_i = n\overline{x}; \quad \sum_{i=1}^{n} n_{xy} y_i = n\overline{y}; \quad \sum_{i=1}^{n} n_{xy} x_i^2 = n\overline{x^2} \quad . \tag{9.7}$$

З урахуванням (9.7) система (9.6) набуває вигляд:

$$\left.\begin{aligned} &\rho \cdot n \cdot \overline{x^2} + b \cdot n \cdot \overline{x} = \sum_{i=1}^{n} n_{xy} x_i y_i; \\ &\rho \cdot n \cdot \overline{x} + b \cdot n = n\overline{y}. \end{aligned}\right\} \tag{9.8}$$

З другого рівняння системи (9.8) отримаємо

$$b = \overline{y} - \rho \overline{x}. \tag{9.9}$$

Підставимо (9.9) в перше рівняння системи (9.8):

$$\rho \cdot n \cdot \overline{x^2} + n \cdot \overline{x}\,\overline{y} - \rho \cdot n \cdot \overline{x}^2 = \sum_{i=1}^{n} n_{xy} x_i y_i,$$

звідки отримаємо:

$$\rho \cdot n \cdot (\overline{x^2} - \overline{x}^2) = \sum_{i=1}^{n} n_{xy} x_i y_i - n\overline{x}\,\overline{y},$$

тобто

$$\rho = \frac{\sum_{i=1}^{n} n_{xy} x_i y_i - n\overline{x}\,\overline{y}}{n\sigma_x^2} \tag{9.10}$$

Щоб підкреслити, що мається на увазі рівняння регресії Y по X, ρ позначають індексом y/x і формули (9.9) і (9.10) записують у вигляді:

$$b = \overline{y} - \rho_{y/x} \overline{x}; \tag{9.11}$$

$$\rho_{y/x} = \frac{\sum_{i=1}^{n} n_{xy} x_i y_i - n\overline{x}\,\overline{y}}{n\sigma_x^2}. \qquad (9.12)$$

Підставивши (9.11) в рівняння (9.4), отримаємо рівняння регресії Y по X у вигляді:

$$\overline{y}_x - \overline{y} = \rho_{y/x}(x - \overline{x}). \qquad (9.13)$$

В отриманому рівнянні $\rho_{y/x}$ визначається за формулою (9.12). Аналогічно записується рівняння регресії X по Y:

$$\overline{x}_y - \overline{x} = \rho_{x/y}(y - \overline{y}). \qquad (9.14)$$

Обидва коефіцієнти регресії $\rho_{y/x}$ і $\rho_{x/y}$ можна виразити через симетричний по x і y ***вибірковий коефіцієнт кореляції :***

$$r = \frac{\sum_{i=1}^{n} n_{xy} x_i y_i - n\overline{x}\,\overline{y}}{n\sigma_x \sigma_y}, \qquad (9.15)$$

а саме

$$\rho_{y/x} = \frac{\sigma_y}{\sigma_x} r; \quad \rho_{x/y} = \frac{\sigma_x}{\sigma_y} r, \qquad (9.16)$$

де σ_x і σ_y – вибіркові середні квадратичні відхилення величин X і Y.

З урахуванням співвідношень (9.16) рівняння регресії (9.13) і (9.14) запишемо у вигляді:

$$\frac{\overline{y}_x - \overline{y}}{\sigma_y} = r\,\frac{x - \overline{x}}{\sigma_x}; \qquad (9.17)$$

$$\frac{\bar{x}_y - \bar{x}}{\sigma_x} = r\,\frac{y - \bar{y}}{\sigma_y}. \tag{9.18}$$

<u>Зауваження.</u> Вибірковий коефіцієнт кореляції задовольняє нерівності $|r| \le 1$. Якщо $r = 0$, то лінійна кореляція між X і Y відсутня. В цьому випадку рівняння (9.17) і (9.18) мають вигляд $\bar{y}_x = \bar{y}$ і $\bar{x}_y = \bar{x}$, тобто умовні середні зберігають постійні значення. Чим ближче $|r|$ до одиниці, тим тісніше експериментальні точки групуються біля прямих регресії (9.17) та (9.18). При $r = 1$ залежність між X і Y стає функціональною, обидві прямі зливаються в одну, розкид точок зникає.

9.3. Обробка кореляційної таблиці

Визначимо рівняння регресії (9.17) і (9.18) за даними табл. 9.1.

Для спрощення розрахунків за новий початку відліку виберемо пару значень $x = 45$, $y = 38$, яка має найбільшу частоту $n_{xy} = 35$, і перейдемо до умовних варіант:

$$u_j = \frac{x_j - c_1}{h_1} = \frac{x_j - 45}{5}; \qquad v_k = \frac{x_j - c_2}{h_2} = \frac{x_j - 38}{10}.$$

Цей перехід не змінює величини коефіцієнта кореляції S. У нових змінних він має вигляд:

$$r = \frac{\sum_{i=1}^{n} n_{uv}uv - n\bar{u}\bar{v}}{n\sigma_u\sigma_v}, \tag{9.19}$$

$$n_{uv} = n_{xy};\; \bar{u} = \frac{1}{n}\sum_{j=1}^{l}\sum_{k=1}^{m} n_{uv}u_j = \frac{1}{n}\sum_{j=1}^{l} n_u u_j;$$

$$\bar{v} = \frac{1}{n}\sum_{k=1}^{n} n_v v_k;\; \sigma_u = \sqrt{\overline{u^2} - \bar{u}^2};\; \sigma_v = \sqrt{\overline{v^2} - \bar{v}^2}, \tag{9.20}$$

l – число стовпців; m – число рядків.

Для визначення величин, що входять в (9.19), обчислимо добутки un_U, $u^2 n_U$, vn_v, $v^2 n_v$, uvn_{uv}. Результати обчислень запишемо в табл. 9.3. У верхньому кутку кожної клітинки таблиці запишемо відповідні добутки uv.

За даними табл. 9.3 відповідно до формул (9.20) отримаємо:

$$\bar{u} = \frac{\sum_{j=1}^{l} n_u u_j}{n} = \frac{-40}{100} = -0{,}4; \quad \bar{v} = \frac{\sum_{k=1}^{m} n_v v_k}{n} = \frac{-4}{100} = -0{,}04; \quad \overline{u^2} = \frac{\sum_{j=1}^{l} n_u u_j^2}{n} = \frac{132}{100} = 1{,}32;$$

$$\overline{v^2} = \frac{\sum_{k=1}^{m} n_v v_k^2}{n} = \frac{104}{100} = 1{,}04;$$

$$\sigma_u = \sqrt{\overline{u^2} - \bar{u}^2} = \sqrt{1{,}32 - 0{,}16} = \sqrt{1{,}16} \approx 1{,}08;$$

$$\sigma_v = \sqrt{\overline{v^2} - \bar{v}^2} = \sqrt{1{,}04 - 0{,}0016} \approx 1{,}00.$$

За формулою (9.19) знаходимо коефіцієнт кореляції:

$$r = \frac{\sum_{i=1}^{n} n_{uv} uv - n\bar{u}\bar{v}}{n\sigma_u \sigma_v} = \frac{90 - 100 \cdot (-0{,}4)(-0{,}04)}{100 \cdot 1{,}8 \cdot 1{,}2} = 0{,}806;$$

Повертаючись до змінних X і Y, отримаємо

$$\bar{x} = h_1 \bar{u} + c_1 = 5(-0{,}4) + 45 = 43;$$

$$\bar{y} = h_2 \bar{v} + c_2 = 10(-0{,}04) + 38 = 37{,}6;$$

$$\sigma_x = h_1 \sigma_u = 5 \cdot 1{,}08 = 5{,}4;$$

$$\sigma_y = h_2 \sigma_v = 10 \cdot 1{,}02 = 10{,}2.$$

Рівняння регресії (9.17) і (9.18) набувають вигляду:

$$\frac{\bar{y}_x - 37{,}6}{10{,}2} = 0{,}806 \frac{x - 43}{5{,}4} \quad \text{і} \quad \frac{\bar{x}_y - 43}{5{,}4} = 0{,}806 \frac{y - 37{,}6}{10{,}2}$$

або

$$\bar{y}_x = 15{,}3x - 28 \qquad \bar{x}_y = 0{,}43y + 27. \qquad (9.21)$$

Таблиця 9.3

v \ u	-3	-2	-1	0	1	2	n_v	vn_v	v^2n_v	uvn_{uv}
-2	$4^{\boxed{6}}$	$6^{\boxed{4}}$	-	-	-	-	10	-20	40	48
-1	-	$8^{\boxed{2}}$	$10^{\boxed{1}}$	-	-	-	18	-18	18	26
0	-	-	$4^{\boxed{0}}$	$35^{\boxed{0}}$	$5^{\boxed{0}}$	-	44	0	0	0
1	-	-	$4^{\boxed{1}}$	$12^{\boxed{0}}$	$6^{\boxed{1}}$	-	22	22	22	2
2	-	-	-	$1^{\boxed{0}}$	$3^{\boxed{2}}$	$2^{\boxed{4}}$	6	12	24	14
n_u	4	14	18	48	14	2	100	$\sum vn_v = -4$	$\sum v^2n_v = 104$	$\sum uvn_{uv} = 90$
un_u	-12	-28	-18	0	14	4	$\sum un_u = -40$			
u^2n_u	36	56	18	0	14	8	$\sum u^2n_u = 132$			
uvn_{uv}	24	40	6	0	12	8	$\sum uvn_{uv} = 90$			

Побудуємо лінії (9.21) і кореляційне поле – точки, задані в таблиці 9.1 (Рис. 9.2).

Побудуємо також теоретичну пряму регресії Y від X за експериментальною ламаною, побудованою за даними таблиці 9.2.

Як видно з рис. 9.3, обидві лінії досить близькі. Це означає, що теоретичний результат добре узгоджується із статичним матеріалом.

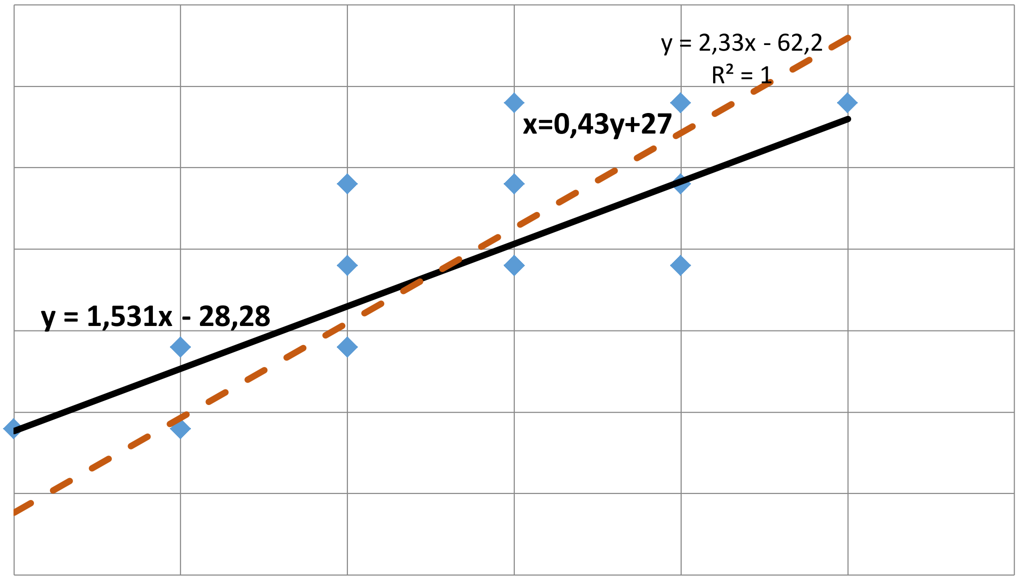


Рис. 9.2. Кореляційне поле та рівняння регресії
суцільна лінія – $\overline{y}_x = 1{,}53x - 28{,}0$; пунктирна - $\overline{x}_y = 0{,}43y + 27{,}0$

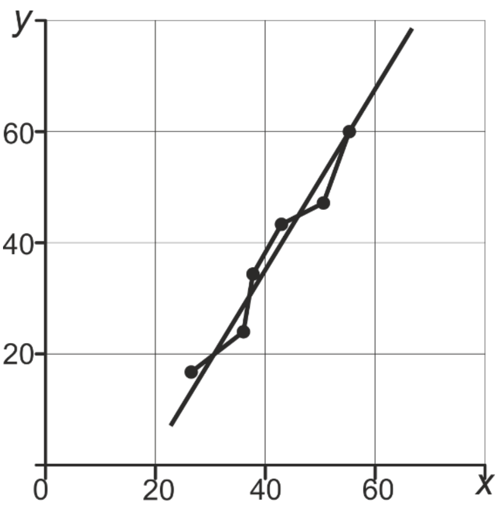


Рис. 9.3. Теоретична пряма регресії Y від X (суцільна лінія) та експериментальна ламана, побудована за даними табл. 9.2 (пунктир)

ДОДАТОК А

Значення нормованої інтегральної функції нормального розподілу

$$F_0(t) = \frac{1}{\sqrt{2\pi}} \cdot \int_{-\infty}^{t} e^{-\frac{t^2}{2}} dt$$

x	*F₀(x)*	*x*	*F₀(x)*	*x*	*F₀(x)*	*x*	*F₀(x)*	*x*	*F₀(x)*	*x*	*F₀(x)*
0	0,5	-0,4	0,3446	-0,8	0,2119	-1,2	0,1151	-1,6	0,0548	-2	0,0228
-0,01	0,496	-0,41	0,3409	-0,81	0,209	-1,21	0,1131	-1,61	0,0537	-2,1	0,0179
-0,02	0,492	-0,42	0,3372	-0,82	0,2061	-1,22	0,1112	-1,62	0,0526	-2,2	0,0139
-0,03	0,488	-0,43	0,3336	-0,83	0,2033	-1,23	0,1093	-1,63	0,0516	-2,3	0,0107
-0,04	0,484	-0,44	0,33	-0,84	0,2005	-1,24	0,1075	-1,64	0,0505	-2,4	0,0082
-0,05	0,4801	-0,45	0,3264	-0,85	0,1977	-1,25	0,1056	-1,65	0,0495	-2,5	0,0062
-0,06	0,4761	-0,46	0,3228	-0,86	0,1949	-1,26	0,1038	-1,66	0,0485	-2,6	0,0047
-0,07	0,4721	-0,47	0,3192	-0,87	0,1922	-1,27	0,102	-1,67	0,0475	-2,7	0,0035
-0,08	0,4681	-0,48	0,3156	-0,88	0,1894	-1,28	0,1003	-1,68	0,0465	-2,8	0,0026
-0,09	0,4641	-0,49	0,3121	-0,89	0,1867	-1,29	0,0985	-1,69	0,0455	-2,9	0,0019
-0,1	0,4602	-0,5	0,3085	-0,9	0,1841	-1,3	0,0968	-1,7	0,0446	-3	0,0013
-0,11	0,4562	-0,51	0,305	-0,91	0,1814	-1,31	0,0951	-1,71	0,0436	-3,1	0,001
-0,12	0,4522	-0,52	0,3015	-0,92	0,1788	-1,32	0,0934	-1,72	0,0427	-3,2	0,0007
-0,13	0,4483	-0,53	0,2981	-0,93	0,1762	-1,33	0,0918	-1,73	0,0418	-3,3	0,0005
-0,14	0,4443	-0,54	0,2946	-0,94	0,1736	-1,34	0,0901	-1,74	0,0409	-3,4	0,0003
-0,15	0,4404	-0,55	0,2912	-0,95	0,1711	-1,35	0,0885	-1,75	0,0401	-3,5	0,0002
-0,16	0,4364	-0,56	0,2877	-0,96	0,1685	-1,36	0,0869	-1,76	0,0392	-3,6	0,0002
-0,17	0,4325	-0,57	0,2843	-0,97	0,166	-1,37	0,0853	-1,77	0,0384	-3,7	0,0001
-0,18	0,4286	-0,58	0,281	-0,98	0,1635	-1,38	0,0838	-1,78	0,0375	-3,8	0,0001
-0,19	0,4247	-0,59	0,2776	-0,99	0,1611	-1,39	0,0823	-1,79	0,0367	-3,9	0
-0,2	0,4207	-0,6	0,2743	-1	0,1587	-1,4	0,0808	-1,8	0,0359	0	0,5
-0,21	0,4168	-0,61	0,2709	-1,01	0,1562	-1,41	0,0793	-1,81	0,0351	0,01	0,504
-0,22	0,4129	-0,62	0,2676	-1,02	0,1539	-1,42	0,0778	-1,82	0,0344	0,02	0,508
-0,23	0,409	-0,63	0,2643	-1,03	0,1515	-1,43	0,0764	-1,83	0,0336	0,03	0,512
-0,24	0,4052	-0,64	0,2611	-1,04	0,1492	-1,44	0,0749	-1,84	0,0329	0,04	0,516
-0,25	0,4013	-0,65	0,2578	-1,05	0,1469	-1,45	0,0735	-1,85	0,0322	0,05	0,5199
-0,26	0,3974	-0,66	0,2546	-1,06	0,1446	-1,46	0,0721	-1,86	0,0314	0,06	0,5239
-0,27	0,3936	-0,67	0,2514	-1,07	0,1423	-1,47	0,0708	-1,87	0,0307	0,07	0,5279
-0,28	0,3897	-0,68	0,2483	-1,08	0,1401	-1,48	0,0694	-1,88	0,0301	0,08	0,5319
-0,29	0,3859	-0,69	0,2451	-1,09	0,1379	-1,49	0,0681	-1,89	0,0294	0,09	0,5359
-0,3	0,3821	-0,7	0,242	-1,1	0,1357	-1,5	0,0668	-1,9	0,0287	0,1	0,5398
-0,31	0,3783	-0,71	0,2389	-1,11	0,1335	-1,51	0,0655	-1,91	0,0281	0,11	0,5438
-0,32	0,3745	-0,72	0,2358	-1,12	0,1314	-1,52	0,0643	-1,92	0,0274	0,12	0,5478
-0,33	0,3707	-0,73	0,2327	-1,13	0,1292	-1,53	0,063	-1,93	0,0268	0,13	0,5517
-0,34	0,3669	-0,74	0,2296	-1,14	0,1271	-1,54	0,0618	-1,94	0,0262	0,14	0,5557
-0,35	0,3632	-0,75	0,2266	-1,15	0,1251	-1,55	0,0606	-1,95	0,0256	0,15	0,5596
-0,36	0,3594	-0,76	0,2236	-1,16	0,123	-1,56	0,0594	-1,96	0,025	0,16	0,5636
-0,37	0,3557	-0,77	0,2206	-1,17	0,121	-1,57	0,0582	-1,97	0,0244	0,17	0,5675
-0,38	0,352	-0,78	0,2177	-1,18	0,119	-1,58	0,0571	-1,98	0,0239	0,18	0,5714
-0,39	0,3483	-0,79	0,2148	-1,19	0,117	-1,59	0,0559	-1,99	0,0233	0,19	0,5753

x	$F_0(x)$	x	$F_0(x)$	x	$F_0(x)$	x	$F_0(x)$	x	$F_0(x)$
0,2	0,5793	0,6	0,7257	1	0,8413	1,4	0,9192	1,8	0,9641
0,21	0,5832	0,61	0,7291	1,01	0,8438	1,41	0,9207	1,81	0,9649
0,22	0,5871	0,62	0,7324	1,02	0,8461	1,42	0,9222	1,82	0,9656
0,23	0,591	0,63	0,7357	1,03	0,8485	1,43	0,9236	1,83	0,9664
0,24	0,5948	0,64	0,7389	1,04	0,8508	1,44	0,9251	1,84	0,9671
0,25	0,5987	0,65	0,7422	1,05	0,8531	1,45	0,9265	1,85	0,9678
0,26	0,6026	0,66	0,7454	1,06	0,8554	1,46	0,9279	1,86	0,9686
0,27	0,6064	0,67	0,7486	1,07	0,8577	1,47	0,9292	1,87	0,9693
0,28	0,6103	0,68	0,7517	1,08	0,8599	1,48	0,9306	1,88	0,9699
0,29	0,6141	0,69	0,7549	1,09	0,8621	1,49	0,9319	1,89	0,9706
0,3	0,6179	0,7	0,758	1,1	0,8643	1,5	0,9332	1,9	0,9713
0,31	0,6217	0,71	0,7611	1,11	0,8665	1,51	0,9345	1,91	0,9719
0,32	0,6255	0,72	0,7642	1,12	0,8686	1,52	0,9357	1,92	0,9726
0,33	0,6293	0,73	0,7673	1,13	0,8708	1,53	0,937	1,93	0,9732
0,34	0,6331	0,74	0,7704	1,14	0,8729	1,54	0,9382	1,94	0,9738
0,35	0,6368	0,75	0,7734	1,15	0,8749	1,55	0,9394	1,95	0,9744
0,36	0,6406	0,76	0,7764	1,16	0,877	1,56	0,9406	1,96	0,975
0,37	0,6443	0,77	0,7794	1,17	0,879	1,57	0,9418	1,97	0,9756
0,38	0,648	0,78	0,7823	1,18	0,881	1,58	0,9429	1,98	0,9761
0,39	0,6517	0,79	0,7852	1,19	0,883	1,59	0,9441	1,99	0,9767
0,4	0,6554	0,8	0,7881	1,2	0,8849	1,6	0,9452	2	0,9772
0,41	0,6591	0,81	0,791	1,21	0,8869	1,61	0,9463	2,1	0,9821
0,42	0,6628	0,82	0,7939	1,22	0,8888	1,62	0,9474	2,2	0,9861
0,43	0,6664	0,83	0,7967	1,23	0,8907	1,63	0,9484	2,3	0,9893
0,44	0,67	0,84	0,7995	1,24	0,8925	1,64	0,9495	2,4	0,9918
0,45	0,6736	0,85	0,8023	1,25	0,8944	1,65	0,9505	2,5	0,9938
0,46	0,6772	0,86	0,8051	1,26	0,8962	1,66	0,9515	2,6	0,9953
0,47	0,6808	0,87	0,8078	1,27	0,898	1,67	0,9525	2,7	0,9965
0,48	0,6844	0,88	0,8106	1,28	0,8997	1,68	0,9535	2,8	0,9974
0,49	0,6879	0,89	0,8133	1,29	0,9015	1,69	0,9545	2,9	0,9981
0,5	0,6915	0,9	0,8159	1,3	0,9032	1,7	0,9554	3	0,9987
0,51	0,695	0,91	0,8186	1,31	0,9049	1,71	0,9564	3,1	0,999
0,52	0,6985	0,92	0,8212	1,32	0,9066	1,72	0,9573	3,2	0,9993
0,53	0,7019	0,93	0,8238	1,33	0,9082	1,73	0,9582	3,3	0,9995
0,54	0,7054	0,94	0,8264	1,34	0,9099	1,74	0,9591	3,4	0,9997
0,55	0,7088	0,95	0,8289	1,35	0,9115	1,75	0,9599	3,5	0,9998
0,56	0,7123	0,96	0,8315	1,36	0,9131	1,76	0,9608	3,6	0,9998
0,57	0,7157	0,97	0,834	1,37	0,9147	1,77	0,9616	3,7	0,9999
0,58	0,719	0,98	0,8365	1,38	0,9162	1,78	0,9625	3,8	0,9999
0,59	0,7224	0,99	0,8389	1,39	0,9177	1,79	0,9633	3,9	1

ДОДАТОК Б

Таблиця значень функції Лапласа: $\Phi(t) = \frac{1}{\sqrt{2\pi}} \cdot \int_0^t e^{-\frac{t^2}{2}} dt$

УВАГА! Дана функція є ***НЕпарною***: $\Phi(-t) = -\Phi(t)$

t	**0**	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**
0,0	0,00000	00399	00798	01197	01595	01994	02392	02790	03188	03586
0,1	03983	04380	04776	05172	05567	05962	06356	06749	07142	07535
0,2	07926	08317	08706	09095	09483	09871	10257	10642	11026	11409
0,3	11791	12172	12552	12930	13307	13683	14058	14431	14803	15173
0,4	15542	15910	16276	16640	17003	17364	17724	18082	18439	18793
0,5	19146	19497	19847	20194	20540	20884	21226	21566	21904	22240
0,6	22575	22907	23237	23565	23891	24215	24537	24857	25175	25490
0,7	25804	26115	26424	26730	27035	27337	27637	27935	28230	28521
0,8	28814	29103	29389	29673	29955	30234	30511	30785	31057	31327
0,9	31594	31859	32121	32381	32639	32894	33147	33398	33646	33891
1,0	34134	34375	34614	34850	35083	35314	35543	35769	35993	36214
1,1	36433	36650	36864	37076	37286	37493	37698	37900	38100	38298
1,2	38493	38686	38877	39065	39251	39435	39617	39796	39973	40147
1,3	40320	40490	40658	40824	40988	41149	41309	41466	41621	41774
1,4	41924	42073	42220	42364	42507	42647	42786	42922	43056	43189
1,5	43319	43448	43574	43699	43822	43943	44062	44179	44295	44408
1,6	44520	44630	44738	44845	44950	45053	45154	45254	45352	45449
1,7	45543	45637	45728	45818	45907	45994	46080	46164	46246	46327
1,8	46407	46485	46562	46638	46712	46784	46856	46926	46995	47062
1,9	47128	47193	47257	47320	47381	47441	47500	47558	47615	47670
2,0	47725	47778	47831	47882	47932	47982	48030	48077	48124	48169
2,1	48214	48257	48300	48341	48382	48422	48461	48500	48537	48574
2,2	48610	48645	48679	48713	48745	48778	48809	48840	48870	48899
2,3	48928	48956	48983	49010	49036	49061	49086	49111	49134	49158
2,4	49180	49202	49224	49245	49266	49286	49305	49324	49343	49361
2,5	49379	49396	49413	49430	49446	49461	49477	49492	49506	49520
2,6	49534	49547	49560	49573	49585	49598	49609	49621	49632	49643
2,7	49653	49664	49674	49683	49693	49702	49711	49720	49728	49736
2,8	49744	49752	49760	49767	49774	49781	49788	49795	49801	49807
2,9	49813	49819	49825	49831	49836	49841	49846	49851	49856	49861

t	**Φ(t)**	**t**	**Φ(t)**	**t**	**Φ(t)**	**t**	**Φ(t)**	**t**	**Φ(t)**
3,0	0,49865	**3,1**	0,49903	**3,2**	0,49931	**3,3**	0,49952	**3,4**	0,49966
3,5	0,49977	**3,6**	0,49984	**3,7**	0,49989	**3,8**	0,49993	**3,9**	0,49995
4,0	0,499968								
4,5	0,499997								
5,0	0,49999997								

ДОДАТОК В

Критичні точки розподілу Стьюдента

Число степенів вільності k	Рівень значущості (двостороння критична область)					
	0,10	0,05	0,02	0,01	0,002	0,001
1	6,31	12,7	31,82	63,7	318,3	637,0
2	2,92	4,30	6,97	9,92	22,33	31,6
3	2,35	3,18	4,54	5,83	10,22	12,9
4	2,13	2,78	3,75	4,60	7,17	8,61
5	2,01	2,57	3,37	4,03	5,89	6,86
6	1,94	2,45	3,14	3,71	5,21	5,96
7	1,89	2,36	3,00	3,50	4,79	5,40
8	1,86	2,31	2,90	3,36	4,50	5,04
9	1,83	2,26	2,82	3,25	4,30	4,78
10	1,81	2,23	2,76	3,17	4,14	4,59
11	1,80	2,20	2,72	3,11	4,03	4,44
12	1,78	2,18	2,68	3,05	3,93	4,32
13	1,77	2,16	2,65	3,01	3,85	4,22
14	1,76	2,14	2,62	2,98	3,79	4,14
15	1,75	2,13	2,60	2,95	3,73	4,07
16	1,75	2,12	2,58	2,92	3,69	4,01
17	1,74	2,11	2,57	2,90	3,65	3,96
18	1,73	2,10	2,55	2,88	3,61	3,92
19	1,73	2,09	2,54	2,86	3,58	3,88
20	1,73	2,09	2,53	2,85	3,55	3,85
21	1,72	2,08	2,52	2,83	3,53	3,82
22	1,72	2,07	2,51	2,82	3,51	3,79
23	1,71	2,07	2,50	2,81	3,49	3,77
24	1,71	2,06	2,49	2,80	3,47	3,74
25	1,71	2,06	2,49	2,79	3,45	3,72
26	1,71	2,06	2,48	2,78	3,44	3,71
27	1,71	2,05	2,47	2,77	3,42	3,69
28	1,70	2,05	2,46	2,76	3,40	3,66
29	1,70	2,05	2,46	2,76	3,40	3,66
30	1,70	2,04	2,46	2,75	3,39	3,65
40	1,68	2,02	2,42	2,70	3,31	3,55
60	1,67	2,00	2,39	2,66	3,23	3,46
120	1,66	1,98	2,36	2,62	3,17	3,37
∞	1,64	1,96	2,33	2,58	3,09	3,29
	0,05	0,025	0,01	0,005	0,001	0,0005
	Рівень значущості (одностороння критична область)					

ДОДАТОК Г

Критичні точки розподілу χ^2

Число ступенів свободи k	Рівень значущості α					
	0,010	**0,025**	**0,050**	**0,950**	**0,975**	**0,980**
1	6,6	5,0	3,8	0,0039	0,00098	0,00016
2	9,2	7,4	6,0	0,103	0,051	0,020
3	11,3	9,4	7,8	0,352	0,216	0,115
4	13,3	11,1	9,5	0,711	0,484	0,297
5	15,1	12,8	11,1	1,15	0,831	0,554
6	16,8	14,4	12,6	1,64	1,24	0,872
7	18,5	16,0	14,1	2,17	1,69	1,24
8	20,1	17,5	15,5	2,73	2,18	1,65
9	21,7	19,0	16,9	3,33	2,70	2,09
10	23,2	20,5	18,3	3,94	3,25	2,56
11	24,7	21,9	19,7	4,57	3,82	3,05
12	26,2	23,3	21,0	5,23	4,40	3,57
13	27,7	24,7	22,4	5,89	5,01	4,11
14	29,1	26,1	23,7	6,57	5,63	4,66
15	30,6	27,5	25,0	7,26	6,26	5,23
16	32,0	28,8	26,3	7,96	6,91	5,81
17	33,4	30,2	27,6	8,67	7,56	6,41
18	34,8	31,5	28,9	9,39	8,23	7,01
19	36,2	32,9	30,1	10,1	8,91	7,63
20	37,6	34,2	31,4	10,9	9,39	8,26
21	38,9	35,5	32,7	11,6	10,3	8,90
22	40,3	36,8	33,9	12,3	11,0	9,54
23	41,6	38,1	35,2	13,1	11,7	10,2
24	43,0	39,4	36,4	13,8	12,4	10,9
25	44,3	40,6	37,7	14,6	13,1	11,5
26	45,6	41,9	38,9	15,4	13,8	12,2
27	47,0	43,2	40,1	16,2	14,6	12,9
28	48,3	44,5	41,3	16,9	15,3	13,6
29	49,6	45,7	42,6	17,7	16,0	14,3
30	50,9	47,0	43,8	18,5	16,8	15,0

ДОДАТОК Д

Варіанти для виконання індивідуального завдання №1

Варіант 1

24,3	23,8	28,4	18,3	12,7
20,6	28	24,2	16,86	34,4
18,7	23,8	17,3	21,5	16,9
23,3	19,6	23,3	28,4	22,4
23,3	21,9	30,3	21	23,3
29,8	28	24,3	22,4	23,8

Варіант 2

14,8	14,5	17,5	10,9	7,3
12,4	17,2	14,8	10	21,4
11,2	14,5	10,3	13	10
14,2	11,8	14,2	17,5	13,6
14,2	13,3	18,7	12,7	14,2
18,4	17,2	14,8	13,6	14,5

Варіант 3

42,5	41,6	50,4	31	28,4
40,7	35,4	49,5	42,5	61,8
41,6	29,3	37,2	28,4	41,6
40,7	33,7	40,7	50,4	39
20,48	40,7	38,1	53,9	36,3
53	49,5	42,5	39	31,9

Варіант 4

32,36	31,7	38,3	23,78	15,86
27,08	37,64	32,36	21,8	46,88
24,44	31,7	22,46	28,4	21,8
31,04	25,76	31,04	38,3	29,72
31,04	29,06	40,94	27,74	31,04
40,28	37,64	32,36	29,72	31,7

Варіант 5

5,6	5,5	6,5	4,3	3,1
4,8	6,4	5,6	4,0	7,8
4,4	5,5	4,1	5,0	4,0
5,4	4,6	5,4	6,5	5,2
5,4	5,1	6,9	4,9	5,4
6,8	6,4	5,6	5,2	5,5

Варіант 6

34,89	`34,18	41,33	25,6	17,01
29,17	40,61	34,89	23,45	50,62
26,31	34,18	24,17	30,6	23,45
33,46	27,74	33,46	41,33	32,03
33,46	31,32	44,19	29,89	33,46
43,47	40,61	34,89	32,03	34,18

Варіант 7

42,4	41,5	50,5	37,7	19,9
35,2	49,6	42,4	28	62,2
31,6	41,5	28,9	37	28
40,6	33,4	40,6	50,5	38,8
40,6	37,9	54,1	36,1	40,6
53,2	49,6	42,4	38,8	41,5

Варіант 8

4,6	4,5	5,5	3,3	2,1
3,8	5,4	4,6	3	6,8
4,4	4,5	3,1	4	3
4,4	3,6	4,4	5,5	4,2
4,4	4,1	5,9	3,9	4,4
5,8	5,4	4,6	4,2	4,5

Варіант 9

37,42	36,65	44,35	27,4	18,17
31,26	43,6	37,42	25,1	54,36
28,18	36,65	25,87	32,8	25,1
35,9	29,72	35,9	44,35	34,34
35,9	33,6	47,43	32,03	35,88
46,66	43,6	37,42	34,34	36,65

Варіант 10

26,3	25,8	30,83	19,7	13,6
22,2	30,3	26,3	18,2	37,4
20,2	25,8	18,7	23,2	18,2
25,3	21,2	25,3	30,8	24,3
25,3	23,8	32,9	22,7	25,3
32,4	30,3	26,3	24,3	25,8

Варіант 11

60,1	49	59,5	36,5	23,9
41,7	58,4	50,1	33,4	73,1
37,5	49	34,4	43,8	33,3
48	39,6	48	59,5	45,9
48	44,9	63,7	42,7	48
62,6	58,4	50	45,9	49

Варіант 12

37,8	37	45	27,4	17,8
31,4	44,2	37,8	25	55,4
28,2	37	25,8	33	25
36,2	29,8	36,2	45	34,6
36,2	33,8	48,2	32,2	36,2
47,4	44,2	37,8	34,6	37

Варіант 13

29,83	29,23	35,28	21,97	14,7
24,99	34,67	29,83	20,15	49,14
22,57	29,23	20,76	26,2	20,15
28,62	23,78	28,62	35,28	27,41
28,62	26,81	35,69	25,59	28,62
37,09	34,67	29,83	27,41	29,23

Варіант 14

39,95	39,13	47,4	29,2	19,3
33,4	46,55	39,9	26,8	58,1
30,1	39,13	27,6	35	26,8
38,3	31,7	38,3	47,4	36,7
38,3	35,8	50,7	34,2	38,3
49,9	46,6	40	36,6	39,1

Варіант 15

52,6	51,5	62,5	38,3	25,1
43,8	61,4	52,6	35	76,8
39,4	51,5	36,1	46	35
50,4	41,6	50,4	62,5	48,2
50,4	47,1	66,9	45	50,4
65,8	61,4	52,6	48,2	51,5

Варіант 16

28,6	28,8	34	20,8	13,6
23,8	33,4	28,6	19	41,8
21,4	28	19,6	25	19
27,4	22,4	27,4	34	26,2
27,4	25,6	36,4	24,4	27,4
35,8	33,4	28,6	26,2	28

Варіант 17

45	44,1	53,4	32,9	21,6
37,5	52,5	45	30,1	65,6
33,8	44,1	31	39,4	30,1
43,1	35,66	43,1	53,4	41,3
43,1	40,3	57,2	38,47	40,1
56,2	52,5	45	41,3	43,14

Варіант 18

19,4	19	23	14,2	9,4
16,2	22,6	19,4	13	28,2
14,6	19	13,4	17	13
18,6	14,5	18,6	23	17,8
18,6	17,4	24,6	16,6	18,6
24,2	22,6	19,4	17,8	19

Варіант 19

47,5	46,6	56,5	34,7	22,8
39,6	55,5	47,5	31,7	69,3
35,7	46,6	32,7	41,6	31,7
45,6	37,6	45,6	56,5	43,6
45,6	42,6	60,4	40,6	45,6
59,4	55,5	47,5	43,6	46,6

Варіант 20

10,2	10	12	7,6	5,2
8,6	11,8	10,2	7	14,6
7,8	10	7,2	9	7
9,8	8,2	9,8	12	9,4
9,8	9,2	12,8	8,8	9,8
12,6	11,8	10,2	9,4	10

Варіант 21

47	46	56	34	22
39	55	47	31	69
35	46	37	41	31
45	37	45	56	43
45	42	60	40	45
59	55	47	43	46

Варіант 22

25,3	24,8	29,6	19	13
21,4	29,1	25,3	17,5	35,9
19,5	24,8	18	22,4	17,5
24,3	20,4	24,3	29,6	23,3
24,3	22,8	31,6	21,9	24,3
31,1	29,1	25,3	23,3	24,8

Варіант 23

33,2	32,5	39,5	24,1	15,7
27,6	38,8	33,2	28	48,6
28,4	32,5	22,7	29	22
31,8	26,2	31,8	39,5	30,4
31,8	29,7	42,3	28,3	31,8
41,6	38,8	33,2	30,4	32,5

Варіант 24

24	23,5	28,5	17,5	11,5
20	28	24	16	35
18	23,5	16,5	21	16
23	19	23	28,5	22
23	21,5	30,5	20,5	23
30	28	24	22	23,5

ДОДАТОК Е

Варіанти для виконання індивідуального завдання №2

1	x	10	10	10	15	15	15	20	20	20	25	25	25	30
	y	25	25	20	35	30	35	45	45	50	55	55	55	60
2	x	10	10	15	20	20	20	25	25	30	35	35	35	40
	y	15	15	15	25	20	25	30	30	30	35	35	35	35
3	x	6	6	6	16	17	18	20	20	20	22	23	22	25
	y	20	21	21	25	24	25	31	29	30	35	35	35	38
4	x	12	12	12	15	15	15	23	21	23	38	27	28	35
	y	14	14	14	22	21	22	30	30	30	36	37	38	42
5	x	20	20	20	22	25	25	30	30	31	25	35	35	41
	y	9	10	9	14	15	15	20	20	19	29	24	24	29
6	x	8	8	8	16	15	16	22	22	23	29	30	30	38
	y	17	47	17	27	26	27	38	37	37	45	47	46	57
7	x	11	12	12	17	17	17	20	22	21	28	27	27	32
	y	20	25	25	34	34	34	36	36	36	40	40	40	41
8	x	10	9	9	15	15	15	18	18	25	22	25	25	25
	y	10	10	10	11	11	11	12	12	15	13	15	15	15
9	x	10	15	15	15	20	21	20	26	27	26	32	32	32
	y	2	3	3	3	4	4	4	5	5	5	6	6	6
10	x	9	9	9	14	14	14	19	19	19	24	24	24	29
	y	24	24	19	34	29	34	44	44	49	54	54	54	59
11	x	9	9	14	19	19	19	24	24	29	34	34	34	39
	y	14	14	14	24	19	24	29	29	29	34	34	34	34
12	x	5	5	5	15	16	17	19	19	19	21	22	21	24
	y	19	20	20	24	23	24	30	28	29	34	34	34	37
13	x	11	11	11	14	14	14	22	20	22	37	26	27	34
	y	13	13	13	21	20	21	29	29	29	35	36	37	41
14	x	19	19	19	21	24	24	29	29	30	24	34	34	40
	y	8	9	8	13	14	14	19	19	18	28	23	23	28
15	x	7	7	7	15	14	15	21	21	22	28	29	29	37
	y	16	46	16	26	25	26	37	36	36	44	46	45	56
16	x	10	11	11	16	16	16	19	21	20	27	26	26	31
	y	19	24	24	33	33	33	35	35	35	39	39	39	40
17	x	9	8	8	14	14	14	17	17	24	21	24	24	24
	y	9	9	9	10	10	10	11	11	14	12	14	14	14
18	x	9	14	14	14	19	20	19	25	26	25	31	31	31
	y	1	2	2	2	3	3	3	4	4	4	5	5	5
19	x	11	11	11	16	16	16	21	21	21	26	26	26	31
	y	26	26	21	36	31	36	46	46	51	56	56	56	61
20	x	11	11	16	21	21	21	26	26	31	36	36	36	41
	y	16	16	16	26	21	26	31	31	31	36	36	36	36

Примітка. У таблиці наведено: «*x*» – навантаження на кріплення (МПа)
«*y*» – зміщення контуру виробки (мм).

Бібліографічний список

1. Вентцель Е. С. Теория вероятностей / Е.С. Вентцель – М.: Наука, 1969. – 572 с.
2. Рыжов П.А. Математическая статистика в горном деле: підручник / Рыжов П.А. – М.: Высшая школа, 1973. – 287 с..
3. Гмурман В.Е Теория вероятностей и математическая статистика: підручник / Гмурман В.Е – М.: Высшая школа, 1972. – 368 с..
4. Карасев Б.В. Статистический подход к изучению природы и некоторые закономерности распределения вещества Земли : підручник / Карасев Б.В. – М.: Наука, 1971. – 151 с..
5. Шашенко А.Н. Некоторые задачи статистической геомеханики: підручник / Шашенко А.Н. , Тулуб С.Б., Сдвижкова Е.А. – К.: Пульсари, 2001. – 243 с..
6. Новикова Л.В. Теорія ймовірностей і математична статистика: підручник / Новикова Л.В., Котляр Б.Д., Бичков В.І. – К. : Техніка, 1996. – 184 с..

Навчальне видання

Сдвижкова Олена Олександрівна
Бугрим Ольга Володимирівна
Бабець Дмитро Володимирович
Іванов Олексій Сергійович

Елементи теорії ймовірностей та математичної статистики в гірництві

Навчальний посібник

Редактор Є.М. Ільченко

Підп. до друку 25.05.2015. Формат 30×42/4.
Папір офсет. Ризографія. Ум. друк. арк. 6,0.
Обл-вид. арк. 8,1. Тираж 100 пр,. Зам. №

Підготовлено до друку та видруковано
у Державному вищому навчальному закладі
«Національний гірничий університет».
Свідоцтво про внесення до Державного реєстру ДК № 1842, від 11.06.2004 р.

49005, м. Дніпропетровськ, просп. К.Маркса, 19.