МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ

ХАРКІВСЬКИЙ НАЦІОНАЛЬНИЙ ЕКОНОМІЧНИЙ УНІВЕРСИТЕТ
ІМЕНІ СЕМЕНА КУЗНЕЦЯ

Е. Ю. Железнякова
Л. О. Норік

ТЕОРІЯ ЙМОВІРНОСТЕЙ ТА МАТЕМАТИЧНА СТАТИСТИКА

Практикум

Харків
ХНЕУ ім. С. Кузнеця
2019

УДК 519.2(075.34)
Ж51

Авторський колектив: канд. фіз.-мат. наук, доцент Е. Ю. Железнякова – вступ, теми 1 – 8; канд. екон. наук, доцент Л. О. Норік – теми 9 – 16.

Рецензенти: провідний науковий співробітник відділу математичного моделювання та оптимального проектування Інституту проблем машинобудування ім. А. М. Підгорного НАН України, д-р техн. наук, професор *Т. Є. Романова*; завідувач кафедри системотехніки Харківського національного університету радіоелектроніки, д-р техн. наук, професор *І. В. Гребеннік*.

Рекомендовано до видання рішенням ученої ради Харківського національного економічного університету імені Семена Кузнеця.

Протокол № 4 від 29.11.2018 р.

Самостійне електронне текстове мережеве видання

Железнякова Е. Ю.

Ж51 Теорія ймовірностей та математична статистика [Електронний ресурс] : практикум / Е. Ю. Железнякова, Л. О. Норік. – Харків : ХНЕУ ім. С. Кузнеця, 2019. – 321 с.

ISBN 978-966-676-726-7

Розглянуто основні питання програми курсу теорії ймовірностей та математичної статистики. Згідно з темами матеріал розподілено на два змістових модулі. До кожної теми подано основні теоретичні відомості, докладні приклади розв'язування типових задач, запропоновано вправи для самостійної роботи, тестові завдання та запитання для самоперевірки, які повністю відповідають програмі навчальної дисципліни.

Рекомендовано для студентів усіх спеціальностей як допоміжний матеріал у процесі самостійного вивчення навчальної дисципліни, а також для використання викладачами під час проведення практичних занять та організації самостійної роботи студентів.

УДК 519.2(075.034)

ISBN 978-966-676-726-7

Зміст

Вступ

Навчальна дисципліна "Теорія ймовірностей та математична статистика" є нормативною навчальною дисципліною, яку вивчають, згідно з навчальним планом підготовки фахівців першого бакалаврського рівня. Предметом цієї навчальної дисципліни є фундаментальні положення теорії ймовірностей щодо дослідження випадкових подій і визначення їхньої ймовірності, основних законів розподілу дискретної та неперервної випадкових величин, їхніх числових характеристик, теорія масового обслуговування, а також принципів формування вибіркової сукупності, реєстрації, опису та аналізу результатів статистичних спостережень, основи теорії кореляції та регресійного аналізу. Під час вивчення цієї навчальної дисципліни основне завдання полягає у формуванні у студента необхідних компетентностей щодо застосування елементів теорії ймовірностей та математичної статистики у процесі дослідження економічних процесів і явищ, де математика відіграє роль інструменту. Із переходом на кредитно-трансферну систему в умовах скорочення аудиторного часу на вивчення дисципліни підвищується роль самостійної роботи студентів. Необхідною умовою успішного засвоєння навчального матеріалу є самостійна робота студентів із додатковою літературою.

Актуальність навчально-методичного та інформаційного забезпечення самостійної роботи студентів закладів вищої освіти підтверджено тим, що в сучасному суспільстві посилено вимоги до учасників системи соціальних взаємин і, як ніколи раніше, зростає роль професійної готовності фахівців.

Цільовим призначенням цього практикуму є формування цілісної системи теоретичних знань математичного апарату, необхідного для розв'язання практичних задач у професійній діяльності фахівця у сфері економіки та менеджменту, уміння аналітичного мислення та навичок у застосуванні математичного апарату до формалізації реальних процесів і явищ, а також до розв'язання економічних задач.

Знання та навички, набуті студентом у результаті вивчення навчальної дисципліни, становлять основу аналітико-дослідницьких компетентностей сучасного фахівця в галузі економіки та менеджменту в контексті Національної рамки кваліфікацій, серед них такі:

уміння застосовувати здобуті теоретичні знання до вирішення ситуаційних вправ та розв'язання реальних практичних задач;

навички самостійно визначати алгоритм математичних обчислень;

уміння відокремлювати незалежні та залежні фактори, визначати числові характеристики розподілу випадкової величини;

здійснювати аналіз результатів, із метою уточнення типу і форми зв'язку між факторами;

уміння аналізувати визначені результати та з їхнім урахуванням робити обґрунтовані висновки на достатньо високому професійному рівні;

здатність автономно і відповідально виконувати завдання;

комунікації щодо взаємодії студентів у процесі спільного вирішення проблем і ухвалення погоджених рішень.

Посібник складається із 16 тем. Кожна тема має 7 частин: мета та компетентності, основні теоретичні відомості, приклади розв'язування задач, вправи для самостійної роботи, тестові завдання, запитання для самоперевірки та висновки за темою. Виклад теоретичного матеріалу за кожною темою супроводжують розв'язанням прикладів, які ілюструють його сутність. Для того щоб процес навчання мав активний характер, у більшості прикладів розглядають ситуації щодо різних галузей економіки. У додатках уміщено основні статистичні таблиці, необхідні для розв'язання задач.

Така структура посібника дозволяє використовувати його як під час аудиторних занять, так і для самостійної роботи студентів. Студенти можуть користуватися викладеним матеріалом також під час написання курсових, дипломних і магістерських робіт.

Розділ 1
Теорія ймовірностей

1. Емпіричні та логічні основи теорії ймовірностей

1.1. Мета та компетентності

Метою вивчення теми є ознайомлення студентів із предметом і завданнями навчальної дисципліни "Теорія ймовірностей та математична статистика", що є теоретичною базою математичного моделювання економічних процесів і явищ; з основними означеннями теорії ймовірностей та формування системи теоретичних знань, практичних умінь і навичок щодо визначення ймовірностей випадкових подій та інтерпретації здобутих результатів за допомогою діаграм Венна – Ейлера.

Компетентності, що формують під час вивчення теми:

знання загальних означень теорії ймовірностей;

уміння визначати умови застосування аксіом, на яких побудовано теорію ймовірностей;

уміння знаходити ймовірності випадкових подій за класичним та геометричним означеннями ймовірності;

уміння знаходити ймовірності випадкових подій за допомогою формул комбінаторики;

надання геометричної інтерпретації випадкових подій та дій над ними за допомогою діаграм Венна – Ейлера.

1.2. Основні теоретичні відомості. Елементи комбінаторики

Перестановками називають комбінації, що складаються з тих самих n різних елементів і відрізняються тільки порядком їхнього розташування. Кількість усіх можливих перестановок дорівнює:

$$P_n = n!, \text{ де } n! = 1 \cdot 2 \cdot 3 \cdot \ldots \cdot n.$$

Розміщеннями називають комбінації, складені з n різних елементів по m елементів, що відрізняються або складом елементів, або їхнім порядком. Кількість усіх можливих розміщень дорівнює:

$$A_n^m = n(n-1)\cdot \ldots \cdot (n-m+1) \text{ або } A_n^m = \frac{n!}{(n-m)!}.$$

Сполученнями називають комбінації, складені з n різних елементів по m елементів, що відрізняються хоча б одним елементом. Кількість сполучень обчислюють за формулою:

$$C_n^m = \frac{n(n-1)\cdot\ldots\cdot(n-m+1)}{1\cdot 2\cdot 3\cdot\ldots\cdot m} \text{ або } C_n^m = \frac{n!}{m!(n-m)!}.$$

Класичне визначення ймовірності події

Нехай n – загальна кількість усіх рівноможливих, несумісних і тих, що утворюють повну групу наслідків випробувань, а m – кількість наслідків, що сприяють появі події A. Тоді ймовірність $P(A)$ дорівнює відношенню m *до* n, тобто $P(A) = \dfrac{m}{n}$.

Згідно з означенням $0 \le P(A) \le 1$.

Геометричне тлумачення ймовірності. Геометрична ймовірність

Геометричне тлумачення ймовірності широко застосовують для більш наочного подання умови задачі в тому разі, якщо простір подій є дискретним, тобто є можливість застосовувати класичне означення ймовірності. Але воно стає незамінним у тому разі, якщо простір подій є неперервним. Слід розглянути діаграму Венна – Ейлера (рис.1.1).

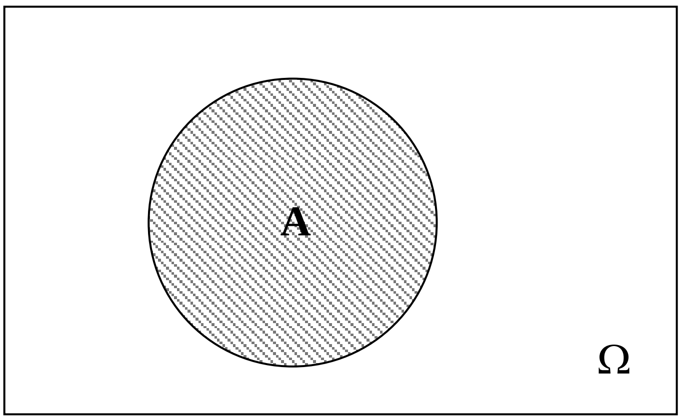


Рис. 1.1. **Діаграма Венна – Ейлера для визначення ймовірності випадкової події**

Простір елементарних подій позначають Ω. Це сукупність усіх можливих несумісних наслідків випробування, і на діаграмі Венна – Ейлера їй відповідає вся площа прямокутника.

Якщо A – випадкова подія і сукупності елементарних подій, що сприяють появі цієї випадкової події, відповідає площі однойменного кола, то ймовірність події A вимірюють як відношення площі кола до площі прямокутника, тобто $P(A)=\dfrac{S_A}{S_\Omega}$. Як і за класичного визначення ймовірності випадкової події, з означення випливає, що $0 \le P(A) \le 1$.

Хоча найбільш поширеним є двовимірне зображення множини подій (тобто діаграми Венна – Ейлера), але геометричне означення ймовірності передбачає використання і геометричних фігур іншої вимірності, наприклад, простору елементарних подій може відповідати відрізок прямої, а випадковій події – частина цього відрізка.

За допомогою діаграм Венна – Ейлера зручно ілюструвати теореми множення та додавання ймовірностей випадкових подій. Так, на рис. 1.2 наведено приклад двох сумісних випадкових подій A та B. Отже, переріз цих подій $A \cap B$ є непорожньою множиною, якій відповідає спільна частина обох кіл, виділена на рис. 1.2 подвійним штрихом. Відношення площі цієї фігури до загальної площі прямокутника і є ймовірністю події добутку подій $A \cdot B$.

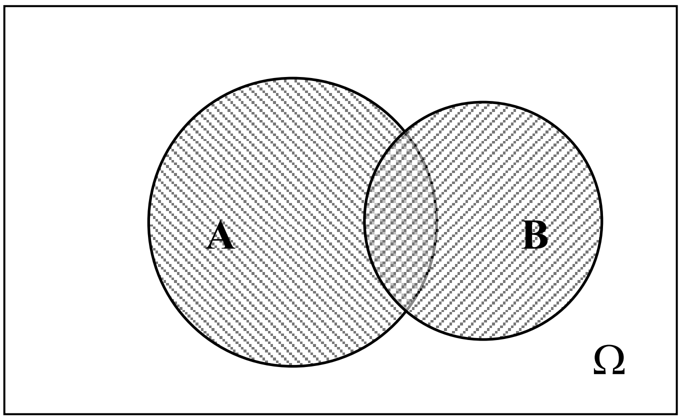


Рис. 1.2. **Діаграма Венна – Ейлера, що визначає сумісні випадкові події A та B**

Сполученню випадкових подій $A \cup B$, що є сумою вихідних випадкових подій, відповідає вся площа, виділена штрихом. Оскільки площа, позначена подвійним штрихом, належить одночасно до обох кіл, то ймовірність суми випадкових подій A та B визначають співвідношенням:

$$P(A+B) = P(A) + P(B) - P(A \cdot B).$$

Слід розглянути поняття "різниця множин". Так, різниці множин $A-B$, що означає $A \setminus B$, відповідає площа кола A за винятком тієї його частини, яка належить одночасно і колу A, і колу B. Отже, різницю множин $A-B$ можна також подати як переріз $A \cap \overline{B}$, де $\overline{B}$ – випадкова подія, яка полягає в тому, що подія B не відбулася. На рис. 1.3 різницю множин $A-B$ позначено подвійним штрихом.

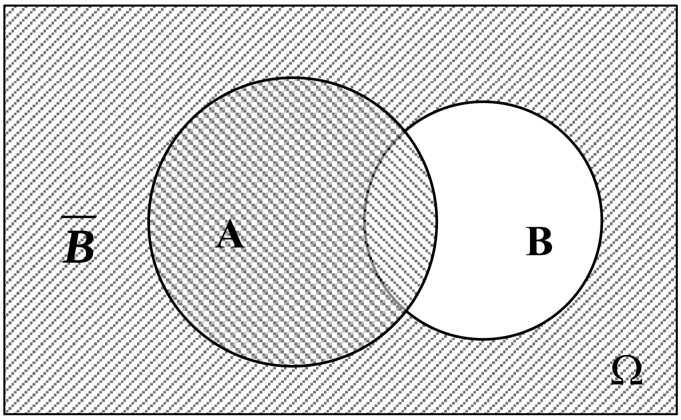


Рис. 1.3. **Діаграма Венна – Ейлера, що ілюструє різницю випадкових подій A та B**

Імовірність події $A \cap \overline{B}$ можна записати в символах умовної ймовірності таким чином: $P(A|\overline{B})$, тобто ймовірність події A за умов, що подія B не відбулася (мала місце подія $\overline{B}$). Тоді ймовірність суми подій записують у вигляді:

$$P(A+B) = P(A|\overline{B}) + P(B|\overline{A}) + P(A \cdot B).$$

1.3. Приклади розв'язування задач

Приклад 1.1. У конкурсі беруть участь 5 студентів. Порядок їхнього виступу визначають жеребкуванням. Скільки існує варіантів порядку їхнього виступу?

Розв'язання. Кожний варіант жеребкування відрізняється від інших тільки порядком виступу конкурсантів, а тому він є перестановкою з 5 елементів. Кількість варіантів дорівнює $P_5 = 5! = 1 \cdot 2 \cdot 3 \cdot 4 \cdot 5 = 120$.

Приклад 1.2. Розклад занять одного дня студентів складається із 3 пар. Визначте кількість варіантів розкладу за вибору з 10 дисциплін.

Розв'язання. Кожен варіант розкладу є набором 3 дисциплін із 10, що відрізняється від інших варіантів або складом дисциплін, або порядком їхнього розташування, тобто він є розміщенням із 10 елементів по 3 елементи. Кількість таких варіантів розкладу занять дорівнює $A_{10}^3 = 10 \cdot 9 \cdot 8 = 720$.

Приклад 1.3. У шаховому турнірі беруть участь 20 юнаків. Скільки партій вони зіграють, якщо між будь-якими двома учасниками має бути зіграно одну партію?

Розв'язання. Кожна партія відрізняється тільки складом пар учасників, тобто вона є сполученням із 20 елементів по 2 елементи. Їхня кількість дорівнює $C_{20}^2 = \frac{20!}{2!18!} = \frac{20 \cdot 19}{1 \cdot 2} = 190$.

Приклад 1.4. Із 50 стандартних і 5 нестандартних деталей для контролю навмання взято 10, що виявилися стандартними. Знайдіть імовірність того, що наступна навмання взята деталь буде стандартною.

Розв'язання. Стандартних деталей залишилося 50 – 10 = 40, а нестандартних 5. Усього залишилося 40 + 5 = 45 деталей. Шукана ймовірність дорівнює $P = \frac{40}{45} = \frac{8}{9}$.

Приклад 1.5. Монету кинуто 2 рази. Знайдіть імовірність того, що хоча б один раз з'явиться "герб".

Розв'язання. За двох кидань монети можливі чотири елементарні результати: "герб" – "герб", "герб" – "герб", "цифра" – "герб", "цифра" – "цифра". Із них сприятливими є три перші, тобто m = 3, а n = 4. Шукана ймовірність $P = \frac{3}{4}$.

Приклад 1.6. На 5 картках написано букви "е", "о", "ч", "н", "в". Після ретельного перемішування беруть по 1 картці та кладуть послідовно поруч. Яка ймовірність того, що в результаті утвориться слово "човен"?

Розв'язання. Кількість усіх можливих наслідків випробувань у цьому разі дорівнює кількості перестановок із п'яти букв, тобто:

$$P_5 = 5! = 1 \cdot 2 \cdot 3 \cdot 4 \cdot 5 = 120.$$

Із них лише одна утворить слово "човен". Тому m = 1, а ймовірність шуканої події $P = \frac{1}{120}$.

Приклад 1.7. Серед 17 студентів, із яких 8 дівчат, розігрують 7 квитків. Яка ймовірність, що серед власників квитків буде 4 дівчини?

Розв'язання. Кількість можливих способів розподілити 7 квитків серед 17 студентів дорівнює кількості сполучень із 17 елементів по 7, тобто C_{17}^7. Кількість відбору чотирьох дівчат із восьми дорівнює C_8^4. Кожну таку четвірку можна сполучити з кожною трійкою з 9 юнаків. Кількість таких трійок дорівнює C_9^3.

Отже, кількість результатів такого розподілу 7 квитків, коли 4 з них виявляться в дівчат, а 3 – у юнаків, дорівнює $C_8^4 \cdot C_9^3$. Тоді шукана ймовірність дорівнює $P = \frac{C_8^4 \cdot C_9^3}{C_{17}^7} = \frac{735}{2431} \approx 0,302.$

Приклад 1.8. Набираючи номер телефону, абонент забув останні три цифри, і, пам'ятаючи лише, що ці цифри різні, набрав їх навмання. Знайдіть імовірність того, що набрані цифри правильні.

Розв'язання. Нехай подія A – набрані три потрібні цифри. Усього можна набрати стільки різних цифр, скільки може бути складено розміщень із десяти цифр по три, тобто $A_{10}^3 = 10 \cdot 9 \cdot 8 = 720$. Це загальна кількість елементарних подій. Сприятливий події *A* тільки один результат – три правильні цифри. Тому шукана ймовірність дорівнює $P = \frac{1}{720}$.

Приклад 1.9. Велике коло радіусом 50 см містить у центрі маленьке коло радіусом 5 см. Визначте ймовірність того, що куля влучить у маленьке коло, якщо відомо, що вона влучила у велике коло.

Розв'язання. У цьому прикладі неможливо скористатися класичним означенням імовірності. Застосуйте її геометричне означення. Оскільки куля влучила у велике коло, то можна вважати, що його площа відповідає множині Ω (повній групі елементарних подій). Тоді ймовірність випадкової події, а саме, улучення кулі в маленьке коло, визначають як відношення площі малого кола до площі великого кола, тобто:

$$P(A) = \frac{\pi r^2}{\pi R^2} = \frac{25}{2500} = 0,01.$$

Приклад 1.10. Квадратну коробку розподілено на секції, три з яких є однаковими за площею квадратами, а інші три – прямокутниками, до того ж площа кожного із прямокутників удвічі більша за площу квадратної секції. До коробки кидають кульку. Яка ймовірність того, що вона потрапить до квадратної секції?

Розв'язання. Імовірність улучення кульки до однієї із квадратних секцій визначають відношенням загальної площі цих трьох секцій до площі всієї квадратної коробки. Нехай один бік коробки має розмір a. Отже, площа всієї коробки дорівнює a^2. Загальна площа кожної квадратної секції вдвічі більша за площу прямокутної, звідси загальна площа всіх трьох квадратних секцій становить $\frac{a^2}{3}$. Тоді ймовірність улучення кульки до квадратної секції дорівнює

$$P(A) = \frac{\frac{a^2}{3}}{a^2} = \frac{1}{3}.$$

Приклад 1.11. Серед 100 студентів потоку 60 добре володіють англійською, 40 – німецькою та 20 – французькою. Визначте ймовірність того, що навмання вибраний студент не володіє іноземною мовою.

Розв'язання. Хоча у прикладі розглядають дискретні множини подій, умову задачі зручно подати за допомогою діаграми Венна – Ейлера (рис. 1.4).

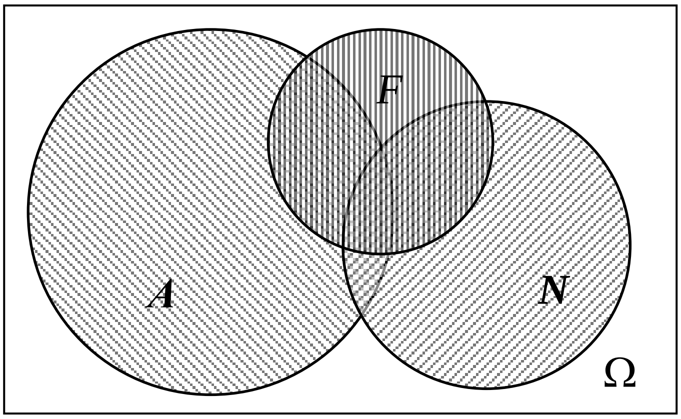


Рис. 1.4. **Діаграма Венна – Ейлера до прикладу 1.11**

На діаграмі множини студентів, які володіють певною іноземною мовою, позначено початковими літерами. Площина поза колом відповідає множині студентів, які не володіють цією мовою, тобто протилежна подія.

Отже, множині студентів, які не володіють жодною іноземною мовою, відповідає частина прямокутника повної групи елементарних подій, вільній від штрихування. Нехай це буде випадкова подія V, імовірність якої слід визначити.

Випадкову подію V, яка полягає в тому, що навмання вибраний студент не володіє жодною іноземною мовою, визначають як переріз множин $\overline{A}$, $\overline{N}$ та $\overline{F}$. Ці події сумісні, але незалежні, тому для визначення ймовірності є таке співвідношення:

$$P(V) = P(\overline{A} \cap \overline{N} \cap \overline{F}) = P(\overline{A}) \cdot P(\overline{N}) \cdot P(\overline{F}) =$$
$$= \left(1 - P(A)\right) \cdot \left(1 - P(N)\right) \cdot \left(1 - P(F)\right) = (1 - 0,6) \cdot (1 - 0,4) \cdot (1 - 0,2) = 0,192.$$

Знайдено відповідь на запитання задачі, але для перевірки розгляньте інший шлях розв'язання. Визначте ймовірність випадкової події, яка полягає в тому, що навмання вибраний студент добре володіє якою-небудь іноземною мовою (оскільки ця подія є протилежною тій, імовірність якої досліджують, то її позначають $\overline{V}$). Такий студент може знати будь-яку, але тільки одну мову, дві мови або навіть три мови. Випадковій події $\overline{V}$ відповідає вся заштрихована площа на рис. 1.4. Події A, N та F незалежні, але сумісні, отже:

$$P(\overline{V}) = P(A) + P(N) + P(F) - P(A \cdot N) - P(A \cdot F) - P(N \cdot F) +$$
$$+ P(A \cdot N \cdot F) = P(A) + P(N) + P(F) - P(A) \cdot P(N) - P(A) \cdot P(F) -$$
$$- P(N) \cdot P(F) + P(A) \cdot P(N) \cdot P(F) = 0,6 + 0,4 + 0,2 - 0,6 \cdot 0,4 -$$
$$- 0,6 \cdot 0,2 - 0,4 \cdot 0,2 + 0,6 \cdot 0,4 \cdot 0,2 = 0,808.$$

Звідси знаходять, що $P(V) = 1 - P(\overline{V}) = 1 - 0,808 = 0,192.$ Зрозуміло, що обидва розв'язання дали той самий результат.

Приклад 1.12. На числовій прямій узято відрізок OA, початок якого знаходиться в початку координат, а точці A відповідає одиниця. Відомо, що точка B, яка є внутрішньою точкою відрізку, ділить його у відношенні $1:2$. На цей відрізок довільно нанесено точку, яку позначено C. Визначте ймовірність того, що відстань між точками B та C буде меншою, ніж відстань від точки C до будь-якого з кінців відрізка.

Розв'язання. Множина випадкових подій, про яку йде мова у прикладі, є неперервною, отже, для обчислення ймовірності слід скористатися її геометричним означенням. Спочатку зробіть рисунок (рис. 1.5).

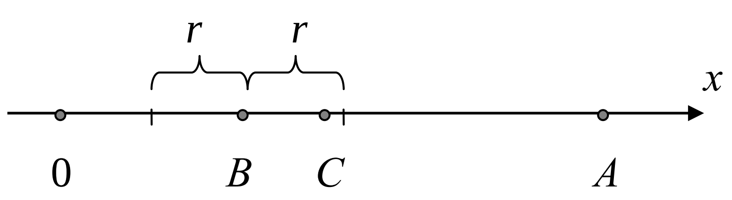


Рис. 1.5. **Ілюстрація до прикладу 1.12**

Оскільки точка B ділить відрізок OA у відношенні $1:2$, то це означає, що вона знаходиться від найближчого до неї кінця відрізка на відстані $\frac{1}{3}$ (наприклад, має координату $\frac{1}{3}$, як показано на рис. 1.5). Те, що відстань між точками B та C є меншою за відстань від точки C до будь-якого з кінців відрізка, передбачає, що точка C має влучити до околу точки B радіусом $r = 0,5 \cdot OB$. Імовірність такої випадкової події визначають як відношення діаметра околу до довжини відрізка OA. Оскільки точка B може мати координату або $\frac{1}{3}$, або $\frac{2}{3}$, то результат треба ще помножити на 2.

Отже, $P(C) = \dfrac{2 \cdot r}{OA} \cdot 2 = \dfrac{2 \cdot 0,5 \cdot \frac{1}{3}}{1} \cdot 2 = \dfrac{2}{3}$.

1.4. Вправи для самостійної роботи

1.1. Обчисліть:

а) $15! \cdot \dfrac{1}{30 \cdot P_{12}}$; б) $\dfrac{P_7 + P_6}{P_5}$; в) $\dfrac{A_5^4 + A_5^3}{A_5^2}$; г) $\dfrac{A_6^2 \cdot P_8}{P_{10}} + \dfrac{A_5^2}{A_4^2}$; д) $\dfrac{C_{100}^{97} \cdot P_5}{66 \cdot A_{50}^2}$.

1.2. Спростіть вираз $\dfrac{A_n^7 - A_n^5}{A_n^5}$ і обчисліть за $n = 15$.

1.3. Спростіть вираз $\dfrac{C_n^2 + C_n^{n-2}}{C_{n+1}^3 + C_{n+1}^{n-2}}$ і обчисліть за $n = 11$.

1.4. Розв'яжіть рівняння $C_{x+3}^3 = 3(x+1)(x+2)$.

1.5. Є 5 прапорців різного кольору. Скільки існує різних сигналів, які можна подавати, піднімаючи по 3 прапорці в будь-якому порядку?

1.6. У конкурсі беруть участь 9 осіб. Скільки існує варіантів списку призерів, якщо кожне із 3 призових місць може посісти тільки одна людина?

1.7. На участь у шкільному конкурсі "Пісня року" подали заявки 12 осіб. Порядок їхнього виступу визначався жеребкуванням. Скільки існує способів скласти програму концерту?

1.8. Для підбиття підсумків конкурсу необхідно із 30 присутніх обрати незалежне журі у складі голови, секретаря та спостерігача. Скільки існує способів такого вибору?

1.9. Учні вивчають 8 різних предметів. Розклад уроків має бути складено так, щоб на день було 3 уроки та предмети не повторювалися. Скільки існує способів скласти розклад на один день?

1.10. Скільки існує способів розмістити 6 осіб за одним столом?

1.11. Слово зашифроване 6 позначками, які не повторюються. Скільки існує варіантів розшифрування, якщо позначка може відповідати одній із 30 букв алфавіту?

1.12. У шаховому турнірі беруть участь 12 осіб. Кожен з учасників має зіграти з кожним із інших по 2 партії. Скільки партій мають зіграти учасники турніру?

1.13. У шухляді лежать 12 олівців однакової форми й розмірів, із яких 4 олівці – кольорові, а інші – прості. Яка ймовірність, що, відкривши шухляду, буде взято простий олівець?

1.14. Із цифр від 1 до 9 включно навмання вибирають одну. Знайдіть імовірність того, що вибране число буде: 1) парним; 2) непарним; 3) простим; 4) більшим за 7.

1.15. У шухляді в 5 разів більше червоних куль, ніж чорних. Навмання виймають 1 кулю. Яка ймовірність того, що вона буде червоною?

1.16. У цеху з виготовлення м'ячів для гольфа в одній коробці було 77 м'ячів правильної форми, а в іншій – 23 м'ячі з дефектами. М'ячі висипали в одну коробку. Яка ймовірність того, що навмання витягнутий м'яч буде бракованим?

1.17. Кинуто дві гральні кістки. Знайдіть імовірності таких подій:

а) сума чисел на гранях, що випали, дорівнює 7;

б) сума чисел дорівнює 8, а різниця – 4;

в) сума чисел дорівнює 5, а добуток – 4;

г) сума чисел більша за 7;

д) сума чисел більша за 3 і не більша за 8.

1.18. Монету кинуто два рази. Яка ймовірність того, що обидва рази випаде "герб"?

1.19. Монету підкидають тричі. Чому дорівнює ймовірність того, що "герб" випаде хоча б один раз?

1.20. На кожній із 5 однакових карток надруковано одну з букв: "а", "м", "р", "т", "ю". Картки ретельно перемішують. Яка ймовірність того, що з 4 послідовно взятих карток складуть слово "юрта"?

1.21. У лотереї 1 000 квитків. Із них 500 є виграшними. Куплено 2 квитки. Яка ймовірність того, що вони обидва виграшні?

1.22. У шухляді 15 деталей, серед яких 10 пофарбованих. Складальник навмання виймає 3 деталі. Яка ймовірність того, що вони виявляться пофарбованими?

1.23. У шухляді 100 деталей, із них 10 бракованих. Навмання вийнято 4 деталі. Знайдіть імовірність того, що серед них немає бракованих.

1.24. Набираючи номер телефону, абонент забув останні дві цифри, і, пам'ятаючи лише, що ці цифри різні, набрав їх навмання. Знайдіть імовірність того, що набрані цифри правильні.

1.25. У партії з 10 деталей 7 стандартних. Знайдіть імовірність того, що серед 6 взятих навмання деталей 4 стандартні.

1.26. У цеху працюють 6 чоловіків і 4 жінки. Навмання відібрано 7 осіб. Яка ймовірність того, що серед них виявляться 3 жінки?

1.27. У групі 12 студентів, із них 8 відмінників. За списком навмання відібрано 9 студентів. Яка ймовірність того, що серед них 5 відмінників?

1.28. На столі лежать 32 екзаменаційні квитки. Яка ймовірність того, що номер навмання взятого квитка буде кратним 7?

1.29. Яка ймовірність того, що в лютому навмання вибраного високосного року буде 5 субот?

1.30. В урні 10 куль: 6 білих, 4 чорні. Вийняли 2 кулі. Яка ймовірність того, що вони обидві білі?

1.31. Із розрізної абетки, що налічує 30 карток із різними буквами алфавіту, виймають 5 карток. Яка ймовірність того, що з 5 букв, розташованих у порядку появи, складуть слово "спорт"?

1.32. Множина A складається із простих додатних чисел, що не перевищують 20, множина B – із цілих додатних чисел, які не перевищують 50 і в записі яких міститься цифра 7. Серед множини чисел $A+B$ навмання вибирають одне. Визначте ймовірність того, що воно буде належати множині $A \setminus B$. Поясніть розрахунок діаграмою Венна – Ейлера.

1.33. Множина A складається з парних тризначних чисел, множина B – із тризначних чисел, які є паліндромами, тобто мають однакове значення, якщо їх читати справа наліво або зліва направо. Визначте ймовірність того, що вибране навмання будь-яке тризначне число буде належати обом множинам одночасно. Поясніть розв'язання діаграмою Венна – Ейлера.

1.34. Два студенти домовилися зустрітися між 18-ю та 19-ю годинами. Той, хто прийде першим, буде чекати другого чверть години, а потім піде. Визначте ймовірність того, що зустріч відбудеться, якщо кожний зі студентів може прийти в довільний момент часу з визначеного проміжку.

1.35. На відрізку AB числової осі, який має довжину L, навмання вибрали дві точки $C(x_1)$ та $D(x_2)$ такі, що $x_2 \geq x_1$. Знайдіть імовірність того, що довжина відрізка CD буде мати довжину, меншу за $0,5L$.

1.5. Тестові завдання

1.1. Кількість способів вибору k елементів із n без урахування порядка відбору дорівнює

А	Б	В	Г
$A_n^k = \dfrac{n!}{k!}$	$C_n^k = \dfrac{n!}{k!(n-k)!}$	$C_n^k = \dfrac{n!}{(n-k)!}$	$A_n^k = \dfrac{n!}{(n-k)!}$

1.2. Кількість способів вибору k елементів із n з урахуванням порядка відбору дорівнює

А	Б	В	Г
$A_n^k = \dfrac{n!}{k!}$	$C_n^k = \dfrac{n!}{k!(n-k)!}$	$C_n^k = \dfrac{n!}{(n-k)!}$	$A_n^k = \dfrac{n!}{(n-k)!}$

1.3. Подія – це

А імовірнісна закономірність

Б наслідок експерименту

В комплекс умов експерименту

Г зчисленна множина подій

Д інша відповідь

1.4. Простір елементарних подій – це

А множина всіх можливих наслідків експерименту

Б зчисленна множина подій

В множина сумісних подій

Г множина подій, які неможливо перелічити

Д інша відповідь

1.5. Імовірнісний простір елементарних подій позначають

А	**Б**	**В**	**Г**	**Д**
Ω	Σ	α	ω_i	інше

1.6. Експеримент – це

А простір елементарних подій

Б множина подій, що виключають появу інших

В випробування за певних умов

Г множина подій, які неможливо перелічити

1.7. Сумісні події

А поява однієї події виключає появу інших

Б подія, яка розкладається на елементарні події

В поява однієї не виключає можливості появи інших

Г множина подій, які неможливо перелічити

Д інша відповідь

1.8. Імовірність події – це

А показник випадковості

Б характеристика настання події

В числова міра, що характеризує ступінь об'єктивної можливості настання події

Г властивість настання випадкової події

Д інша відповідь

1.9. Для довільної випадкової події

А	Б	В	Г	Д
$P(A)<1$	$0<P(A)<1$	$P(A)\geq 0$	$0\leq P(A)\leq 1$	інше

1.10. Геометричну ймовірність обчислюють за формулою

А	Б	В	Г	Д
$P(A)=\dfrac{m(A)}{m(\Omega)}$	$m(\Omega)\cdot m(A)$	$P(A)=\dfrac{m(\Omega)}{m(A)}$	$m(A)$	інше

1.6. Запитання для самоперевірки

1.1. Дайте означення таких основних понять теорії ймовірностей: "випробування", "випадкова подія", "імовірність випадкової події".

1.2. Назвіть аксіоми, на яких ґрунтується теорія ймовірностей.

1.3. Сформулюйте властивості, притаманні ймовірності події.

1.4. Дайте визначення вірогідної, неможливої та випадкової подій.

1.5. Назвіть критерій, за яким визначають малоймовірні події.

1.6. Сформулюйте критерій, за яким розрізняють сумісні й несумісні події. Наведіть приклад їхнього зображення за допомогою діаграми Венна – Ейлера.

1.7. Дайте означення повної групи несумісних випадкових подій.

1.8. Поясніть, чи відрізняється повна група несумісних випадкових подій від простору елементарних подій.

1.9. Дайте означення суми випадкових подій.

1.10. Наведіть приклад зображення суми несумісних подій за допомогою діаграми Венна – Ейлера.

1.11. Наведіть приклад зображення суми сумісних подій за допомогою діаграми Венна – Ейлера.

1.12. Дайте означення добутку випадкових подій. Наведіть приклад його зображення за допомогою діаграми Венна – Ейлера.

1.13. Дайте означення різниці двох випадкових подій. Наведіть приклад її зображення за допомогою діаграми Венна – Ейлера.

1.14. Поясніть, чому існує кілька означень імовірності випадкової події.

1.15. Дайте означення статистичної ймовірності.

1.16. Дайте означення класичної ймовірності. Поясніть це тлумачення за допомогою діаграми Венна – Ейлера.

1.17. Дайте означення геометричної ймовірності. Поясніть це означення за допомогою геометричної побудови.

1.18. Визначте межі використання класичного означення ймовірності.

1.19. Поясніть, яких умов необхідно дотримуватися під час побудови діаграм Венна – Ейлера для того, щоб їх можна було застосовувати для безпосереднього обчислення ймовірності.

1.7. Висновки за темою

Після вивчення теми студент має:

вільно володіти формулами комбінаторики та розуміти, у яких випадках треба використовувати той чи той вид сполучень;

уміти розв'язувати задачі, використовуючи класичне та геометричне визначення ймовірності;

уміти визначати умови застосування аксіом, на яких побудовано теорію ймовірностей;

надавати геометричну інтерпретацію випадкових подій та дій над ними за допомогою діаграм Венна – Ейлера.

2. Основні теореми теорії ймовірностей, їхня економічна інтерпретація

2.1. Мета та компетентності

Метою вивчення теми є формування системи теоретичних знань, практичних умінь і навичок щодо основних теорем теорії ймовірностей та умов їхнього використання для обчислення ймовірності випадкових подій.

Компетентності, що формують під час вивчення теми:

знання сутності теорем множення та додавання ймовірностей;

знання означення умовної ймовірності;

знання означення повної ймовірності події та апостеріорної ймовірності гіпотез за формулою Байєса;

уміння застосовувати основні формули комбінаторики та теореми теорії ймовірностей для обчислення ймовірності випадкової події;

уміння здійснювати класифікацію умов застосування теорем та аксіом теорії ймовірностей і ухвалювати рішення щодо застосування певного підходу до обчислення ймовірності.

2.2. Основні теоретичні відомості

Під час побудови будь-якої теорії після того, як названі об'єкти дослідження, їхні основні відношення, а також аксіоми, яким підлягають ці об'єкти та їхні відношення, можна переходити до розгляду теорем.

Слід розглянути основні теореми теорії ймовірностей.

Теореми множення ймовірностей

Імовірність одночасної появи двох подій (імовірність добутку подій) дорівнює добутку ймовірності появи однієї з подій на умовну ймовірність іншої події у припущенні, що перша подія вже відбулася:

$$P(AB) = P(A) \cdot P(B \mid A),$$

де $P(A)$ – імовірність події A (безумовна ймовірність події A);

$P(B \mid A)$ – імовірність появи події B за умови, що подія A вже відбулася (умовна ймовірність події B).

Або $$P(AB) = P(B) \cdot P(A \mid B).$$

Якщо ймовірність появи події B не залежить від того, відбулася чи ні подія A, то такі події є взаємно незалежними, і для кожної випадкової події ймовірність її появи за умови, що перша з подій уже відбулася, дорівнює безумовній імовірності: $P(A \mid B) = P(A)$ та $P(B \mid A) = P(B)$ і теорема множення ймовірностей набирає такої форми: $P(AB) = P(A) \cdot P(B)$.

Наслідком теореми множення ймовірностей є теорема, що дозволяє визначити ймовірність появи хоча б однієї з випадкових подій.

Теорема. Імовірність появи хоча б однієї з випадкових подій $A_1, A_2, \ldots, A_n$, які є незалежними в сукупності, дорівнює різниці між одиницею та добутком імовірностей протилежних подій $\overline{S_1}, \overline{S_2}, ..., \overline{S_n}$.

Якщо $P(A_i) = p_i$ $(i = \overline{1,n})$, то, відповідно, імовірність появи протилежної події позначають $P(\overline{S_i}) = 1 - p_i = q_i$. Згідно із цими позначеннями теорема про ймовірність появи хоча б однієї з випадкових подій набирає такої форми:

$$P(A) = 1 - q_1 \cdot q_2 \cdot ... \cdot q_n.$$

Окремим випадком теореми про визначення ймовірності появи хоча б однієї події є формула для обчислення ймовірності появи хоча б однієї події в разі, якщо для всіх подій, що розглядають, імовірність їхньої появи є однаковою.

Нехай для кожної з випадкових подій $A_1, A_2, \ldots, A_n$ ймовірність появи становить $P\left(A_i\right) = p$. Тоді для протилежної події $P(\overline{S}) = 1 - p = q$ і $P(A) = 1 - q^n$.

Теореми додавання ймовірностей

Імовірність суми несумісних випадкових подій A та B дорівнює сумі їхніх безумовних імовірностей:

$$P\left(A + B\right) = P\left(A\right) + P\left(B\right).$$

Імовірність появи однієї із двох сумісних подій дорівнює сумі їхніх імовірностей за виключенням імовірності їхнього добутку:

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

Формула повної ймовірності

Нехай подія A може наступати лише за умови появи однієї з несумісних випадкових подій (гіпотез) $B_1, B_2, \ldots, B_n$, що утворюють повну групу подій, і ймовірність кожної з подій цієї групи $P\left(B_i\right)$ відома, а також відомі умовні ймовірності події A в разі, якщо реалізують певну гіпотезу, тобто $P\left(A \middle| B_i\right)$, де $i = \overline{1,n}$. Тоді ймовірність події A визначають за такою формулою:

$$P\left(A\right) = P\left(B_1\right)P\left(A \middle| B_1\right) + P\left(B_2\right)P\left(A \middle| B_2\right) + \ldots + P\left(B_n\right)P\left(A \middle| B_n\right).$$

Формула Байєса

Нехай подія A настає лише за умови появи однієї з несумісних випадкових подій (гіпотез) $B_1, B_2, \ldots, B_n$, що утворюють повну групу випадкових подій. Якщо припустити, що подія A вже відбулася, і необхідно визначити ймовірність того, що подія A відбулася саме завдяки реалізації гіпотези B_i, тоді:

$$P\left(B_i \middle| A\right) = \frac{P\left(B_i\right) \cdot P\left(A \middle| B_i\right)}{\sum\limits_{i=1}^{n} P\left(B_i\right) \cdot P\left(A \middle| B_i\right)}, \ i = \overline{1,n}.$$

2.3. Приклади розв'язування задач

Приклад 2.1. Імовірність складання іспиту студентом на "5" дорівнює 0,3; на "4" – 0,45; на "2" – 0,1; що він не з'явиться на іспит – 0,05. Яка ймовірність того, що студент отримає позитивну оцінку?

Розв'язання. Очевидно, щоб отримати позитивну оцінку, студент має скласти іспит на "5", "4" або на "3". Імовірність отримати "3" дорівнює $P = 1 - \left(0,3 + 0,45 + 0,1 + 0,05\right) = 0,1$. За теоремою додавання ймовірностей імовірність отримання позитивної оцінки на іспиті дорівнює $P = 0,3 + 0,45 + 0,1 = 0,85$.

Приклад 2.2. Імовірність улучення в мішень за одного пострілу першим стрільцем дорівнює 0,8, а другим – 0,9. Знайдіть імовірність того що: а) обидва стрільці влучать у мішень; б) тільки перший стрілець улучить у мішень; в) тільки другий стрілець улучить у мішень; г) один стрілець улучить у мішень; д) жоден стрілець не влучить у мішень; е) хоча б один стрілець улучить у мішень.

Розв'язання. Нехай подія A – це ціль, уражена першим стрільцем, а подія B – ціль, уражена другим стрільцем.

За умовою $P\left(A\right) = 0,8$, $P\left(B\right) = 0,9$, тоді ймовірності протилежних подій дорівнюють $P\left(\overline{A}\right) = 0,2$; $P\left(\overline{B}\right) = 0,1$. Отже:

а) $P\left(A \text{ та } B\right) = P\left(AB\right) = P\left(A\right) \cdot P\left(B\right) = 0,8 \cdot 0,9 = 0,72$;

б) $P\left(A \text{ та } \overline{B}\right) = P\left(A\overline{B}\right) = P\left(A\right) \cdot P\left(\overline{B}\right) = 0,8 \cdot 0,1 = 0,08$;

в) $P\left(\overline{A} \text{ та } B\right) = P\left(\overline{A}B\right) = P\left(\overline{A}\right) \cdot P\left(B\right) = 0,2 \cdot 0,9 = 0,18$;

г) $P\left(\bar{A}\cdot B \text{ або } A\cdot\bar{B}\right)=P\left(\bar{A}\cdot B\right)+P\left(A\cdot\bar{B}\right)=0{,}18+0{,}08=0{,}26$;

д) $P\left(\bar{A} \text{ та } \bar{B}\right)=P\left(\bar{A}\cdot\bar{B}\right)=P\left(\bar{A}\right)\cdot P\left(\bar{B}\right)=0{,}2\cdot 0{,}1=0{,}02$;

е) $P\left(\text{хоча б один}\right)=1-P\left(\bar{A}\cdot\bar{B}\right)=1-0{,}02=0{,}98$.

Приклад 2.3. Для деякої місцевості середня кількість ясних днів у липні дорівнює 25. Яка ймовірність того, що перші два дні липня будуть ясними?

Розв'язання. Імовірність того, що першого липня буде ясний день (подія A), дорівнює $P(A)=\frac{25}{31}$. Імовірність того, що другого липня буде ясний день (подія B) за умови, що першого липня був ясний день, тобто умовна ймовірність події B дорівнює $P\left(B|A\right)=\frac{24}{30}=\frac{4}{5}$.

Шукана ймовірність за теоремою множення ймовірностей залежних подій дорівнює $P\left(AB\right)=P\left(A\right)\cdot P\left(B|A\right)=\frac{25}{31}\cdot\frac{4}{5}=\frac{20}{31}$.

Приклад 2.4. Імовірність того, що подія з'явиться хоча б один раз у трьох незалежних випробуваннях, дорівнює 0,936. Знайдіть імовірність появи події в одному випробуванні.

Розв'язання. Імовірність появи події хоча б один раз у трьох незалежних випробуваннях дорівнює $P\left(A\right)=1-q^3$, де $q=1-p$.

За умовою $P\left(A\right)=0{,}936$.

Тоді $0{,}936=1-q^3$, звідки $q^3=0{,}064$, а $q=0{,}4$ і шукана ймовірність $p=1-0{,}4=0{,}6$.

Приклад 2.5. Імовірність того, що деталь виявиться нестандартною, дорівнює 0,1. Визначте, яку кількість деталей необхідно буде взяти, щоб з імовірністю 0,998 можна було стверджувати, що хоча б одна з них буде стандартною.

Розв'язання. Нехай n – кількість деталей. Тоді ймовірність того, що всі деталі нестандартні, дорівнює $\left(0{,}1\right)^n$. Імовірність наявності хоча б однієї стандартної деталі дорівнює $1-\left(0{,}1\right)^n$ і дорівнює за умовою 0,998, звідки $1-\left(0{,}1\right)^n=0{,}998 \Rightarrow \left(0{,}1\right)^n=0{,}002 \Rightarrow n=2{,}7\approx 3$ (n – ціле число).

Приклад 2.6. На складання надходять деталі із трьох автоматів. Перший автомат дає 20 %, другий – 30 %, третій – 50 % усіх деталей, що надходять на складання. Перший автомат дає 0,2 % браку, другий – 0,3 %, третій – 0,1 %. Знайдіть імовірність того, що до складального цеху надійшла бракована деталь.

Розв'язання. Позначте через B_1 подію "деталь виготовлено першим автоматом", B_2 – "деталь виготовлено другим автоматом", B_3 – "деталь виготовлено третім автоматом". Тоді за умовою задачі:

$$P(B_1)=0,2;\ P(B_2)=0,3;\ \ P(B_3)=0,5.$$

Перевірка: $0,2+0,3+0,5=1$.

Нехай подія A – "на складання надійшла бракована деталь".

Тоді $P(A|B_1)=0,002$; $P(A|B_2)=0,003$; $P(A|B_3)=0,001$.

За формулою повної ймовірності буде:

$$P(A)=P(B_1)P(A|B_1)+P(B_2)P(A|B_2)+P(B_3)P(A|B_3).$$

$$P(A)=0,2\cdot 0,002+0,3\cdot 0,003+0,5\cdot 0,001=0,0018.$$

Приклад 2.7. У першій урні 5 білих і 10 чорних куль, у другій – 3 білі і 7 чорних. Із другої урни до першої переклали 1 кулю, а потім із першої урни навмання вийняли 1 кулю. Яка ймовірність того, що вона біла?

Розв'язання. Можливі дві гіпотези: B_1 – "перекладено білу кулю"; B_2 – "перекладено чорну кулю". Тоді:

$$P(B_1)=\frac{3}{10};\ \ P(B_2)=\frac{7}{10}.$$

Перевірка: $\frac{3}{10}+\frac{7}{10}=1$.

Нехай подія A – "вийнято білу кулю з першої урни" після перекладання до неї кулі з другої урни. Очевидно, що:

$$P(A|B_1)=\frac{6}{16},\ P(A|B_2)=\frac{5}{16}.$$

За формулою повної ймовірності буде:

$$P(A)=\frac{3}{10}\cdot\frac{6}{16}+\frac{7}{10}\cdot\frac{5}{16}=\frac{53}{160}.$$

Приклад 2.8. У першій шухляді 12 свердел, із них 8 використаних, а у другій шухляді 8 свердел, із них 4 використані. Із першої шухляди навмання перекладають до другої 2 свердла. Знайдіть імовірність того, що витягнуте після цього із другої шухляди свердло ще жодного разу не використовували.

Розв'язання. Можливі три гіпотези: B_1 – "із першої шухляди до другої перекладено 2 невикористані свердла"; B_2 – "із першої шухляди до другої перекладено 1 використане й 1 невикористане свердло"; B_3 – "із першої шухляди до другої перекладено 2 використані свердла". Імовірності цих гіпотез дорівнюють:

$$P(B_1)=\frac{4}{12}\cdot\frac{3}{11}=\frac{1}{11},\ P(B_2)=\frac{4}{12}\cdot\frac{8}{11}+\frac{8}{12}\cdot\frac{4}{11}=\frac{16}{33},\ P(B_3)=\frac{8}{12}\cdot\frac{7}{11}=\frac{14}{33}.$$

Перевірка: $P(B_1)+P(B_2)+P(B_3)=\frac{1}{11}+\frac{16}{33}+\frac{14}{33}=\frac{3+16+14}{33}=1.$

Нехай подія A – "свердло, що витягнули із другої шухляди після перекладання, виявиться жодного разу не використаним". Тоді умовні ймовірності події A за умови, що відбулися події B_1, B_2, B_3, відповідно, дорівнюють:

$$P(A|B_1)=\frac{6}{10},\ P(A|B_2)=\frac{5}{10},\ P(A|B_3)=\frac{4}{10}.$$

За формулою повної ймовірності буде:

$$P(A)=\frac{1}{11}\cdot\frac{6}{10}+\frac{16}{33}\cdot\frac{5}{10}+\frac{14}{33}\cdot\frac{4}{10}=\frac{77}{165}.$$

Приклад 2.9. У першій шухляді 20 білих куль, у другій – 10 білих і 10 чорних куль, у третій – 20 чорних куль. Із вибраної навмання шухляди вийняли кулю. Вона виявилася білою. Яка ймовірність того, що ця куля з першої шухляди?

Розв'язання. Нехай гіпотеза B_1 – "кулю вийнято з першої шухляди", B_2 – "кулю вийнято із другої шухляди", B_3 – "кулю вийнято із третьої шухляди". Тоді, за умовою, що ці події рівноймовірні, буде:

$$P(B_1)=P(B_2)=P(B_3)=\frac{1}{3}.$$

Подія A – "вийнято білу кулю", тоді умовні ймовірності дорівнюють:

$$P\left(A\middle|B_1\right)=1,\ P\left(A\middle|B_2\right)=\frac{10}{20}=\frac{1}{2},\ P\left(A\middle|B_3\right)=0.$$

За формулою повної ймовірності буде:

$$P(A)=\frac{1}{3}\cdot 1+\frac{1}{3}\cdot\frac{1}{2}+\frac{1}{3}\cdot 0=\frac{1}{2}.$$

Шукану ймовірність $P\left(B_1\middle|A\right)$ слід знаходити за формулою Байєса:

$$P\left(B_1\middle|A\right)=\frac{P(B)P\left(A\middle|B_1\right)}{P(A)}=\frac{\frac{1}{3}\cdot 1}{\frac{1}{2}}=\frac{2}{3}.$$

Як видно, після переоцінки, тобто після того, як було враховано факт появи події $A,$ імовірність гіпотези B_1 зросла удвічі.

2.4. Вправи для самостійної роботи

2.1. В урні 10 білих, 15 червоних, 20 синіх і 25 чорних куль. Вийняли одну кулю. Яка ймовірність, що вийнята куля мала колір: а) білий або синій; б) синій або червоний; в) білий, чорний або синій?

2.2. У шухляді 6 білих і 8 чорних куль. Узяли 2 кулі. Яка ймовірність того, що вони обидві білі?

2.3. Студент знає 20 питань із 25. Знайдіть імовірність, що він відповість на три питання.

2.4. Серед виробів, виготовлених на окремому заводі, 98 % відповідають стандарту. Із них 60 % – першого ґатунку і 40 % – другого. Узято один виріб. Яка ймовірність того, що взятий виріб виявиться виробом першого ґатунку?

2.5. На 10 картках надруковано цифри 0, 1, 2, ..., 9. Знайдіть імовірність того, що 3 навмання взяті й поставлені в ряд картки становлять число 125.

2.6. У першій шухляді 2 білі й 10 чорних куль, у другій – 8 білих і 2 чорні. Із кожної шухляди взяли по кулі. Яка ймовірність того, що: а) обидві кулі білі; б) одна куля біла, а друга – чорна; в) кулі одного кольору?

2.7. Імовірність виграшу одного квитка лотереї дорівнює 1/20. Яка ймовірність того, що маючи 5 квитків, можна виграти: а) за всіма квитками; б) за жодним із квитків; в) хоча б за одним квитком?

2.8. Білет містить три питання. Імовірність того, що студент відповість на перше і друге питання дорівнює 0,9; на третє – 0,8. Знайдіть імовірність того, що студент складе іспит, якщо для цього необхідно відповісти: а) на всі питання; б) хоча б на 2 питання.

2.9. Імовірність улучення в ціль за одного пострілу дорівнює 0,2. Зроблено 10 пострілів. Знайдіть імовірність улучення в ціль.

2.10. У майстерні стоять 4 токарські верстати. Імовірність того, що в заданий момент кожен із верстатів працює, дорівнює 0,7. Визначте ймовірність того, що в заданий момент працює хоча б один верстат.

2.11. Імовірність улучення в ціль за одного залпу із двох гармат дорівнює 0,38. Знайдіть імовірність улучення в ціль за одного залпу першої гармати, якщо для другої гармати ця ймовірність дорівнює 0,8.

2.12. Вироби перевіряються експертом на нестандартність. Імовірність того, що виріб стандартний, дорівнює 0,9. Знайдіть імовірність того, що із двох перевірених виробів тільки один стандартний.

2.13. У шухляді 10 куль, серед яких 6 білих. Навмання витягають 4 кулі. Яка ймовірність того, що всі кулі виявляться білими?

2.14. Імовірність хоча б одного влучення стрільцем за трьох пострілів дорівнює 0,875. Знайдіть імовірність улучення за одного пострілу.

2.15. Імовірність улучення в ціль для окремого стрільця дорівнює 60 %. Зроблено 5 пострілів. Яка ймовірність того, що: а) усі 5 пострілів виявляться вдалими; б) стрілець жодного разу не влучить у ціль; в) стрілець улучить у ціль хоча б один раз; г) скільки пострілів достатньо зробити, щоб імовірність улучення в ціль досягла 95 %?

2.16. Є 4 урни. У першій урні 1 біла і 1 чорна куля, у другій – 2 білі та 3 чорні кулі, у третій – 3 білі й 5 чорних, у четвертій – 4 білі та 7 чорних куль. Вибирають навмання урну та виймають із неї 1 кулю. Яка ймовірність того, що вийнята куля біла?

2.17. Дві однакові на вигляд коробки містять: одна – 3 червоні й 2 чорні олівці, друга – 2 червоні та 4 чорні олівці. Із навмання взятої коробки виймають 1 олівець. Яка ймовірність того, що він виявиться чорним?

2.18. На заводі болти виготовляють на 3 верстатах, що виробляють, відповідно, 25, 30 та 45 % усієї кількості болтів. У продукції верстатів брак становить 4, 3 і 2 %, відповідно. Яка ймовірність того, що болт, випадково взятий з усієї продукції, що надійшла, виявиться бракованим?

2.19. На складання надходять деталі із трьох автоматів. Перший автомат дає 0,3 % браку, другий – 0,2 % браку, третій – 0,4 % браку.

Знайдіть імовірність потрапляння на складання бракованої деталі, якщо з першого автомата надходить 1 000 деталей, із другого – 2 000 деталей, із третього – 2 500 деталей.

2.20. Під час розриву снаряда утворюються великі, середні та дрібні осколки у відношенні 1 : 3 : 6. У разі влучення в танк великий осколок пробиває броню з імовірністю 0,9; середній – 0,3; дрібний – 0,1. Яка ймовірність того, що осколок, потрапивши у броню, проб'є її?

2.21. З урни, що містить 4 білі та 2 чорні кулі, перекладено 1 кулю в урну, що містить 5 білих і 8 чорних куль. Після цього із другої урни було взято 1 кулю. Яка ймовірність того, що вона виявиться чорною?

2.22. У відділі А інституту працюють 5 інженерів і 3 старші інженери, а у відділі В – 8 інженерів і 2 старші інженери. Із відділу А у відділ В перевели 1 співробітника. Знайдіть імовірність того, що 3 співробітники, випадково вибрані з нового складу відділу А, є старшими інженерами.

2.23. У першій урні 10 куль, із них 2 чорні; у другій урні 20 куль, із них 16 чорних. Із кожної урни витягли по 1 кулі, а потім із цих 2 куль навмання взяли одну. Яка ймовірність того, що вона біла?

2.24. У першій бригаді 5 робітників мають I розряд і 4 – II розряд. У другій бригаді 6 робітників мають I розряд і 7 – II розряд. Із першої бригади до другої переведено 2 робітників. Знайдіть імовірність того, що 2 робітників, навмання узятих із нового складу другої бригади, мають I розряд?

2.25. У партії 600 лампочок: 200 виготовлено на першому заводі, 250 – на другому, 150 – на третьому. Імовірність того, що лампочка є стандартною, для першого заводу дорівнює 0,97; для другого – 0,91; для третього – 0,93. Навмання взята 1 лампочка, що виявилася стандартною. Яка ймовірність того, що її виготовлено на першому заводі?

2.26. Серед деталей, що надходять до складального цеху, із першого автомата 0,1 % бракованих, із другого – 0,2 %, із третього – 0,25 %, із четвертого – 0,5 %. Продуктивність їх відноситься як 4 : 3 : 2 : 1, відповідно. Узята навмання деталь виявилася бракованою. Яка ймовірність того, що її виготовлено: а) на першому автоматі; б) на другому автоматі; в) на третьому автоматі; г) на четвертому автоматі? Як перевірити правильність обчислень?

2.27. Є 5 урн: дві з них містять по 2 білі та 3 чорні кулі, дві – по 1 білій і 3 чорні кулі й одна – 4 білі та 1 чорну кулю. Із навмання взятої урни вийняли кулю. Вона виявилася білою. Яка ймовірність того, що кулю вийняли з урни з 4 білими й 1 чорною кулею?

2.28. Група складається із 2 відмінних, 2 гарних і 4 посередніх стрільців. Імовірність улучення в мішень (за одного пострілу) для відмінного, гарного та посереднього стрільців дорівнює, відповідно, 0,9; 0,8 і 0,7. Навмання вибраний стрілець вистрілив двічі; було зазначено 1 влучення й 1 промах. Яким найімовірніше був цей стрілець: відмінним, гарним чи посереднім?

2.5. Тестові завдання

2.1. Формула додавання ймовірностей несумісних подій має такий вигляд

А $P(A+B)=P(A)+P(B)$

Б $P(A+B)=P(A)+P(B)-P(AB)$

В $P(A\cdot B)=P(A)\cdot P(B)$

Г інша відповідь

2.2. Формула додавання ймовірностей сумісних подій має такий вигляд

А $P(A+B)=P(A)+P(B)$

Б $P(A+B)=P(A)+P(B)-P(AB)$

В інша відповідь

2.3. Дві події називають протилежними, якщо

А вони несумісні

Б вони сумісні та становлять повну групу

В вони несумісні та становлять повну групу

2.4. Дві події називають незалежними, якщо

А імовірність сумісної появи цих подій дорівнює добутку ймовірності однієї з них та умовної ймовірності другої

Б імовірність появи однієї з таких подій дорівнює сумі їхніх ймовірностей

В імовірність однієї з них не залежить від появи другої

Г імовірність однієї з них змінюється під час появи другої

2.5. Імовірність сумісної появи двох залежних подій дорівнює

А	Б	В
$P(AB) = P(A) \cdot P\left(B\middle|A\right)$	$P\left(AB\right) = P\left(A\right) + P\left(B\right)$	$P(AB) = P(A) \cdot P\left(B\right)$

2.6. Імовірність появи хоча б однієї з випадкових подій обчислюють за такою формулою

А $P(A) = 1 - q_1 \cdot q_2 \cdot ... \cdot q_n$

Б $P\left(A\right) = 1 + q_1 \cdot q_2 \cdot \ldots \cdot q_n$

В $P\left(A\right) = 1 - q_1 \cdot q_2$

2.7. Формулу повної ймовірності записують у такому вигляді

А $P\left(A\right) = P\left(B_i\right) \cdot P\left(A\middle|B_i\right),\ i = \overline{1,n}$

Б $P\left(A\right) = \sum\limits_{i=1}^{n} P\left(B_i\right) \cdot P\left(A\middle|B_i\right),\ i = \overline{1,n}$

В $P\left(A\right) = \sum\limits_{i=1}^{n} P\left(B_i\right) \cdot P\left(B_i\middle|A\right),\ i = \overline{1,n}$

Г інша відповідь

2.8. Формула Байєса має такий вигляд

А $P\left(B_i\middle|A\right) = \dfrac{\sum\limits_{i=1}^{n} P(B_i) \cdot P(A|B_i)}{P(B_i) \cdot P(A|B_i)},\ i = \overline{1,n}$

Б $P\left(B_i\middle|A\right) = \dfrac{P\left(B_i\right) \cdot P\left(A\middle|B_i\right)}{\sum\limits_{i=1}^{n} P\left(B_i\right) \cdot P\left(A\middle|B_i\right)},\ i = \overline{1,n}$

В $P\left(A\right) = \sum\limits_{i=1}^{n} P\left(B_i\right) \cdot P\left(A\middle|B_i\right),\ i = \overline{1,n}$

Г інша відповідь

2.9. Формули Байєса дозволяють

А переоцінити ймовірність гіпотези за умови, що випадкова подія A здійснилася

Б переоцінити ймовірність гіпотези за умови, що випадкова подія A ще не здійснилася

В інше

2.10. Після переоцінювання всіх гіпотез $\sum_{i=1}^{n} P\left(B_i \middle| A\right)$ дорівнює

А	Б	В	Г
0	1	2	інша відповідь

2.6. Запитання для самоперевірки

2.1. Наведіть означення умовної ймовірності та поясніть, як її визначають.

2.2. Поясніть, чим відрізняється умовна ймовірність від безумовної.

2.3. Сформулюйте теорему множення ймовірностей.

2.4. Наведіть формулу для визначення ймовірності добутку двох сумісних подій.

2.5. Сформулюйте теорему додавання ймовірностей.

2.6. Визначте, чи можна формулу для додавання ймовірностей несумісних випадкових подій уважати теоремою, та обґрунтуйте висновок.

2.7. Наведіть формулу для визначення суми трьох сумісних подій і розкрийте зміст кожної події, що входить до складу цієї суми.

2.8. За формулою визначення ймовірності появи хоча б однієї події поясніть, які події слід було б ураховувати як сприятливі.

2.9. Поясніть, які події під час визначення повної ймовірності вважають гіпотезами. Яким вимогам відповідають гіпотези?

2.10. Наведіть формулу повної ймовірності.

2.11. Назвіть теоретичні положення, на базі яких виводять формулу повної ймовірності.

2.12. Поясніть різницю між апріорною та апостеріорною ймовірностями. Як ви вважаєте, чи завжди їхні значення мають відрізнятися одне від одного?

2.13. Сформулюйте теорему гіпотез (для визначення апостеріорної ймовірності).

2.14. Доведіть, що сума апостеріорних імовірностей має дорівнювати одиниці.

2.15. Поясніть, за яким принципом будують дерево станів, яку інформацію воно відображає і як за його допомогою знаходять розв'язок задач щодо визначення повної ймовірності події.

2.7. Висновки за темою

Після вивчення теми студент має:

знати умови використання теорем додавання ймовірностей;

знати умови використання теорем множення ймовірностей;

уміти розв'язувати задачі, використовуючи теореми додавання та множення ймовірностей,

чітко розрізняти випадки, коли треба застосовувати формулу повної ймовірності та формули Байєса;

уміти переоцінити ймовірність появи події після того, як стає відомим результат випробування;

уміти зробити перевірку результатів розрахунків;

уміти обчислювати ймовірність того, що випадкова подія відбудеться хоча б один раз.

3. Схема незалежних випробувань

3.1. Мета та компетентності

Метою вивчення теми є надання уявлення про повторні випробування за схемою Бернуллі та набуття навичок в обчисленні ймовірності появи події в цих випробуваннях певну кількість разів.

Компетентності, що формують під час вивчення теми:

уміння здійснювати обчислення за формулою Бернуллі та застосовувати довідкові таблиці значень функцій Гаусса та Лапласа для визначення ймовірностей за асимптотичними теоремами;

застосовування формули Бернуллі, локальної та інтегральної теореми Муавра – Лапласа, теореми Пуассона для обчислення ймовірності появи випадкової події певну кількість разів під час випробувань за схемою Бернуллі;

здатність здійснювати аналіз умов, у яких проводять повторні випробування, та рекомендувати використання формули Бернуллі та асимптотичних теорем для обчислення ймовірності того, що випадкова подія відбудеться певну кількість разів у процесі незалежних випробувань.

3.2. Основні теоретичні відомості

Формула Бернуллі

Якщо ймовірність появи події A в кожному з n незалежних випробувань постійна і дорівнює P, то ймовірність $P_n(m)$ того, що подія з'явиться саме m разів, знаходять за формулою Бернуллі:

$$P_n(m) = C_n^m p^m q^{n-m},$$

де $q = 1 - p$.

Імовірність появи події A принаймні один раз під час проведення n незалежних випробувань за схемою Бернуллі, дорівнює

$$P_n(m \geq 1) = 1 - q^n,$$

де $q = 1 - p$.

Імовірність того, що подія A під час проведення n незалежних випробувань за схемою Бернуллі, з'явиться не менше ніж m_1 і не більше ніж m_2 разів обчислюють за такою формулою:

$$P_n(m_1 \leq m \leq m_2) = \sum_{m_1}^{m_2} P_n(m).$$

Число m_0 появи події A в n незалежних випробуваннях називають *найімовірнішим*, якщо ймовірність появи події A є найбільшою. Це число знаходять із такої нерівності:

$$np - q \leq m_0 \leq np + p.$$

Якщо ймовірність $P(A)$ появи події A в кожному з n незалежних випробувань постійна і дорівнює p, то можна знайти ймовірність $P_n(m)$ того, що подія A настане саме m разів.

Формула Пуассона

У разі, якщо $n \to \infty$, а $p \to 0$ для обчислення $P_n(m)$ застосовують асимптотичну формулу Пуассона:

$$P_n(m) \approx \frac{\lambda^m}{m!} e^{-\lambda},$$

де $\lambda = np$.

Значення функції $P_n(m) \approx \dfrac{\lambda^m}{m!} e^{-\lambda}$ для різних значень $\lambda = np$ наведено в додатку Б.

Рекомендують застосувати формулу Пуассона, якщо $\lambda \leq 10$.

Локальна теорема Лапласа

Якщо ймовірність p появи події A в кожному з n незалежних випробуваннях постійна, а число n достатньо велике, то ймовірність $P_n(m)$ того, що подія A відбувається m разів у цих випробуваннях, обчислюють за такою формулою:

$$P_n(m) \approx \frac{1}{\sqrt{npq}} \varphi(x),$$

де $x = \dfrac{m - np}{\sqrt{npq}}$, а $\varphi(x) = \dfrac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$ – функція Гаусса.

Таблицю значень функції $\varphi(x)$ наведено в додатку А, причому $\varphi(-x) = \varphi(x)$. Треба вважати, що для значень $x > 4$ $\varphi(x) \approx 0$.

За допомогою *інтегральної теореми Лапласа* знаходять наближено ймовірність того, що в n незалежних випробуваннях подія A відбувається від m_1 до m_2 разів (включно):

$$P_n(m_1 \leq m \leq m_2) \approx \Phi(x_2) - \Phi(x_1),$$

де $x_1 = \frac{m_1 - np}{\sqrt{npq}}$; $x_2 = \frac{m_2 - np}{\sqrt{npq}}$, а $\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_0^x e^{-\frac{t^2}{2}} dt$.

Значення функції Лапласа $\Phi(x)$ наведено в додатку В, причому $\Phi(-x) = -\Phi(x)$. Треба вважати, що для $x > 5$ $\Phi(x) \approx 0,5$.

Інтегральна теорема Лапласа дає незначну похибку, якщо $npq \geq 20$. Інакше треба застосовувати формулу Бернуллі та теорему додавання ймовірностей, а якщо p або q наближається до нуля, то застосовують наближену формулу Пуассона.

Наслідком інтегральної формули Лапласа є формула ймовірності відхилення відносної частоти $\frac{m}{n}$ появи події A в n незалежних випробуваннях від її ймовірності p на величину $\varepsilon > 0$ (за абсолютною величиною):

$$P\left(\left|\frac{m}{n} - p\right| \leq \varepsilon\right) \approx 2\Phi(t),$$

де $t = \varepsilon\sqrt{\frac{n}{pq}}$.

Нерівність $\left|\frac{m}{n} - p\right| \leq \varepsilon$ є рівносильною такій нерівності:

$$p - \varepsilon \leq \frac{m}{n} \leq p + \varepsilon \text{ або } n(p - \varepsilon) < m \leq n(p + \varepsilon).$$

За допомогою формули $P\left(\left|\frac{m}{n} - p\right| \leq \varepsilon\right) \approx 2\Phi(t)$ можна також знаходити кількість випробувань n, за яких із заданою ймовірністю p відхилення відносної частоти $\frac{m}{n}$ події A від імовірності p її появи за абсолютною величиною не буде перевищувати задане додатне число ε:

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

Із тієї самої формули можна за заданої ймовірності p і кількості випробувань n знаходити похибку ε:

$$\varepsilon = t\sqrt{\frac{pq}{n}}.$$

Формула $p-\varepsilon \le \frac{m}{m} \le p+\varepsilon$ означає, що відносна частота $\frac{m}{n}$ перебуває в симетричному інтервалі. Якщо частота $\frac{m}{n}$ події A перебуває в межах від α до β (включно), то:

$$P_n(\alpha \le \frac{m}{n} \le \beta) \approx \Phi(x_2) - \Phi(x_1),$$

де $x_1 = \frac{(\alpha - p)\sqrt{n}}{\sqrt{pq}}$; $\quad x_2 = \frac{(\beta - p)\sqrt{n}}{\sqrt{pq}}$.

3.3. Приклади розв'язування задач

Приклад 3.1. Схожість насіння заданої рослини становить 90 %. Знайдіть імовірність того, що з 4 посіяних насінин зійдуть: а) 3; б) не менше від 3.

Розв'язання:

а) за умовою $n=4$; $m=3$; $p=0,9$. За формулою Бернуллі:

$$P_4(3) = C_4^3 \cdot 0,9^3 \cdot 0,1 \approx 0,2916;$$

б) слід визначити ймовірність події, яка полягає в тому, що з 4 насінин зійде 3 або 4. Шукана ймовірність буде дорівнювати:

$$P_4(3) + P_4(4) \approx 0,2916 + 0,9^4 \approx 0,9477.$$

Приклад 3.2. Імовірність виготовлення нестандартної деталі дорівнює 0,05. Знайдіть імовірність того, що з 5 взятих навмання деталей 4 є стандартними.

Розв'язання. За умовою $n=5; m=4; p=0,95; q=0,05$.

Імовірність того, що з 5 взятих навмання деталей 4 є стандартними $P_5(4)$ можна обчислити за формулою Бернуллі:

$$P_5(4) = C_5^4 \cdot 0,95^4 \cdot 0,05 \approx 0,2036.$$

Приклад 3.3. Імовірність виробництва стандартної деталі дорівнює 0,98. Знайдіть найімовірніше число стандартних деталей серед 625.

Розв'язання. Якщо $n = 625;\ p = 0,98;\ q = 0,02$, то

$$612,48 \le m_0 \le 613,48,$$

звідки $m_0 = 613$.

Приклад 3.4. Імовірність улучення в ціль за одного пострілу дорівнює 0,2. Знайдіть найімовірніше число влучень у ціль за 14 пострілів.

Розв'язання. Якщо $n = 14$; $p = 0,2$; $q = 0,8$, тоді найімовірніше число влучень у ціль за 14 пострілів обчислюють за допомогою нерівності:

$$14 \cdot 0,2 - 0,8 \le m_0 \le 14 \cdot 0,2 + 0,2,$$

$$2 \le m_0 \le 3.$$

Отже, визначено два найімовірніші числа $m_0 = 2$ і $m_0 = 3$.

Приклад 3.5. Імовірність того, що автомат спрацює правильно під час опускання в нього монети, дорівнює 0,97. Скільки необхідно опустити в нього монет, щоб найімовірніше число випадків правильної роботи автомата становило 100?

Розв'язання. За умовою $p = 0,97$; $q = 0,03;\ m_0 = 100$.

Число n знаходять із подвійної нерівності:

$$n \cdot 0,97 - 0,03 \le 100 \le n \cdot 0,97 + 0,97,$$

або $\begin{cases} n \cdot 0,97 \le 100,03, \\ n \cdot 0,97 \ge 99,03, \end{cases}$ тоді $\begin{cases} n \le 103,12, \\ n \ge 102,09. \end{cases}$

Звідки $n = 103$.

Приклад 3.6. Імовірність народження хлопчика дорівнює 0,515. Знайдіть імовірність того, що із 200 народжених дітей хлопчиків і дівчаток буде порівно.

Розв'язання. За умовою $n = 200; m = 100; p = 0,515; q = 0,485$.

За локальною теоремою Лапласа:

$$P_{200} \approx \frac{1}{\sqrt{200 \cdot 0,515 \cdot 0,485}} \varphi(x) = \frac{1}{7,068} \varphi(x);$$

$$x = \frac{100 - 200 \cdot 0,515}{7,068} = -0,42; \quad \varphi(-0,42) = \varphi(0,42) = 0,3653.$$

І тоді $P_{200}(100) \approx \dfrac{1}{7,068} \cdot \varphi(-0,42) \approx \dfrac{0,3653}{7,068} \approx 0,052.$

Приклад 3.7. Імовірність того, що буде виявлено помилку на сторінці книги, дорівнює 0,002. Перевірено 500 сторінок. Знайдіть імовірність того, що з помилкою виявляться: а) 3 сторінки; б) від 3 до 5 сторінок.

Розв'язання:

а) за умовою $p = 0,002$; $n = 500$.

Звідси $\lambda = n \cdot p = 0,002 \cdot 500 = 1$.

Тоді шукана ймовірність $P(3) \approx \frac{1^3}{3!} e^{-1} \cong 0,0613$;

б) для обчислення ймовірності $P\left(3 \le m \le 5\right)$ слід використати теорему додавання:

$$P\left(3 \le m \le 5\right) = P\left(3\right) + P\left(4\right) + P\left(5\right) \approx 0,0613 + \frac{1^4}{4!} e^{-1} + \frac{1^5}{5!} e^{-1} = 0,0613 +$$
$$+ 0,0153 + 0,0031 = 0,0797.$$

Приклад 3.8. Знайдіть імовірність того, що серед 1 000 новонароджених хлопчиків буде: а) від 465 до 550 включно; б) не менше за 550; в) не більше за 465, якщо ймовірність народження хлопчика дорівнює 0,5.

Розв'язання:

а) за умовою $n = 1\,000$; $p = 0,5$; $q = 0,5$; $m_1 = 465$; $m_2 = 550$

$$P(465 \le m \le 550) \approx \Phi(x_2) - \Phi(x_1);$$

$$x_2 = \frac{550 - 1\,000 \cdot 0,5}{\sqrt{1\,000 \cdot 0,5 \cdot 0,5}} = 3{,}16; \quad x_1 = \frac{465 - 1\,000 \cdot 0,5}{\sqrt{1\,000 \cdot 0,5 \cdot 0,5}} = -2,21;$$

$$P(465 \le m \le 550) \approx \Phi(3{,}16) - \Phi(-2,21) = \Phi(3{,}16) + \Phi(2,21) = 0,9856;$$

б) за умовою $m_1 = 550$; $m_2 = 1\,000$

$$P(550 \le m \le 1000) \approx \Phi(x_2) - \Phi(x_1);$$

$$x_2 = \frac{1000 - 1\,000 \cdot 0,5}{\sqrt{1\,000 \cdot 0,5 \cdot 0,5}} \approx 32; \quad x_1 = \frac{550 - 1\,000 \cdot 0,5}{\sqrt{1\,000 \cdot 0,5 \cdot 0,5}} = 3{,}16;$$

$$P(550 \le m \le 1\,000) \approx \Phi(32) - \Phi(3{,}16) = 0,5 - 0,4992 = 0,0008;$$

в) за умовою $m_1 = 0$; $m_2 = 464$.

Очевидно, що

$$P(0 \le m \le 464) = 1 - P(465 \le m \le 1\,000) \approx 1 - (0,9856 + 0,0008) = 0,0136.$$

Приклад 3.9. Імовірність улучення в мішень за одного пострілу дорівнює 0,7. Яка ймовірність того, що за 500 пострілів частота влучення в мішень відхилиться від імовірності р не більше ніж на 0,04 (за абсолютною величиною)?

Розв'язання. За умовою $n = 500$; $p = 0,7$; $q = 0,3$; $\varepsilon = 0,04$.

Тоді обчислюють:

$$P\left(\left|\frac{m}{500} - 0,7\right| \le 0,04\right) \approx 2\Phi\left(0,04\sqrt{\frac{500}{0,3 \cdot 0,7}}\right) = 2\Phi(1,95) = 0,9488.$$

Таким чином, те, що частота влучення в мішень буде перебувати в межах від 0,66 до 0,74 (0,7 ± 0,04), можна стверджувати з імовірністю 0,9488.

Приклад 3.10. Імовірність улучення в мішень за одного пострілу дорівнює 0,3. Скільки потрібно зробити пострілів, щоб з імовірністю 0,996 відхилення частоти влучень від імовірності $p = 0,3$ не перевищувало 0,04 (за абсолютною величиною)?

Розв'язання. За умовою $2\Phi(t) = 0,996,$ а $\Phi(t) = 0,498$.

Для розв'язання задачі необхідно використати таку формулу:

$$P\left(\left|\frac{m}{n} - p\right| \le \varepsilon\right) = 2\Phi(t), \quad t = \varepsilon\sqrt{\frac{n}{pq}}.$$

Тоді $2\Phi(t) = 0,996,$ а $\Phi(t) = 0,498$. Значення t слід знаходити за таблицею значень функції $\Phi(x)$ (додаток В): $t = 2,88$.

Звідки $\varepsilon\sqrt{\frac{n}{pq}} = 2,88$ або $0,04\sqrt{\frac{n}{0,3 \cdot 0,7}} = 2,88$, тоді n буде дорівнювати: $n = 0,21 \cdot 72^2 = 1\,088,64 \approx 1\,089$.

Отже, необхідно зробити 1 089 пострілів.

Приклад 3.11. Імовірність улучення в мішень за кожного з 800 пострілів дорівнює 0,3. У яких межах буде перебувати частота влучень, щоб імовірність невиходу за межі становила 0,9624?

Розв'язання. За умовою $2\Phi(t) = 0,9624$. За таблицею значень функції $\Phi(x)$ (додаток В) t набуває значення: $t = 2,08$.

Як і в попередньому прикладі для знаходження ε слід розв'язати рівняння $\varepsilon\sqrt{\frac{n}{pq}} = 2{,}08$.

Звідки за $n = 800$, $p = 0{,}3$, $q = 0{,}7$ $\varepsilon = 0{,}0337$.

Отже, за 800 пострілів з імовірністю 0,9264 можна стверджувати, що частота влучення в мішень буде перебувати в межах від 0,2663 до 0,3337 (0,3 ± 0,0337).

3.4. Вправи для самостійної роботи

3.1. У партії кількість великих деталей удвічі більша від кількості дрібних. Яка ймовірність того, що серед узятих навмання 10 деталей виявиться 6 великих?

3.2. Імовірність улучення в мішень за кожного окремого пострілу $p = 0{,}8$. Яка ймовірність того, що за 5 пострілів у мішень буде влучено?

3.3. Бавовна містить 10 % коротких волокон. Яка ймовірність того, що в навмання взятому пучку з 5 волокон виявиться не більше від 2 коротких?

3.4. Схожість насіння деякої рослини оцінюють імовірністю $p = 0{,}8$. Знайдіть імовірність того, що з 8 посіяних насінин зійде не менше від 6.

3.5. Імовірність появи події A в окремому випробуванні дорівнює 0,75. Яка ймовірність того, що за восьмикратного повторення випробування ця подія з'явиться понад 6 разів?

3.6. Монету кидають 5 разів. Знайдіть ймовірність того, що "герб" випаде хоча б 2 рази.

3.7. 30 % виробів підприємства – це продукція вищого сорту. Хтось придбав 6 виробів. Яка ймовірність того, що 4 вироби з вибраних є виробами вищого сорту?

3.8. Імовірність своєчасного прибуття потяга на станцію дорівнює 0,8. Яка ймовірність того, що з 4 очікуваних потягів: а) 3 прибудуть вчасно; б) не менше від 3 потягів прибудуть вчасно; в) не більше від 3 потягів прибудуть вчасно; г) принаймні 1 із потягів прибуде вчасно?

3.9. Що ймовірніше: виграти в рівносильного суперника 3 партії з 4 чи 5 із 8?

3.10. Із 10 пострілів стрілець у середньому влучає в мішень 8 разів. Яка ймовірність того, що із 3 незалежних пострілів він точно 2 рази влучить у ціль?

3.11. Як, використовуючи формулу Бернуллі, знайти ймовірність того, що в n випробуваннях Бернуллі деяка подія не настане жодного разу? Чи можна обчислити цю ймовірність, не звертаючись до формули Бернуллі?

3.12. Як, використовуючи формулу Бернуллі, знайти ймовірність того, що в n випробуваннях Бернуллі "успіх" відбудеться хоча б один раз? Чи можна обчислити цю ймовірність, не звертаючись до формули Бернуллі?

3.13. У результаті багаторічних спостережень установлено, що ймовірність випадання дощу 1 жовтня в заданому місті дорівнює $\frac{1}{7}$. Визначте найімовірніше число дощових днів 1 жовтня в заданому місті за 40 років.

3.14. Дані тривалої перевірки якості стандартних деталей показали, що брак у середньому становить 7,5 %. Визначте найімовірніше число цілком справних деталей у партії із 39 штук.

3.15. За якої кількості пострілів найімовірніше число влучень дорівнює 16, якщо ймовірність улучення за одного пострілу є 0,7?

3.16. Чому дорівнює ймовірність p появи події в кожному із 39 незалежних випробувань, якщо найімовірніше число появи події в цих випробуваннях дорівнює 25?

3.17. Зроблено 6 пострілів по об'єкту. Імовірність улучень в об'єкт за одного пострілу дорівнює 0,8. Знайдіть: а) найімовірніше число влучень; б) імовірність найімовірнішого числа влучень; г) імовірність того, що об'єкт буде зруйновано, якщо для цього досить хоча б 2 улучень.

3.18. Імовірність улучення за одного пострілу дорівнює 0,6. Знайдіть: а) імовірність того, що за 5 400 пострілів мішень буде вражено точно 3 240 разів; б) найімовірніше число влучень за 5 400 пострілів.

3.19. Верстат-автомат дає 75 % виробів I сорту. Знайдіть найімовірніше число виробів I сорту з партії із 320 штук і ймовірність цього числа.

3.20. У шухляді знаходяться 3 чорні та 4 білі кулі. Кулі виймають таким чином, що кожна вийнята куля повертається назад у шухляду. Знайдіть імовірність того, що за 250 спроб біла куля з'явиться 100 разів.

3.21. Визначте ймовірність одночасної зупинки 30 машин зі 100 тих, що працюють, якщо ймовірність зупинки однієї машини дорівнює 0,2.

3.22. Імовірність бракованого виробу дорівнює 0,2. Знайдіть імовірність того, що в партії з 400 виробів буде 104 браковані.

3.23. Імовірність збити літак гвинтівковим пострілом дорівнює 0,004. Яка ймовірність знищення літака під час залпу із 250 гвинтівок?

3.24. Частка браку за деякого технічного процесу становить 0,2 %. Визначте ймовірність того, що в партії з 1 000 виробів бракованих буде 25.

3.25. Імовірність улучення в літак за пострілу із гвинтівки дорівнює 0,001. Визначте ймовірність того, що під час залпу з 5 000 гвинтівок ціль буде вражено 2 і більше кулями.

3.26. Завод відправив на базу 5 000 доброякісних виробів. Імовірність того, що в дорозі виріб зіпсується, дорівнює 0,0002. Знайдіть імовірність того, що: а) на базу надійдуть тільки 3 зіпсовані вироби; б) на базу надійдуть від 2 до 3 зіпсованих виробів.

3.27. Робітниця обслуговує 800 веретен. Імовірність обриву пряжі на кожному з веретен протягом деякого проміжку часу t дорівнює 0,005. Знайдіть: а) найімовірніше число обривів; б) імовірність того, що протягом проміжку часу t відбудеться не більше ніж 10 обривів.

3.28. Імовірність появи події в кожному із 2 100 незалежних випробувань дорівнює 0,7. Знайдіть імовірність того, що подія з'явиться: а) від 1 470 до 1 500 разів; б) не менше ніж 1 470 разів; в) не більше ніж 1 469 разів.

3.29. Імовірність того, що деталь не пройшла перевірку ВТК, дорівнює 0,3. Знайдіть імовірність того, що серед 100 деталей пройдуть перевірку ВТК не більше ніж 79.

3.30. Стрілець улучає в ціль із ймовірністю $p = 0,75$. Яка ймовірність того, що за 100 пострілів кількість улучень буде: а) не менша за 71 і не більша за 80; б) не менша за 70 і не більша за 80; в) не менша за 81; г) не більша за 70.

3.31. На кожні 20 штампованих виробів із пластмаси припадає в середньому 3 дефектні. Знайдіть імовірність того, що з 50 взятих навмання виробів понад 42 будуть без дефекту.

3.32. Стрілець уражає мішень з імовірністю $p = 0,75$. Яка ймовірність того, що за 400 пострілів відносна частота влучень відхилиться від величини р менше ніж на 0,035?

3.33. Стрілець уражає ціль з імовірністю $p = 0,75$. У яких межах має бути кількість m улучень за 400 пострілів, щоб імовірність невиходу за ці межі дорівнювала 0,97?

3.34. Стрілець уражає мішень з імовірністю $p = 0,75$. Скільки пострілів має зробити стрілець, щоб імовірність того, що відхилення відносної частоти влучення від величини p буде меншим ніж 0,035; дорівнювала 0,95?

3.35. Імовірність появи події в кожному з 10 000 незалежних випробувань дорівнює $\frac{3}{4}$. Знайдіть імовірність того, що відносна частота появи події відхилиться від імовірності за абсолютною величиною не більше ніж на 0,0001.

3.36. В урні знаходяться білі й чорні кулі у відношенні 4 : 1. Після виймання кулі реєструють її колір, а куля повертається в урну. Чому дорівнює кількість виймань, за якої з імовірністю 0,9722 можна чекати, що абсолютна величина відхилення відносної частоти появи білої кулі від її ймовірності буде не більша ніж 0,01?

3.37. Імовірність появи події в кожному з 900 незалежних випробувань дорівнює 0,5. Знайдіть таке додатне число ε, щоб з імовірністю 0,7698 абсолютна величина відхилення відносної частоти появи події від її ймовірності p не перевищувала ε.

3.38. Імовірність появи події в кожному з 10 000 незалежних випробувань дорівнює 0,75. Знайдіть таке додатне число ε, щоб з імовірністю 0,979 абсолютна величина відхилення відносної частоти появи події від її ймовірності p не перевищувала ε.

3.39. Імовірність улучення в мішень за одного пострілу дорівнює 0.6. Знайдіть імовірність таких подій:

а) за 12 пострілів ціль буде вражено 7 разів;

б) за 200 пострілів ціль буде вражено: 1) не менше ніж 111 і не більше ніж 130 разів; 2) не менше ніж 110 і не більше ніж 130 разів; 3) не більше ніж 110 разів; 4) не менше ніж 115 разів;

в) за 600 пострілів частота влучення в мішень відхилиться від імовірності 0,6 за абсолютною величиною не більше ніж на 0,03;

г) знайдіть межі влучень у мішень за 600 пострілів, щоб імовірність невиходу за ці межі дорівнювала 0,993;

д) знайдіть таку кількість пострілів по мішені, за якою з імовірністю 0,993 можна чекати, що відхилення частоти влучень від імовірності 0,6 за абсолютною величиною не перевищать 0,03.

3.40. Гральну кістку кидають 80 разів. Знайдіть з імовірністю 0,9973 межі, у яких буде з'являтися кількість m появи цифри "6".

3.5. Тестові завдання

3.1. Формула Бернуллі має такий вигляд

А	Б	В	Г
$P_n(k) \approx \dfrac{\lambda^k}{k!} e^{-\lambda}$	$P_n(k) = p^k q^{n-k}$	$P_n(k) = C_n^k p^k q^{n-k}$	$P_n(k) \approx \dfrac{\lambda^k}{k!}$

3.2. Формула Пуассона має такий вигляд

А	Б	В	Г
$P_n(k) \approx \dfrac{\lambda^k}{k!} e^{-\lambda}$	$P_n(k) = p^k q^{n-k}$	$P_n(k) = C_n^k p^k q^{n-k}$	$P_n(k) \approx \dfrac{\lambda^k}{k!}$

3.3. Імовірність того, що в n незалежних випробуваннях подія відбудеться від k_1 до k_2 разів, обчислюють за допомогою

А локальної теореми Муавра – Лапласа
Б інтегральної теореми Лапласа
В формули Бернуллі
Г формули Пуассона
Д формули Байєса

3.4. Подію вважають малоймовірною, якщо

А	Б	В	Г
$n \to 0, \quad p \to 0$	$n \to \infty, \quad p \to 0$	$n \to 0, \quad p \to 1$	$n \to \infty, \quad p > 0{,}1$

3.5. У випадках, коли $n \le 30$, а $p > 0{,}1$, слід використовувати

А формулу Пуассона
Б формулу Бернуллі
В формулу повної ймовірності
Г інтегральну формулу Лапласа
Д локальну формулу Муавра – Лапласа

3.6. Які із зазначених властивостей функції $\varphi(x)$ є правильними

А $\varphi(-x) = \varphi(x)$ та якщо $x > 4$, то $\varphi(x) \approx 0$

Б $\varphi(-x)=\varphi(x)$ та якщо $x>0$, то $\varphi(x)\approx 1$

В $\varphi(-x)=\varphi(x)$ та якщо $x>0$, то $\varphi(x)\approx 0$

Г $\varphi(-x)=-\varphi(x)$ та якщо $x>4$, то $\varphi(x)\approx 0$

3.7. Які із зазначених властивостей функції $\Phi(x)$ є правильними

А $\Phi(-x)=\Phi(x)$ та якщо $x>5$, то $\Phi(x)\approx 0{,}5$

Б $\Phi(-x)=-\Phi(x)$ та якщо $x>0$, то $\Phi(x)\approx 0{,}5$

В $\Phi(-x)=-\Phi(x)$ та якщо $x>5$, то $\Phi(x)\approx 0{,}5$

Г $\Phi(-x)=-\Phi(x)$ та якщо $x>5$, то $\Phi(x)\approx 0$

3.8. За локальною теоремою Муавра – Лапласа ймовірність того, що в n незалежних випробуваннях подія відбудеться k разів, приблизно дорівнює

А $P_n(k)\approx\dfrac{\varphi(x)}{\sqrt{npq}}$, де $x=\dfrac{k-np}{\sqrt{npq}}$

Б $P_n(k)\approx\dfrac{\varphi(x)}{\sqrt{npq}}$, де $x=\dfrac{k}{\sqrt{npq}}$

В $P_n(k)\approx\varphi(x)$, де $x=\dfrac{k-np}{\sqrt{npq}}$

Г $P_n(k)\approx\varphi(x)$, де $x=k-np$

3.9. Найімовірніше число появи події A в n незалежних випробуваннях m_0 перебуває в інтервалі

А $np-q\le m_0\le np+p$

Б $np-p\le m_0\le np+q$

В $0\le m_0\le 1$

Г $np+q\le m_0\le np-p$

3.10. Кількість випробувань n, за яких із заданою ймовірністю p відхилення відносної частоти $\dfrac{m}{n}$ події A від імовірності p її появи за абсолютною величиною не буде перевищувати задане додатне число ε, дорівнює

А	Б	В	Г
$n=\dfrac{\varepsilon^2 pq}{t^2}$	$n=\dfrac{t^2 pq}{\varepsilon^2}$	$n=\dfrac{tp^2 q}{\varepsilon^2}$	$n=\dfrac{t^2 p^2 q}{\varepsilon^2}$

3.6. Запитання для самоперевірки

3.1. Поясніть, що таке "повторні випробування", наведіть їхні приклади.

3.2. Пояснісіть, яким умовам відповідає схема Бернуллі.

3.3. Наведіть формулу Бернуллі та поясніть, за яких умов її можна застосовувати.

3.4. Поясніть, чи має обмеження формула Бернуллі за кількістю випробувань.

3.5. Дайте означення поняття "довірча ймовірність".

3.6. Наведіть формулу для визначення кількості випробувань, що із заданою надійністю забезпечує появу певної випадкової події хоча б один раз.

3.7. Пояснісіть, чому обидві формули (Муавра – Лапласа та Пуассона) називають асимптотичними.

3.8. Сформулюйте локальну теорему Муавра – Лапласа.

3.9. Як впливає ймовірність появи події в окремому випробуванні на точність обчислень за локальною теоремою Муавра – Лапласа?

3.10. Сформулюйте інтегральну теорему Муавра – Лапласа.

3.11. Наведіть формулу, за допомогою якої можна визначити ймовірність відхилення частоти появи події від її ймовірності на величину, що не перевищує заздалегідь обумовленого значення.

3.12. Порівняйте можливості застосування формули Пуассона та локальної формули Муавра – Лапласа.

3.13. Наведіть формулу Пуассона.

3.14. Поясніть, чому закон, для характеристики якого застосовують формулу Пуассона, називають законом рідкісних подій.

3.15. Наведіть приклади повторних випробувань, коли для визначення ймовірності появи події в серії випробувань не можна застосовувати формулу Бернуллі.

3.7. Висновки за темою

Після вивчення теми студент має:

знати ознаки, за якими здійснюють класифікацію схеми повторних випробувань та особливості застосування асимптотичних теорем;

знати умови застосування формули Бернуллі;

знати умови застосування формули Пуассона;

знати умови застосування локальної та інтегральної формул Лапласа;

уміти виконувати обчислення за формулами Бернуллі та Пуассона;

застосовувати довідкові таблиці значень функцій Гаусса та Лапласа для визначення ймовірностей за локальною та інтегральною теоремами Лапласа.

4. Випадкові величини та їхня економічна інтерпретація

4.1. Мета та компетентності

Метою вивчення теми є надання уявлення про типи випадкових величин і способи визначення їхніх законів розподілу; про функцію розподілу, її властивості та особливості побудови.

Компетентності, що формуються під час вивчення теми:

знання означення випадкової величини;

уміння будувати функцію розподілу випадкової величини;

знання основних числових характеристик випадкової величини та їхніх властивостей;

уміння знаходити початкові й центральні теоретичні моменти.

4.2. Основні теоретичні відомості

Випадковою величиною називають величину, яка в результаті випробування може набути того чи того значення, але якого саме – невідомо. Випадкові величини позначають великими літерами, а їхні можливі значення – відповідними малими літерами латинського алфавіту, наприклад, $X(x_1, x_2, ..., x_n)$.

Дискретною (перервною) випадковою величиною називають випадкову величину, яка набуває окремих ізольованих можливих значень із певними ймовірностями. Кількість можливих значень дискретної випадкової величини може бути скінченною або зліченною.

Неперервною називають випадкову величину, яка може набувати всіх значень із деякого скінченного чи нескінченного проміжку.

Законом розподілу (рядом розподілу) дискретної випадкової величини називають відповідність між можливими значення та їхніми ймовірностями. Його можна задати таблично (табл. 4.1):

Таблиця 4.1

Закон розподілу дискретної випадкової величини

X	x_1	x_2	x_3	…	x_n
p	p_1	p_2	p_3	…	p_n

У першому рядку табл. 4.1 знаходяться можливі значення x_i, а у другому – відповідні ймовірності p_i, де $\sum_{i=1}^{n} p_i = 1$.

Закон розподілу дискретної випадкової величини можна задати аналітично (у вигляді формули): $P(X = x_i) = \varphi(x_i)$, а також графічно.

Для цього у прямокутній системі координат будують точки $M_1(x_1; p_1)$, $M_2(x_2; p_2)$, …, $M_n(x_n; p_n)$ (x_i – можливі значення X, p_i – відповідні ймовірності) і з'єднують їх відрізками прямих.

Побудовану фігуру називають **многокутником розподілу** (рис. 4.1).

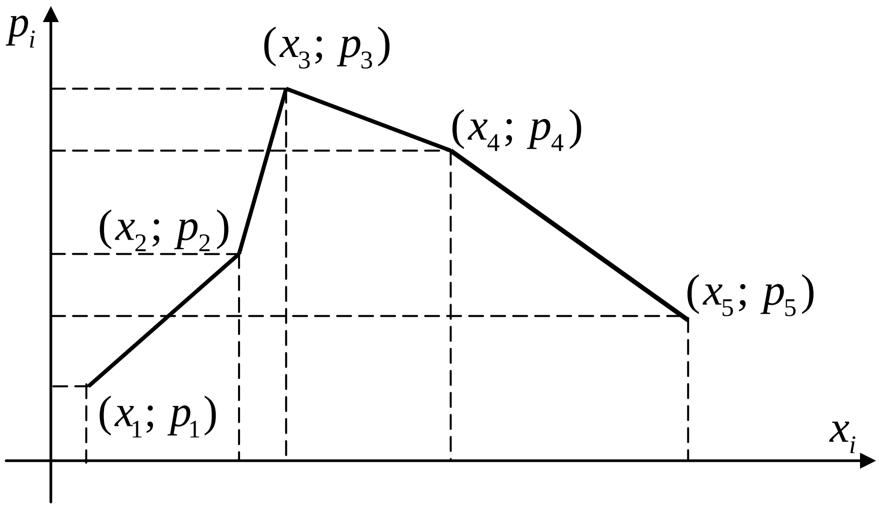


Рис. 4.1. **Многокутник розподілу**

Дискретну випадкову величину можна задати й функцією розподілу.

Функцією розподілу ймовірностей випадкової величини називають функцію $F(x)$, яка визначає ймовірність того, що випадкова величина X у результаті випробування набуде значення, яке менше ніж x:

$$F(x) = P(X < x) = \sum_{x_i < x} p_i \, .$$

Тут для кожного значення x підсумовують імовірності тих значень x_i, які знаходяться ліворуч від точки x.

4.2.1. Числові характеристики дискретної випадкової величини

Характеристикою середнього значення випадкової величини є математичне сподівання.

Математичним сподіванням дискретної випадкової величини називають суму добутків усіх її можливих значень на відповідні ймовірності:

$$M(X) = x_1 p_1 + x_2 p_2 + ... + x_n p_n = \sum_{i=1}^{n} x_i p_i .$$

Якщо дискретна випадкова величина X набуває зчисленної множини можливих значень, то $M(X) = \sum_{i=1}^{\infty} x_i p_i$, за умови, що ряд збігається абсолютно.

Математичне сподівання показує, яке значення випадкової величини слід очікувати в середньому під час випробувань.

Властивості математичного сподівання:

1) математичне сподівання сталої величини дорівнює цій сталій:

$$M(C) = C;$$

2) сталий множник можна виносити за знак математичного сподівання:

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

3) математичне сподівання суми скінченного числа випадкових величин дорівнює сумі їхніх математичних сподівань:

$$M(X_1 + X_2 + ... + X_n) = M(X_1) + M(X_2) + ... + M(X_n).$$

Наслідок – математичне сподівання різниці двох випадкових величин дорівнює різниці їхніх математичних сподівань:

$$M(X - Y) = M(X) - M(Y);$$

4) математичне сподівання добутку скінченного числа взаємно незалежних випадкових величин дорівнює добутку їхніх математичних сподівань:

$$M(X_1 \cdot X_2 \cdot ... \cdot X_n) = M(X_1) \cdot M(X_2) \cdot ... \cdot M(X_n).$$

Іншою важливою характеристикою випадкової величини є дисперсія, яка характеризує міру розсіювання значень випадкової величини навколо її математичного сподівання.

Дисперсією випадкової величини X називають математичне сподівання квадрата відхилення цієї величини від її математичного сподівання:

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

Для розрахунку використовують ще й таку формулу:

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

Властивості дисперсії:

1) дисперсія сталої дорівнює нулю:

$$D(C) = 0;$$

2) сталий множник можна виносити за знак дисперсії, підносячи його до квадрата:

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

3) дисперсія суми попарно незалежних випадкових величин дорівнює сумі їхніх дисперсій:

$$D(X_1 + X_2 + ... + X_n) = D(X_1) + D(X_2) + ... + D(X_n);$$

4) дисперсія різниці двох незалежних випадкових величин дорівнює сумі їхніх дисперсій:

$$D(X - Y) = D(X) + D(Y).$$

Дисперсія має розмірність квадратних одиниць вимірюваної величини. Щоб знайти розмірність міри розсіювання в одиницях вимірюваної величини, розглядають квадратний корінь із дисперсії:

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

який називають **середнім квадратичним відхиленням** (стандартним відхиленням).

В економічних задачах часто середнє квадратичне відхилення σ використовують як міру ризику.

4.2.2. Додаткові числові характеристики розподілу дискретної випадкової величини

Початковим моментом k-го порядку випадкової величини X називають математичне сподівання величини X^k:

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

Зокрема, початковий момент першого порядку дорівнює математичному сподіванню:

$$\nu_1 = M(X);$$

початковий момент другого порядку:

$$\nu_2 = M(X^2).$$

Центральним моментом k-го порядку випадкової величини X називають математичне сподівання величини $(X - M(X))^k$.

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

Зокрема, центральний момент першого порядку дорівнює нулю:

$$\mu_1 = M(X - M(X));$$

центральний момент другого порядку дорівнює дисперсії:

$$\mu_2 = M(X - M(X))^2.$$

Центральні моменти зручніше обчислювати за формулами, що виражають центральні моменти через початкові:

$$\mu_2 = \nu_2 - \nu_1^2;$$

$$\mu_3 = \nu_3 - 3\nu_1\nu_2 + 2\nu_1^3;$$

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

Однією з поширених додаткових характеристик розподілу є **коефіцієнт варіації**. Він відображає у відсотках частку середнього квадратичного відхилення, що припадає на одиницю математичного сподівання:

$$V = \frac{\sigma}{M\left(X\right)} \cdot 100\,\%.$$

Коефіцієнт варіації є зручною характеристикою, за якою можна оцінювати, за яким саме законом розподілено випадкову величину.

Разом із математичним сподіванням показниками центральних тенденцій є мода та медіана.

Модою випадкової величини ($Mo(X)$) називають те її значення, якому відповідає найбільше значення ймовірності.

Медіаною випадкової величини $Me\left(X\right)$ називають таке її значення, яке за кількістю значень випадкової величини ділить навпіл ряд розподілу.

Додатковими числовими характеристиками розподілу є також асиметрія та ексцес випадкової величини. Формально їх можна визначати і для дискретної випадкової величини, однак в основному їх застосовують як характеристики розподілу неперервної випадкової величини. Саме для визначення цих характеристик застосовують теоретичні моменти вищих порядків. Так, **асиметрію**, або **коефіцієнт асиметрії** розподілу, визначають співвідношенням:

$$A_s = \frac{\mu_3}{\sigma^3},$$

де μ_3 – центральний момент третього порядку;

σ – середнє квадратичне відхилення випадкової величини.

Ексцесом, або **коефіцієнтом ексцесу** випадкової величини, називають величину, що визначають за такою формулою:

$$E_k = \frac{\mu_4}{\sigma^4} - 3,$$

де μ_4 – центральний момент четвертого порядку;

σ – середнє квадратичне відхилення випадкової величини.

4.3. Приклади розв'язування задач

Приклад 4.1. Дискретну випадкову величину X задано законом розподілу (табл. 4.2):

Таблиця 4.2

Закон розподілу випадкової величини X

X	1	5	8	10
p	0,3	0,4	0,2	0,1

Побудуйте многокутник розподілу.

Розв'язання. У прямокутній системі координат на осі абсцис слід відкласти можливі значення випадкової величини, а на осі ординат – відповідні ймовірності.

Таким чином будують точки $M_1(1; 0,3)$, $M_2(5; 0,4)$, $M_3(8; 0,2)$, $M_4(10; 0,1)$.

Якщо з'єднати ці точки відрізками прямих, то буде многокутник розподілу (рис. 4.2).

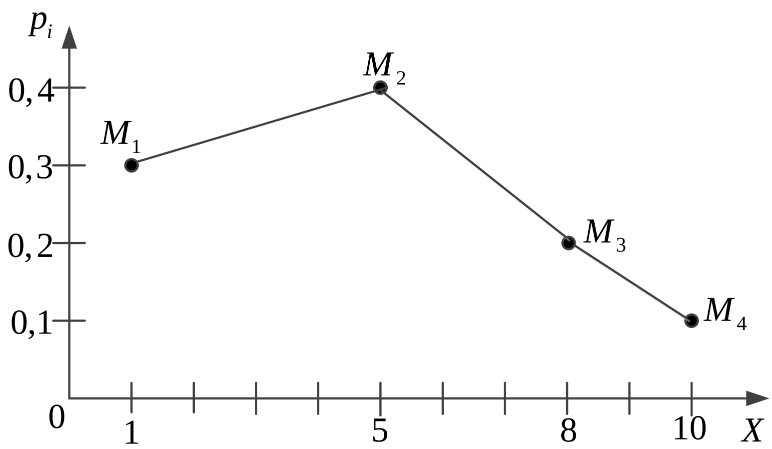


Рис. 4.2. **Многокутник розподілу**

Приклад 4.2. Два стрільці виконують по одному пострілу в мішень. Імовірність улучення для першого дорівнює 0,6, а для другого – 0,5. Складіть закон розподілу кількості влучень у мішень.

Розв'язання. Дискретна випадкова величина X (кількість улучень у мішень) має такі значення: $x_1 = 0$ (обидва стрільці не влучили в мішень), $x_2 = 1$ (один улучив), $x_3 = 2$ (обидва влучили).

Знайдіть відповідні ймовірності.

Імовірність улучення для першого $p_1 = 0,6$, а ймовірність промаху: $q_1 = 1 - 0,6 = 0,4$;

імовірність улучення для другого $p_2 = 0,5$, а ймовірність промаху: $q_2 = 1 - 0,5 = 0,5$.

Тоді ймовірність того, що обидва стрільці не влучили в мішень:

$$P(X = 0) = q_1 \cdot q_2 = 0,4 \cdot 0,5 = 0,2;$$

імовірність того, що один улучив:

$$P(X = 1) = p_1 q_2 + p_2 q_1 = 0,6 \cdot 0,5 + 0,6 \cdot 0,4 = 0,5;$$

імовірність того, що обидва влучили:

$$P(X = 2) = p_1 \cdot p_2 = 0,6 \cdot 0,5 = 0,3.$$

Перевірка: $0,2 + 0,5 + 0,3 = 1$.

Ряд розподілу кількості влучень у мішень наведено в табл. 4.3.

Таблиця 4.3

Закон розподілу випадкової величини X

X	0	1	2
p	0,2	0,5	0,5

Приклад 4.3. Стрілець стріляє в ціль до першого влучення, але встигає зробити не більше від 4 пострілів. Складіть закон розподілу кількості пострілів, зроблених стрільцем, якщо ймовірність улучення в ціль за одного пострілу дорівнює 0,7.

Розв'язання. Кількість пострілів, зроблених стрільцем, може набувати таких значень: 1 (перший постріл улучає в ціль), 2 (перший постріл – промах, другий – улучення), 3 (два перші постріли – промахи, третій – улучення), 4 (перші три постріли – промахи, а четвертий – улучення або промах).

За умовою задачі $P_4(1) = 0,7$.

За теоремою множення ймовірностей буде:

$$P_4(2)=0,3\cdot 0,7=0,21;\ \ P_4(3)=0,3\cdot 0,3\cdot 0,7=0,063;$$

$$P_4(4)=0,3\cdot 0,3\cdot 0,3=0,027$$

(тому що результат четвертого пострілу не має значення).

Закон розподілу шуканої випадкової величини X – кількість пострілів зручно надати у вигляді таблиці (табл. 4.4):

Таблиця 4.4

Закон розподілу випадкової величини X

Кількість пострілів	1	2	3	4
p_i	0,700	0,210	0,063	0,027

Перевірка: $0,700+0,210+0,063+0,027=1.$

Приклад 4.4. Знайдіть функцію розподілу випадкової величини X, заданої законом розподілу (див. табл. 4.4).

Знайдіть імовірність того, що випадкова величина X набуде значення не меншого за 1 і не більшого за 4.

Розв'язання. Випадкова величина X не набуває значень, менших за 1. Тому для $x\leq 1$ події $X<x$ неможливі, а $F(x)=0$.

Якщо $1<x\leq 2$, то $F(x)=0,700$, тому що X може набути тільки значення $x=1$ з імовірністю $p=0,700$.

Якщо $2<x\leq 3$, то $F(x)=0{,}700+0{,}210=0{,}910$, тому що X може набути одного зі значень $x=1$ або $x=2$ з імовірностями $p=0,7$ та $p=0,210$, що потребує застосування теореми додавання ймовірностей для обчислення шуканої ймовірності.

Аналогічно, якщо $3<x\leq 4$, то $F(x)=0{,}700+0{,}210+0{,}063=0{,}973$.

Якщо $x>4$, то $F(x)=1$, тому що подія $X\leq 1$ достовірна і ймовірність її дорівнює 1.

Отже, за $x>4$ $F(x)=0{,}700+0{,}210+0{,}063+0{,}270=1.$

Таким чином, шукана функція розподілу випадкової величини X, а також її графік (рис. 4.3) мають такий вигляд:

$$F(x)=\begin{cases}0, & \text{якщо} \quad x\le 1,\\ 0{,}700, & \text{якщо} \quad 1<x\le 2,\\ 0{,}910, & \text{якщо} \quad 2<x\le 3,\\ 0{,}973, & \text{якщо} \quad 3<x\le 4,\\ 1, & \text{якщо} \quad x>4.\end{cases}$$

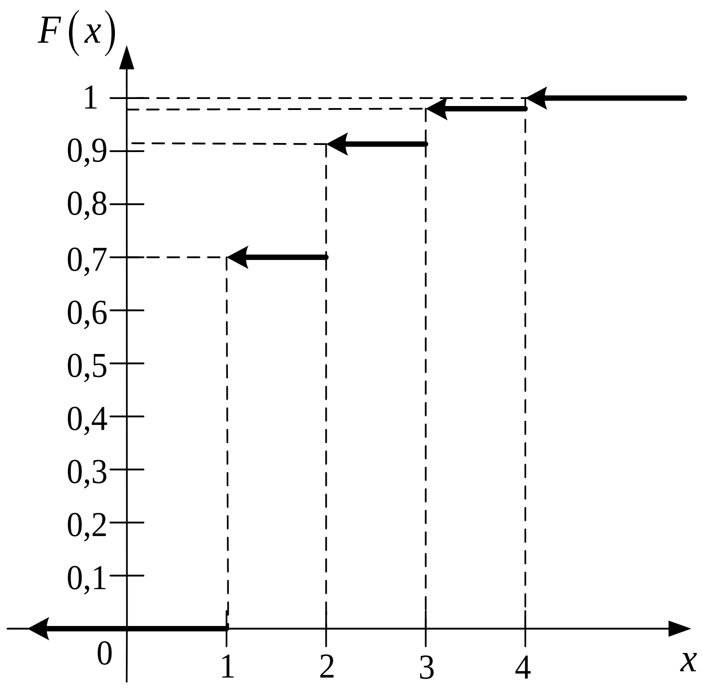


Рис. 4.3. **Графік інтегральної функції розподілу**

Імовірність того, що випадкова величина X набуде значення не меншого за 1 і не більшого за 4 дорівнює:

$$P(1\le X\le 4)=F(4)-F(1)=0{,}973-0=0{,}973.$$

Приклад 4.5. Для випадкової величини X, заданої законом розподілу (табл. 4.5), обчисліть $M(X)$, $M(X^2), M^2(X), D(X)$.

Таблиця 4.5

Закон розподілу випадкової величини X

x_i	5	10	15
p_i	0,2	0,6	0,2

Розв'язання. $M(X)=\sum_{i=1}^{n} x_i p_i$.

Тоді $M(X)=5\cdot 0,2+10\cdot 0,6+15\cdot 0,2=10$.

Далі слід записати закон розподілу випадкової величини X^2. Це зручно зробити у вигляді таблиці (табл. 4.6):

Таблиця 4.6

Закон розподілу випадкової величини X^2

x_i^2	25	100	225
p_i	0,2	0,6	0,2

Тоді $M(X^2)=25\cdot 0,2+100\cdot 0,6+225\cdot 0,2=110.$

Дисперсію слід обчислити за формулою:

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

Таким чином, $D(X)=110-10^2=10.$

Приклад 4.6. Випадкову величину X задано законом розподілу (табл. 4.7).

Таблиця 4.7

Закон розподілу випадкової величини X

x_i	1	2	3
p_i	p_1	p_2	p_3

Також дано $M\left(X\right)=2,3;\ M\left(X^2\right)=5,9$.

Знайдіть $p_1,\ p_2,\ p_3$.

Розв'язання. Очевидно, що $p_1+p_2+p_3=1$;

$$M(X)=1\cdot p_1+2\cdot p_2+3\cdot p_3;$$

$$M(X^2)=1\cdot p_1+4\cdot p_2+9\cdot p_3.$$

Тоді необхідно розв'язати систему рівнянь:

$$\begin{cases} p_1 + p_2 + p_3 = 1, \\ p_1 + 2 \cdot p_2 + 3 \cdot p_3 = 2{,}3, \\ p_1 + 4 \cdot p_2 + 9 \cdot p_3 = 5{,}9. \end{cases}$$

Звідки $p_1 = 0{,}2;\ p_2 = 0{,}3;\ p_3 = 0{,}5.$

Приклад 4.7. Знайдіть математичне сподівання і дисперсію випадкової величини X , заданої законом розподілу (табл. 4.8):

Таблиця 4.8

Закон розподілу випадкової величини X

X	2	5	7	10	12
p	0,1	0,2	0,4	0,2	0,1

Розв'язання. Математичне сподівання випадкової величини X буде дорівнювати $M(X) = 2 \cdot 0{,}1 + 5 \cdot 0{,}2 + 7 \cdot 0{,}4 + 10 \cdot 0{,}2 + 12 \cdot 0{,}1 = 7{,}2.$

Для обчислення дисперсії за означенням необхідно використати таку формулу:

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

Розрахунки зручно оформити у вигляді таблиці (табл. 4.9).

Таблиця 4.9

Розрахункова таблиця

x_i	p_i	$x_i p_i$	$x_i^2 p_i$	$x_i - 7{,}2$	$(x_i - 7{,}2)^2$	$(x_i - 7{,}2)^2 p_i$
1	2	3	4	5	6	7
2	0,1	0,2	0,4	−5,2	27,04	2,704
5	0,2	1,0	5,0	−2,2	4,84	0,968
7	0,4	2,8	19,6	−0,2	0,04	0,016
10	0,2	2,0	20,0	2,8	7,84	1,568
12	0,1	1,2	14,1	4,8	23,04	2,304
–	–	7,2	59,4	–	–	7,56

У перший стовпець таблиці запишіть значення випадкової величини, у другий – відповідні ймовірності. Для знаходження математичного сподівання треба спочатку знайти добуток $x_i p_i$ (третій стовпець). Для цього числа першого стовпця слід помножити на відповідні числа другого стовпця, а сума чисел $\left(\sum_{i=1}^{n} x_i p_i\right)$ цього стовпця і є математичним сподіванням, а саме: $M(X) = \sum_{i=1}^{n} x_i p_i = 7,2.$

Для обчислення дисперсії спочатку знайдіть $M(X^2) = \sum_{i=1}^{n} x_i^2 p_i$, для цього заповнюйте четвертий стовпець, де числа третього стовпця слід помножити на відповідні числа першого стовпця. Знайдіть суму чисел цього стовпця і визначте $M(X^2) = 59,4$. Для знаходження дисперсії за означенням спочатку заповніть п'ятий стовпець, де від кожного x_i відніміть $M(X) = 7,2$. Далі знайдіть $(x_i - 7,2)^2$, тобто кожне число з п'ятого стовпця піднесіть до квадрата і запишіть результат у шостий стовпець. Тепер знайдіть $(x_i - 7,2)^2 p_i$, для цього кожне число із шостого стовпця помножте на відповідні числа другого стовпця і результат запишіть у сьомий стовпець. Сума чисел сьомого стовпця і є дисперсією, тобто $D(X) = 7,56$.

Тепер обчисліть дисперсію за формулою:

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

За табл. 4.9 $M(X^2) = \sum_{i=1}^{n} x_i^2 p_i = 59,4.$

Тоді $D(X) = 59,4 - (7,2)^2 = 7,56.$

Отже, за двома формулами дисперсії збіглися.

Приклад 4.8. Дискретну випадкову величину X задано законом розподілу (табл. 4.10). Знайдіть початкові моменти першого, другого, третього і четвертого порядку.

Таблиця 4.10

Закон розподілу випадкової величини X

X	2	4	5
P	0,3	0,6	0,1

Розв'язання:

початковий момент першого порядку $\nu_1 = M(X)$,

$$M(X) = x_1 \cdot p_1 + x_2 \cdot p_2 + x_3 \cdot p_3,$$

тому $\nu_1 = 2 \cdot 0{,}3 + 4 \cdot 0{,}6 + 5 \cdot 0{,}1 = 3{,}5$;

початковий момент другого порядку $\nu_2 = M(X^2)$,

$$M(X^2) = x_1^2 \cdot p_1 + x_2^2 \cdot p_2 + x_3^2 \cdot p_3,$$

тому $\nu_1 = 2^2 \cdot 0{,}3 + 4^2 \cdot 0{,}6 + 5^2 \cdot 0{,}1 = 13{,}3$;

початковий момент третього порядку $\nu_3 = M(X^3)$,

$$M(X^3) = x_1^3 \cdot p_1 + x_2^3 \cdot p_2 + x_3^3 \cdot p_3,$$

звідки $\nu_1 = 2^3 \cdot 0{,}3 + 4^3 \cdot 0{,}6 + 5^3 \cdot 0{,}1 = 53{,}3$;

початковий момент четвертого порядку $\nu_4 = M(X^4)$,

$$M(X^4) = x_1^4 \cdot p_1 + x_2^4 \cdot p_2 + x_3^4 \cdot p_3,$$

звідки $\nu_1 = 2^4 \cdot 0{,}3 + 4^4 \cdot 0{,}6 + 5^4 \cdot 0{,}1 = 220{,}9$.

Приклад 4.9. Дискретну випадкову величину X задано законом розподілу (табл. 4.11).

Таблиця 4.11

Закон розподілу випадкової величини X

X	−1	2	4
P	0,4	0,3	0,3

Знайдіть центральні моменти першого, другого, третього та четвертого порядку.

Розв'язання. Центральний момент першого порядку дорівнює

$$\mu_1 = M(X - M(X)).$$

Спочатку знайдіть $M(X)$: $M(X) = -1 \cdot 0,4 + 2 \cdot 0,3 + 4 \cdot 0,3 = 1,4$.

$$\mu_1 = (x_1 - M(X)) \cdot p_1 + (x_2 - M(X)) \cdot p_2 + (x_3 - M(X)) \cdot p_3$$

Ураховуючи дані задачі, буде:

$$\mu_1 = (-1 - 1,4) \cdot 0,4 + (2 - 1,4) \cdot 0,3 + (4 - 1,4) \cdot 0,3 = 0,$$

що й підтвердили розрахунки.

Далі знайдіть центральний момент другого порядку:

$$\mu_2 = M(X - M(X))^2 \text{ – дисперсія.}$$

$$\mu_2 = (x_1 - M(X))^2 \cdot p_1 + (x_2 - M(X))^2 \cdot p_2 + (x_3 - M(X))^2 \cdot p_3.$$

Підставте дані задачі:

$$\mu_2 = (-1 - 1,4)^2 \cdot 0,4 + (2 - 1,4)^2 \cdot 0,3 + (4 - 1,4)^2 \cdot 0,3 = 4,44.$$

Тепер обчисліть центральний момент третього порядку:

$$\mu_3 = M(X - M(X))^3.$$

$$\mu_3 = (x_1 - M(X))^3 \cdot p_1 + (x_2 - M(X))^3 \cdot p_2 + (x_3 - M(X))^3 \cdot p_3.$$

Використовуючи умови задачі, буде:

$$\mu_3 = (-1 - 1,4)^3 \cdot 0,4 + (2 - 1,4)^3 \cdot 0,3 + (4 - 1,4)^3 \cdot 0,3 = 0,192.$$

Центральний момент четвертого порядку:

$$\mu_4 = M(X - M(X))^4,$$

або $\mu_4 = (x_1 - M(X))^4 \cdot p_1 + (x_2 - M(X))^4 \cdot p_2 + (x_3 - M(X))^4 \cdot p_3$.

Для цієї задачі буде:

$$\mu_4 = (-1 - 1,4)^4 \cdot 0,4 + (2 - 1,4)^4 \cdot 0,3 + (4 - 1,4)^4 \cdot 0,3 = 27,0192.$$

4.4. Вправи для самостійної роботи

4.1. У студентській групі 4 юнаки і 6 дівчат. На чергування випадково відібрано двох студентів. Складіть закон розподілу кількості дівчат серед відібраних.

4.2. Імовірність виготовлення нестандартного виробу під час деякого технологічного процесу дорівнює 0,05. Із партії беруть виріб і перевіряють його якість. Якщо він буде нестандартним, випробування припиняють, а партію затримують. Якщо виріб буде стандартним, беруть наступний і так далі, але перевіряють не більше ніж п'ять виробів. Складіть закон розподілу кількості перевірюваних виробів.

4.3. Дискретну випадкову величину задано законом розподілу (табл. 4.12).

Таблиця 4.12

Закон розподілу випадкової величини X

X	0	2	4	5
p	0,1	0,6	?	0,1

Знайдіть інтегральну функцію $F(x)$ і побудуйте її графік.

4.4. З урни, що містить 4 білі та 3 чорні кулі, навмання виймають 2 кулі. Нехай X – кількість вийнятих білих куль. Знайдіть функцію розподілу випадкової величини X, побудуйте графік цієї функції.

4.5. Кожна із 3 гармат стріляє по цілі 1 раз. Імовірності влучення в ціль за одного пострілу з 1, 2 і 3-ї гармати дорівнюють, відповідно, 0,7; 0,8; 0,9. Знайдіть функцію розподілу випадкової величини X – кількості влучень у ціль.

4.6. Напишіть закон розподілу випадкової величини X – кількості появ "герба" за двох кидань монети.

4.7. З урни, що містить 2 білі та 2 чорні кулі, навмання виймають кулю. Вийняту кулю повертають в урну, після чого з неї виймають навмання другу кулю. Нехай X – кількість білих куль, вийнятих за обох спроб. Складіть закон розподілу, функцію розподілу випадкової величини X; побудуйте графік функції розподілу.

4.8. Пару гральних кісток кидають 10 разів. Яка ймовірність того, що: а) 12 очок з'являться 2 рази; б) 3 очки з'являться 2 рази; в) 7 очок з'являться 4 рази?

4.9. Дано випадкову величину X (табл. 4.13).

Таблиця 4.13

Закон розподілу випадкової величини X

x_i	5	10	15	20
p_i	0,2	0,3	0,4	0,1

Обчисліть $M(X)$ і $D(X)$.

4.10. Дано випадкову величину X (табл. 4.14).

Таблиця 4.14

Закон розподілу випадкової величини X

x_i	4	6	x_3
p_i	0,5	0,3	p_3

Знайдіть x_3 і p_3 якщо $M(X)=8$.

4.11. Дано випадкову величину X (табл. 4.15).

Таблиця 4.15

Закон розподілу випадкової величини X_2

x_i	−1	0	1
p_i	p_1	p_2	p_3

Відомо, що $M(X)=0{,}1$; $M(X^2)=0{,}9$. Знайдіть p_1, p_2, p_3.

4.12. Випадкова величина X може набувати лише двох значень: 1 і 2. Знайдіть імовірність цих значень, якщо: а) $M(X)=\frac{4}{3}$; б) $D(X)=\frac{1}{4}$.

4.13. Випадкова величина X може набувати тільки таких значень: − 1, 0 і 1. Знайдіть імовірності цих значень, якщо відомо, що: а) $M(X)=0$, $D(X)=\frac{2}{3}$; б) $M(X)=\frac{1}{6}$, $D(X)=\frac{17}{36}$.

4.14. Визначте математичне сподівання і дисперсію:

а) кількості очок, що випали за одного кидання гральної кістки;

б) суми очок, що випали за кидання 2 гральних кісток.

4.15. Визначте математичне сподівання і середнє квадратичне відхилення добутку очок, що випали за кидання 2 гральних кісток.

4.16. Виконують ряд пострілів з імовірністю влучення 0,8 за кожного пострілу. Стрілянину ведуть до першого влучення, але виконують не більше від 4 пострілів. Визначте математичне сподівання кількості зроблених пострілів.

4.17. За табл. 4.16 розподілу випадкової величини X визначте її математичне сподівання і середнє квадратичне відхилення.

Таблиця 4.16

Закон розподілу випадкової величини X

x_i	0	1	2	3	4
p_i	0,1	0,3	0,3	0,2	0,1

4.18. З урни, що містить 3 білі та 3 чорні кулі, одночасно навмання витягають 3 кулі. Визначте математичне сподівання і дисперсію кількості вийнятих білих куль.

4.19. Із колоди із 32 карт навмання витягають 2 карти. Знайдіть математичне сподівання кількості вийнятих тузів.

4.20. В одній урні лежать 4 білі та 1 чорна куля, в іншій – 2 білі та 3 чорні кулі. Із навмання вибраної урни витягають а) 1 кулю; б) 2 кулі. Визначте математичне сподівання кількості вийнятих білих куль для випадків а) і б).

4.21. Із першої урни, що містить 1 білу і 4 чорні кулі, навмання перекладають 1 кулю у другу урну, що містить 3 білі та 1 чорну кулі. Потім кулі другої урни перемішують і навмання виймають 2 кулі. Визначте математичне сподівання, дисперсію і середнє квадратичне відхилення кількості вийнятих білих куль.

4.22. Робітник обслуговує 4 верстати. Імовірність того, що протягом години верстат не потребує нагляду робітника, дорівнює для першого верстата 0,80; для другого – 0,85; для третього – 0,70; для четвертого – 0,75. Знайдіть математичне сподівання і дисперсію кількості верстатів, що не потребують нагляду робітника протягом години.

4.23. Дискретну випадкову величину X задано законом розподілу:

а)

X	1	3	5
P	0,3	0,5	0,2

;

б)

X	−2	0	2
P	0,3	0,6	0,1

.

Знайдіть початкові моменти першого, другого, третього та четвертого порядків.

4.24. Доведіть, що центральний момент четвертого порядку пов'язано з початковими моментами таким співвідношенням:

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

4.5. Тестові завдання

4.1. Дискретною випадковою величиною називають випадкову величину, яка

А набуває окремих ізольованих можливих значень із певними ймовірностями

Б набуває всіх можливих значень із деякого інтервалу

В інше

4.2. Сума ймовірностей у таблиці закону розподілу дорівнює

А	Б	В	Г
0	1	2	інше

4.3. Математичне сподівання дискретної випадкової величини обчислюють за такою формулою

А	Б	В	Г
$M(X) = \sum_{i=1}^{n} x_i$	$M(X) = \sum_{i=1}^{n} x_i p_i$	$M(X) = \sum_{i=1}^{n} x_i^2 p_i$	$M(X) = \sum_{i=1}^{n} p_i$

4.4. Дисперсію дискретної випадкової величини обчислюють за такою формулою

А $D(X) = \sum_{i=1}^{n} x_i^2 \cdot p_i$

Б $D(X) = \sum_{i=1}^{n} x_i \cdot p_i - M(X)$

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

Г $D(X) = \sum_{i=1}^{n} x_i \cdot p_i$

4.5. Функція розподілу $F(x)$ визначає ймовірність події

А	Б	В	Г
$F(x)=P(X=x)$	$F(x)=P(X>x)$	$F(x)=P(X<x)$	$F(x)=P(x_1<X<x_2)$

4.6. Початковий момент k-го порядку обчислюють за такою формулою

А	Б	В	Г
$\nu_k=M(X^2)$	$\nu_k=M(X^k)$	$\nu_k=M(X_k)$	$\nu_k=M(X)$

4.7. Центральний момент k-го порядку обчислюють за такою формулою

А	Б	В	Г
$\mu_k=(X)^k$	$\mu_k=(X-M(X))^2$	$\mu_k=(X-M(X^k))$	$\mu_k=(X-M(X))^k$

4.8. Початковий момент першого порядку дорівнює

А	Б	В	Г
$D(X)$	0	$M(X)$	$\sigma(X)$

4.9. Центральний момент 2-го порядку дорівнює

А	Б	В	Г
$D(X)$	0	$M(X)$	$\sigma(X)$

4.10. Коефіцієнт варіації обчислюють за такою формулою

А	Б	В	Г
$V=\dfrac{M(X)}{\sigma}\cdot 100\,\%$	$V=\dfrac{\sigma}{M(X)}\cdot 100\,\%$	$V=\dfrac{\sigma}{M(X)}$	$V=M(X)\cdot\sigma$

4.6. Запитання для самоперевірки

4.1. Що називають законом розподілу дискретної випадкової величини та як його можна задати?

4.2. Як побудувати многокутник розподілу?

4.3. Як перевірити правильність складеного закону розподілу випадкової величини?

4.4. Які висновки можна зробити за складеним законом розподілу випадкової величини?

4.5. Які числові характеристики випадкової величини вам відомі?

4.6. Що таке "математичне сподівання"?

4.7. Назвіть властивості математичного сподівання.

4.8. Дайте означення дисперсії.

4.9. Назвіть властивості дисперсії. Що вона характеризує?

4.10. Який вигляд має формула для обчислення дисперсії?

4.11. Яких значень може набувати дисперсія?

4.12. Як обчислити середнє квадратичне відхилення випадкової величини?

4.13. Запишіть формули для обчислення початкового і центрального моментів k-го порядку.

4.14. Що називають модою випадкової величини?

4.15. Що називають медіаною випадкової величини?

4.16. Що характеризує коефіцієнт варіації та як його обчислити?

4.17. Як обчислити коефіцієнт ексцесу?

4.18. Як обчислити коефіцієнт асиметрії?

4.7. Висновки за темою

Після вивчення теми студент має вміти:

складати закон розподілу дискретної випадкової величини та робити контрольну перевірку;

будувати многокутник розподілу;

створювати функцію розподілу та будувати її графік;

обчислювати основні числові характеристики дискретної випадкової величини;

обчислювати додаткові числові характеристики;

знаходити початкові та центральні теоретичні моменти;

робити висновки.

5. Закони розподілу та числові характеристики дискретної випадкової величини

5.1. Мета та компетентності

Метою вивчення теми є ознайомлення студентів з основними законами розподілу дискретних випадкових величин, навчання їх складати закони розподілу дискретних випадкових величин та обчислювати їхні числові характеристики.

Компетентності, що формують під час вивчення теми:

знання означення законів розподілу дискретних випадкових величин;

уміння будувати функцію розподілу випадкової величини за певним законом розподілу;

знання основних числових характеристик випадкової величини та вміння їх обчислювати;

уміння робити висновки.

5.2. Основні теоретичні відомості

Біноміальним називають закон розподілу дискретної випадкової величини X – кількості появи події в n незалежних випробуваннях, у кожному з яких імовірність появи події дорівнює p, якщо відповідні значення ймовірностей обчислюють за формулою Бернуллі:

$$P_n(X=k)=C_n^k p^k q^{n-k},\ (k=0,1,2...,n).$$

Біноміальний закон розподілу є двопараметричним, оскільки визначається параметрами n (кількість випробувань) та p (імовірність появи події в окремому випробуванні).

Математичне сподівання і дисперсію випадкової величини, розподіленої за біноміальним законом, визначають за такими формулами:

$$M\left(X\right)=np,\ D\left(X\right)=npq.$$

Випадкову величину X розподілено за **пуассонівським законом**, якщо її ймовірності визначають за формулою Пуассона:

$$P_n(X=k)=\frac{\lambda^k e^{-\lambda}}{k!},$$

де $\lambda = np$.

Закон розподілу Пуассона є однопараметричним, єдиним його параметром є $\lambda = np$.

Для випадкової величини, розподіленої за законом Пуассона

$$M\left(X\right)\approx D\left(X\right)=\lambda.$$

Ще одним прикладом закону розподілу дискретної випадкової величини, що можна задати в аналітичній формі, є **геометричний розподіл**. Нехай здійснюють незалежні й однорідні випробування, для кожного з яких імовірність p появи певної події є сталою величиною. У цих випробуваннях спостерігають за появою певної події. Випробування тривають доти, доки з'явиться ця подія. Випадковою величиною є кількість випробувань, яка передує появі цієї події. Випадкова величина X має **геометричний закон розподілу**, якщо відповідні значення її ймовірностей обчислюють за такою формулою:

$$P_n(X=k)=(1-p)^k\cdot p,\ \ k=0,1,2...,k.$$

Геометричний розподіл є однопараметричним і його визначають параметром p.

Нехай є партія з N одиниць товару, із яких M одиниць є якісними. Із цієї партії вибирають n одиниць товару таким чином, що кожна одиниця має однакову ймовірність потрапити до вибірки. Якщо випадковою величиною вважати кількість m одиниць якісного товару серед множини n, то закон розподілу цієї випадкової величини називають гіпергеометричним. Випадкова величина X має **гіпергеометричний закон розподілу**, якщо відповідні значення її ймовірностей обчислюють за такою формулою:

$$P_N(X=k)=\frac{C_n^k C_{N-n}^{m-k}}{C_N^m},$$

де C_N^n – кількість сполучень із N по n, тобто загальна кількість можливостей вибрати n елементів із множини, що містить N елементів;

C_M^m – кількість сполучень із M по m, тобто кількість можливостей вибрати m якісних елементів із їхньої множини, що містить M елементів;

C_{N-M}^{n-m} – кількість сполучень із $N-M$ по $n-m$, тобто кількість можливостей заповнити $n-m$ вільних місць (ці місця залишилися у вибірковій сукупності після того, як m місць було зайнято якісними елементами) неякісними елементами, загальна кількість яких становить $N-M$.

Відповідно, добуток $C_M^m \cdot C_{N-M}^{n-m}$ дорівнює кількості сприятливих елементарних подій на просторі елементарних подій, загальна кількість яких становить C_N^n. Гіпергеометричний закон розподілу характеризують за трьома параметрами: N, M та n.

5.3. Приклади розв'язування задач

Приклад 5.1. Імовірність скласти іспит на "відмінно" для кожного із 6 студентів дорівнює 0,4. Складіть закон розподілу кількості відмінних оцінок, отриманих студентами на іспиті. Знайдіть $M(X)$ і $D(X)$.

Розв'язання. Нехай випадкова величина X – кількість відмінних оцінок, отриманих студентами на іспиті.

Вона може набувати значень: 0, 1, 2, 3, 4, 5, 6.

За умовою $p=0{,}4; q=0{,}6.$

Імовірності, із якими величина X набуває своїх можливих значень, слід обчислювати за формулою Бернуллі:

$$P_6(0)=C_6^{\,0}\cdot 0{,}4^0\cdot 0{,}6^6\approx 0{,}047, \quad (C_6^{\,0}=1);$$

$$P_6(1)=C_6^{\,1}\cdot 0{,}4^1\cdot 0{,}6^5\approx 0{,}187;$$

$$P_6(2)=C_6^{\,2}\cdot 0{,}4^2\cdot 0{,}6^4\approx 0{,}311;$$

$$P_6(3)=C_6^{\,3}\cdot 0{,}4^3\cdot 0{,}6^3\approx 0{,}276;$$

$$P_6(4)=C_6^{\,4}\cdot 0{,}4^4\cdot 0{,}6^2\approx 0{,}138;$$

$$P_6(5)=C_6^{\,5}\cdot 0{,}4^5\cdot 0{,}6^1\approx 0{,}037;$$

$$P_6(6)=0{,}4^6\approx 0{,}004.$$

Закон розподілу цієї випадкової величини X зручно навести у вигляді таблиці (табл. 5.1).

Таблиця 5.1

Закон розподілу випадкової величини X

x_i	0	1	2	3	4	5	6
p_i	0,047	0,187	0,311	0,276	0,138	0,037	0,004

Перевірка: $0,047+0,187+0,311+0,276+0,138+0,037+0,004=1.$

Із табл. 5.1 видно: найбільш імовірно, що цими студентами буде отримано 2 оцінки "відмінно", тому що $P_6(2)=0,311$ є найбільшою ймовірністю. Практично неможливо отримати 6 оцінок "відмінно" тому, що $P_6(6)=0,004$. Малоймовірно отримати 5 оцінок "відмінно" або жодної ($P_6(0)=0,047$ і $P_6(5)=0,037$). Але практично вірогідно отримати від 1 до 4 оцінок "відмінно" тому, що

$$P_6(1\le k\le 4)=0,187+0,311+0,276+0,138=0,912.$$

$$M(X)=n\cdot p=6\cdot 0,4=2,4;\ D(X)=n\cdot p\cdot q=6\cdot 0,4\cdot 0,6=1,44.$$

Приклад 5.2. Припускаючи однаковими ймовірності народження хлопчика і дівчинки, складіть закон розподілу випадкової величини X, яка обчислює кількість хлопчиків у родині, що має 5 дітей.

Розв'язання. Випадкова величина X може набувати 6 значень: 0, 1, 2, 3, 4 і 5. Імовірності того, що вона набуде кожне з них, слід шукати за формулою Бернуллі $P_n(m)=C_n^m p^m q^{n-m}$, за $n=5,\ p=0,5$ і $m=0,\dots,5$. Після цього слід скласти такий закон розподілу випадкової величини X (табл. 5.2):

Таблиця 5.2

Закон розподілу випадкової величини X

x_i	0	1	2	3	4	5
p_i	0,03125	0,15623	0,31250	0,31250	0,15625	0,03125

Очевидно, що $\sum_{i=1}^{6} p_i=1$.

Із табл. 5.2 видно, що найбільш імовірно, що в родині буде 2 хлопчики та 3 дівчинки або навпаки.

Приклад 5.3. Обладнання складається зі 100 елементів, що працюють незалежно один від одного. Імовірність відмови будь-якого елемента $p = 0,01$. Складіть закон розподілу кількості відмов.

Розв'язання. Нехай випадкова величина X – кількість відмов елементів, яка може набувати значень 0, 1, 2, 3, ..., 100.

Імовірності цих значень знаходять за формулою Пуассона тому, що $n = 100$ – велика, а $p = 0,01$ – мала ймовірність.

За умовою $\lambda = 100 \cdot 0,01 = 1$.

$$P_n(X = k) = \frac{\lambda^k}{k!} \cdot e^{-\lambda}, \quad (e^{-1} \approx 0,3679);$$

$$P(X = 0) = \frac{e^{-1}}{0!} \approx 0,3679; \quad P(X = 1) = \frac{e^{-1}}{1!} \approx 0,3679;$$

$$P(X = 2) = \frac{e^{-1}}{2!} \approx 0,1839; \quad P(X = 3) = \frac{e^{-1}}{3!} \approx 0,0613;$$

$$P(X = 4) = \frac{e^{-1}}{4!} \approx 0,0153; \quad P(X = 5) = \frac{e^{-1}}{5!} \approx 0,0031;$$

$$P(X = 6) = \frac{e^{-1}}{6!} \approx 0,0005.$$

За $k \geq 7$ імовірності $P(X \geq 7) \approx 0$. Отже, закон розподілу заданої випадкової величини має такий вигляд (табл. 5.3):

Таблиця 5.3

Закон розподілу випадкової величини X

x_i	0	1	2	3	4	5	6
p_i	0,3679	0,3679	0,1839	0,0613	0,0153	0,0031	0,0005

Перевірка:

$$0,3679 + 0,3679 + 0,1839 + 0,0613 + 0,0153 + 0,0031 + 0,0005 = 1.$$

Аналіз табл. 5.3 дозволяє зробити деякі висновки. Найбільш імовірно, що відмовить 1 або жодний елемент. Практично неможливо, що відмовлять 5 або більше елементів.

Приклад 5.4. Звітність містить певні неточності, які за експрес-перевірки виявляють з імовірністю 0,7. Здійснюють перевірку всієї документації до виявлення першого порушення. Випадковою величиною вважають кількість документів, у яких не було помічено порушень. Складіть закон розподілу цієї випадкової величини.

Розв'язання. Оскільки випробування здійснюють до першої появи події (виявлення порушення) і ймовірність появи події в кожному випробуванні є сталою величиною, то закон розподілу випадкової величини є геометричним. Випадкова величина може набувати значень 1, 2, 3, …, отже, це нескінчена дискретна випадкова величина. Для обчислення ймовірності кожного її значення слід користуватися формулою:

$$P_n(X=k)=(1-p)^k\cdot p,\;\; k=0,1,2...,k.$$

Так,

$$P_1(X=0)=(1-0,7)^0\cdot 0,3=0,7;$$

$$P_2(X=1)=(1-0,7)^1\cdot 0,7=0,21 \text{ і далі.}$$

Результати обчислення наведено в табл. 5.4. Видно, що випадкова подія відбувається в першому ж випробуванні з імовірністю 0,7. Зі збільшенням значення випадкової величини його ймовірність швидко спадає, і можливість пропустити помилки навіть у 2 документах є подією малоймовірною.

Таблиця 5.4

Ряд розподілу дискретної випадкової величини, розподіленої за геометричним законом

X	0	1	2	3	4	5
p_i	0,7000	0,2100	0,0630	0,0189	0,0057	0,0017

Імовірність події "пропустити не більше ніж п'ять помилок" дорівнює сумі ймовірностей перших шести подій (ураховуючи подію "не пропустити жодної помилки"), у цьому разі вона становить 0,9993.

Приклад 5.5. Магазин отримав партію товару, що складається із 20 опалювальних котлів. Під час транспортування 5 із них зазнали незначних пошкоджень корпусу. Із цієї партії для філіалу було передано

4 котли. Складіть ряд розподілу випадкової величини, якою вважають кількість пошкоджених котлів серед тих, що відібрані для філіалу.

Розв'язання. Оскільки з партії товару вибирають її частину, де кожний з об'єктів має однакову ймовірність потрапити до вибірки, то мова йде про гіпергеометричний розподіл. Значення випадкової величини можуть змінюватися від 0 до 4.

За такою формулою:

$$P_N(X=k)=\frac{C_n^k C_{N-n}^{m-k}}{C_N^m}$$

буде:

$$P_4(X=0)=\frac{C_5^0\cdot C_{15}^4}{C_{20}^4}=\frac{5!}{5!\cdot 0!}\cdot\frac{15!}{11!\cdot 4!}\cdot\frac{16!\cdot 4!}{20!}=0,2817;$$

$$P_4(X=1)=\frac{C_5^1\cdot C_{15}^3}{C_{20}^4}=\frac{5!}{4!\cdot 1!}\cdot\frac{15!}{12!\cdot 3!}\cdot\frac{16!\cdot 4!}{20!}=0,4696;$$

$$P_4(X=2)=\frac{C_5^2\cdot C_{15}^2}{C_{20}^4}=\frac{5!}{3!\cdot 2!}\cdot\frac{15!}{13!\cdot 2!}\cdot\frac{16!\cdot 4!}{20!}=0,2167;$$

$$P_4(X=3)=\frac{C_5^3\cdot C_{15}^1}{C_{20}^4}=\frac{5!}{2!\cdot 3!}\cdot\frac{15!}{14!\cdot 1!}\cdot\frac{16!\cdot 4!}{20!}=0,0310;$$

$$P_4(X=4)=\frac{C_5^4\cdot C_{15}^0}{C_{20}^4}=\frac{5!}{1!\cdot 4!}\cdot\frac{15!}{15!\cdot 0!}\cdot\frac{16!\cdot 4!}{20!}=0,0010.$$

Для перевірки того, що перелічено всі можливі значення випадкової величини, слід обчислити суму ймовірностей. Дійсно, $\sum_{i=1}^{n} p_i=1$.

Зверніть увагу, що 1 котел із 4 становить ті самі 25 %, що і 5 котлів із 20, тобто випадкова подія $X=1$ має бути найімовірнішою, але хоча вона і є найімовірнішою, імовірність того, що випадкова величина набуде значення $X=1$ становить лише 0,4696.

5.4. Вправи для самостійної роботи

5.1. Складіть закон розподілу кількості влучення в мішень за 4 постріли, якщо ймовірність улучення за кожного пострілу дорівнює 0,3.

Обчисліть математичне сподівання і дисперсію цієї випадкової величини, користуючись тільки їхніми означеннями, а результати перевірте за формулами цих характеристик випадкової величини, розподіленої за біноміальним законом.

5.2. Виконують 8 пострілів у ціль. Імовірність улучення в ціль за кожного пострілу дорівнює 0,6. Нехай X – кількість улучень у ціль. Знайдіть функцію розподілу випадкової величини X. Визначте ймовірність того, що буде не менше ніж 7 улучень у ціль. Знайдіть $M(X)$, $D(X)$.

5.3. Імовірність виграти за лотерейним квитком дорівнює 0,4. Придбано 30 квитків. Визначте математичне сподівання і дисперсію кількості квитків, на які дістануть виграші.

5.4. Знайдіть функцію розподілу кількості влучень у ціль, якщо зроблено 6 пострілів, а ймовірність улучення за одного пострілу дорівнює 0,2. Користаючись цією функцією, знайдіть імовірність того, що ціль буде вражено не менше від 1 разу, але менше ніж 5 разів.

5.5. Складіть закон розподілу кількості влучення в мішень за 4 постріли, якщо ймовірність улучення за кожного пострілу дорівнює 0,3. Обчисліть математичне сподівання і дисперсію цієї випадкової величини, користуючись тільки їхніми означеннями.

5.6. У місті 10 комерційних банків. У кожного ризик банкрутства протягом року становить 10 %. Складіть ряд розподілу кількості банків, які можуть збанкрутувати протягом наступного року; побудуйте його графік. Знайдіть числові характеристики його розподілу. Запишіть функцію розподілу ймовірностей і побудуйте її графік. Обчисліть імовірність того, що протягом року збанкрутує не більше від 1 банку.

5.7. Деякий пристрій містить 8 деталей. У пристрої застосовують деталі з імовірністю браку 0,01. Знайдіть імовірність того, що пристрій виявиться придатним, знаючи, що він виходить із ладу, коли кількість зіпсованих деталей не менша від 2.

5.8. Підручник видано тиражем 100 000 примірників. Імовірність того, що підручник зброшуровано неправильно, дорівнює 0,0001. Знайдіть імовірність того, що тираж містить тільки 5 бракованих книг.

5.9. Випадкову величину X розподілено за законом Пуассона з параметром $\lambda = 1$. Знайдіть імовірності подій: а) $X = 1$; б) $X \geq 1$; в) $X \leq 2$; г) $X = 0$; д) $X > 1$; е) $1 \leq X \leq 3$.

5.10. Верстат-автомат штампує деталі. Імовірність того, що виготовлена деталь буде бракованою, дорівнює 0,01. Знайдіть імовірність того, що серед 200 деталей буде 3 браковані. Визначте математичне сподівання і дисперсію кількості бракованих деталей.

5.11. Монету кидають 5 разів. Складіть закон розподілу кількості появ "герба". Знайдіть імовірність того, що "герб" випаде не менше від 3 разів.

5.12. Імовірність того, що аудитор зробить помилку під час перевірки бухгалтерського балансу, дорівнює 0,05. Аудитор має зробити висновок про 2 баланси. Складіть закон розподілу кількості правильних висновків за балансами, що перевіряють. Обчисліть основні числові характеристики. Зробіть висновки.

5.13. На телефонну станцію надходять у середньому 2 виклики за 1 хв. Знайдіть імовірність того, що на телефонну станцію: а) протягом 1 хв надійде 3 виклики; б) протягом 0,5 хв точно надійде один виклик; в) протягом 24 с не надійде жодного виклику; г) протягом 2 хв надійде хоча б 1 виклик.

5.14. Середня кількість замовлень на таксі, що надходять на диспетчерський пункт за 1 хв, дорівнює 3. Знайдіть імовірність того, що за 2 хв надійде: а) 4 виклики; б) менше від 4 викликів; в) не менше від 4 викликів.

5.15. Мисливець стріляє по дичині до першого влучення, але встигає зробити не більше ніж 5 пострілів. Імовірність улучення в ціль за 1 пострілу дорівнює 0,8. Складіть закон розподілу кількості промахів. Обчисліть математичне сподівання і середнє квадратичне відхилення цієї випадкової величини.

5.5. Тестові завдання

5.1. Значення ймовірностей біноміального закону розподілу обчислюють за такою формулою

А	Б	В	Г
$\dfrac{\lambda^k e^{-\lambda}}{k!}$	$(1-p)^k \cdot p$	$C_n^k p^k q_n^{n-k}$	$\dfrac{C_n^k C_{N-n}^{m-k}}{C_N^m}$

5.2. Біноміальний закон розподілу є

А однопараметричним

Б двопараметричним

В трипараметричним

5.3. Значення ймовірностей закону розподілу Пуассона обчислюють за такою формулою

А	Б	В	Г
$\dfrac{\lambda^k e^{-\lambda}}{k!}$	$(1-p)^k \cdot p$	$C_n^k p^k q_n^{n-k}$	$\dfrac{C_n^k C_{N-n}^{m-k}}{C_N^m}$

5.4. Значення ймовірностей геометричного закону розподілу обчислюють за такою формулою

А	Б	В	Г
$\dfrac{\lambda^k e^{-\lambda}}{k!}$	$(1-p)^k \cdot p$	$C_n^k p^k q_n^{n-k}$	$\dfrac{C_n^k C_{N-n}^{m-k}}{C_N^m}$

5.5. Значення ймовірностей гіпергеометричного закону розподілу обчислюють за такою формулою

А	Б	В	Г
$\dfrac{\lambda^k e^{-\lambda}}{k!}$	$(1-p)^k \cdot p$	$C_n^k p^k q_n^{n-k}$	$\dfrac{C_n^k C_{N-n}^{m-k}}{C_N^m}$

5.6. Параметрами гіпергеометричного закону розподілу є

А	Б	В	Г
$n,\ p$	N, M, n	N, n	λ

5.7. Параметрами геометричного закону розподілу є

А	Б	В	Г
$n,\ p$	N, M, n	p	λ

5.8. Математичне сподівання біноміального закону розподілу дорівнює

А	Б	В	Г
np	npq	p	λ

5.9. Математичне сподівання приблизно дорівнює дисперсії для випадкової величини заданої

А біноміальним законом

Б законом розподілу Пуассона

В геометричним законом

Г гіпергеометричним законом

5.10. Дисперсія біноміального закону розподілу дорівнює

А	Б	В	Г
np	npq	q	λ

5.6. Запитання для самоперевірки

5.1. Які існують способи задавання дискретної випадкової величини?

5.2. У чому полягає різниця між скінченими та нескінченими випадковими величинами?

5.3. Сформулюйте принципи, яким має відповідати ряд розподілу випадкової величини.

5.4. Які основні числові характеристики використовуються для дискретних випадкових величин? Який сенс має кожна з основних числових характеристик розподілу і як їх визначають?

5.5. Який закон розподілу дискретної випадкової величини називають біноміальним? Які параметри його характеризують?

5.6. Сформулюйте умови застосування формули Пуассона, поясніть, чому цей розподіл називають законом виключних подій. Скількома параметрами його задають?

5.7. Який закон розподілу дискретної випадкової величини називають геометричним? Скількома параметрами його задають?

5.8. Який закон розподілу дискретної випадкової величини називають гіпергеометричним? Скількома параметрами його задають?

5.7. Висновки за темою

Після вивчення теми студент має вміти:

чітко визначати закони розподілу випадкових дискретних величин;

складати закон розподілу й обчислювати його числові характеристики;

надавати геометричну інтерпретацію побудованого закону розподілу;

робити висновки за побудованим розподілом.

6. Основні закони розподілу неперервної випадкової величини

6.1. Мета та компетентності

Метою вивчення теми є ознайомлення з основними законами розподілу неперервної випадкової величини та формування системи теоретичних знань, практичних умінь і навичок у визначенні функції розподілу та основних числових характеристик неперервної випадкової величини.

Компетентності, що формують під час вивчення теми:

здатність обчислювати основні числові характеристики неперервної (одновимірної) випадкової величини, визначати зв'язок між функціями розподілу та щільності ймовірностей;

знання основних законів розподілу неперервної випадкової величини, функції розподілу щільності ймовірностей, розуміння зв'язку між розподілами, заданими різними законами;

уміння за функцією щільності ймовірностей визначати ймовірність улучення значень випадкової величини до заданого інтервалу та інтервал, до якого значення випадкової величини будуть належати із заздалегідь заданою надійністю;

навички у застосовуванні функцій розподілу та щільності ймовірностей для визначення характеристик розподілу неперервної випадкової величини, використанні табличних даних для здійснення обчислень.

6.2. Основні теоретичні відомості

Для неперервної випадкової величини функція розподілу є неперервною, отже, для завдання неперервної випадкової величини разом із функцією розподілу можна використовувати першу похідну від цієї функції:

$$f(x) = F'(x),$$

де $f(x)$ – **щільність імовірностей**, або **диференціальна функція розподілу** (на відміну від функції $F(x)$, яку іноді називають інтегральною функцією розподілу).

Відповідно, функція розподілу є невласним інтегралом зі змінною верхньою межею від щільності ймовірностей:

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

Властивості диференціальної функції розподілу

1. Щільність імовірностей є невід'ємною:

$$f(x) \geq 0.$$

2. Невласний інтеграл від щільності ймовірностей у межах від $-\infty$ до $+\infty$ дорівнює одиниці:

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

3. Імовірність того, що неперервна випадкова величина набуде значення, яке належить до інтервалу $(a; b)$, дорівнює визначеному інтегралу від диференціальної функції розподілу, межами якого є межі цього інтервалу:

$$P(a < X < b) = \int_{a}^{b} f(x)dx = F(b) - F(a).$$

Основні числові характеристики неперервної випадкової величини

Математичне сподівання неперервної випадкової величини X визначають за співвідношенням:

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

а дисперсія за означенням має такий вигляд:

$$DX = \int_{-\infty}^{+\infty} (x - MX)^2 \cdot f(x)dx,$$

де $f(x)$ – щільність імовірностей.

Формула для обчислення дисперсії має такий вигляд:

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

Поняття початкового та центрального теоретичних моментів, а також моди, медіани, асиметрії та ексцесу для неперервної випадкової величини формально визначають тими самими співвідношеннями, що й для дискретної випадкової величини.

6.2.1. Закони розподілу неперервної випадкової величини

Слід розглянути найбільш поширені закони розподілу неперервної випадкової величини.

Рівномірним називають закон розподілу неперервної випадкової величини, диференціальна функція розподілу якого має такий вигляд:

$$f(x)=\begin{cases}0, & \text{якщо} \quad x\le a;\\ \dfrac{1}{b-a}, & \text{якщо} \quad a<x\le b;\\ 0, & \text{якщо} \quad x>b.\end{cases}$$

Отже, на проміжку $(a;b]$, до якого належать усі можливі значення випадкової величини, щільність імовірностей є сталою величиною.

Рівномірний закон розподілу визначають за двома параметрами: початком (a) і кінцем (b) інтервалу, до якого належать усі значення неперервної випадкової величини. Отже, він є двопараметричним.

Функція розподілу випадкової величини, розподіленої за рівномірним законом, має такий вигляд:

$$F(x)=\begin{cases}0, & \text{якщо} \quad x\le a;\\ \dfrac{x-a}{b-a}, & \text{якщо} \quad a<x\le b;\\ 1, & \text{якщо} \quad x>b.\end{cases}$$

На рис. 6.1. наведено графіки щільності ймовірностей та функції розподілу випадкової величини, розподіленої за рівномірним законом.

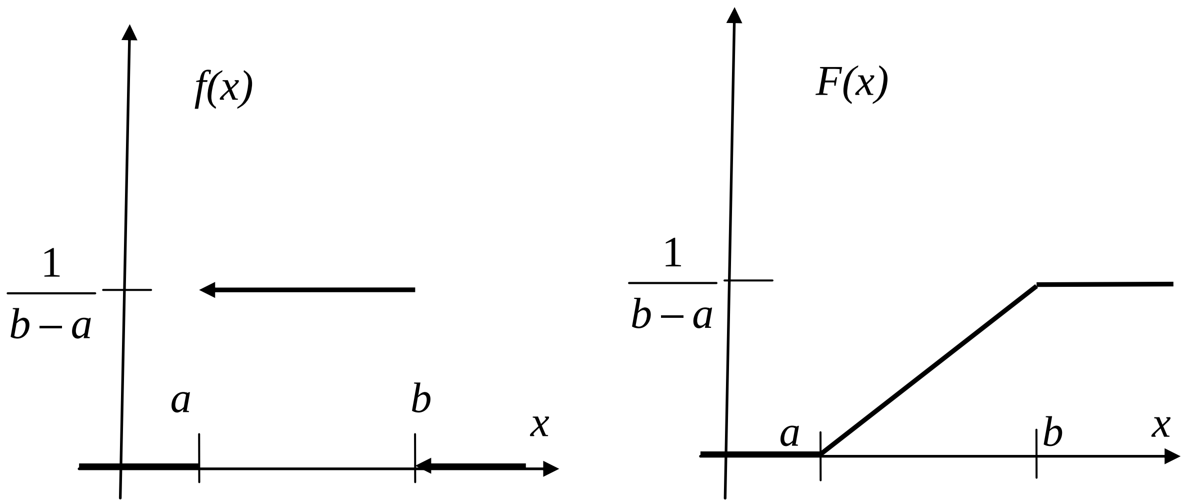


Рис. 6.1. **Щільність імовірностей** $f(x)$ **та функція розподілу** $F(x)$ **випадкової величини, розподіленої за рівномірним законом**

Основні числові характеристики випадкової величини, розподіленої за рівномірним законом такі:

математичне сподівання:

$$M\left(X\right)=\frac{a+b}{2};$$

дисперсія:

$$D\left(X\right)=\frac{(b-a)^2}{12};$$

середнє квадратичне відхилення:

$$\sigma=\frac{b-a}{2\sqrt{3}}.$$

Неперервну випадкову величину ξ називають розподіленою за **показниковим**, або **експоненціальним законом**, якщо її щільність імовірності має такий вигляд:

$$f(x)=\begin{cases}0, & \text{якщо } x\le 0;\\ \lambda e^{-\lambda x}, & \text{якщо } x>0,\end{cases}$$

де λ – параметр функції показникового розподілу.

Для цього закону функція розподілу має такий вигляд:

$$F(x)=\begin{cases}0, & \text{якщо } x\le 0;\\ 1-e^{-\lambda x}, & \text{якщо } x>0,\end{cases}$$

Показниковий розподіл є однопараметричним із параметром λ.

На рис. 6.2. наведено графіки щільності ймовірностей та функції розподілу неперервної випадкової величини, розподіленої за експоненціальним законом.

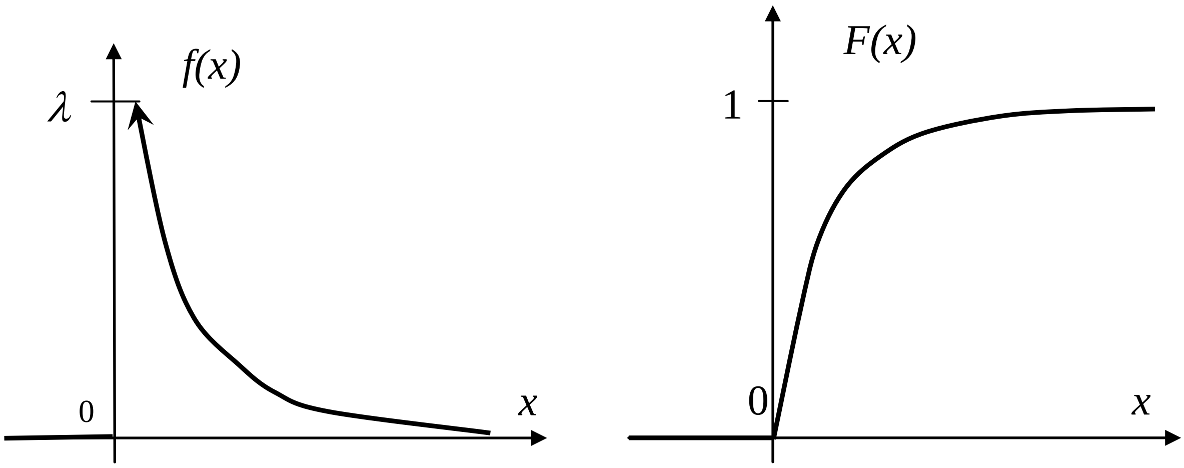


Рис. 6.2. **Щільність імовірностей $f(x)$ та функція розподілу $F(x)$ випадкової величини, розподіленої за експоненціальним законом**

Основні числові характеристики випадкової величини, розподіленої за експоненціальним законом такі:

математичне сподівання:

$$M(X)=\frac{1}{\lambda};$$

дисперсія:

$$D(X)=\frac{1}{\lambda^2};$$

середнє квадратичне відхилення:

$$\sigma=\frac{1}{\lambda}.$$

Отже, для експоненціального закону має місце співвідношення:

$$M(X)=\sigma.$$

Відповідно, коефіцієнт варіації дорівнює 100 %:

$$V = \frac{\sigma}{M(X)} \cdot 100\,\% = 100\,\%.$$

Найбільш поширеним законом розподілу неперервної випадкової величини є нормальний розподіл.

Випадкову величину X уважають розподіленою за **нормальним законом**, якщо її диференціальна функція розподілу (щільність імовірності) має такий вигляд:

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

де a – математичне сподівання випадкової величини X;

σ – середнє квадратичне відхилення випадкової величини.

Основні числові характеристики випадкової величини, розподіленої за нормальним законом такі:

математичне сподівання:

$$M(X) = a;$$

дисперсія:

$$D(X) = \sigma^2.$$

Функція розподілу для нормального закону має такий вигляд:

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

Отже, нормальний закон розподілу є двопараметричним, його параметрами є математичне сподівання a та середнє квадратичне відхилення σ. Випадкову величину, розподілену за нормальним законом, позначають так:

$$X \sim N(a; \sigma^2),$$

де a і σ – параметри розподілу.

На рис. 6.3 наведено графіки інтегральної та диференціальної функцій розподілу для нормального закону.

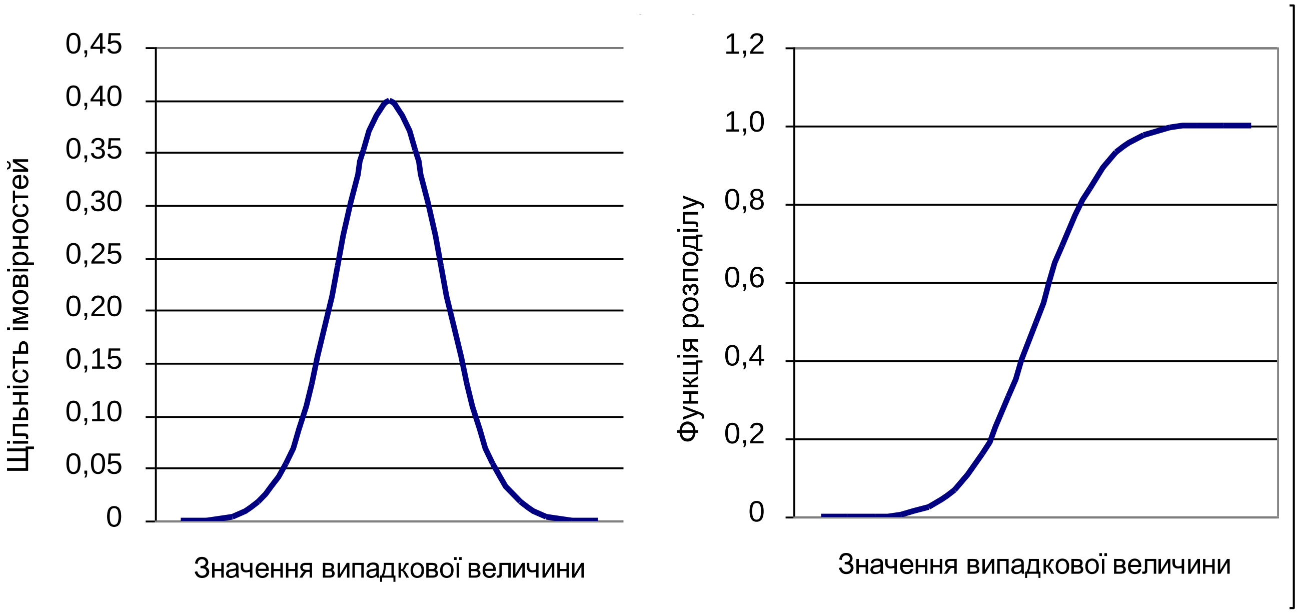


Рис. 6.3. **Графіки функції щільності ймовірностей $f(x)$ та функції розподілу $F(x)$ для нормального закону**

Головна особливість нормального розподілу полягає в тому, що цей закон є граничним, тобто до нього за певних умов наближаються інші закони розподілу.

Оскільки крива є симетричною, то для нормального закону мода випадкової величини дорівнює медіані й обидві вони збігаються з математичним сподіванням:

$$Mo(X) = Me(X) = M(X) = a.$$

Зміна параметра a призводить до зсуву кривої вздовж осі OX.

Параметр σ характеризує розпорошення випадкової величини навколо математичного сподівання. Середнє квадратичне відхилення дорівнює відстані від точок перегину на кривій $f(x)$ до точки максимуму, яка відповідає значенню математичного сподівання. Отже, чим більше значення має σ, тим ширша крива і, відповідно, тим нижчий максимум, оскільки площа під кривою $f(x)$ є сталою величиною й дорівнює одиниці.

За $a = 0$ та $\sigma = 1$ функція щільності ймовірностей має такий вигляд:

$$\varphi(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}.$$

Таким чином, функція $\varphi(x)$ є диференціальною функцією розподілу випадкової величини X, розподіленої за нормальним законом, що має числові характеристики $M\left(X\right)=0$ та $D\left(X\right)=1$. Вона ще має назву **нормованої** кривої, або **функції Гаусса**.

Функція розподілу випадкової величини, розподіленої за нормальним законом, має такий вигляд:

$$F(x)=0{,}5+\varPhi(x).$$

Для випадкової величини, розподіленої за нормальним законом, коефіцієнт асиметрії $A_s=0$ і ексцес $E_k=0$.

У тому разі, якщо закон розподілу є наближеним до нормального, за коефіцієнтом асиметрії порівнюють внески значень випадкової величини, які менші та більші за математичне сподівання.

Якщо $A_s>0$, то внесок значень випадкової величини, більших за математичне сподівання, більш вагомий, отже, крива розподілу розмита в бік більших значень випадкової величини. Якщо $A_s<0$, то крива розподілу розмита в бік менших значень.

На рис. 6.4 показано вплив асиметрії на вигляд диференціальної функції розподілу.

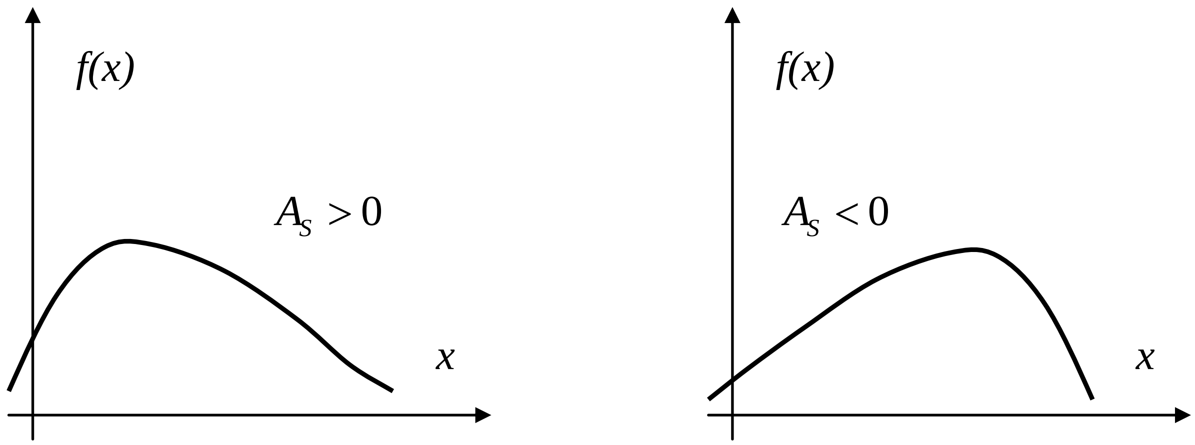


Рис. 6.4. **Вигляд диференціальної функції розподілу, залежно від знака асиметрії**

Ексцес характеризує гостровершинність кривої розподілу, порівняно із кривою Гаусса. Якщо $E_k>0$, то вершина кривої розподілу є більш гострою, ніж для кривої Гаусса, якщо $E_k<0$, то є більш пологою.

Імовірність того, що значення випадкової величини, розподіленої за нормальним законом, буде належати певному проміжку $(\alpha; \beta)$, визначають за такою формулою:

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

Імовірність відхилення випадкової величини від свого математичного сподівання на величину, що не перевищує δ, обчислюють за такою формулою:

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

Відповідно, відхилення випадкової величини від свого математичного сподівання на величину, що більша за δ, становить:

$$P(|X - a| \geq \delta) = 1 - 2\Phi\left(\frac{\delta}{\sigma}\right).$$

Нехай $\delta = 3\sigma$. Тоді за правилом трьох сигм:

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

Отже, 99,73 % значень випадкової величини належать проміжку $(a - 3\sigma; a + 3\sigma)$.

Згідно із цим правилом, відхилення випадкової величини, розподіленої за нормальним законом, від свого математичного сподівання на величину, більшу за 3σ, є подією малоймовірною. Звідси випливає важливий практичний висновок. Якщо невідомо, який саме закон розподілу має випадкова величина, то, коли правило трьох сигм порушено, є підстави вважати, що цей закон не відповідає нормальному.

На рис. 6.5 наведено графік диференціальної функції розподілу випадкової величини, розподіленої за нормальним законом. Площа заштрихованої фігури відповідає ймовірності випадкової події, яка полягає в тому, що відхилення значення випадкової величини від її математичного сподівання не перевищує її середнього квадратичного відхилення.

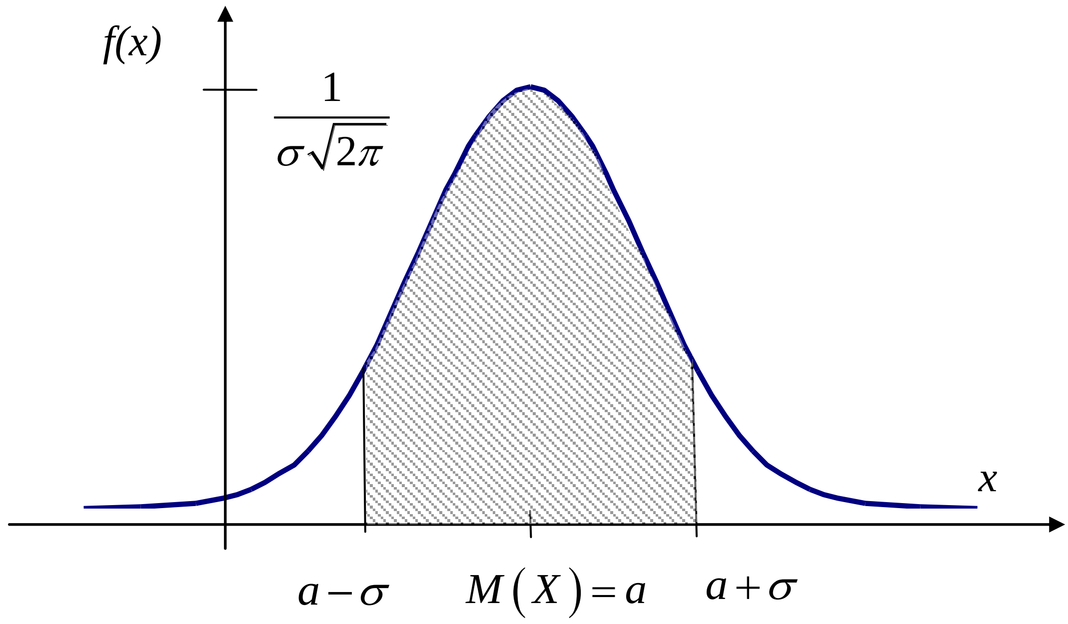


Рис. 6.5. **Загальний вигляд функції щільності ймовірностей для нормального закону розподілу**

6.3. Приклади розв'язування задач

Приклад 6.1. Дано інтегральну функцію випадкової величини X:

$$F(x)=\begin{cases}0, & \text{якщо} \quad x\le 0,\\ \sin 2x, & \text{якщо} \quad 0<x\le \dfrac{\pi}{4},\\ 1, & \text{якщо} \quad x>\dfrac{\pi}{4}.\end{cases}$$

Знайдіть диференціальну функцію $f(x)$.

Розв'язання.

$$f(x)=F'(x)=\begin{cases}0, & \text{якщо} \quad x\le 0,\\ 2\cos 2x, & \text{якщо} \quad 0<x\le \dfrac{\pi}{4},\\ 0, & \text{якщо} \quad x>\dfrac{\pi}{4}.\end{cases}$$

Приклад 6.2. Випадкову величину X задано функцією щільності ймовірностей $f(x)=\dfrac{a}{1+x^2}$, $\quad -\infty<x<+\infty$.

Знайдіть коефіцієнт a функції $f(x)$ і $F(x)$, імовірність того, що випадкова величина набуде будь-якого значення, не меншого за 0, але не більшого за 5.

Розв'язання. Із рівності $\int\limits_{-\infty}^{+\infty}\dfrac{a}{1+x^2}dx=1$ слід знайти a.

$a\cdot\operatorname{arctg}x\Big|_{-\infty}^{+\infty}=1$ або $a\cdot\left(\dfrac{\pi}{2}+\dfrac{\pi}{2}\right)=1$, звідки $a=\dfrac{1}{\pi}$, отже, $f(x)=\dfrac{1}{\pi(1+x^2)}$.

$$F(x)=\int\limits_{-\infty}^{x}f(x)dx=\int\limits_{-\infty}^{x}\frac{1}{\pi(1+x^2)}dx=\frac{\operatorname{arctg}x}{\pi}\Bigg|_{-\infty}^{x}=\frac{\operatorname{arctg}x}{\pi}+\frac{1}{\pi}\cdot\frac{\pi}{2}=$$

$$=\frac{1}{\pi}\operatorname{arctg}x+\frac{1}{2}.$$

$$P(0\le X\le 5)=\int\limits_{0}^{5}f(x)dx=\int\limits_{0}^{5}\frac{1}{\pi(1+x^2)}dx=\frac{\operatorname{arctg}x}{\pi}\Bigg|_{0}^{5}\approx 0,435.$$

Приклад 6.3. Дано диференціальну функцію неперервної випадкової величини X:

$$f(x)=\begin{cases}0, & \text{якщо} \quad x\le\dfrac{\pi}{6},\\ 3\sin 3x, & \text{якщо} \quad \dfrac{\pi}{6}<x\le\dfrac{\pi}{3},\\ 0, & \text{якщо} \quad x>\dfrac{\pi}{3}.\end{cases}$$

Знайдіть інтегральну функцію $F(x)$.

Розв'язання. Для знаходження функції $F(x)$ слід використати таку формулу:

$$F(x)=\int\limits_{-\infty}^{x}f(x)dx.$$

Якщо $x \le \frac{\pi}{6}$, то $f(x) = 0$, а, значить,

$$F(x) = \int_{-\infty}^{x} 0 \cdot dx = 0.$$

Якщо $\frac{\pi}{6} < x \le \frac{\pi}{3}$, то

$$F(x) = \int_{-\infty}^{\frac{\pi}{6}} 0 \cdot dx + \int_{\frac{\pi}{6}}^{x} 3\sin 3x dx = \left.\frac{-3\cos 3x}{3}\right|_{\frac{\pi}{6}}^{x} = -\cos 3x.$$

Якщо $x > \frac{\pi}{3}$, то

$$F(x) = \int_{-\infty}^{\frac{\pi}{6}} 0 \cdot dx + \int_{\frac{\pi}{6}}^{\frac{\pi}{3}} 3\sin 3x dx + \int_{\frac{\pi}{3}}^{x} 0 \cdot dx = \left.\frac{-3\cos 3x}{3}\right|_{\frac{\pi}{6}}^{\frac{\pi}{3}} = -\cos 3x\Big|_{\frac{\pi}{6}}^{\frac{\pi}{3}} = 1.$$

Отже, $F(x) = \begin{cases} 0, & \text{якщо} \quad x \le \frac{\pi}{6}, \\ -\cos 3x, & \text{якщо} \quad \frac{\pi}{6} < x \le \frac{\pi}{3}, \\ 0, & \text{якщо} \quad x > \frac{\pi}{3}. \end{cases}$

Приклад 6.4. Знайдіть математичне сподівання, дисперсію випадкової величини X та побудуйте графіки інтегральної та диференціальної функцій розподілу, якщо

$$F(x) = \begin{cases} 0, & \text{якщо} \quad x \le -1, \\ \frac{x+1}{3}, & \text{якщо} \quad -1 < x \le 2, \\ 1, & \text{якщо} \quad x > 2. \end{cases}$$

Розв'язання. Диференціальна функція має такий вигляд:

$$f(x) = F'(x) = \begin{cases} 0, & \text{якщо} \quad x \le -1, \\ \dfrac{1}{3}, & \text{якщо} \quad -1 < x \le 2, \\ 0, & \text{якщо} \quad x > 2. \end{cases}$$

Для обчислення математичного сподівання та дисперсії слід застосувати такі формули:

$$M(X) = \int_{-\infty}^{+\infty} x \cdot f(x)dx \text{ та } D(X) = \int_{-\infty}^{+\infty} x^2 f(x)dx - M^2(X);$$

$$M(X) = \int_{-\infty}^{-1} x \cdot 0dx + \int_{-1}^{2} x\frac{1}{3}dx + \int_{2}^{\infty} x \cdot 0dx = \left.\frac{x^2}{6}\right|_{-1}^{2} = \frac{4}{6} - \frac{1}{6} = \frac{1}{2};$$

$$D(X) = \int_{-1}^{2} x^2 \frac{1}{3}dx - \frac{1}{4} = \left.\frac{x^3}{9}\right|_{-1}^{2} - \frac{1}{4} = \left(\frac{8}{9} - \frac{1}{9}\right) - \frac{1}{4} = \frac{19}{36}.$$

На рис. 6.6 та 6.7 зображено графіки інтегральної функції розподілу $F(x)$ і щільності розподілу $f(x)$.

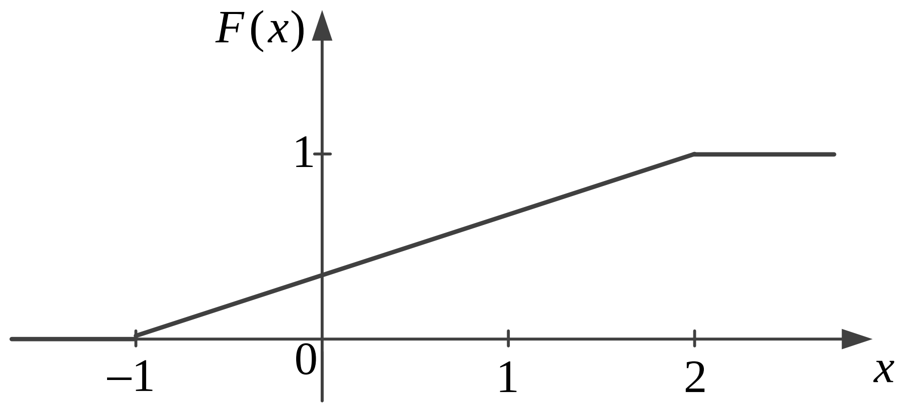


Рис. 6.6. **Графік інтегральної функції розподілу**

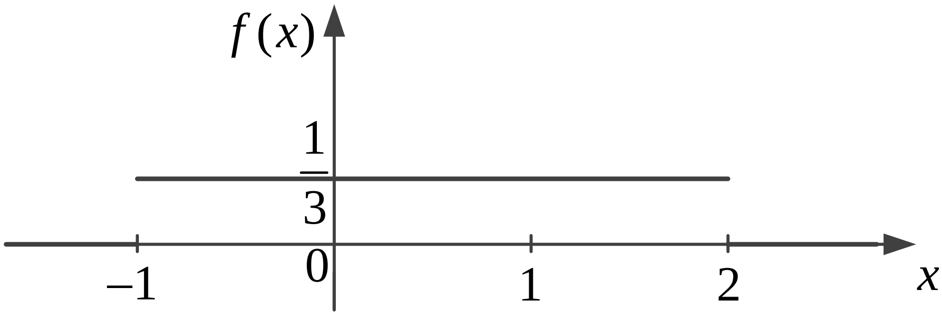


Рис. 6.7. **Графік диференціальної функції розподілу**

Приклад 6.5. Випадкову величину розподілено на проміжку $(0; 24]$ за рівномірним законом. Визначте її основні числові характеристики.

Розв'язання. За умовою $a = 0$ та $b = 24$. Обчисліть основні числові характеристики розподілу:

$$M(X) = \frac{a+b}{2} = \frac{0+24}{2} = 12;$$

$$D(X) = \frac{(b-a)^2}{12} = \frac{(24-0)^2}{12} = 48;$$

$$\sigma = \sqrt{D(X)} = \sqrt{48} \approx 6{,}93.$$

Приклад 6.6. Усі значення рівномірно розподіленої випадкової величини належать відрізку $[2; 8]$. Знайдіть імовірність того, що випадкова величина набуде значення з інтервалу $(3; 5)$.

Розв'язання. Для знаходження ймовірностей застосовуйте умову:

$$P(\alpha < X < \beta) = F(b) - F(a).$$

Для рівномірно розподіленої випадкової величини:

$$F(x) = \begin{cases} 0, & \text{якщо} \quad x \le a, \\ \dfrac{x-a}{b-a}, & \text{якщо} \quad a < x \le b, \\ 1, & \text{якщо} \quad x > b. \end{cases}$$

Тому в цьому прикладі:

$$F(x)=\begin{cases}0, & \text{якщо} \quad x\le 2,\\ \dfrac{x-2}{6}, & \text{якщо} \quad 2<x\le 8,\\ 1, & \text{якщо} \quad x>8.\end{cases}$$

звідки $P(3<X<5)=F(5)-F(3)=\dfrac{5-2}{6}-\dfrac{3-2}{6}=\dfrac{3}{6}-\dfrac{1}{6}=\dfrac{1}{2}.$

Приклад 6.7. У деякий час потяги метро рухаються з інтервалом 5 хв. Пасажир приходить на станцію в деякий час. Яка ймовірність появи пасажира не раніше ніж через 1 хв після відходу попереднього, але не пізніше ніж за 2 хв до підходу наступного потяга?

Розв'язання. Час прибуття пасажира на станцію можна розглядати як випадкову величину X, розподілену рівномірно в інтервалі руху, тривалістю 5 хв:

$$f(x)=\frac{1}{5}.$$

Треба знайти $P(1<X<3)$.

За формулою $P(\alpha<X<\beta)=\int\limits_{\alpha}^{\beta} f(x)dx$ буде:

$$P(1<X<3)=\int\limits_1^3 f(x)dx=5\int\limits_1^3\frac{1}{5}dx=\frac{1}{5}x\Big|_1^3=\frac{1}{5}(3-1)=\frac{2}{5}.$$

Приклад 6.8. Неперервну випадкову величину X задано інтегральною функцією розподілу:

$$f(x)=\begin{cases}0, & \text{якщо } x<0,\\ 1-4e^{-5x}, & \text{якщо } x\ge 0.\end{cases}$$

Знайдіть параметр λ показникового розподілу.

Розв'язання. Для знаходження параметра λ слід використати таку формулу:

$$f(x)=\begin{cases}0, & \text{якщо } x\le 0;\\ \lambda e^{-\lambda x}, & \text{якщо } x>0.\end{cases}$$

Тоді $\lambda=5$.

Приклад 6.9. Неперервна випадкова величина X має показниковий розподіл із функцією:

$$f(x)=\begin{cases}0, & \text{якщо} \quad x\le 0;\\ 0,04\cdot e^{-0,04x}, & \text{якщо} \quad x>0.\end{cases}$$

Знайдіть: $F(x), M(X), \sigma(X), P(1<X<2)$.

Розв'язання.

$$F(x)=\begin{cases}0, & \text{якщо} \quad x\le 0;\\ 1-e^{-0,04x}, & \text{якщо} \quad x>0.\end{cases}$$

За умовою $\lambda=0,04$, тому $M(X)=\sigma(X)=\dfrac{1}{0,04}=25.$

$$P(1<X<2)=e^{-0,04\cdot 2}-e^{-0,04\cdot 1}\approx 0,038.$$

Приклад 6.10. Випадкову величину розподілено за показниковим законом із параметром $\lambda=2$. Визначте ймовірність того, що випадкова величина набуде значення, меншого за її математичне сподівання.

Розв'язання. Оскільки закон є показниковим, то $M\left(X\right)=\dfrac{1}{\lambda}=0,5$.

Необхідно визначити ймовірність події, яка полягає в тому, що значення випадкової величини будуть належати відрізку $[0;0,5]$.

$$P(0<X<0,5)=F(0,5)-F(0)=1-e^{-0,5}-(1-e^{-0})=1-e^{-0,5}\approx 0,3935.$$

Отже, майже 40 % значень випадкової величини є меншими за її математичне сподівання.

Приклад 6.11. Випадкову величину X розподілено за показниковим законом, до того ж $\lambda=1$. Знайдіть імовірність потрапляння випадкової величини X в інтервали: 1) (0; 2); 2) (3; 4).

Розв'язання. Для знаходження ймовірності потрапляння випадкової величини X в інтервал слід застосовувати таку формулу:

$$P(\alpha<X<\beta)=e^{-\lambda\alpha}-e^{-\lambda\beta}.$$

Тоді:

1) $P(0<X<2)=e^{-1\cdot 0}-e^{-1\cdot 2}=1-e^{-2}=1-0,1353=0,8647$.

2) $P(3<X<4)=e^{-1\cdot 3}-e^{-1\cdot 4}=e^{-3}-e^{-4}=0,0498-0,0183=0,0315$.

Приклад 6.12. Час безвідмовної роботи елемента розподілено за показниковим законом $f(x) = 0,02e^{-0,02t}$ за $t \geq 0$ (t – час у годинах).

Знайдіть імовірність того, що елемент пропрацює безвідмовно 100 год.

Розв'язання. За формулою

$$F(x) = \begin{cases} 0, & \text{якщо } x \leq 0; \\ 1 - e^{-\lambda x}, & \text{якщо } x > 0 \end{cases}$$

знайдіть імовірність того, що елемент пропрацює безвідмовно 100 год, для цього замість t підставте 100, $F(t) = 1 - e^{-0,02t}$, тоді

$$F(100) = 1 - e^{-0,02 \cdot 100} = 1 - e^{-2} = 0,1353.$$

Приклад 6.13. Математичне сподівання нормально розподіленої випадкової величини X дорівнює $a = 15$ і середнє квадратичне відхилення $\sigma = 2$. Напишіть диференціальну функцію випадкової величини X.

Розв'язання. Для того щоб записати диференціальну функцію нормально розподіленої випадкової величини X, застосуйте таку формулу:

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

Отже, $f(x) = \dfrac{1}{2\sqrt{2\pi}} e^{-\frac{(x-15)^2}{8}}$.

Приклад 6.14. Деталі, що випускаються цехом, за розміром діаметра розподіляють за нормальним законом за такими параметрами: математичне сподівання дорівнює 5 см, дисперсія – 0,81 см^2.

Знайдіть: а) імовірність того, що діаметр навмання взятої деталі від 4 до 7 см; б) імовірність того, що діаметр деталі відрізняється від математичного сподівання не більше ніж на 2 см; в) межі, у яких варто очікувати розмір діаметра деталі, щоб імовірність невиходу за ці межі дорівнювала 0,95.

Розв'язання. Використовуючи наведені формули, знаходять:

1. $P(4 < X < 7) = \Phi\left(\dfrac{7-5}{\sqrt{0,81}}\right) - \Phi\left(\dfrac{4-5}{\sqrt{0,81}}\right) \approx \Phi(2,22) + \Phi(1,11) = 0,8533.$

2. $P(|X-5|<2)=2\Phi\left(\dfrac{2}{0,9}\right)\approx 2\Phi(2,22)=0,9736.$

3. $2\Phi\left(\dfrac{\varepsilon}{\sigma}\right)=0,95;\quad \Phi\left(\dfrac{\varepsilon}{\sigma}\right)=0,475.$

За таблицею значень функції $\Phi(x)$ (додаток В) знаходять, що $\Phi(x)=0,475$ за $x=1,96$. Тоді, $\dfrac{\varepsilon}{\sigma}=1,96$, звідси $\varepsilon=0,9\cdot 1,96\approx 1,8$ см.

Отже, з імовірністю 0,95 можна стверджувати, що діаметр деталі перебуває в межах 5 ± 1,8 см або від 3,2 до 6,8 см.

Приклад 6.15. Розглядають можливість придбання акцій 2 видів – I та II. Їхні прибутковості розподіляють за нормальним законом з однаковими середніми значення 20 %. Середнє квадратичне відхилення для першої акції – σ = 1,5 % для другої акції – σ = 2 %.

Визначте інтервал прибутковості кожної акції з надійністю 95 %. Яка акція є кращою і чому?

Розв'язання. Нехай випадкова величина X_1 – прибутковість акцій I виду, а випадкова величина X_2 – прибутковість акцій II виду. Для знаходження інтервалу прибутковості використайте таку формулу:

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

Звідки $|X_1-a|<\delta$, або $-\delta+a<X_1<\delta+a$,

$|X_2-a|<\delta$, або $-\delta+a<X_2<\delta+a$.

За умовою a = 20 %, отже, треба знайти δ.

Відомо, що $P=2\Phi\left(\dfrac{\delta}{\sigma}\right)=0,95$, звідки $\Phi\left(\dfrac{\delta}{\sigma}\right)=0,475$ та $\dfrac{\delta}{\sigma}=1,96$ (див. додаток В); за умовою для випадкової величини X_1 σ = 1,5 %, тоді $\sigma=1,96\cdot 1,5=2,94$ та $-2,94\,\%+20\,\%<X<20\,\%+2,94\,\%$, отже, $17,06\,\%<X<22,94\,\%$.

Таким чином, з імовірністю 95 % можна стверджувати, що прибутковість акцій I виду перебуває в межах (17 %; 23 %).

За умовою для випадкової величини X_2 σ = 2 %, тоді $\sigma = 1{,}96 \cdot 2 = 3{,}92$ та $-3{,}92\,\% + 20\,\% < X < 20\,\% + 3{,}92\,\%$.

Отже, $16{,}08\,\% < X < 23{,}92\,\%$. Таким чином, прибутковість акцій II виду від 16 до 24 % можна гарантувати з імовірністю 95 %.

Відповідь на друге питання: кращою є перша акція, бо вона є менш ризиковою (середнє квадратичне відхилення в неї менше).

Приклад 6.16. Величину X відхилення довжини виготовлюваних деталей від стандарту розподілено за нормальним законом із математичним сподіванням a = 40 см, і середнім квадратичним відхиленням σ = 0,4 см. Яку точність довжини виробу можна гарантувати з імовірністю 0,8?

Розв'язання. Знайдіть додатне δ, для якого

$$P(|X-40| \le \delta) = 0{,}8.$$

Із умови $P(|X-40| \le \delta) = 2\Phi\left(\dfrac{\delta}{0{,}4}\right) = 0{,}8$, буде $\Phi\left(\dfrac{\delta}{0{,}4}\right) = 0{,}4$.

Із додатка В $\dfrac{\delta}{0{,}4} = 1{,}28$, звідки $\delta = 0{,}512$.

Отже, з імовірністю 80 % можна гарантувати, що довжина виробу відхиляється від 40 см менше ніж на 0,512 см.

Приклад 6.17. Випадкову величину розподілено за нормальним законом з математичним сподіванням a = 20. Імовірність потрапляння її в інтервал (20; 30) дорівнює 0,4772. Визначте ймовірність потрапляння випадкової величини в інтервал (10; 25).

Розв'язання. Для знаходження ймовірності потрапляння випадкової величини в інтервал (a; β) використайте формулу:

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

тобто $P(10 < X < 25) = \Phi\left(\dfrac{25-20}{\sigma}\right) - \Phi\left(\dfrac{10-20}{\sigma}\right) = \Phi\left(\dfrac{5}{\sigma}\right) + \Phi\left(\dfrac{10}{\sigma}\right).$

Отже, треба знайти σ.

За умовою $P(20 < X < 30) = 0,4772$,

Тоді

$$P(20 < X < 30) = \Phi\left(\frac{30-20}{\sigma}\right) - \Phi\left(\frac{20-20}{\sigma}\right) = \Phi\left(\frac{10}{\sigma}\right) - \Phi(0) = \Phi\left(\frac{10}{\sigma}\right) - 0.$$

Тепер буде $\Phi\left(\frac{10}{\sigma}\right) = 0,4772$.

Із додатка В $\frac{10}{\sigma} = 2$, звідки $\sigma = 5$.

Обчисліть $P(10 < X < 25) = \Phi\left(\frac{5}{5}\right) - \Phi\left(\frac{10}{5}\right) = \Phi(1) + \Phi(2) =$

$$= 0,3413 + 0,4772 = 0,8185.$$

Таким чином, випадкова величина потрапить в інтервал (10; 25) з імовірністю 81,85 %.

6.4. Вправи для самостійної роботи

6.1. Функцію розподілу $F(x)$ випадкової величини X задано такою рівністю:

$$F(x) = \begin{cases} 0, & \text{якщо} \quad x \le 1, \\ \dfrac{(x-1)^2}{4}, & \text{якщо} \quad < x \le 3, \\ 1, & \text{якщо} \quad x > 3. \end{cases}$$

Знайдіть:

а) щільність імовірності $F(x)$;

б) $P(1 < X \le 2,5)$;

в) $P(2,5 < X < 3,5)$.

Побудуйте графіки $f(x)$ і $F(x)$.

6.2. Відомо, що щільність імовірності $f(x)$ величини X визначено за допомогою таких рівностей:

$$f(x)=\begin{cases} A\cos x, & \text{якщо } |x|\le\dfrac{\pi}{2},\\ 0, & \text{якщо } |x|>\dfrac{\pi}{2}.\end{cases}$$

Знайдіть коефіцієнт A. Визначте ймовірність потрапляння X в інтервал $\left(-\dfrac{\pi}{4};\dfrac{\pi}{4}\right)$.

6.3. Щільність імовірності $f(x)$ випадкової величини X задано рівністю: $f(x)=\dfrac{a}{e^x+e^{-x}}$, де a – стала. Знайшовши цю сталу, визначте функцію розподілу $F(x)$ величини X.

6.4. Щільність імовірності $f(x)$ випадкової величини X задано рівністю: $f(x)=\dfrac{ae^x}{(e^x+1)^2}$, де a – стала. Знайшовши цю сталу, визначте функцію розподілу $F(x)$ величини X. Знайдіть імовірність виконання нерівності $|X|<1$.

6.5. Диференціальна функція випадкової величини X $f(x)=ax$ в інтервалі (0; 2); поза цим інтервалом $f(x)=0$. Визначте a, знайдіть математичне сподівання і медіану X.

6.6. Щільність імовірності випадкової величини X дорівнює 0 за $|x|>4$ і дорівнює $\dfrac{1}{\pi\sqrt{16-x^2}}$ за $|x|<4$. Знайдіть математичне сподівання і медіану X.

6.7. Щільність імовірностей випадкової величини X задано рівностями:

$$f(x)=\begin{cases} a(x^2+2x), & \text{якщо } 0\le x\le 1,\\ 0, & \text{якщо } x<0,\ x>1.\end{cases}$$

Визначивши параметр a, знайдіть $M(X)$.

6.8. Знайдіть математичне сподівання випадкової величини X за заданою функцією розподілу:

а) $F(x)=\begin{cases} 0, & \text{якщо} \quad x\le 0, \\ \dfrac{x}{4}, & \text{якщо} \quad 0<x\le 4, \\ 1, & \text{якщо} \quad x>4. \end{cases}$ б) $F(x)=\begin{cases} 0, & \text{якщо} \quad x\le -2, \\ \dfrac{x}{4}+\dfrac{1}{2}, & \text{якщо} \quad -2<x\le 2, \\ 1, & \text{якщо} \quad x>2. \end{cases}$

6.9. Щільність імовірностей випадкової величини X $f(x)=a\sin x$ в інтервалі $(0;\pi)$; поза цим інтервалом $f\left(x\right)=0$. Визначивши параметр a, знайдіть математичне сподівання, моду, медіану, дисперсію величини X.

6.10. Випадкову величину X задано диференціальною функцією $f(x)=a\cos 2x$ в інтервалі $\left(0;\dfrac{\pi}{4}\right)$; поза цим інтервалом $f(x)=0$. Визначивши параметр a, знайдіть математичне сподівання, моду, медіану, дисперсію величини X.

6.11. Випадкову величину X в інтервалі (2; 4) задано диференціальною функцією $f(x)=-\dfrac{3}{4}x^2+\dfrac{9}{2}x-6$, поза цим інтервалом $f\left(x\right)=0$. Знайдіть математичне сподівання, моду і медіану величини X.

6.12. Усі значення рівномірно розподіленої випадкової величини належать відрізку (12; 18). Знайдіть щільність розподілу цієї випадкової величини, а також імовірність того, що випадкова величина набуде значення з інтервалу (14; 17).

6.13. Знайдіть математичне сподівання випадкової величини X, яка є рівномірно розподіленою в інтервалі (4; 10).

6.14. Знайдіть дисперсію і середнє квадратичне відхилення випадкової величини X, рівномірно розподіленої в інтервалі (2; 10).

6.15. Неперервну випадкову величину X рівномірно розподілено на відрізку (1; 3). Необхідно: 1) знайти диференціальну й інтегральну функції розподілу, а також побудувати їхні графіки; 2) обчислити математичне сподівання і середнє квадратичне відхилення; 3) знайти ймовірність того, що X набуде якого-небудь значення з інтервалу (15; 2).

6.16. Автобуси деякого маршруту йдуть за розкладом з інтервалом руху 6 хв. Знайдіть імовірність того, що пасажир, який підійде до зупинки, буде очікувати автобус менше ніж 2 хв.

6.17. Випадкова величина X має рівномірний розподіл із математичним сподіванням $M(X)=7$ і дисперсією $D(X)=\frac{4}{3}$. Знайдіть інтегральну функцію розподілу випадкової величини X.

6.18. Випадкова величина X має рівномірний розподіл із математичним сподіванням $M(X)=9$ і середнім квадратичним відхиленням $\sigma(X)=\frac{2\sqrt{3}}{3}$. Знайдіть щільність рівномірного розподілу.

6.19. Знайдіть математичне сподівання, дисперсію, середнє квадратичне відхилення випадкової величини X, якщо інтегральна функція показникового розподілу така:

$$f(x)=\begin{cases}0, & \text{якщо } x<0,\\ 1-e^{-2x}, & \text{якщо } x\geq 0.\end{cases}$$

6.20. Випадкову величину X, дисперсія якої $D(X)=\frac{1}{25}$, розподілено за показниковим законом. Знайдіть імовірність потрапляння випадкової величини X в інтервал: а) (0; 1); б) (1; 3).

6.21. Неперервну випадкову величину X розподілено за показниковим законом із функцією розподілу $F(x)=1-e^{-0,2x}$ за $x\geq 0$; $F(x)=0$ за $x<0$. Знайдіть імовірність виконання нерівності $2<X<4$.

6.22. Неперервну випадкову величину розподілено за показниковим законом зі щільністю ймовірності $f(x)=0,03e^{-0,03x}$ за $x\geq 0$; $f(x)=0$ за $x<0$. Знайдіть: а) функцію розподілу $F(x)$; б) імовірність потрапляння випадкової величини X в інтервал (1; 3).

6.23. Знайдіть математичне сподівання, дисперсію, середнє квадратичне відхилення випадкової величини зі щільністю ймовірності $f(x)=3e^{-3x}$ за $x\geq 0$; $f(x)=0$ за $x<0,4$.

6.24. Випадкову величину X розподілено за показниковим законом із функцією розподілу $F(x)=1-e^{-0.2x}$ за $x\geq 0$; $F(x)=0$ за $x<0$. Знайдіть: а) математичне сподівання величини X; б) дисперсію величини X; в) середнє квадратичне відхилення величини X; г) імовірність виконання нерівності $1\leq X\leq 2$.

6.25. Випробовують 2 елементи, що незалежно працюють. Тривалість безвідмовної роботи першого має показників розподіл $F_1(t)=1-e^{-0,1t}$, другого – $F_2(t)=1-e^{-0,05t}$. Знайдіть імовірність того, що протягом 20 год: а) обидва елементи будуть працювати; б) відмовить тільки 1 елемент; в) відмовить хоча б 1 елемент; г) відмовлять обидва елементи.

6.26. Математичне сподівання нормально розподіленої випадкової величини X a = 2, а дисперсія $D(X)=9$. Напишіть диференціальну функцію випадкової величини.

6.27. Диференціальна функція нормально розподіленої випадкової величини X така:

$$f(x)=\frac{1}{6\sqrt{2\pi}}e^{-\frac{(x-10)^2}{72}}.$$

Знайдіть математичне сподівання та дисперсію.

6.28. Випадкову величину X розподілено за нормальним законом із математичним сподіванням a = 40 і дисперсією $D(X)=200$. Обчисліть імовірність потрапляння випадкової величини в інтервал (30; 80).

6.29. Вагу риб, що виловлюють у ставку, підпорядковано нормальному закону розподілу з математичним сподіванням a = 1 000 г і середнім квадратичним відхиленням $\sigma=150$ г. Знайдіть імовірність того, що вага спійманої риби буде: 1) від 900 до 1 300 г; 2) не менше ніж 800 г; 3) не більше ніж 1 500 г; 4) відрізняється від середньої ваги за модулем не більше ніж на 200 г.

6.30. Керівник банку встановив, що тривалість обслуговування клієнта в черзі підпорядковано нормальному закону. Математичне сподівання 3 хв, середнє квадратичне відхилення – 1 хв. Визначте ймовірність того, що клієнт перебуває в черзі: 1) від 2 до 3,5 хв; 2) менше ніж 1 хв; 3) понад 5 хв.

6.31. Установлено, що діаметром деталі є нормально розподілена випадкова величина з дисперсією 4 мм^2. Яка ймовірність браку, якщо бракують деталі, діаметр яких відхиляється від норми (математичного сподівання) більш ніж на 3,5 мм?

6.32. Автомат виготовляє деталь, яку вважають стандартною, якщо відхилення діаметра деталі від проектного розміру не перевищує 2 мм.

Випадкове відхилення діаметра підпорядковано нормальному закону із середнім квадратичним відхиленням $\sigma = 1,6$мм і математичним сподіванням $a = 0$. Скільки відсотків стандартних деталей виготовляє автомат?

6.33. Деталі за розміром розподілено за нормальним законом із середнім значенням 20 см і дисперсією 0,04 см^2. Знайдіть точність виробу, яку можна гарантувати з імовірністю 0,95.

6.34. Математичне сподівання та середнє квадратичне відхилення нормально розподіленої випадкової величини X, відповідно, дорівнюють 20 і 5. Знайдіть імовірність того, що величина X набуде значення у проміжку (15; 25).

6.35. Зріст юнаків, яких забирають до війська за віком, передбачають нормально розподіленим із середнім значенням $a = 170$ см і середнім квадратичним відхиленням $\sigma = 5$см. Знайдіть відсоток юнаків, що мають зріст: а) нижчий за 160 см; б) вищий за 180 см; в) від 160 до 175 см.

6.36. Діаметр деталі є випадковою величиною, розподіленою за нормальним законом з $a = 40$см і $\sigma = 0,4$см. Яку точність ε діаметра можна гарантувати з імовірністю $P = 0,8$?

6.37. Діаметр деталей, виготовлених цехом, є випадковою величиною з нормальним законом розподілу. Її математичне сподівання дорівнює 2,5 см, а дисперсія – 0,0001 см^2. У яких межах можна практично гарантувати діаметр деталі з імовірністю $P = 0,9973$?

6.38. Випадкову величину X розподілено нормально з $a = 25$. Імовірність $P(35 < X < 40) = 0,2$. Знайдіть імовірність $P(10 < X < 15)$.

6.39. Математичне сподівання і дисперсія нормально розподіленої випадкової величини X, відповідно, дорівнюють 1 і 4. Знайдіть імовірності таких нерівностей: а) $|X - 1| < 2$; б) $1 < X < 4$; в) $X > 2$; г) $X < 0$.

6.40. Випадкову величину розподілено за нормальним законом із параметрами $a = 5$, $\sigma = 0,3$. Визначте ймовірність того, що випадкова величина набуде значення, що буде перевищувати $x = 5,5$.

6.41. Випадкову величину розподілено за нормальним законом із параметрами $a = 10$, $\sigma = 2$. Визначте ймовірність того, що значення, котрі набуде випадкова величина, будуть належати відрізку $x \in [8; 12]$.

6.42. Кількість покупців, яких обслуговує продавець протягом години, є випадковою величиною, розподіленою за нормальним законом із параметрами $a = 7$, $\sigma = 2$. Визначте проміжок, до якого з імовірністю 0,9 будуть належати значення цієї випадкової величини.

6.43. Випадкову величину розподілено за нормальним законом із параметрами $a=5$, $\sigma=1$. Порівняйте ймовірності потрапляння значення випадкової величини до інтервалів $x\in[2;4]$ та $x\in[5;6]$.

6.5. Тестові завдання

6.1. Щільністю розподілу неперервної випадкової величини X називають функцію

А	Б	В	Г
$f(x)=F(x)$	$f(x)=F'(x)$	$F(x)=f'(x)$	$f(x)=F''(x)$

6.2. Математичне сподівання $M(X)$ неперервної випадкової величини X, можливі значення якої належать інтервалу $(x_1;x_2)$, дорівнює

А	Б	В	Г
$\int_{x_1}^{x_2} x^2 f(x)dx$	$\int_{x_1}^{x_2} xdx$	$\int_{x_1}^{x_2} f(x)dx$	$\int_{x_1}^{x_2} xf(x)dx$

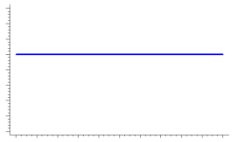
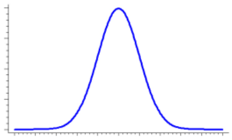
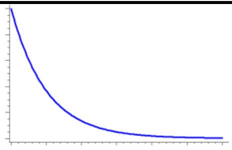
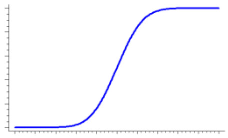
6.3. Дисперсія $D(X)$ неперервної випадкової величини X, можливі значення якої належать інтервалу $(x_1;x_2)$, дорівнює

А	Б	В	Г
$\int_{x_1}^{x_2} xf(x)dx-M(X)$	$\int_{x_1}^{x_2} x^2 f(x)dx-(M(X))^2$	$\int_{x_1}^{x_2} x^2 f(x)dx$	$\int_{x_1}^{x_2} f(x)dx-(M(X))^2$

6.4. Щільність розподілу нормально розподіленої випадкової величини X дорівнює

А	Б	В	Г
$f(x)=\dfrac{1}{\sigma\sqrt{2\pi}}e^{-\frac{(x-a)^2}{2\sigma^2}}$	$f(x)=e^{-\frac{(x-a)^2}{2\sigma^2}}$	$f(x)=\dfrac{1}{\sigma\sqrt{2\pi}}e^{-(x-a)^2}$	$f(x)=e^{-\frac{(x-a)^2}{\sigma}}$

6.5. На якому рисунку зображено графік щільності розподілу нормально розподіленої випадкової величини X

А	Б	В	Г
			

6.6. Числові характеристики нормально розподіленої випадкової величини X дорівнюють

А	Б	В	Г
$M(X)=a$, $D(X)=\sigma^2$	$M(X)=a$, $D(X)=\sigma^2$	$M(X)=a$, $D(X)=\sigma$	$M(X)=a$, $D(X)=\sqrt{\sigma}$

6.7. Якщо $f(x)$ – щільність розподілу випадкової величини X, то ймовірність того, що випадкова величина X набуває значення в інтервалі $(x_1;\, x_2)$, дорівнює

А $P(x_1 < X < x_2) = \int\limits_{x_1}^{x_2} f(x)dx$

Б $P(x_1 < X < x_2) = \int\limits_{x_1}^{x_2} x^2 f(x)dx$

В $P(x_1 < X < x_2) = \int\limits_{x_1}^{x_2} xf(x)dx$

Г $P(x_1 < X < x_2) = \int\limits_{x_1}^{x_2} x^2 f(x)dx - (M(X))^2$

6.8. Імовірність $P(x_1 < X < x_2)$ того, що випадкова величина, розподілена за нормальним законом, набуває значення в інтервалі $(x_1;\, x_2)$, визначають за такою формулою

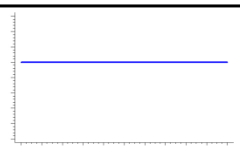
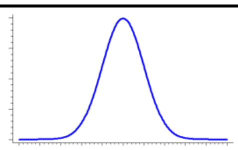
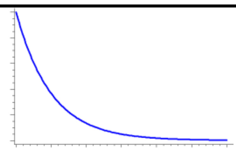
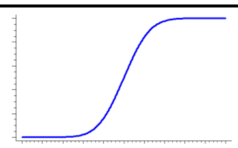
А $P(x_1 < X < x_2) = \Phi(x_2) - \Phi(x_1)$

Б $P(x_1 < X < x_2) = \Phi\left(\dfrac{x_2 - a}{\sigma}\right) - \Phi\left(\dfrac{x_1 - a}{\sigma}\right)$

В $P(x_1 < X < x_2) = \Phi\left(\dfrac{x_1 - a}{\sigma}\right) - \Phi\left(\dfrac{x_2 - a}{\sigma}\right)$

Г $P(x_1 < X < x_2) = \Phi\left(\dfrac{x_2 - a}{\sigma}\right) + \Phi\left(\dfrac{x_1 - a}{\sigma}\right)$

6.9. На якому рисунку зображено графік щільності розподілу випадкової величини X, розподіленої за рівномірним законом

А	Б	В	Г
			

6.10. Якщо $a < x \le b$, то щільність розподілу випадкової величини, розподіленої за рівномірним законом дорівнює

А	Б	В	Г
$f(x) = b - a$	$f(x) = \dfrac{1}{a - b}$	$f(x) = \dfrac{1}{b - a}$	$f(x) = \dfrac{a}{b - a}$

6.11. Числові характеристики рівномірно розподіленої випадкової величини X дорівнюють

А	Б	В	Г
$M(X) = \dfrac{a+b}{2}$, $D(X) = \dfrac{(b-a)^2}{12}$	$M(X) = a + b$, $D(X) = \dfrac{b-a}{12}$	$M(X) = \dfrac{a+b}{2}$, $D(X) = \dfrac{(b-a)^2}{12}$	$M(X) = \dfrac{a+b}{2}$, $D(X) = \dfrac{b-a}{12}$

6.12. Для випадкової величини, розподіленої за показниковим законом, виконано умову

А	Б	В	Г
$M(X) = \sigma(X) = \dfrac{1}{\lambda}$	$M(X) = \sigma^2(X)$	$D(X) = \lambda$	$M(X) = \sigma(X) = \lambda$

6.6. Запитання для самоперевірки

6.1. Які існують способи задавання неперервної випадкової величини?

6.2. Поясніть, чому неможливо побудувати диференціальну функцію розподілу для дискретної випадкової величини.

6.3. Яка відмінність між визначенням числових характеристик розподілу дискретної та неперервної випадкових величин?

6.4. Назвіть властивості диференціальної функції розподілу.

6.5. Укажіть зв'язок між функцією розподілу і щільністю ймовірностей.

6.6. Який закон розподілу неперервної випадкової величини називають рівномірним? Поясніть сенс цієї назви. Назвіть параметри рівномірного закону та наведіть формули для обчислення його основних числових характеристик.

6.7. Який закон розподілу неперервної випадкової величини називають експоненціальним? Укажіть кількість параметрів експоненціального закону і наведіть формули для обчислення його основних числових характеристик.

6.8. Який закон розподілу неперервної випадкової величини називають нормальним? Поясніть сенс цієї назви. Назвіть параметри нормального закону і поясніть їхній вплив на вигляд диференціальної функції розподілу.

6.9. Які значення мають додаткові числові характеристики розподілу випадкової величини, розподіленої за нормальним законом?

6.10. Як визначають імовірність потрапляння значення випадкової величини до певного інтервалу за нормального закону розподілу?

6.11. Сформулюйте правило трьох сигм і поясніть його застосування.

6.7. Висновки за темою

Після вивчення теми студент має вміти:

чітко визначати закони розподілу неперервних випадкових величин;

застосовувати функції розподілу та функції щільності ймовірностей для визначення характеристик розподілу неперервної випадкової величини;

надавати геометричну інтерпретацію побудованого закону розподілу;

робити висновки за побудованим розподілом.

7. Багатовимірні випадкові величини

7.1. Мета та компетентності

Метою вивчення теми є ознайомлення з основними законами розподілу двовимірної дискретної випадкової величини та формування системи теоретичних знань, практичних умінь і навичок у визначенні умовних законів розподілу та обчислення їхніх основних числових характеристик.

Компетентності, що формують під час вивчення теми:

уявлення про закон розподілу та умовні закони розподілу її компонентів;

знання основних числових характеристик системи двох випадкових величин, означення кореляційного моменту, його властивостей, числових характеристик умовних розподілів;

уміння будувати умовні закони розподілу двовимірної дискретної випадкової величини, обчислювати основні числові характеристики двовимірних випадкових величин, визначати щільність кореляційного зв'язку;

здатність самостійно визначати етапи дослідження двовимірної випадкової величини для обчислення щільності кореляційного зв'язку її компонентів.

7.2. Основні теоретичні відомості

Випадкові величини, можливі значення яких визначають двома числами, називають **двовимірними**. Двовимірну випадкову величину $(X;Y)$ визначають за двома складовими X і Y, які утворюють систему двох випадкових величин.

Геометрична інтерпретація двовимірної величини – це випадкова точка $M(X;Y)$ на площині. Якщо X і Y дискретні, то двовимірна величина є дискретною.

Законом розподілу двовимірної випадкової величини називають відповідність між можливими значеннями цієї величини, тобто парами чисел $(x_i;\ y_i)$, та ймовірністю $p(x_i;\ y_i)$, із якою випадкова величина набуває цих значень.

Закон розподілу задають:

а) аналітично, наприклад, у вигляді інтегральної функції;

б) таблицею з подвійним входом (табл. 7.1), де $x_1 < x_2 < ... < x_n$, $y_1 < y_2 < ... < y_m$, $p(x_i; y_j)$ – імовірність події, яка полягає в одночасному виконанні рівностей $X = x_i, Y = y_j$.

Водночас $\sum_{i=1}^{n}\sum_{j=1}^{m} p(x_i; y_j) = 1.$

Таблиця 7.1

Закон розподілу двовимірної випадкової величини

$Y \backslash X$	x_1	x_2	...	x_i	...	x_n
y_1	$p(x_1; y_1)$	$p(x_2; y_1)$	...	$p(x_i; y_1)$	...	$p(x_n; y_1)$
y_2	$p(x_1; y_2)$	$p(x_2; y_2)$	...	$p(x_i; y_2)$	...	$p(x_n; y_2)$
...	...	...	...	...	...	...
y_j	$p(x_1; y_j)$	$p(x_2; y_j)$	...	$p(x_i; y_j)$	...	$p(x_n; y_j)$
...	...	...	...	...	...	...
y_m	$p(x_1; y_m)$	$p(x_2; y_m)$	...	$p(x_i; y_m)$	...	$p(x_n; y_m)$

Аналогічно тому, як це визначали для одновимірної випадкової величини, закон розподілу двовимірної випадкової величини будь-якої природи (як дискретну, так і неперервну) можна задати за допомогою функції розподілу $F(x; y)$, яка визначає ймовірність такої випадкової події, що компонента X двовимірної випадкової величини набуде значення, меншого за певне значення x, яке не обов'язково належить множині значень цієї випадкової величини, одночасно з тим, що компонента Y набуде значення, меншого за певне значення y. Цю функцію визначають співвідношенням:

$$F(x; y) = P(X < x; Y < y),$$

де $F(x; y)$ – функції розподілу двовимірної випадкової величини, визначеної на всій множині дійсних чисел, що називають **інтегральною функцією розподілу ймовірностей двовимірної випадкової величини** $(X; Y)$ (як для дискретної, так і неперервної).

Функції розподілу $F(x;\, y)$ можна надати геометричне тлумачення (рис. 7.1).

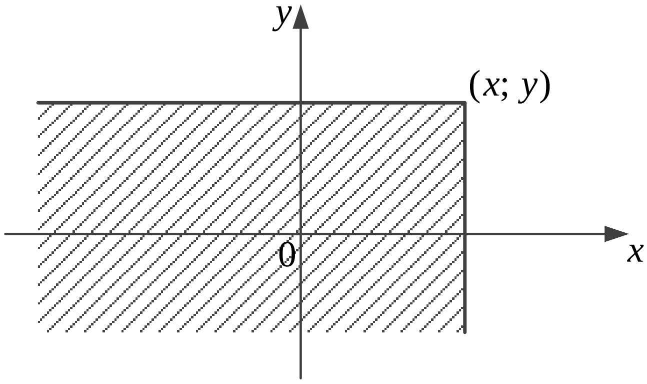


Рис. 7.1. **Геометричне тлумачення функції розподілу** $F(x;\, y)$

Властивості інтегральної функції розподілу:

1. Значення інтегральної функції задовольняють подвійній нерівності:

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

2. Інтегральна функція є неспадною функцією за кожним аргументом:

$$F(x_2;\, y) \ge F(x_1;\, y)\text{, якщо } x_2 > x_1;$$
$$F(x;\, y_2) \ge F(x;\, y_1)\text{, якщо } y_2 > y_1.$$

3. Виконуються граничні співвідношення:

1) $F(-\infty;\, y) = 0$; 2) $F(x; -\infty) = 0$; 3) $F(-\infty; -\infty) = 0$; 4) $F(\infty; \infty) = 1$.

4. а) якщо $y = \infty$, то інтегральна функція системи стає інтегральною функцією складової X: $F(x;\infty) \le F_1(x)$;

б) якщо $x = \infty$, то інтегральна функція системи стає інтегральною функцією складової Y: $F(\infty;\, y) \le F_2(y)$.

Використовуючи інтегральну функцію, можна знайти ймовірність потрапляння випадкової точки $(X;Y)$ у напівсмугу $x_1 < X < x_2$ і $Y < y$ (рис. 7.2):

$$P\left(x_1 < X < x_2; Y < y\right) = F\left(x_2;\, y\right) - F\left(x_1;\, y\right),$$

або в напівсмугу $X < x$ і $y_1 < Y < y_2$ (рис. 7.3):

$$P(X < x; y_1 < Y < y_2) = F(x; y_2) - F(x; y_1).$$

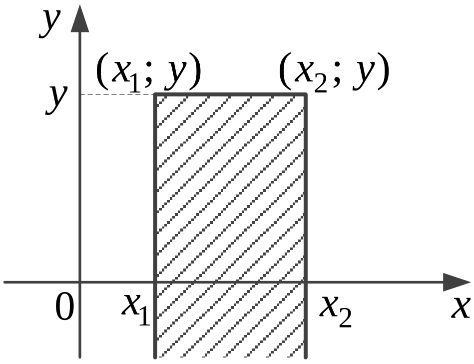


Рис. 7.2. **Імовірність потрапляння випадкової точки** $(X; Y)$ **у напівсмугу** $x_1 < X < x_2$ **і** $Y < y$

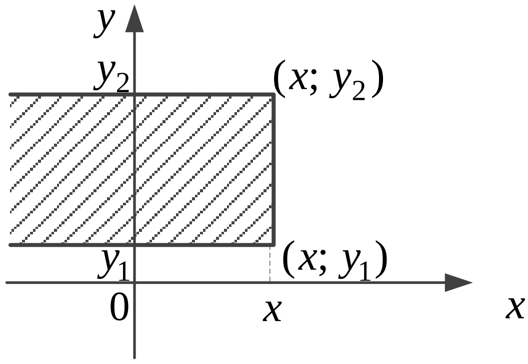


Рис. 7.3. **Імовірність потрапляння випадкової точки** $(X; Y)$ **у напівсмугу** $X < x$ **і** $y_1 < Y < y_2$

Імовірність потрапляння випадкової точки у прямокутник (рис. 7.4) $x_1 < X < x_2$, $y_1 < Y < y_2$ дорівнює:

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

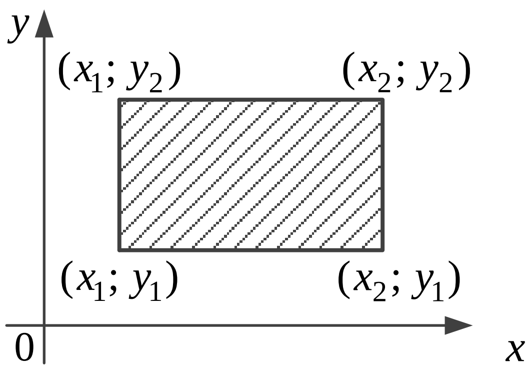


Рис. 7.4. **Імовірність потрапляння випадкової точки у прямокутник** $x_1 < X < x_2$, $y_1 < Y < y_2$

Нехай складові X і Y мають такі значення:

$$x_1, x_2, ..., x_n; \quad y_1, y_2, ..., y_m.$$

Умовним розподілом складової X за $Y = y_j$ називають сукупність значень X і відповідні їм умовні ймовірності $P(x_1 | y_j),\ P(x_2 | y_j), \ldots,$ $P(x_n | y_j)$, які обчислюють за такою формулою:

$$P(x_i | y_j) = \frac{P(x_i;\, y_j)}{P(y_j)} \quad (i = 1, 2, \ldots, n),$$

де $P(y_j)$ – сума ймовірностей у *j*-му рядку.

Умовним розподілом складової Y за $X = x_j$ називають сукупність значень Y і відповідні їм умовні ймовірності $P(y_1 | x_i),\ P(y_2 | x_i), \ldots,$ $P(y_m | x_i)$, які обчислюють за такою формулою:

$$P(y_j | x_i) = \frac{P(x_i;\, y_j)}{P(x_i)} \quad (j = 1, 2, \ldots, m),$$

де $P(x_i)$ – сума ймовірностей в *i*-му рядку.

7.2.1. Способи задавання неперервної випадкової двовимірної величини

Двовимірну випадкову величину можна задати за допомогою інтегральної функції. Неперервну двовимірну випадкову величину також можна задати, використовуючи диференціальну функцію розподілу (щільність імовірностей).

Диференціальною функцією розподілу $f(x;\, y)$ **двовимірної неперервної випадкової величини** $(X;\, Y)$ називають другу мішану частинну похідну від інтегральної функції:

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

Графічне зображення функції $f(x;\, y)$ має назву **поверхні розподілу**.

Функцію розподілу $F(x; y)$ визначають за щільністю ймовірностей таким співвідношенням:

$$F(x;\, y) = \int_{-\infty}^{x} \int_{-\infty}^{y} f(x;\, y) dxdy.$$

Властивості диференціальної функції двовимірної величини:

1. Диференціальна функція невід'ємна: $f(x;\, y) \geq 0$.

2. Подвійний невласний інтеграл на нескінчених проміжках від диференціальної функції дорівнює одиниці:

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x;\, y)dxdy = 1.$$

Якщо $(X;\, Y) \subset D$, властивість (2) набуває такого вигляду:

$$\iint_D f(x;\, y)dxdy = 1.$$

Імовірність потрапляння випадкової величини $(X;\, Y)$ в область D визначають за рівністю:

$$P((X;\, Y) \subset D) = \iint_D f(x;\, y)dxdy\,.$$

Геометрично цю формулу можна розглядати як об'єм тіла, обмеженого зверху поверхнею $z = f(x;\, y)$, в основі якого лежить проекція цієї поверхні на площину $x0y$.

Диференціальна функція однієї зі складових дорівнює невласному інтегралу на нескінченному проміжку від диференціальної функції системи зі зміною інтегрування, що відповідає другій складовій:

$$f_1(x) = \int_{-\infty}^{\infty} f(x;\, y)dy;\;\; f_2(y) = \int_{-\infty}^{\infty} f(x;\, y)dx\,.$$

Умовною диференціальною функцією $\varphi(x|y)$ **складових** X за заданого значення $Y = y$ називають відношення диференціальної функції $f(x;\, y)$ до диференціальної функції $f_2(y)$ складової Y:

$$\varphi(x|\,y) = \frac{f(x;\, y)}{f_2(y)} = \frac{f(x;\, y)}{\int_{-\infty}^{\infty} f(x;\, y)dx}\,.$$

Умовною диференціальною функцією $\psi(y|x)$ **складової** Y за заданого значення $X = x$ називають відношення диференціальної функції $f(x;\, y)$ до диференціальної функції $f_1(x)$ складової X:

$$\psi(y|x) = \frac{f(x;\, y)}{f_1(x)} = \frac{f(x;\, y)}{\int_{-\infty}^{\infty} f(x;\, y)dy}\,.$$

Властивості умовних диференціальних функцій:

1. $\varphi(x|y) \geq 0, \quad \psi(y|x) \geq 0.$

2. $\int_{-\infty}^{\infty} \varphi(x|y)dx = 1, \quad \int_{-\infty}^{\infty} \psi(y|x)dx = 1.$

Теорема. Для того щоб випадкові величини X і Y були незалежними, необхідно і достатньо, щоб інтегральна функція системи $(X;Y)$ дорівнювала добутку інтегральних функцій складових:

$F(x; y) = F_1(x) \cdot F_2(y)$, де $F_1(x) = F(x; \infty), \quad F_2(y) = F(\infty; y)$.

Висновок: $f(x; y) = f_1(x) \cdot f_2(y)$.

7.2.2. Основні числові характеристики розподілу двовимірної випадкової величини

Математичні сподівання дискретних випадкових величин X і Y, що входять до системи, визначають за такими формулами:

$$M(X) = \sum_{i=1}^{n}\sum_{j=1}^{m} x_i p_{ij} \text{ та } M(Y) = \sum_{i=1}^{n}\sum_{j=1}^{m} y_j p_{ij},$$

а математичні сподівання неперервних випадкових величин X і Y визначають за такими формулами:

$$M(X) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} xf(x;y)dxdy \text{ та } M(Y) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} yf(x;y)dxdy,$$

а також:

$$M(X) = \int_{-\infty}^{\infty} xf_1(x)dx \text{ та } M(Y) = \int_{-\infty}^{\infty} yf_2(y)dy.$$

Точку $(M(X); M(Y))$ називають **центром розсіювання системи випадкових величин** $(X;Y)$.

Для обчислення дисперсії дискретних випадкових величин X і Y застосовують такі формули:

$$D(X) = \sum_{i=1}^{n}\sum_{j=1}^{m} p_{ij}(x_i - M(X))^2 \text{ та } D(Y) = \sum_{i=1}^{n}\sum_{j=1}^{m} p_{ij}(y_j - M(Y))^2,$$

а дисперсії неперервних випадкових величин X і Y, що входять до системи, знаходять за такими формулами:

$$D(X)=\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}(x-M(X))^2 f(x;y)dxdy$$

та $$D(Y)=\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}(y-M(Y))^2 f(x;y)dxdy.$$

Середні квадратичні відхилення випадкових величин X і Y:

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

Умовним математичним сподіванням дискретної випадкової величини Y за $X=x$ (x – певне можливе значення X) називають суму добутків можливих значень Y на їхні умовні ймовірності:

$$M\left(Y\middle|X=x\right)=\sum_{j=1}^{m} y_j p\left(y_j\middle|x\right).$$

Умовним математичним сподіванням дискретної випадкової величини X за $Y=y$ називають суму добутків можливих значень X на їхні умовні ймовірності:

$$M(X|Y=y)=\sum_{i=1}^{n} x_i p\left(x_i\middle|y\right).$$

Умовним математичним сподіванням неперервної випадкової величини Y за $X=x$ називають невласний інтеграл із нескінченними межами інтегрування від добутку змінної y на умовну щільність випадкової величини Y за $X=x$:

$$M\left(Y\middle|X=x\right)=\int_{-\infty}^{\infty} y\psi(y|x)dy.$$

Аналогічно

$$M\left(X\,|Y=y\right)=\int_{-\infty}^{\infty}x\varphi(x|\,y)dx.$$

Початковим моментом $\nu_{k.s}$ порядку $k.s$ системи $(X;Y)$ називають математичне сподівання добутку X^kY^s: $\nu_{k.s}=M(X^kY^s)$.

Окремим випадком є $\nu_{1.0}=M(X),\ \nu_{0.1}=M(Y).$

Центральним моментом $\mu_{k.s}$ порядку $k.s$ системи $(X;Y)$ називають математичне сподівання добутку відхилень, відповідно, k -го і μ-го ступеня:

$$\mu_{k.s}=M((X-M(X))^k(Y-M(Y))^s).$$

Окремим випадком є $\mu_{2.0}=M(X-M(X))^2=D(X),$

$$\mu_{0.2}=M(Y-M(Y))^2=D(Y).$$

Кореляційним моментом μ_{xy} системи $(X;Y)$ називають центральний момент $\mu_{1,1}$ порядку 1,1, що дорівнює математичному сподіванню добутку відхилень цих величин: $\mu_{xy}=M((X-M(X))(Y-M(Y)))$.

Для дискретних величин формула набуває такого вигляду:

$$\mu_{xy}=\sum_{i=1}^{n}\sum_{j=1}^{m}(x_i-M(X))(y_i-M(Y))p(x_i;\,y_i),$$

для неперервних величин:

$$\mu_{xy}=\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}(x-M(X))(y-M(Y))f(x;\,y)dxdy=$$

$$=\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}xyf(x;\,y)dxdy-M(X)\cdot M(Y).$$

Кореляційний момент також має назву **коефіцієнта коваріації**, тобто одночасної варіації випадкових величин X і Y. Відповідно, його позначають K_{xy}, або $\operatorname{cov}(X;Y)$.

Властивості коефіцієнта коваріації:

1. Коефіцієнт коваріації не змінюється, якщо випадкові величини поміняти місцями: $\operatorname{cov}(X; Y) = \operatorname{cov}(Y; X)$.

2. Кореляційний момент між однаковими компонентами двовимірної випадкової величини дорівнює дисперсії цього компонента:

$$\mu_{xx} = M\Big(\big(X - M(X)\big)\big(X - M(X)\big)\Big) = D\big(X\big);$$

$$\mu_{yy} = M\Big(\big(Y - M(Y)\big)\big(Y - M(Y)\big)\Big) = D\big(Y\big).$$

3. **Коваріаційною матрицею двовимірної випадкової величини** називають матрицю $C = \left(c_{ij}\right)_{n \times n}$, де $c_{ij} = \operatorname{cov}(X_i; X_j)$, тобто для двовимірної випадкової величини коваріаційна матриця має вигляд:

$$C = \begin{pmatrix} D\big(X\big) & \operatorname{cov}(X; Y) \\ \operatorname{cov}(Y; X) & D\big(Y\big) \end{pmatrix}.$$

Коваріаційна матриця є симетричною відносно головної діагоналі. Вона є невід'ємно визначеною, звідки випливає, що

$$D\big(X\big) \cdot D\big(Y\big) - \operatorname{cov}(X; Y) \cdot \operatorname{cov}(Y; X) \geq 0.$$

Отже, кореляційний момент є обмеженим за модулем:

$$\left|\mu_{xy}\right| \leq \sqrt{D\big(X\big) \cdot D\big(Y\big)}.$$

4. Для залежних випадкових величин X і Y буде співвідношення для дисперсії суми випадкових величин:

$$D(X + Y) = D\big(X\big) + D\big(Y\big) + 2\operatorname{cov}(X; Y).$$

Якщо випадкові величини є незалежними, то $\operatorname{cov}(X; Y) = 0$, і визначають співвідношення:

$$D(X + Y) = D\big(X\big) + D\big(Y\big).$$

У ході обчислення кореляційного моменту замість означення зручно використовувати таку перетворену формулу:

$$\mu_{xy} = M\big(XY\big) - M\big(X\big) \cdot M\big(Y\big).$$

Теорема. Кореляційний момент двох незалежних випадкових величин X і Y дорівнює нулю.

Таким чином, якщо кореляційний момент не дорівнює нулю, то це є ознакою наявності зв'язку між випадковими величинами X і Y.

Коефіцієнтом кореляції називають числову характеристику, яку обчислюють за такою формулою:

$$r_{xy} = \frac{\mu_{xy}}{\sigma_x \sigma_y}.$$

Властивості коефіцієнта кореляції:

1. Для незалежних випадкових величин $r_{xy} = 0$, оскільки в цьому випадку $\mu_{xy} = 0$.

2. За абсолютною величиною коефіцієнт кореляції є не більшим за одиницю:

$$-1 \le r_{xy} \le 1.$$

3. Оскільки для кореляційного моменту $\mathrm{cov}(X;Y) = \mathrm{cov}(Y;X)$, то для коефіцієнта кореляції теж

$$\rho_{xy} = \rho_{yx}.$$

Випадкові величини, для яких $r_{xy} = 0$, називають **некорельованими**.

7.3. Приклади розв'язування задач

Приклад 7.1. Задано закон розподілу ймовірностей дискретної двовимірної випадкової величини (табл. 7.2). Знайдіть закони розподілу X і Y.

Таблиця 7.2

Закон розподілу ймовірностей дискретної двовимірної випадкової величини

Y \ X	10	15	10
2	0,03	0,09	0,11
4	0,10	0,15	0,24
5	0,12	0,13	0,03

Розв'язання. Слід знайти закон розподілу для складової X. Вона набуває таких значень: 10, 15, 20. Тепер обчисліть відповідні ймовірності, для цього треба підсумувати ймовірності за стовпцями:

$$P(10) = 0,03 + 0,10 + 0,12 = 0,25;$$

$$P(15) = 0,09 + 0,15 + 0,13 = 0,37;$$

$$P(20) = 0,11 + 0,24 + 0,03 = 0,39.$$

Перевірка: $0,25 + 0,37 + 0,38 = 1$.

Отже, закон розподілу складової X має такий вигляд (табл. 7.3):

Таблиця 7.3

Закон розподілу для складової X

X	10	15	20
P	0,25	0,37	0,38

Знайдіть закон розподілу для складової Y, значення якої 2, 4, 6. Для обчислення відповідних імовірностей підсумуйте ймовірності за рядками:

$$P(2) = 0,03 + 0,09 + 0,11 = 0,23;$$

$$P(4) = 0,10 + 0,15 + 0,24 = 0,49;$$

$$P(6) = 0,12 + 0,13 + 0,03 = 0,28.$$

Перевірка: $0,23 + 0,49 + 0,28 = 1$.

Отже, закон розподілу складової Y має такий вигляд (табл. 7.4):

Таблиця 7.4

Закон розподілу для складової Y

Y	2	3	6
P	0,23	0,49	0,28

Приклад 7.2. Задано закон розподілу ймовірностей дискретної двовимірної випадкової величини (табл. 7.5). Знайдіть: а) безумовні закони розподілу складових X і Y; б) умовний закон розподілу складової X

за умови, що складова Y набула значення $y_1 = 10$; в) умовний закон розподілу складової Y за умови, що складова X набула значення $x_2 = 7$.

Таблиця 7.5

Закон розподілу ймовірностей
дискретної двовимірної випадкової величини

Y \ X	5	7	10
10	0,31	0,12	0,19
20	0,11	0,15	0,12

Розв'язання:

а) складіть закон розподілу для складової X. Для цього підрахуйте ймовірності:

$$P(5) = 0,31 + 0,11 = 0,42;$$
$$P(7) = 0,12 + 0,15 = 0,27;$$
$$P(10) = 0,19 + 0,12 = 0,31.$$

Перевірка: $0,42 + 0,27 + 0,31 = 1.$

Отже, закон розподілу складової X має такий вигляд (табл. 7.6):

Таблиця 7.6

Закон розподілу для складової X

X	5	7	10
P	0,42	0,27	0,31

Для складової Y імовірності:

$$P(10) = 0,31 + 0,12 + 0,19 = 0,62;$$
$$P(20) = 0,11 + 0,15 + 0,12 = 0,38.$$

Перевірка: $0,62 + 0,38 = 1.$

Отже, закон розподілу складової Y має такий вигляд (табл. 7.7):

Таблиця 7.7

Закон розподілу для складової Y

Y	10	20
P	0,62	0,38

б) тепер знайдіть умовний закон розподілу складової X за умови, що складова Y набула значення $y_1 = 10$. Для цього обчисліть умовні ймовірності можливих значень X за такою формулою:

$$P(x_i | y_i) = \frac{P(x_i; y_i)}{P(y_i)};$$

$$P(y_1) = P(10) = 0,62.$$

Тоді

$$P(x_1 | y_1) = \frac{P(x_1; y_1)}{P(y_1)} = \frac{0,31}{0,62} = \frac{1}{2};$$

$$P(x_2 | y_1) = \frac{P(x_2; y_1)}{P(y_1)} = \frac{0,12}{0,62} = \frac{6}{31};$$

$$P(x_3 | y_1) = \frac{P(x_3; y_1)}{P(y_1)} = \frac{0,19}{0,62} = \frac{19}{62}.$$

Отже, умовний закон розподілу складової X за умови, що складова Y набула значення $y_1 = 10$, має такий вигляд (табл. 7.8):

Таблиця 7.8

Закон розподілу для складової X

X	5	7	10
$P(X \| y_1)$	$\frac{1}{2}$	$\frac{6}{31}$	$\frac{19}{62}$

Перевірка: $\frac{1}{2} + \frac{6}{31} + \frac{19}{62} = 1;$

в) для знаходження умовного закону розподілу складової Y за умови, що складова X набула значення $x_2 = 7$, обчисліть умовні ймовірності можливих значень Y за такою формулою:

$$P(y_i | x_i) = \frac{P(x_i;\, y_i)}{P(x_i)};$$

$$P(x_2) = P(7) = 0{,}12 + 0{,}15 = 0{,}27.$$

Далі

$$P(y_1 | x_2) = \frac{P(x_2;\, y_1)}{P(x_2)} = \frac{0{,}12}{0{,}27} = \frac{12}{27} = \frac{4}{9};$$

$$P(y_2 | x_2) = \frac{P(x_2;\, y_2)}{P(x_2)} = \frac{0{,}15}{0{,}27} = \frac{15}{27} = \frac{5}{9}.$$

Таким чином, умовний закон розподілу складової Y за умови, що складова X набула значення $x_2 = 7$, має такий вигляд (табл. 7.9):

Таблиця 7.9

Закон розподілу для складової Y

Y	10	20
$P(Y \| x_2)$	$\frac{4}{9}$	$\frac{5}{9}$

Перевірка: $\frac{4}{9} + \frac{5}{9} = 1.$

Приклад 7.3. Задано інтегральну функцію розподілу ймовірностей двовимірної випадкової величини $(X; Y)$:

$$F(x;\, y) = \left(\frac{1}{4\pi} arctg \frac{x}{4} + \frac{1}{8}\right)\left(\frac{1}{5\pi} arctg \frac{y}{5} + \frac{1}{10}\right).$$

Знайдіть імовірність того, що в результаті випробування складова X набуде значення $X \le 4$, а складова Y – значення $Y < 5$.

Розв'язання. Шукану ймовірність знаходять за такою формулою:

$$F(x; y) = P(X < x; Y < y).$$

Тому

$$P(X < x; Y < y) = F(4; 5) = \left(\frac{1}{4\pi} arctg\frac{4}{4} + \frac{1}{8}\right)\left(\frac{1}{5\pi} arctg\frac{5}{5} + \frac{1}{10}\right) =$$

$$= \left(\frac{1}{4\pi}\cdot\frac{\pi}{4} + \frac{1}{8}\right)\left(\frac{1}{5\pi}\cdot\frac{\pi}{4} + \frac{1}{10}\right) = \left(\frac{1}{16} + \frac{1}{8}\right)\left(\frac{1}{20} + \frac{1}{10}\right) = \frac{3}{16}\cdot\frac{3}{20} = \frac{9}{320}.$$

Приклад 7.4. Задано інтегральну функцію двовимірної випадкової величини:

$$F(x; y) = \begin{cases} \frac{1}{54}(x^2 y + xy^2), & \text{за } 0 \le x \le 3,\ 0 \le y \le 3, \\ 0, & \text{за межами квадрата.} \end{cases}$$

Знайдіть імовірність того, що в результаті випробування випадкова точка $(X; Y)$ потрапить у квадрат, який обмежено лініями $x_1 = 1,\ x_2 = 2,$ $y_1 = 2,\ y_2 = 3.$

Розв'язання. Для знаходження ймовірності використайте таку формулу:

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

Підставте $x_1 = 1, x_2 = 2, y_1 = 2, y_2 = 3$:

$$P(1 < X < 2; 2 < Y < 3) = F(2; 3) - F(1; 3) - F(2; 2) + F(1; 2) =$$

$$= \frac{1}{54}(2^2\cdot 3 + 2\cdot 3^2 - 1^2\cdot 3 - 1\cdot 3^2 - 2^2\cdot 2 - 2\cdot 2^2 + 1^2\cdot 2 + 1\cdot 2^2) =$$

$$= \frac{1}{54}(12 + 18 - 3 - 9 - 8 - 8 + 2 + 4) = \frac{8}{54} = \frac{4}{27}.$$

Приклад 7.5. Задано інтегральну функцію двовимірної випадкової величини $(X;Y)$:

$$F(x;\,y)=\begin{cases}(1-e^{-5x})(1-e^{-2y}), & \text{за } x>0,\ y>0,\\ 0, & \text{за } x\le 0 \text{ або } y\le 0.\end{cases}$$

Знайдіть диференціальну функцію розподілу величини $(X;Y)$.

Розв'язання. За означенням

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

Спочатку, уважаючи, що $y=const$, знайдіть частинну похідну за X від функції $F(x;\,y)=(1-e^{-5x})(1-e^{-2y})$:

$$\frac{\partial F(x;\,y)}{\partial x}=(1-e^{-2y})(0-e^{-5x}(-5))=5e^{-5x}(1-e^{-2y}).$$

Тепер знайдіть частинну похідну за y від функції $\dfrac{\partial F(x;\,y)}{\partial x}$, уважаючи, що $x=const$:

$$\frac{\partial^2 F\left(x;\,y\right)}{\partial x\partial y}=5e^{-5x}\left(0-e^{-2y}\left(-2\right)\right)=10e^{-5x}\cdot e^{-2y}=10e^{-\left(5x+2y\right)}.$$

Отже,

$$f\left(x;\,y\right)=\begin{cases}10e^{-\left(5x+2y\right)}, & \text{за } x>0,\ y>0,\\ 0, & \text{за } x\le 0 \text{ або } y\le 0.\end{cases}$$

Приклад 7.6. Знайдіть a, якщо диференціальна функція системи випадкових величин $\left(X;Y\right)$ має такий вигляд:

$$f(x;\ y)=\begin{cases}a\cos x\cos y, & \text{за } 0\le x\le \dfrac{\pi}{4},\ 0\le y\le \dfrac{\pi}{4},\\ 0, & \text{за межами квадрата.}\end{cases}$$

Розв'язання. Для того щоб функція $f\left(x;\,y\right)$ була диференціальною функцією системи випадкових величин, необхідно виконання двох умов:

а) $f\left(x;\,y\right)\ge 0$; б) $\iint\limits_{D} f\left(x;\,y\right)dxdy=1$.

Тоді:

а) $f\left(x;\,y\right)=a\cos x\cos y$, $\cos x\ge 0$, $\cos y\ge 0$, отже $a\ge 0$;

б) $\iint\limits_{D} f\left(x;\,y\right)dxdy=1$;

$$a\iint\limits_{D}\cos x\cos y dxdy=a\int\limits_{0}^{\frac{\pi}{4}}\int\limits_{0}^{\frac{\pi}{4}}\cos x\cos y dxdy=a\int\limits_{0}^{\frac{\pi}{4}}\cos xdx\int\limits_{0}^{\frac{\pi}{4}}\cos ydy=$$

$$=a\int\limits_{0}^{\frac{\pi}{4}}\cos xdx\sin y\Big|_{0}^{\frac{\pi}{4}}=a\int\limits_{0}^{\frac{\pi}{4}}\cos x\left(\sin\frac{\pi}{4}-\sin 0\right)dx=\frac{a\sqrt{2}}{2}\int\limits_{0}^{\frac{\pi}{4}}\cos xdx=$$

$$=\frac{a\sqrt{2}}{2}\sin x\Big|_{0}^{\frac{\pi}{4}}=\frac{a\sqrt{2}}{2}\left(\sin\frac{\pi}{4}-0\right)=\frac{a\sqrt{2}}{2}\cdot\frac{\sqrt{2}}{2}=\frac{a}{2}.$$

Отже, $\frac{a}{2}=1$, звідки $a=2$.

Приклад 7.7. Закон розподілу дискретної двовимірної випадкової величини задано таблицею (табл. 7.10).

Таблиця 7.10

Закон розподілу двовимірної дискретної випадкової величини

Y \ X	1	2	$\sum\limits_{i=1}^{n} p_{ij}$
2	0,4	0,3	0,7
4	0,2	0,1	0,3
$\sum\limits_{j=1}^{m} p_{ij}$	0,6	0,4	1 = 1

Визначте кореляційний момент компонентів X і Y двовимірної випадкової величини.

Розв'язання. Обчисліть математичні сподівання випадкових величин X і Y за їхніми розподілами. Так, для випадкової величини X за першим і четвертим рядками табл. 7.10 визначте:

$$M\left(X\right)=\sum_{i=1}^{n}x_i p_i=1\cdot 0,6+2\cdot 0,4=1,4.$$

Для випадкової величини Y за першим і останнім стовпчиками табл. 7.10 обчисліть математичне сподівання:

$$M\left(Y\right)=\sum_{j=1}^{m}y_j p_j=2\cdot 0,7+4\cdot 0,3=2,6.$$

Визначте математичне сподівання двовимірної випадкової величини, яка є добутком випадкових величин X і Y:

$$M\left(XY\right)=\sum_{i=1}^{n}\sum_{j=1}^{m}x_i y_j p_{ij}=1\cdot 2\cdot 0,4+2\cdot 2\cdot 0,3+4\cdot 1\cdot 0,2+4\cdot 2\cdot 0,1=3,6.$$

За формулою $\mu_{xy}=M\left(XY\right)-M(X)\cdot M(Y)$ обчисліть кореляційний момент:

$$\mu_{xy}=3,6-1,4\cdot 2,6=-0,04.$$

Отже, кореляційний момент не дорівнює нулю, відповідно, компоненти X і Y двовимірної випадкової величини не можна вважати незалежними.

7.4. Вправи для самостійної роботи

7.1. Задано закон розподілу ймовірностей дискретної двовимірної випадкової величини (табл. 7.11). Складіть закон розподілу складових X і Y.

Таблиця 7.11

Закон розподілу двовимірної дискретної випадкової величини

Y \ X	5	15	0
2	0,15	0,21	0,05
3	0,16	0,30	0,13

7.2. Складіть закон розподілу складових X і Y, якщо закон розподілу ймовірностей дискретної двовимірної випадкової величини має такий вигляд (табл. 7.12):

Таблиця 7.12

Закон розподілу двовимірної дискретної випадкової величини

Y \ X	10	15	20	25
−1	0,05	0,08	0,11	0,12
0	0,04	0,12	0,10	0,08
1	0,06	0,09	0,07	0,08

7.3. Задано закон розподілу ймовірностей дискретної двовимірної випадкової величини (табл. 7.13).

Таблиця 7.13

Закон розподілу двовимірної дискретної випадкової величини

Y \ X	40	45	50	55
1,5	0,07	0,08	0,09	0,12
2,8	0,06	0,12	0,10	0,08
3,3	0,08	0,02	0,06	0,02

Знайдіть:

а) безумовні закони розподілу складових X і Y;

б) умовний закон розподілу складової X за умови, що складова Y набуде значення $y_2 = 2{,}8$;

в) умовний закон розподілу складової Y за умови, що складова X набуде значення $x_2 = 45$.

7.4. Задано закон розподілу ймовірностей дискретної двовимірної випадкової величини (табл. 7.14).

Таблиця 7.14

Закон розподілу двовимірної дискретної випадкової величини

Y \ X	11	15	17
2	0,06	0,05	0,04
9	0,09	0,09	0,11
14	0,10	0,12	0,06
17	0,07	0,09	0,12

Знайдіть:

а) безумовні закони розподілу складових X і Y;

б) умовний закон розподілу складової X за умови, що складова Y набуде значення $y_3 = 14$;

в) умовний закон розподілу складової Y за умови, що складова X набуде значення $x_1 = 11$.

7.5. На біржі продають акції п'яти компаній, для яких складено прогноз можливого підвищення їхньої вартості. Перші три компанії є будівельними. Для акцій першої компанії ймовірність зростання вартості становить 0,2; для другої – 0,1; для третьої – 0,4. Четверта та п'ята компанії є виробниками продовольчих товарів. Для четвертої компанії ймовірність зростання вартості становить 0,3; для п'ятої – 0,1. Клієнт придбав 2 акції четвертої компанії, 5 акцій третьої компанії та по 1 акції другої та п'ятої компаній. Складіть закон розподілу, розглядаючи двовимірну випадкову величину, одним компонентом якої є кількість акцій будівельних компаній, ціна на які може зрости, а другим компонентом – кількість акцій компаній із виготовлення харчових продуктів, вартість яких буде зростати, якщо підвищення вартості акції різних компаній є незалежними подіями.

7.6. Інтегральна функція розподілу ймовірностей двовимірної випадкової величини $(X;Y)$ має такий вигляд:

$$F(x;y) = \left(\frac{1}{\pi} arctg\, 2x + \frac{1}{2}\right)\left(\frac{1}{\pi} arctg\, 3y + \frac{1}{2}\right).$$

Знайдіть імовірність того, що складова X набуде значення $X < \frac{1}{2}$ та водночас складова Y – значення $Y < \frac{1}{3}$.

7.7. Перевірте, чи становлять дані, наведені в таблиці (табл. 7.15), закон розподілу двовимірної випадкової величини:

Таблиця 7.15

Вихідні дані

Y \ X	0	2	4	6
5	0,15	0,10	0,10	0,05
10	0,20	0,25	0,10	0,05

Якщо в таблиці дійсно наведено закон розподілу двовимірної випадкової величини $(X;Y)$, побудуйте умовний розподіл випадкової величини X, якщо $Y=10$.

7.8. За умовами задачі 7.7 обчисліть основні числові характеристики двовимірної випадкової величини $(X;Y)$. За значенням коефіцієнта кореляції зробіть висновок про наявність зв'язку між її компонентами.

7.9. За умовами задачі 7.8 складіть коваріаційну матрицю випадкової величини $(X;Y)$ і перевірте, чи є ця матриця невід'ємно визначеною.

7.10. Задано розподіл імовірностей дискретної двовимірної випадкової величини (табл. 7.16):

Таблиця 7.16

Закон розподілу двовимірної дискретної випадкової величини

Y \ X	1,4	1,6	2,0
0	0,11	0,12	0,07
4,2	0,18	0,11	0,11
6,2	0,10	0,08	0,12

Знайдіть:

а) математичне сподівання складових;

б) дисперсії випадкових;

в) середні квадратичні відхилення;

г) коефіцієнт кореляції.

7.11. Задано розподіл імовірностей дискретної двовимірної випадкової величини (табл. 7.17):

Таблиця 7.17

Закон розподілу двовимірної дискретної випадкової величини

$Y \backslash X$	10	15	20	25
0	0,10	0,15	0,14	0,11
4	0,09	0,08	0,07	0,06
6	0,04	0,06	0,08	0,02

Знайдіть:

1) умовні математичні сподівання складової X за $Y = y_2 = 4$, а також складової Y за $X = x_3 = 20$;

2) умовні математичні сподівання складової X за $Y = y_1 = 0$, а також складової Y за $X = x_2 = 15$.

7.12. Систему двох випадкових величин $(X; Y)$ рівномірно розподілено у трикутнику, обмеженому прямими $y = x$, $y = 0$, $x = 2$.

Знайдіть:

1) $f(x; y)$;

2) математичні сподівання складових X і Y;

3) дисперсії складових X і Y;

4) коефіцієнт кореляції.

7.13. Диференціальна функція системи двох випадкових величин $(X; Y)$:

$$f(x; y) = \begin{cases} a(x+y), & \text{в області } D, \\ 0, & \text{за межами } D, \end{cases}$$

де область D обмежено лініями $x = 0, x = 3, y = 0, y = 3$.

Знайдіть:

1) a;

2) математичні сподівання складових X і Y;

3) дисперсії складових X і Y;

4) коефіцієнт кореляції.

7.14. Задано щільність сумісного розподілу неперервної випадкової величини $(X;Y)$:

$$f(x;y)=\begin{cases}2\cos x\cos y, & \text{за } 0\le x\le\frac{\pi}{2},\ 0\le y\le\frac{\pi}{2},\\ 0, & \text{за межами квадрата.}\end{cases}$$

Знайдіть математичні сподівання і дисперсії складових X і Y.

7.15. Задано щільність сумісного розподілу неперервної випадкової величини $(X;Y)$:

$$f(x;y)=\begin{cases}2\sin(x+y), & \text{за } 0\le x\le\frac{\pi}{2},\ 0\le y\le\frac{\pi}{2},\\ 0, & \text{за межами квадрата.}\end{cases}$$

Знайдіть:

1) коефіцієнт a;

2) $M(X)$, $M(Y)$;

3) $D(X)$, $D(Y)$;

4) коефіцієнт кореляції.

7.16. Задано диференціальну функцію неперервної двовимірної випадкової величини $(X;Y)$:

$$f(x;y)=\begin{cases}\frac{1}{2}\sin x\sin y, & \text{за } 0\le x\le\pi,\ 0\le y\le\pi,\\ 0, & \text{за межами квадрата.}\end{cases}$$

Знайдіть:

1) математичні сподівання складових X і Y;

2) дисперсії складових X і Y;

3) коефіцієнт кореляції.

7.5. Тестові завдання

7.1. Інтегральною функцією розподілу ймовірностей двовимірної випадкової величини $(X;Y)$ називають функцію $F(x;\, y)$ таку, що

А $F(x;\, y) = P(X \le x;\; Y < y)$

Б $F(x;\, y) = P(X < x;\; Y \le y)$

В $F(x;\, y) = P(X < x;\; Y < y)$

Г $F(x;\, y) = P(X \le x;\; Y \le y)$

7.2. Значення інтегральної функції задовольняють подвійну нерівність

А	Б	В	Г
$0 \le F(x;\, y) < 1$	$0 \le F(x;\, y) \le 1$	$0 < F(x;\, y) \le 1$	$0 < F(x;\, y) < 1$

7.3. Умовним розподілом складової X за $Y = y_j$ називають сукупність значень X і відповідні їм умовні ймовірності, які обчислюють за такою формулою

А $P(x_i | y_j) = \dfrac{P(x_i;\, y_j)}{P(y_j)}$

Б $P(y_j | x_i) = \dfrac{P(x_i;\, y_j)}{P(x_i)}$

В $P(x_i | y_j) = \dfrac{P(x_i;\, y_j)}{P(x_i)}$

Г $P(y_j | x_i) = \dfrac{P(x_i;\, y_j)}{P(y_j)}$

7.4. Умовним розподілом складової Y за $X = x_j$ називають сукупність значень Y і відповідні їм умовні ймовірності, які обчислюють за такою формулою

А $P(x_i | y_j) = \dfrac{P(x_i;\, y_j)}{P(y_j)}$

Б $P(y_j|x_i)=\dfrac{P(x_i;\,y_j)}{P(x_i)}$

В $P(x_i|y_j)=\dfrac{P(x_i;\,y_j)}{P(x_i)}$

Г $P(y_j|x_i)=\dfrac{P(x_i;\,y_j)}{P(y_j)}$

7.5. Математичні сподівання дискретних випадкових величин X і Y, що входять до системи, визначають за такими формулами

А $M(X)=\sum_{i=1}^{n}\sum_{j=1}^{m}x_i p_{ij},\ M(Y)=\sum_{i=1}^{n}\sum_{j=1}^{m}y_j p_{ij}$

Б $M(X)=\sum_{i=1}^{n}\sum_{j=1}^{m}x_j p_{ij},\ M(Y)=\sum_{i=1}^{n}\sum_{j=1}^{m}y_j p_{ij}$

В $M(X)=\sum_{i=1}^{n}\sum_{j=1}^{m}x_i p_{ij},\ M(Y)=\sum_{i=1}^{n}\sum_{j=1}^{m}y_i p_{ij}$

Г $M(X)=\sum_{j=1}^{n}\sum_{i=1}^{m}x_i p_{ij},\ M(Y)=\sum_{i=1}^{n}\sum_{j=1}^{m}y_j p_{ij}$

7.6. Для обчислення дисперсії дискретних випадкових величин X і Y застосовують формули

А $D(X)=\sum_{i=1}^{n}\sum_{j=1}^{m}p_{ij}\left(x_j-M(X)\right)^2,\ D(Y)=\sum_{i=1}^{n}\sum_{j=1}^{m}p_{ij}\left(y_j-M(Y)\right)^2$

Б $D(X)=\sum_{i=1}^{n}\sum_{j=1}^{m}p_{ij}\left(x_i-M(X)\right)^2,\ D(Y)=\sum_{i=1}^{n}\sum_{j=1}^{m}p_{ij}\left(y_j-M(Y)\right)^2$

В $D(X)=\sum_{i=1}^{n}\sum_{j=1}^{m}p_{ij}\left(x_i-M(X)\right)^2,\ D(Y)=\sum_{i=1}^{n}\sum_{j=1}^{m}p_{ij}\left(y_i-M(Y)\right)^2$

Г $D(X)=\sum_{j=1}^{n}\sum_{i=1}^{m}p_{ij}\left(x_i-M(X)\right)^2,\ D(Y)=\sum_{i=1}^{n}\sum_{j=1}^{m}p_{ij}\left(y_j-M(Y)\right)^2$

7.7. Кореляційним моментом μ_{xy} системи $(X;Y)$ називають

А $\mu_{xy} = M\big((Y - M(X))(X - M(Y))\big)$

Б $\mu_{xy} = M\big((X - M(Y))(Y - M(X))\big)$

В $\mu_{xy} = M\big((X - M(X))(X - M(Y))\big)$

Г $\mu_{xy} = M\big((X - M(X))(Y - M(Y))\big)$

7.8. Коефіцієнт кореляції обчислюють за такою формулою

А	Б	В	Г
$r_{xy} = \dfrac{\mu_{xy}}{D_x D_y}$	$r_{xy} = \dfrac{\mu_{xy}}{\sigma_x \sigma_y}$	$r_{xy} = \dfrac{\rho_{xy}}{\sigma_x \sigma_y}$	$r_{xy} = \dfrac{\nu_{xy}}{\sigma_x \sigma_y}$

7.9. Коефіцієнт кореляції набуває таких значень

А	Б	В	Г
$\lvert r_{xy} \rvert \ge 1$	$\lvert r_{xy} \rvert \le 1$	$\lvert r_{xy} \rvert \le 0$	$\lvert r_{xy} \rvert \ge 0$

7.10. Для двовимірної випадкової величини коваріаційна матриця має такий вигляд

А $C = \begin{pmatrix} D(Y) & \operatorname{cov}(X;Y) \\ \operatorname{cov}(Y;X) & D(X) \end{pmatrix}$

Б $C = \begin{pmatrix} D(X) & \operatorname{cov}(Y;X) \\ \operatorname{cov}(X;Y) & D(Y) \end{pmatrix}$

В $C = \begin{pmatrix} D(X) & \operatorname{cov}(X;Y) \\ \operatorname{cov}(Y;X) & D(Y) \end{pmatrix}$

Г $C = \begin{pmatrix} D(X) & \operatorname{cov}(Y;X) \\ \operatorname{cov}(Y;X) & D(Y) \end{pmatrix}$

7.6. Запитання для самоперевірки

7.1. Дайте означення багатовимірної випадкової величини.

7.2. Якими бувають багатовимірні величини?

7.3. Наведіть геометричну інтерпретацію двовимірної випадкової величини.

7.4. Як можна задати закон розподілу двовимірної дискретної величини?

7.5. Що таке "умовний розподіл"? Як побудувати умовний розподіл, користуючись табличними даними про розподіл випадкової величини?

7.6. Як за табличного задавання розподілу дискретної двовимірної величини отримати інформацію про розподіл її компонентів?

7.7. Наведіть способи задавання двовимірної неперервної випадкової величини.

7.8. Як можна надати геометричну інтерпретацію розподілу неперервної випадкової величини?

7.9. Як визначають щільність імовірностей і які властивості вона має?

7.10. Які існують основні числові характеристики розподілу двовимірної випадкової величини?

7.11. Якому теоретичному моменту відповідає коефіцієнт коваріації?

7.12. Що таке "коваріаційна матриця"?

7.13. Які властивості має коваріаційна матриця?

7.14. У чому полягають переваги використання коефіцієнта кореляції як характеристики щільності кореляційного зв'язку перед кореляційним моментом?

7.15. Чи може коефіцієнт кореляції бути від'ємним? Якщо так, то що це означає?

7.16. Чи можна вважати некорельовані випадкові величини незалежними?

7.7. Висновки за темою

Після вивчення теми студент має вміти:

складати закон розподілу двовимірної дискретної випадкової величини;

визначати умовні закони розподілу дискретних і неперервних двовимірних випадкових величин та вміти обчислювати їхні основні числові характеристики.

8. Функції випадкового аргументу

8.1. Мета та компетентності

Метою вивчення теми є ознайомлення студентів із поняттям функції випадкового аргументу, принципів побудови її закону розподілу, особливостями побудови розподілів функції одного випадкового аргументу за відомим розподілом.

Компетентності, що формують під час вивчення теми:

розуміння сутності поняття функції, аргументом якої є випадкова величина;

уміння будувати закон розподілу функції одного випадкового аргументу за відомим розподілом її аргументу та визначати основні числові характеристики цієї функції;

формування схильності до самостійного пошуку засобів до розв'язання задачі;

відповідальність за точність і коректність розрахунків імовірності значень функції випадкового аргументу та її числових характеристик.

8.2. Основні теоретичні відомості

Якщо кожному можливому значенню випадкової величини X відповідає одне можливе значення випадкової величини Y, то Y називають **функцією випадкового аргументу** X і записують так: $Y = f(X)$.

Якщо аргумент X – дискретна випадкова величина і різним можливим значенням аргументу X відповідають різні можливі значення функції Y, то ймовірності відповідних значень X і Y однакові між собою, тобто $P(Y = y_i) = P(X = x_i)$.

Якщо $x_1, x_2, \dots, x_n$ – можливі значення дискретної випадкової величина X, імовірності яких дорівнюють, відповідно, $p_1, p_2, \dots, p_n$, то для випадкової величини X закон розподілу записують у такому вигляді (табл. 8.1):

Таблиця 8.1

Закон розподілу дискретної випадкової величини X

X	x_1	x_2	$\cdots$	x_n
$P(X)$	p_1	p_2	$\cdots$	p_n

Якщо різним значенням X відповідають однакові значення Y, то ймовірності значення Y, що повторюються, додаються.

Нехай $Y = f(X)$, де аргумент X – дискретна випадкова величина, тоді випадкова величина Y також є дискретною випадковою величиною з можливими значеннями $y_1 = f(x_1)$, $y_2 = f(x_2)$, …, $y_n = f(x_n)$. Для того щоб випадкова величина Y набула значення $y_1 = f(x_1)$ необхідно, щоб випадкова величина X набула значення x_1. Випадкова величина X набуває можливого значення x_1 з імовірністю p_1. Аналогічно, для того щоб випадкова величина Y набула значення $y_2 = f(x_2)$ необхідно, щоб випадкова величина X набула значення x_2. Випадкова величина X набуває можливого значення x_2 з імовірністю p_2. Таким чином, можна зробити висновок, що закон розподілу випадкової величини Y має такий вигляд (табл. 8.2):

Таблиця 8.2

Закон розподілу випадкової величини Y

$Y = f(X)$	$f(x_1)$	$f(x_2)$	$\cdots$	$f(x_n)$
$P(Y)$	p_1	p_2	$\cdots$	p_n

Якщо аргумент X – це неперервна випадкова величина, задана диференціальною функцією (щільністю розподілу) $f(x)$, і якщо $y = \varphi(x)$ – диференційована строго зростаюча або строго спадна функція, обернена функція якої $x = \psi(y)$, то диференціальну функцію $g(y)$ випадкової величини Y знаходять із такої рівності:

$$g(y) = f(\psi(y)) \cdot |\psi'(y)|.$$

Примітка. Якщо функція $y = \varphi(x)$ не монотонна в інтервалі можливих значень X, то потрібно цей інтервал розподілити на такі інтервали, у яких функція $y = \varphi(x)$ буде монотонною, і знайти диференціальні функції $g_i(y)$ у кожному із цих інтервалів, а потім обчислити їхню суму:

$$g(y) = \sum_{i=1}^{n} g_i(y) = \sum_{i=1}^{n} f(\psi_i(y)) \cdot |\psi_i'(y)|.$$

8.2.1. Числові характеристики функції дискретного випадкового аргументу

До числових характеристик дискретної випадкової величини Y належать математичне сподівання, дисперсія та середнє квадратичне відхилення. Їх обчислюють за такими формулами:

а) математичне сподівання:

$$M(Y)=\sum_{k=1}^{n} f(x_k)p_k\,;$$

б) дисперсію:

$$D(Y)=M\left(Y^2\right)-M^2(Y)=\sum_{k=1}^{n}\left(f(x_k)\right)^2 p_k-M^2(Y);$$

в) середнє квадратичне відхилення:

$$\sigma(Y)=\sqrt{D(Y)}\,.$$

8.2.2. Числові характеристики функції неперервного випадкового аргументу

Нехай аргумент X – це неперервна випадкова величина, тоді математичне сподівання функції Y обчислюють за такою формулою:

$$M(Y)=M(\varphi(X))=\int_{-\infty}^{\infty}\varphi(X)\,f(x)dx,$$

якщо $x\in(a;b)$, то:

$$M(Y)=\int_{a}^{b}\varphi(X)\,f(x)dx.$$

Дисперсію функції Y обчислюють за такою формулою:

$$D(Y)=D\left(\varphi(X)\right)=\int_{-\infty}^{\infty}\left(\varphi(x)-M\left(\varphi(X)\right)\right)^2 f(x)dx=$$

$$=\int_{-\infty}^{\infty}\varphi^2(x)\,f(x)dx-\left(M\left(\varphi(X)\right)\right)^2.$$

Якщо $x \in (a;b)$, то:

$$D(Y) = \int_a^b \left(\varphi(x) - M\left(\varphi(X)\right)\right)^2 f(x)dx = \int_a^b \varphi^2(x) f(x)dx - \left(M\left(\varphi(X)\right)\right)^2.$$

Примітка. Якщо використовувати диференціальну функцію $g(y)$ випадкової величини Y, то математичне сподівання знаходять за такою формулою:

$$M(Y) = \int_c^d y \cdot g(y)dy,$$

де c і d – межі інтегрування для Y,

а дисперсію – за такою формулою:

$$D(Y) = \int_c^d y^2 g(y)dy - \left(M(Y)\right)^2.$$

8.8.3. Функції двох випадкових аргументів

Означення. Якщо кожній парі можливих значень випадкових величин X та Y відповідає одне можливе значення випадкової величини Z, то Z є **функцією двох випадкових аргументів** X **та** Y**:** $Z = \varphi(X;\ Y)$.

Якщо X та Y є дискретними випадковими величинами, то й Z буде дискретною, а якщо X та Y є неперервними випадковими величинами, то й Z буде неперервною.

Диференціальну функцію суми незалежних випадкових величин називають **композицією**.

Якщо X та Y – це неперервні випадкові величини, то щільність розподілу $g(z)$ суми $Z = X + Y$ може бути знайдено за допомогою такої рівності:

$$g(z) = \int_{-\infty}^{\infty} f_1(x) f_2(z-x)dx \text{ або } g(z) = \int_{-\infty}^{\infty} f_1(z-y) f_2(y)dy,$$

де f_1, f_2 – щільності розподілу аргументів.

Якщо можливі значення аргументів невід'ємні, то $g(z)$ знаходять за такою формулою:

$$g(z) = \int_0^z f_1(x) f_2(z-x)dx \text{ або } g(z) = \int_0^z f_1(z-y) f_2(y)dy.$$

Якщо обидві диференціальні функції $f_1(x)$ і $f_2(y)$ задано на обмежених інтервалах, то для обчислення диференціальної функції $Z = X + Y$ доцільно спочатку знайти інтегральну функцію $\zeta(z)$, а потім продиференціювати її за Z:

$$g(z) = \zeta'(z),$$

де $\zeta(z) = \iint f(x;y)dxdy = \int_{-\infty}^{\infty} dy \int_{-\infty}^{z-x} f(x; z-x)dx = \int_{-\infty}^{\infty} dx \int_{-\infty}^{z-y} f(z-y; y)dy,$

а $f(x;y)$ – диференціальна функція системи випадкових величин $(X;Y)$.

Якщо X і Y незалежні випадкові величини, що мають диференціальні функції $f_1(x)$ і $f_2(y)$, то ймовірність потрапляння випадкової точки в область D дорівнює:

$$P((X;Y) \subset D) = \iint_D f_1(x) f_2(y)dxdy.$$

8.3. Приклади розв'язування задач

Приклад 8.1. Задано закон розподілу дискретної випадкової величини X (табл. 8.3):

Таблиця 8.3

Закон розподілу дискретної випадкової величини X

X	1	3	4	6
$P(X)$	0,1	0,3	0,2	0,4

Побудуйте закон розподілу ймовірностей для випадкової величини $Y = 4X^2 + 3$.

Розв'язання. Оскільки $Y = 4X^2 + 3$, то закон розподілу дискретної випадкової величини Y має такий вигляд (табл. 8.4):

Таблиця 8.4

Закон розподілу дискретної випадкової величини Y

$Y=4X^2+3$	7	39	67	147
$P(Y)$	0,1	0,3	0,2	0,4

Приклад 8.2. Напишіть закон розподілу функції $Z=X+Y$, якщо

X	1	2	3
p	0,2	0,5	0,3

та

Y	4	5	6
p	0,1	0,6	0,3

.

Розв'язання. Можливі значення випадкової величини Z дорівнюють сумі кожного можливого значення випадкової величини X з усіма значеннями Y, тобто:

$$z_1=1+4=5,\quad z_2=1+5=6,\quad z_3=1+6=7,$$
$$z_4=2+4=6,\quad z_5=2+5=7,\quad z_6=2+6=8,$$
$$z_7=3+4=7,\quad z_8=3+5=8,\quad z_9=3+6=9.$$

Знайдіть імовірності цих можливих значень. Для того щоб $Z=5$ достатньо, щоб випадкова величина X набула значення $x_1=1$, а випадкова величина Y – значення $y_1=4$. Імовірності цих можливих значень, відповідно, дорівнюють 0,2 та 0,1.

Величини X та Y – незалежні, тому, за теоремою множення, імовірність їх сумісної появи дорівнює $0,2\cdot 0,1=0,02$. Аналогічно знаходять:

$$P(Z=1+5)=0,2\cdot 0,6=0,12;$$
$$P(Z=1+6)=0,2\cdot 0,3=0,06;$$
$$P(Z=2+4)=0,5\cdot 0,1=0,05;$$
$$P(Z=2+5)=0,5\cdot 0,6=0,30;$$
$$P(Z=2+6)=0,5\cdot 0,3=0,15;$$
$$P(Z=3+4)=0,3\cdot 0,1=0,03;$$
$$P(Z=3+5)=0,3\cdot 0,6=0,18;$$
$$P(Z=3+6)=0,3\cdot 0,3=0,09.$$

Таким чином, шуканий розподіл має такий вигляд (табл. 8.5):

Таблиця 8.5

Закон розподілу функції Z

Z	5	6	7	6	7	8	7	8	9
p	0,02	0,12	0,06	0,05	0,30	0,15	0,03	0,18	0,09

Додавання ймовірності несумісних подій:

$$Z = z_2,\ Z = z_4 \qquad 0{,}12 + 0{,}05 = 0{,}17;$$

$$Z = z_3,\ Z = z_5,\ Z = z_7 \qquad 0{,}06 + 0{,}3 + 0{,}03 = 0{,}39;$$

$$Z = z_6,\ Z = z_8 \qquad 0{,}15 + 0{,}18 = 0{,}33.$$

Остаточно, закон розподілу функції $Z = X + Y$ набуває такого вигляду (табл. 8.6):

Таблиця 8.6

Закон розподілу функції Z

Z	5	6	7	8	9
p	0,02	0,17	0,39	0,33	0,09

Перевірка: $0{,}02 + 0{,}17 + 0{,}39 + 0{,}33 + 0{,}09 = 1.$

Приклад 8.3. Випадкову величину X розподілено за нормальним законом із математичним сподіванням, що дорівнює нулеві. Напишіть закон розподілу функції $Y = X^5$.

Розв'язання. Диференціальна функція нормального закону розподілу неперервної величини X має такий вигляд:

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

Функція $Y = X^5$ є диференційованою, $Y' = 5X^4 > 0$, тому вона зростає для всіх $x \in (-\infty; +\infty)$.

Оскільки $Y = X^5$, то $X = \sqrt[5]{Y}$, тоді $\psi(y) = \sqrt[5]{y} = y^{\frac{1}{5}}$.

Отже, буде

$$g(y) = \frac{1}{\sigma\sqrt{2\pi}} \cdot e^{\frac{y^{\frac{2}{5}}}{2\sigma^2}} \cdot \left| \left(y^{\frac{1}{5}} \right)' \right| = \frac{e^{\frac{y^{\frac{2}{5}}}{2\sigma^2}}}{5\sigma\sqrt{2\pi} \cdot y^{\frac{4}{5}}}.$$

Приклад 8.4. Дискретну випадкову величину задано законом розподілу (табл. 8.7):

Таблиця 8.7

Закон розподілу дискретної випадкової величини X

X	0	1	4	9
$P(X)$	0,729	0,243	0,027	0,001

Знайдіть математичне сподівання випадкової величини $Y = X^2 + 1$.

Розв'язання. Закон розподілу випадкової величини $Y = X^2 + 1$ має такий вигляд (табл. 8.8):

Таблиця 8.8

Закон розподілу дискретної випадкової величини Y

$Y = X^2 + 1$	$0^2 + 1 = 1$	$1^2 + 1 = 2$	$4^2 + 1 = 17$	$9^2 + 1 = 82$
$P(Y)$	0,729	0,243	0,027	0,001

Математичне сподівання випадкової величини $Y = X^2 + 1$ дорівнює:

$$M(Y) = 1 \cdot 0{,}729 + 2 \cdot 0{,}243 + 17 \cdot 0{,}027 + 82 \cdot 0{,}001 = 1{,}756.$$

Для обчислення дисперсії побудуйте закон розподілу випадкової величини Y^2 (табл. 8.9):

Таблиця 8.9

Закон розподілу дискретної випадкової величини Y^2

Y^2	1	4	289	6 724
$P(Y)$	0,729	0,243	0,027	0,001

Тоді

$$M\left(Y^2\right)=1\cdot 0,729+4\cdot 0,243+289\cdot 0,027+6\,724\cdot 0,001=$$
$$=0,729+0,972+7,803+6,724=16,228.$$

Таким чином,

$$D(Y)=M\left(Y^2\right)-M\left(Y\right)=16,228-(1,756)^2=13,144.$$

Середнє квадратичне відхилення дорівнює:

$$\sigma(Y)=\sqrt{D(Y)}=\sqrt{13,1445}\approx 3,625.$$

Приклад 8.5. Неперервну випадкову величину X задано щільністю розподілу ймовірностей:

$$f(x)=\begin{cases}0, & \text{якщо} \quad x<0,\\ \sin x, & \text{якщо} \quad 0\le x\le \dfrac{\pi}{2},\\ 0, & \text{якщо} \quad x>\dfrac{\pi}{2}.\end{cases}$$

Знайдіть числові характеристики випадкової величини $Y=3X^2$.

Розв'язання. За умовою $f\left(x\right)=\sin x$, тоді

$$M\left(Y\right)=M\left(3X^2\right)=\int_0^{\frac{\pi}{2}}3x^2\sin x dx;$$

$$\int_0^{\frac{\pi}{2}} 3x^2 \sin x dx = 3\left(-x^2 \cos x\Big|_0^{\frac{\pi}{2}} + 2\int_0^{\frac{\pi}{2}} x\cos x dx \right) = 3\left(x\sin x\Big|_0^{\frac{\pi}{2}} - \int_0^{\frac{\pi}{2}} \sin x dx \right) =$$

$$= 3\left(\pi + 2\cos x\Big|_0^{\frac{\pi}{2}} \right) = 3(\pi - 2) = 3\pi - 6.$$

Таким чином, $M(Y) = 3\pi - 6$.

Обчисліть дисперсію функції $Y = 3X^2$:

$$D(Y) = M\left(Y^2\right) - M^2(Y) = \int_0^{\frac{\pi}{2}} 9x^4 \sin x dx - M^2(Y),$$

$$M\left(Y^2\right) = \int_0^{\frac{\pi}{2}} 9x^4 \sin dx = 9\left(-x^4 \cos x\Big|_0^{\frac{\pi}{2}} + 4\int_0^{\frac{\pi}{2}} x^3 \cos x dx \right) =$$

$$= 36\left(x^3 \sin x\Big|_0^{\frac{\pi}{2}} - 3\int_0^{\frac{\pi}{2}} x^2 \sin x dx \right) = 36\left(\frac{\pi^3}{8} - 3(\pi - 2) \right) = \frac{36}{8}\pi^3 - 108\pi + 216.$$

$$D(Y) = \frac{36}{8}\pi^3 - 108\pi + 216 - (3\pi - 6)^2 = \frac{36}{8}\pi^3 - 108\pi + 216 - 9\pi^2 + 36\pi -$$

$$-36 = \frac{36}{8}\pi^3 - 9\pi^2 - 72\pi + 180 \approx 4,4997.$$

$$\sigma(Y) = \sqrt{D(Y)} = \sqrt{4,4997} \approx 2,1213.$$

Приклад 8.6. Випадкові величини X та Y мають такі закони розподілу:

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

та

Y	0	1	2
P_{II}	0,5	0,3	0,2

.

Складіть закон розподілу для: 1) $X + Y$; 2) $X \cdot Y$.

Розв'язання.

Випадкова величина $X + Y$ набуває значення 2 з імовірністю 0,15;

3 – з імовірністю 0,19;

4 – якщо $X+Y=2+2$ або $X+Y=4+0$ з імовірністю 0,06 + 0,20 = = 0,26 (для несумісних подій імовірності додають);

5 – з імовірністю 0,12;

6 – якщо $X+Y=4+2$ або $X+Y=6+0$ з імовірністю 0,08 + 0,05 = = 0,13;

7 – з імовірністю 0.03;

8 – якщо $X+Y=6+2$ або $X+Y=8+0$ з імовірністю 0,10 + 0,02 = = 0,12;

9 – з імовірністю 0,06;

10 – з імовірністю 0,04.

Запишіть ці розрахунки у вигляді таблиці (табл. 8.10):

Таблиця 8.10

Розрахункова таблиця

X	Y	$X+Y$	$P_I \cdot P_{II}$
2	0	2 + 0 = 2	0,3 · 0,5 = 0,15
2	1	2 + 1 = 3	0,3 · 0,3 = 0,09
2	2	2 + 2 = 4	0,3 · 0,2 = 0,06
4	0	4 + 0 = 4	0,4 · 0,5 = 0,20
4	1	4 + 1 = 6	0,4 · 0,3 = 0,12
4	2	4 + 2 = 2	0,4 · 0,2 = 0,08
6	0	6 + 0 = 6	0,1 · 0,5 = 0,05
6	1	6 + 1 = 7	0,1 · 0,3 = 0,03
6	2	6 + 2 = 8	0,1 · 0,2 = 0,02
8	0	8 + 0 = 8	0,2 · 0,5 = 0,10
8	1	8 + 1 = 9	0,2 · 0,3 = 0,06
8	2	8 + 2 = 10	0,2 · 0,2 = 0,04

Перевірка:

$$0{,}15+0{,}09+0{,}26+0{,}12+0{,}13+0{,}03+0{,}12+0{,}06+0{,}04=1.$$

Отже, остаточно буде такий ряд розподілу для випадкової величини $X+Y$ (табл. 8.11):

Таблиця 8.11

Закон розподілу випадкової величини $X+Y$

$X+Y$	2	3	4	5	6	7	8	9	10
p	0,15	0,09	0,26	0,12	0,13	0,03	0,12	0,06	0,04

Складіть закон розподілу випадкової величини $X \cdot Y$.

Випадкова величина $X \cdot Y$ набуває значення 0, якщо $X \cdot Y = 2 \cdot 0$, або $X \cdot Y = 4 \cdot 0$, або $X \cdot Y = 6 \cdot 0$, або $X \cdot Y = 8 \cdot 0$ з імовірністю 0,15 + + 0,20 + 0,05 + 010 = 0,50 (для несумісних подій імовірності додають);

2 – з імовірністю 0,09;

4 – якщо $X \cdot Y = 2 \cdot 2$ або $X \cdot Y = 4 \cdot 1$ з імовірністю 0,06 + 0,12 = 0,18;

6 – з імовірністю 0,03;

8 – якщо $X \cdot Y = 4 \cdot 2$ або $X \cdot Y = 8 \cdot 1$ з імовірністю 0,08 + 0,06 = 0,14;

12 – з імовірністю 0,02;

16 – з імовірністю 0,04.

Запишіть ці розрахунки у вигляді таблиці (табл. 8.12):

Таблиця 8.12

Розрахункова таблиця

X	Y	$X \cdot Y$	$P_I \cdot P_{II}$
2	0	2 · 0 = 0	0,3 · 0,5 = 0,15
2	1	2 · 1 = 2	0,3 · 0,3 = 0,09
2	2	2 · 2 = 4	0,3 · 0,2 = 0,06
4	0	4 · 0 = 0	0,4 · 0,5 = 0,20
4	1	4 · 1 = 4	0,4 · 0,3 = 0,12
4	2	4 · 2 = 8	0,4 · 0,2 = 0,08
6	0	6 · 0 = 0	0,1 · 0,5 = 0,05
6	1	6 · 1 = 6	0,1 · 0,3 = 0,03
6	2	6 · 2 = 23	0,1 · 0,2 = 0,02
8	0	8 · 0 = 0	0,2 · 0,5 = 0,10
8	1	8 · 1 = 8	0,2 · 0,3 = 0,06
8	2	8 · 2 = 16	0,2 · 0,2 = 0,04

Перевірка: $0,50+0,09+0,18+0,03+0,14+0,02+0,04=1.$

Отже, буде такий ряд розподілу для випадкової величини $X\cdot Y$ (табл. 8.13):

Таблиця 8.13

Закон розподілу випадкової величини $X\cdot Y$

$X\cdot Y$	0	2	4	6	8	12	16
p	0,50	0,09	0,18	0,03	0,14	0,02	0,04

Приклад 8.7. Незалежні випадкові величини X та Y задано за допомогою щільностей розподілу:

$$f(x)=\frac{1}{4}e^{-\frac{x}{4}}\ (0\le x<\infty);\ \ f(y)=\frac{1}{3}e^{-\frac{y}{3}}\ (0\le y<\infty).$$

Знайдіть композицію цих законів.

Розв'язання. Знайдіть щільність розподілу випадкової величини $Z=X+Y$.

Можливі значення аргументів невід'ємні, тому можна застосувати таку формулу:

$$g(z)=\int_0^z f_1(x)\,f_2(z-x)dx.$$

Таким чином,

$$g(z)=\int_0^z f_1(x)\,f_2(z-x)dx=\int_0^z\left(\frac{1}{4}e^{-\frac{x}{4}}\right)\left(\frac{1}{3}e^{-\frac{z-x}{3}}\right)dx=\frac{1}{12}e^{-\frac{z}{3}}\int_0^z e^{-\frac{x}{12}}dx=$$

$$=e^{-\frac{z}{3}}\left(1-e^{-\frac{z}{12}}\right).$$

8.4. Вправи для самостійної роботи

8.1. Дискретну випадкову величину X задано законом розподілу (табл. 8.14):

Таблиця 8.14

Закон розподілу дискретної випадкової величини X

X	−1	1	2	3
$P(X)$	0,4	0,3	0,2	0,1

Знайдіть закон розподілу випадкової величини Y:

1) $Y = 3X$; 2) $Y = 3X - 1$; 3) $Y = X^2$; 4) $Y = |X|$; 5) $Y = 2^X$.

8.2. Задано розподіли випадкових величин X та Y:

X	3	4	6
p	0,2	0,3	0,5

та

Y	1	8
p	0,7	0,3

.

Складіть розподіл величини $Z = X + Y$ і знайдіть математичне сподівання та дисперсію.

8.3. Випадкову величину X розподілено рівномірно на інтервалі $(0; 3)$. Знайдіть математичне сподівання та дисперсію випадкової величини $Y = 3 + 2X^2$.

8.4. Випадкову величину X розподілено нормально з параметрами $a = 2,\ \sigma = 3$ на інтервалі $(0; 3)$. Знайдіть щільність розподілу випадкової величини $Y = 2X^2$.

8.5. Випадкову величину X розподілено рівномірно на інтервалі $(1; 4)$. Знайдіть $M\left(3X^2 - 2X + 4\right)$.

8.6. Диференціальна функція випадкової величини X:

$$f(x) = \begin{cases} 0, & \text{якщо} \quad x < 0, \\ \cos x, & \text{якщо} \quad 0 \le x \le \dfrac{\pi}{2}, \\ 0, & \text{якщо} \quad x > \dfrac{\pi}{2}. \end{cases}$$

Знайдіть математичне сподівання і дисперсію випадкової величини $Y = X^2$: 1) не знаходячи диференціальної функції Y; 2) використовуючи диференціальну функцію $g(y)$.

8.7. Неперервну випадкову величину X розподілено за показниковим законом:

$$f(x) = \begin{cases} 0, & \text{якщо } x < 0, \\ \lambda e^{-\lambda x}, & \text{якщо } x \geq 0. \end{cases}$$

Знайдіть математичне сподівання і дисперсію випадкової величини $Y = e^{-X}$.

8.8. Випадкова величина X має рівномірний розподіл в інтервалі $(1; 2)$. Знайдіть математичне сподівання і дисперсію випадкової величини $Y = \frac{1}{X}$.

8.9. Неперервні випадкові величини X та Y мають нормальний розподіл із такими математичними сподіваннями та дисперсіями: $a_1 = 4$; $a_2 = 8$; $D_1 = 0,5$; $D_2 = 0,8$ та $a_1 = 5$; $a_2 = 6$; $D_1 = 0,6$; $D_2 = 1,2$, відповідно. Знайдіть математичне сподівання і дисперсію випадкової величини $Z = X + Y$.

8.10. Незалежні випадкові величини X та Y задано за допомогою щільностей розподілу:

$$f_1 = \frac{1}{3} e^{-\frac{x}{3}} \ (0 \leq x < \infty); \ \ f_2 = \frac{1}{5} e^{-\frac{y}{5}} \ (0 \leq y < \infty).$$

Знайдіть композицію цих законів, тобто щільність розподілу випадкової величини $Z = X + Y$.

8.5. Тестові завдання

8.1. Якщо аргумент X – це дискретна випадкова величина і різним можливим значенням аргументу X відповідають однакові можливі значення функції Y, то ймовірності значення Y, що повторюються

А	Б	В	Г
додають	віднімають	множать	ділять

8.2. Якщо аргумент X – це неперервна випадкова величина, задана щільністю розподілу $f(x)$, то диференціальну функцію $g(y)$ випадкової величини Y знаходять із рівності

А	Б	В
$g(y)=f(\psi(y))\cdot\lvert\psi'(y)\rvert$	$g(y)=f(\psi(y))\cdot\lvert\psi(y)\rvert$	$g(y)=\lvert\psi'(y)\rvert$

8.3. Нехай аргумент X – це дискретна випадкова величина, а задана функція $Y=f(x)$ випадкового аргументу X, тоді математичне сподівання функції Y дорівнює

А	Б	В	Г
$M(f(x))=\sum\limits_{i=1}^{n} f(x_i)p_i$	$M(Y)=\sum\limits_{i=1}^{n} f(x_i)$	$M(f(x))=\sum\limits_{i=1}^{n} p_i$	інше

8.4. Нехай аргумент X – це неперервна випадкова величина, причому $x\in(a;b)$, тоді математичне сподівання функції Y дорівнює

А	Б	В
$M(Y)=\int\limits_{a}^{b} f(x)dx$	$M(Y)=\int\limits_{a}^{b}\varphi(x)f(x)dx$	$M(Y)=\int\limits_{a}^{b}\varphi(x)f(y)dx$

8.5. Якщо використовувати диференціальну функцію $g(y)$ випадкової величини Y, то математичне сподівання знаходять за формулою

А	Б	В
$M(Y)=\int\limits_{d}^{c} yg(y)dy$	$M(Y)=\int\limits_{c}^{d} yg(y)dy$	$M(Y)=\int\limits_{c}^{d} g(y)dy$

8.6. Диференціальну функцію суми незалежних випадкових величин називають

А	Б	В	Г
композицією	комбінацією	сполученням	інакше

8.6. Запитання для самоперевірки

8.1. Дайте означення функції розподілу випадкової величини.

8.2. Сформулюйте закон розподілу функції дискретного випадкового аргументу.

8.3. Назвіть основні числові характеристики функції дискретного випадкового аргументу.

8.4. Який вигляд має закон розподілу функції неперервного випадкового аргументу?

8.5. Назвіть основні числові характеристики функції неперервного випадкового аргументу.

8.6. Дайте означення функції двох випадкових аргументів.

8.7. Що називають композицією двох функцій?

8.7. Висновки за темою

Після вивчення теми студент має розуміти особливості побудови розподілів функції одного випадкового аргументу (дискретного чи неперервного) і двох випадкових аргументів, а також визначати основні числові характеристики цих функцій та робити висновки.

9. Елементи теорії випадкових процесів і теорії масового обслуговування

9.1. Мета та компетентності

Метою вивчення теми є ознайомлення з основами теорії випадкових процесів і теорії масового обслуговування та формування компетентностей щодо розрахунків параметрів систем масового обслуговування (СМО).

Компетентності, що формують під час вивчення теми:

знання основних понять теорії випадкових процесів і СМО;

знання класифікації СМО;
розуміння характеристик СМО;
навички в обчисленні основних параметрів СМО.

9.2. Основні теоретичні відомості

Теорія випадкових процесів вивчає закономірності випадкових явищ у динаміці їхнього розвитку.

Випадковим процесом $X(t)$ називають процес, значення якого є випадковою величиною за будь-якого значення аргументу t.

Реалізацією випадкового процесу є функція $x(t)$, на яку перетворюється випадковий процес $X(t)$ унаслідок випробування, тобто його траєкторія. Випадковий процес можна розглядати як сукупність усіх його реалізацій.

Значення випадкового процесу $X(t)$ у заданий фіксований момент часу називають **перерізом випадкового процесу (одновимірним законом розподілу процесу** $X(t)$).

Сукупність можливих перерізів випадкового процесу утворює **простір перерізів**. Ця множина може бути **дискретною** або **неперервною**.

Випадковий процес називають **процесом із дискретним часом**, якщо система, у якій він здійснюється, може змінювати свій стан тільки в моменти часу $t_1, t_2, ..., t_k, ...$, кількість яких є обмеженою або зліченною. Випадковий процес називають **процесом із неперервним часом**, якщо перехід системи з одного стану до іншого може бути здійснений у будь-який можливий момент часу *t* за певний спостережуваний період. У цьому разі множина можливих станів системи є незліченною.

Випадковий процес називають **процесом із неперервним станом**, якщо його переріз у будь-який можливий момент часу *t* утворює неперервну випадкову величину, множина значень якої є незліченною.

Випадковий процес називають **процесом із дискретним станом**, якщо в будь-який можливий момент часу *t* множина його станів є скінченною або зліченною.

Якщо значення змінних, притаманних одному перерізу випадкового процесу, змінюються на такі, які характерні іншому перерізу, це означає, що має місце **перехід** випадкового процесу з одного стану в інший.

Закон розподілу вектора $(X(t_1), X(t_2), ..., X(t_n))$, компонентами якого є перерізи випадкового процесу $X(t)$, що відповідають значенням $t_1, t_2, ..., t_n$ аргументу t, називають ***n*-вимірним законом розподілу випадкового процесу** $X(t)$.

Його можна задати ***n*-вимірною функцією розподілу**:

$$F_n(x_1, x_2, ..., x_n; t_1, t_2, ..., t_n) = P(X(t_1) < x_1, ..., X(t_n) < x_n),$$

або ***n*-вимірною щільністю** для процесів із неперервними перерізами:

$$f_n(x_1, x_2, ..., x_n; t_1, t_2, ..., t_n) = \frac{\partial^n F(x_1, x_2, ..., x_n; t_1, t_2, ..., t_n)}{\partial x_1 \partial x_2 ... \partial x_n}.$$

Основними числовими характеристиками випадкового процесу є його математичне сподівання $M(X(t)) = m_x(t)$ і дисперсія $D(X(t))$, які є функціями від t.

Математичним сподіванням і **дисперсією випадкового процесу** $X(t)$ називають, відповідно, такі невипадкові функції $m_x(t)$ та $D_x(t)$, які для кожного t дорівнюють, відповідно, математичному сподіванню і дисперсії перерізу процесу в цей момент часу. *Математичне сподівання* – це середня траєкторія для всіх реалізацій процесу, а *дисперсія* характеризує можливе розсіювання реалізацій випадкового процесу навколо його середньої траєкторії.

Якщо $X(t_1)$ та $X(t_2)$ – два будь-які перерізи випадкового процесу $X(t)$, то **кореляційною функцією** $K_x(t_1; t_2)$, яка характеризує розкид перерізів відносно математичного сподівання $m_x(t)$, є така функція:

$$K_x(t_1; t_2) = M[[X(t_1) - m_x(t_1)] \cdot [X(t_2) - m_x(t_2)]].$$

Властивості кореляційної функції випадкового процесу:

симетрія:

$$K_x(t_1; t_2) = K_x(t_2; t_1),$$

де t_1 і t_2 – певні моменти часу;

$$K_x(t; t) = D_x(t);$$

якщо до випадкового процесу $X(t)$ додати невипадковий процес $\varphi(t)$, то кореляційна функція не зміниться:

$$K_{X(t)+\varphi(t)}(t_1; t_2) = K_{X(t)}(t_1; t_2);$$

$$\left|K_x(t_1; t_2)\right| \le \sqrt{K_x(t_1; t_1) \cdot K_x(t_2; t_2)}.$$

Тісноту лінійної залежності двох перерізів випадкового процесу $X(t)$ вимірюють **нормованою кореляційною функцією**:

$$r_x(t_1; t_2) = \frac{K_x(t_1; t_2)}{\sigma_x(t_1) \cdot \sigma_x(t_2)}.$$

Властивості нормованої кореляційної функції:

$$r_x(t; t) = 1,$$

якщо

$$t_1 = t_2 = t;\ r_x(t_1; t_2) = r_x(t_2; t_1);\ \left|r_x(t_1; t_2)\right| \le 1.$$

Певні ситуації, такі як: черга покупців у касах магазинів; колона автомобілів, рух яких зупинено світлофором; ряд верстатів, що вийшли з ладу і чекають на ремонт, та ін., – об'єднує перебування у стані очікування. Очікування є наслідком реалізації випадкового процесу щодо обслуговування заявки.

Відповідні системи, призначені для багаторазового використання в ході такого процесу, називають **системами масового обслуговування** (**СМО**).

Залежно від характеру формування черги заявок СМО, розрізняють:

системи з відмовами, у яких за умови зайнятості всіх каналів обслуговування заявка не стає в чергу і систему не обслуговують;

системи з необмеженими очікуваннями, у яких заявка стає в чергу, якщо в момент її надходження всі канали були зайняті.

Також існують і *системи змішаного типу* з очікуванням та обмеженою довжиною черги: заявка дістає відмову, якщо надходить у мить, коли всі місця в черзі зайняті. Заявку, що потрапила в чергу, обслуговують обов'язково.

За кількістю каналів обслуговування СМО розподіляють на *одноканальні* й *багатоканальні*.

Залежно від розташування джерела заявок системи можуть бути *розімкненими* (джерело заявок перебуває поза системою) і *замкнутими* (джерело перебуває безпосередньо в самій системі).

Основними елементами СМО є **джерела заявок**, їхній **вхідний потік**, **канали обслуговування** і **вихідний потік**.

Найбільш поширеним вхідним потоком є **найпростіший потік** заявок, що характеризується стаціонарністю, ординарністю і відсутністю післядії.

Стаціонарність визначає, що ймовірність надходження певної кількості вимог (заявок) протягом певного проміжку часу залежить тільки від довжини цього проміжку.

Ординарність потоку означає неможливість одночасної появи двох або більше заявок.

Відсутність післядії характеризується тим, що надходження заявки не залежить від того, коли та скільки заявок надійшло до цього моменту.

Якщо потік подій найпростіший, то ймовірність того, що кількість заявок, які надійшли на обслуговування за інтервал часу t, дорівнює k і її визначають за законом Пуассона:

$$P_k(t) = \frac{(\lambda t)^k e^{-\lambda t}}{k!},$$

де λ – **інтенсивність потоку заявок**, тобто середня кількість заявок за одиницю часу: $\lambda = \dfrac{1}{\tau}$,

де τ – середнє значення інтервалу часу, який минає між двома послідовними надходженнями заявок.

Для такого потоку заявок час між надходженнями двох послідовних заявок є неперервною випадковою величиною, розподіленою за показниковим законом зі щільністю ймовірностей:

$$f(t) = \lambda e^{-\lambda t}.$$

Випадковий час очікування в черзі до початку обслуговування теж уважають розподіленим за показниковим законом зі щільністю ймовірностей:

$$f(t) = \nu e^{-\nu t},$$

де ν – **інтенсивність руху черги**, тобто середня кількість заявок, що надходять на обслуговування за одиницю часу: $\nu = (t_{оч.})^{-1}$,
де $t_{оч.}$ – середній час очікування.

Вихідний потік заявок пов'язаний із потоком обслуговування в каналі, де тривалість обслуговування $t_{обсл.}$, є випадковою величиною, розподіленою за показниковим законом зі щільністю ймовірностей:

$$f(t_{обсл.}) = \mu e^{-\mu t},$$

де μ – **інтенсивність потоку обслуговування**, тобто середня кількість заявок, що обслуговують за одиницю часу: $\mu = (t_{обсл.})^{-1}$;

$t_{обсл.}$ – середній час обслуговування.

Важливою характеристикою СМО, яка пов'язує між собою λ і μ, є **інтенсивність навантаження** $\rho = \dfrac{\lambda}{\mu}$.

СМО з відмовами. Заявка, що надійшла в систему з відмовами тоді, коли всі канали виявилися зайнятими, отримує відмову, унаслідок чого система залишається без обслуговування. Показником якості обслуговування є ймовірність відмови. Передбачають, що всі канали доступні однаковою мірою для всіх заявок, вхідний потік є найпростішим, час обслуговування однієї заявки ($t_{обсл.}$) розподілено за показниковим законом.

Характеристики, що визначають сталий режим роботи СМО з відмовами:

імовірність простою каналів обслуговування в очікуванні заявок ($k = 0$):

$$P_0 = \frac{1}{\sum_{k=0}^{n} \frac{\rho^k}{k!}}; \quad (9.1)$$

імовірність відмови в обслуговуванні, коли заявка, що надійшла на обслуговування, застане всі канали зайнятими ($k = n$):

$$P_{відм.} = P_n = \frac{P_0 \rho^n}{n!}; \qquad (9.2)$$

імовірність обслуговування:

$$P_{обсл.} = 1 - P_{відм.}; \qquad (9.3)$$

середня кількість зайнятих обслуговуванням каналів:

$$n_з = \rho \cdot P_{обсл.}; \qquad (9.4)$$

частка каналів, зайнятих обслуговуванням заявок:

$$k_з = \frac{n_з}{n}; \qquad (9.5)$$

абсолютна пропускна спроможність СМО:

$$A = \lambda \cdot P_{обсл.}. \qquad (9.6)$$

СМО з необмеженим очікуванням. Заявка, що надійшла в систему з необмеженим очікуванням і застала всі канали системи зайнятими, стає в чергу й очікує звільнення одного з каналів. Основною характеристикою якості обслуговування для такої системи є *час очікування*, протягом якого заявка перебуває в черзі. Для таких систем характерна відсутність відмови в обслуговуванні, отже, $P_{відм.} = 0$ і $P_{обсл.} = 1$.

Для систем з очікуванням існує **дисципліна черги**:

обслуговування за принципом "першим прийшов – першим обслуговують";

випадкове неорганізоване обслуговування за принципом "останнім прийшов – першим обслуговують";

обслуговування із пріоритетами за принципом "генералів і полковників обслуговують поза чергою".

Характеристики, що визначають сталий режим роботи СМО з необмеженим очікуванням:

імовірність простою каналів, коли немає заявок ($k = 0$):

$$P_0 = \frac{1}{\sum\limits_{k=0}^{n} \frac{\rho^k}{k!}} + \frac{\rho^{n+1}}{n!\left(n-\rho\right)}, \tag{9.7}$$

де $\frac{\rho}{n} < 1$;

імовірність одночасного обслуговування k заявок:

$$P_k = \frac{P_0 \rho^k}{k!}, \quad 1 \le k \le n; \tag{9.8}$$

імовірність зайнятості в обслуговуванні одночасно всіх каналів:

$$P_n = \frac{P_0 \rho^n}{n!}; \tag{9.9}$$

імовірність того, що заявка буде очікувати на обслуговування:

$$P_{оч.} = \frac{\rho^{n+1}}{n!\left(n-\rho\right)} P_0; \tag{9.10}$$

середня кількість заявок у черзі:

$$L_{оч.} = \frac{\rho^{n+1}}{\left(n-1\right)!\left(n-\rho\right)^2} P_0; \tag{9.11}$$

середній час очікування заявки в черзі:

$$t_{оч.} = \frac{L_{оч.}}{\lambda}; \tag{9.12}$$

середній час перебування заявки в СМО:

$$t_{СМО} = t_{оч.} + t_{обсл.}; \tag{9.13}$$

середня кількість зайнятих обслуговуванням каналів:

$$n_з = \rho; \tag{9.14}$$

середня кількість вільних каналів:

$$n_в = n - n_з; \tag{9.15}$$

коефіцієнт зайнятості каналів обслуговування:

$$k_з = \frac{n_з}{n}; \tag{9.16}$$

середня кількість заявок у СМО:

$$z = L_{оч.} + n_з. \tag{9.17}$$

СМО з очікуванням і з обмеженою довжиною черги. Якщо заявка, що надійшла в систему з очікуванням і з обмеженою довжиною черги, знайшла всі канали цієї системи зайнятими, а також зайнятою виявилася обмежена черга заявок, у такому разі система залишається без обслуговування.

Існування обмеження на довжину черги може бути викликано такими причинами: обмеженням верхньої межі часу перебування заявки в черзі; обмеженням верхньої межі довжини черги; обмеженням загального часу перебування заявки в системі. Основною характеристикою якості такої системи є *ймовірність* того, що для заявки буде мати місце *відмова в обслуговуванні*.

Характеристики, що визначають сталий режим роботи СМО з очікуванням і з обмеженою довжиною черги:

імовірність простою каналів обслуговування, коли немає заявок ($k = 0$):

$$P_0 = \frac{1}{\sum_{k=0}^{n} \frac{\rho^k}{k!} + \frac{\rho^{n+1}}{n!(n-\rho)}\left[1 - \left(\frac{\rho}{n}\right)^m\right]}; \tag{9.18}$$

імовірність відмови в обслуговуванні:

$$P_{відм.} = \frac{P_0 \cdot \rho^{n+m}}{n! n^m}; \tag{9.19}$$

імовірність обслуговування:

$$P_{обсл.} = 1 - P_{відм.}; \tag{9.20}$$

абсолютна пропускна спроможність:

$$A = \lambda \cdot P_{обсл.}; \tag{9.21}$$

середня кількість зайнятих каналів:

$$n_з = \frac{A}{\mu}; \tag{9.22}$$

середня кількість заявок у черзі:

$$L_{оч.} = \frac{\rho^{n+1}}{n \cdot n!} \cdot \frac{1 - (\rho / n)^m (m + 1 - m\rho / n)}{(1 - \rho / n)^2} \cdot P_0; \tag{9.23}$$

середній час очікування обслуговування:

$$t_{оч.} = \frac{L_{оч.}}{\lambda}; \tag{9.24}$$

середня кількість заявок, які одночасно перебувають у системі:

$$z = L_{оч.} + n_з; \tag{9.25}$$

середній час перебування заявки в системі:

$$t_{СМО} = \frac{z}{\lambda}. \tag{9.26}$$

Серед випадкових процесів особливе місце посідають марковські процеси, які становлять основу теорії масового обслуговування.

Марковський процес – це випадковий процес із дискретними станами, що ґрунтується на **принципі Маркова**. Згідно із цим принципом, імовірнісні характеристики процесу в майбутні моменти часу $t+\tau$ залежать лише від стану в заданий момент часу t і не залежать від того, які значення ці характеристики мали в попередні моменти часу.

Нехай система перебуває в одному зі станів $S_1, S_2, ..., S_k, ...$ і в певні фіксовані моменти часу $t_1, t_2, ..., t_n, ...$ під впливом випадкових факторів вона може переходити з одного стану в інший. У будь-який момент часу t_n імовірність для системи стати в наперед заданому стані S_j цілком визначено тим станом, у якому вона перебувала безпосередньо перед переходом і не залежить від усіх попередніх станів, у яких система перебувала до моменту часу t_{n-1}. Поведінку цієї системи описують простим **ланцюгом Маркова**. Отже, ланцюг Маркова є випадковим процесом із дискретним часом.

Імовірність переходу зі стану S_i у стан S_j у момент часу t позначають через $p_{ij}(t)$.

Повну картину всіх можливих переходів систем з одного стану в інший за умови, що кількість усіх станів дорівнює N, безпосередньо описують такою матрицею ймовірностей переходу:

$$\pi = \begin{pmatrix} p_{11}(t) & p_{12}(t) & ... & p_{1N}(t) \\ p_{21}(t) & p_{22}(t) & ... & p_{2N}(t) \\ ... & ... & ... & ... \\ p_{N1}(t) & p_{N2}(t) & ... & p_{NN}(t) \end{pmatrix}, \qquad (9.27)$$

де $\sum_{j=1}^{N} p_{ij}(t)=1.$

Перебуваючи в будь-якому можливому стані S_i у фіксований момент часу t, система обов'язково перейде з певною ймовірністю в будь-який інший можливий стан S_j або залишиться в цьому самому стані. А ці події є несумісними й утворюють повну групу.

Якщо $p_{ij}(t)$ не залежить від часу, то ланцюг Маркова називають **однорідним**. Імовірність переходу системи зі стану S_i у стан S_j за n кроків обчислюють за такою формулою:

$$p_{ij}^{(n)} = \sum_{k=1}^{N} p_{ik}\, p_{kj}^{(n-1)}, \tag{9.28}$$

де $p_{ij}^{(n)}$ – елемент матриці π^n.

Матрицю π^n називають ***n*-кроковою матрицею переходу** з одного стану в інший.

Якщо в кожний можливий стан система може перейти через певний інтервал часу, то такий ланцюг Маркова, що моделює цей процес, називають **циклічним**.

Ланцюг Маркова називають **регулярним**, якщо існує таке ціле число *k* (*k* > 0), що перехід системи з будь-якого можливого стану в інший може здійснитися за *k* кроків.

Для регулярних ланцюгів Маркова доведено, що незалежно від початкового стану, у якому перебувала система, імовірність перебування її в певному можливому стані прямує до деякої сталої величини за умови збільшення кількості кроків *k:* за $k \to \infty$ $p_{ij}^{(n)} \to b_j = \mathrm{const}$.

Імовірності b_j (j = 1, 2, ..., N) називають **стаціонарними**.

Отже,

$$\lim_{n\to\infty} \pi^n = B = \begin{pmatrix} b_1 & b_2 & \ldots & b_N \\ b_1 & b_2 & \ldots & b_N \\ \ldots & \ldots & \ldots & \ldots \\ b_1 & b_2 & \ldots & b_N \end{pmatrix},$$

де $\sum_{j=1}^{N} b_j = 1$, *B* – матриця стаціонарних імовірностей.

Для визначення стаціонарних імовірностей регулярного ланцюга Маркова необхідно розв'язати систему рівнянь:

$$\begin{cases} \bar{b} = \pi \bar{b} \\ \sum_{j=1}^{N} b_j = 1. \end{cases} \tag{9.29}$$

Під час переходу системи з одного стану в інший за один крок її функціонування, а також у разі, коли система залишиться в тому самому стані, у якому вона перебувала, уводять **коефіцієнт вартості** r_{ij}, що інформує про пов'язані, у зв'язку з переходом системи зі стану S_i у стан S_j за один крок, витрати. Щоб визначити загальні витрати, які можна очікувати за *n* кроків роботи системи, використовують **матрицю вартостей**:

$$R = \begin{pmatrix} r_{11} & r_{12} & r_{13} & \dots & r_{1N} \\ r_{21} & r_{22} & r_{23} & \dots & r_{2N} \\ \dots & \dots & \dots & \dots & \dots \\ r_{N1} & r_{N2} & r_{N3} & \dots & r_{NN} \end{pmatrix}, \tag{9.30}$$

елементи якої можуть набувати додатних, нульових та від'ємних числових значень.

Якщо система досягла S_j стану, то очікувану винагороду після всіх переходів можна подати як $r_{ij} + v_i(n-1)$, де r_{ij} – це винагорода, яку дістають під час переходу системи зі стану S_i у стан S_j за один крок, а $v_i(n-1)$ – це очікувана винагорода, яку будуть мати за решту *n* – 1 кроків.

Ураховуючи те, що ймовірність переходу системи зі стану S_i у стан S_j за один крок дорівнює p_{ij}, то **сумарна очікувана вартість** (винагорода) за *n* переходів системи, починаючи зі стану S_i, буде:

$$v_i(n) = \sum_{j=1}^{N} r_{ij} p_{ij} + \sum_{j=1}^{N} v_j(n-1) p_{ij} , \tag{9.31}$$

або

$$v_i(n) = g_i + \sum_{j=1}^{N} v_j(n-1)\, p_{ij} \;, \tag{9.32}$$

де $g_i = \sum_{j=1}^{N} r_{ij}\, p_{ij}$.

У векторно-матричній формі можна записати:

$$\bar{v}(n) = \bar{g} + \pi \bar{v}(n-1), \tag{9.33}$$

де $\bar{g} = (g_1,\, g_2, ...,\, g_N)$, g_i – елементи головної діагоналі матриці $\pi \cdot R^T$;

R^T – транспонована матриця вартостей *R*.

Компоненти вектора $\bar{v}(n)$ визначають із векторно-матричного рівняння:

$$\bar{v}(n) = (E + \pi + \pi^2 + \pi^3 + ... + \pi^{(n-1)})\bar{g} + \pi^{(n)}\bar{v}(0), \tag{9.34}$$

де E – одинична матриця;

$\bar{v}(0)$ – вектор винагороди за *n* = 0 кроків.

9.3. Приклади розв'язування задач

Приклад 9.1. СМО з відмовами. У відділі технічного контролю (ВТК) цеху працюють три контролери. Якщо деталь надходить до ВТК, коли всі контролери зайняті обслуговуванням деталей, що надійшли раніше, то вона буде неперевіреною. Середня кількість деталей, що надходять до ВТК протягом години дорівнює 24, середній час, який витрачає один контролер на обслуговування однієї деталі, дорівнює 5 хв.

Обчисліть основні характеристики СМО. Визначте ймовірність того, що деталь пройде ВТК не обслуженою; наскільки завантажені контролери та скільки їх необхідно поставити, щоб $P_{обсл.} \geq 0{,}95$.

Розв'язання. За умовою задачі інтенсивність потоку заявок $\lambda = 24$ деталей/год = $0{,}4$ деталей/хв, $t_{обсл.} = 5$ хв.

Тоді $\mu = \dfrac{1}{t_{обсл.}} = 0{,}2,\ \rho = \dfrac{\lambda}{\mu} = 2$.

Основні характеристики СМО з відмовами такі.

Імовірність простою трьох каналів обслуговування визначають за формулою (9.1):

$$P_0 = \frac{1}{\frac{2^0}{0!} + \frac{2^1}{1!} + \frac{2^2}{2!} + \frac{2^3}{3!}} = 0{,}1587.$$

Імовірність відмови в обслуговуванні – за формулою (9.2):

$$P_{відм.} = \frac{0{,}1587 \cdot 2^3}{3!} = 0{,}21.$$

Імовірність обслуговування – за формулою (9.3):

$$P_{обсл.} = 1 - 0{,}21 = 0{,}79.$$

Середню кількість зайнятих обслуговуванням каналів – за формулою (9.4):

$$n_з = 2 \cdot 0{,}79 = 1{,}58.$$

Частку каналів, зайнятих обслуговуванням, – за формулою (9.5):

$$k_з = \frac{n_з}{n} = \frac{1{,}58}{3} = 0{,}526.$$

Абсолютну пропускну спроможність – за формулою (9.6):

$$A = 0{,}4 \cdot 0{,}79 = 0{,}316.$$

Отже, обчислено основні характеристики СМО та визначено, що за умови трьох контролерів ($n = 3$) $P_{обсл.} = 0{,}79 \le 0{,}95$, деталь пройде ВТК необслуженою з імовірністю 21 %, контролери будуть зайняті обслуговуванням на 53 %.

Зробіть аналогічні розрахунки за умови, що $n = 4$:

$P_0 = 0{,}14$; $P_{відм.} = 0{,}093$; $P_{обсл.} = 0{,}907$.

Оскільки $P_{обсл.} = 0{,}907 \le 0{,}95$, то роблять обчислення для $n = 5$.

Знаходять: $P_0 = 0{,}137$; $P_{відм.} = 0{,}035$; $P_{обсл.} = 0{,}965 \ge 0{,}95$.

Отже, щоб забезпечити ймовірність обслуговування понад 95 %, необхідно не менше від п'яти контролерів.

Приклад 9.2. СМО з необмеженим очікуванням. Ощадкаса має трьох контролерів-касирів ($n = 3$) для обслуговування вкладників. Потік вкладників надходить в ощадкасу з інтенсивністю $\lambda = 30$ осіб на годину. Середня тривалість обслуговування контролером-касиром одного вкладника $t_{обсл.} = 3$ хв.

Визначте характеристики ощадкаси як об'єкта СМО.

Розв'язання. Інтенсивність потоку обслуговування:

$$\mu = \frac{1}{t_{обсл.}} = \frac{1}{3} = 0{,}333,$$

інтенсивність навантаження:

$$\rho = \frac{\lambda}{\mu} = \frac{30/60}{1/3} = 1{,}5.$$

Тоді основні характеристики СМО з необмеженим очікуванням є такі.

Імовірність простою контролерів-касирів протягом робочого дня визначають за формулою (9.7):

$$P_0 = \frac{1}{\dfrac{1{,}5^0}{0!} + \dfrac{1{,}5^1}{1!} + \dfrac{1{,}5^2}{2!} + \dfrac{1{,}5^3}{3!}} + \frac{1{,}5^4}{3!(3-1{,}5)} = 0{,}21.$$

Імовірність застати всіх контролерів-касирів зайнятими – за формулою (9.9):

$$P_n = \frac{1{,}5^3}{3!} \cdot 0{,}21 = 0{,}118.$$

Імовірність черги – за формулою (9.10):

$$P_{оч.} = \frac{1{,}5^4}{3!(3-1{,}5)} \cdot 0{,}21 = 0{,}118.$$

Середню кількість заявок у черзі – за формулою (9.11):

$$L_{оч.} = \frac{1{,}5^4}{(3-1)!(3-1{,}5)^2} \cdot 0{,}21 = 0{,}236.$$

Середній час очікування заявки в черзі – за формулою (9.12):

$$t_{оч.} = \frac{0{,}236}{0{,}5} = 0{,}472 \text{ (хв).}$$

Середній час перебування заявки у СМО – за формулою (9.13):

$$t_{СМО} = 0{,}472 + 3 = 3{,}472 \text{ (хв).}$$

Середню кількість вільних каналів – за формулою (9.15):

$$n_в = 3 - 1{,}5 = 1{,}5.$$

Коефіцієнт зайнятості каналів обслуговування – за формулою (9.16):

$$k_з = \frac{1{,}5}{3} = 0{,}5.$$

Середню кількість відвідувачів в ощадкасі – за формулою (9.17):

$$z = 0{,}235 + 1{,}5 = 1{,}736 \text{ (особи).}$$

Таким чином, імовірність простою контролерів-касирів дорівнює 21 % робочого часу, імовірність потрапити відвідувачеві в чергу становить 11,8 %, середня кількість відвідувачів у черзі 0,236 особи, середній час очікування відвідувачами обслуговування 0,472 хв.

Приклад 9.3. СМО з очікуванням і з обмеженою довжиною черги. Магазин отримує ранні овочі із приміських теплиць. Машини з вантажем прибувають у різний час з інтенсивністю $\lambda = 6$ машин на день. Підсобні приміщення й устаткування для підготовки овочів до продажу дозволяють обробляти та зберігати товар, привезений двома машинами ($m = 2$). У магазині працюють три фасувальники ($n = 3$), кожен із яких у середньому може обробляти товар з однієї машини протягом $t_{обсл.} = 4$ год. Тривалість робочого дня за змінної роботи становить 12 год.

Визначте, яка має бути місткість підсобних приміщень, щоб імовірність повної обробки товарів була $P_{обсл.} \geq 0{,}97$.

Обчисліть основні характеристики СМО.

Розв'язання. Інтенсивність завантаження фасувальників:

$$\mu = \frac{1}{t_{обсл.}} = 1 \cdot 12/4 = 3 \text{ (машини за день).}$$

Тоді

$$\rho = \frac{\lambda}{\mu} = \frac{6}{3} = 2.$$

Основні характеристики СМО з очікуванням і з обмеженою довжиною черги такі.

Імовірність простою фасувальників за відсутності машин (заявок) визначають за формулою (9.18):

$$P_0 = \frac{1}{\frac{2^0}{0!} + \frac{2^1}{1!} + \frac{2^2}{2!} + \frac{2^3}{3!} + \frac{2^4}{3!(3-2)}\left[1 - \left(\frac{2}{3}\right)^2\right]} = 0{,}128.$$

Імовірність відмови в обслуговуванні – за формулою (9.19):

$$P_{відм.} = \frac{0{,}128 \cdot 2^{3+2}}{3!3^2} = 0{,}075.$$

Імовірність обслуговування – за формулою (9.20):

$$P_{обсл.} = 1 - 0{,}075 = 0{,}925.$$

Оскільки $P_{обсл.} = 0{,}925 \le 0{,}97$, виконують обчислення для $m = 3$, знаходять $P_0 = 0{,}122; P_{відм.} = 0{,}048; P_{обсл.} = 0{,}952$.

Оскільки $P_{обсл.} = 0{,}952 \le 0{,}97$, беруть $m = 4$. Для цього випадку $P_0 = 0{,}12; P_{відм.} = 0{,}028; P_{обсл.} = 0{,}972 > 0{,}97$, тобто місткість підсобних приміщень необхідно збільшити до $m = 4$.

Таким чином, місткість підсобних приміщень магазину має вміщати товар, привезений 4 машинами ($m = 4$), водночас імовірність повної обробки товару буде $P_{обсл.} = 0{,}972$. Для досягнення заданої ймовірності обслуговування можна збільшувати кількість фасувальників, виконуючи послідовно обчислення СМО для $n = 4, 5$тощо. Задачу можна також розв'язати, збільшуючи місткість підсобних приміщень, кількість фасувальників, зменшуючи час оброблення товарів.

Знайдіть решту параметрів СМО для розрахованого випадку за $P_0 = 0{,}12;\ P_{відм.} = 0{,}028;\ P_{обсл.} = 0{,}972$.

Абсолютна пропускна спроможність – за формулою (9.21):

$$A = 0{,}972 \cdot 6 = 5{,}832 \text{ (машини за день).}$$

Середню кількість зайнятих обслуговуванням каналів (фасувальників) – за формулою (9.22):

$$n_з = \frac{5{,}832}{3} = 1{,}944.$$

Середню кількість заявок у черзі – за формулою (9.23):

$$L_{оч.} = \frac{2^4}{3 \cdot 3!} \cdot \frac{1 - (2/3)^4 (4 + 1 - 4 \cdot 2/3)}{(1 - 2/3)^2} \cdot 0{,}12 = 0{,}548.$$

Середній час очікування обслуговування – за формулою (9.24):

$$t_{оч.} = \frac{0{,}548}{6} = 0{,}09 \text{ (дня).}$$

Середню кількість машин в магазині – за формулою (9.25):

$$z = 0{,}548 + 1{,}944 = 2{,}492.$$

Середній час перебування машини в магазині – за формулою (9.26):

$$t_{CMO} = \frac{2{,}492}{6} = 0{,}415 \text{ (дня).}$$

Приклад 9.4. Знайдіть стаціонарні ймовірності для регулярної матриці:

$$\pi = \begin{pmatrix} 0{,}3 & 0{,}1 & 0{,}6 \\ 0{,}2 & 0{,}5 & 0{,}3 \\ 0{,}1 & 0{,}4 & 0{,}5 \end{pmatrix}.$$

Розв'язання. Використовуючи систему (9.29), знаходять:

$$\begin{cases} \begin{pmatrix} b_1 \\ b_2 \\ b_3 \end{pmatrix} = \begin{pmatrix} 0{,}3 & 0{,}1 & 0{,}6 \\ 0{,}2 & 0{,}5 & 0{,}3 \\ 0{,}1 & 0{,}4 & 0{,}5 \end{pmatrix} \begin{pmatrix} b_1 \\ b_2 \\ b_3 \end{pmatrix} \\ b_1 + b_2 + b_3 = 1 \end{cases} \to \begin{cases} \begin{pmatrix} b_1 \\ b_2 \\ b_3 \end{pmatrix} = \begin{pmatrix} 0{,}3b_1 & 0{,}1b_2 & 0{,}6b_3 \\ 0{,}2b_1 & 0{,}5b_2 & 0{,}3b_3 \\ 0{,}1b_1 & 0{,}4b_2 & 0{,}5b_3 \end{pmatrix} \\ b_1 + b_2 + b_3 = 1 \end{cases} \to$$

$$\to \begin{cases} b_1 = 0{,}3b_1 + 0{,}1b_2 + 0{,}6b_3, \\ b_2 = 0{,}2b_1 + 0{,}5b_2 + 0{,}3b_3, \\ b_3 = 0{,}1b_1 + 0{,}4b_2 + 0{,}5b_3, \\ b_1 + b_2 + b_3 = 1 \end{cases} \to \begin{cases} -0{,}7b_1 + 0{,}1b_2 + 0{,}6b_3 = 0, \\ 0{,}2b_1 - 0{,}5b_2 + 0{,}3b_3 = 0, \\ 0{,}1b_1 + 0{,}4b_2 - 0{,}5b_3 = 0, \\ b_1 + b_2 + b_3 = 1. \end{cases}$$

Розв'язком визначеної системи є:

$$\bar{b} = \begin{pmatrix} \frac{1}{3} \\ \frac{1}{3} \\ \frac{1}{3} \end{pmatrix}, \text{ тобто } b_1 = \frac{1}{3},\ b_2 = \frac{1}{3},\ b_3 = \frac{1}{3}.$$

Приклад 9.5. Виробник придбав новий верстат, який може перебувати в одному із трьох несумісних станів: S_1– верстат працює добре, не потребує жодних утручань, пов'язаних із витратами; S_2 – верстат може бути використаний у роботі, але потребує дрібного ремонту, який, своєю чергою, потребує певних витрат; S_3 – верстат перебуває в неробочому стані, що відображається на прибутках виробника.

Матриця ймовірностей переходу верстата з одного стану в інший за один крок (за 1 місяць) має такий вигляд:

$$\pi = \begin{pmatrix} 0{,}9 & 0{,}06 & 0{,}04 \\ 0{,}3 & 0{,}6 & 0{,}1 \\ 0{,}2 & 0{,}02 & 0{,}78 \end{pmatrix}.$$

Оцініть ефективність роботи верстата за три робочі місяці за заданою матрицею вартостей (грн):

$$R = \begin{pmatrix} 200 & 50 & -10 \\ 150 & 40 & -20 \\ 100 & 10 & -50 \end{pmatrix}.$$

Задано, що $(\bar{v}(0))^T = (0\ 0\ 0)$.

Розв'язання. Із рівняння (9.34) для *n* = 3 знайдіть

$$\bar{v}(3) = (E + \pi + \pi^2)\bar{g} + \pi^3\bar{v}(0).$$

Визначте координати вектора $\bar{g}$:

$$\pi R' = \begin{pmatrix} 0{,}9 & 0{,}06 & 0{,}04 \\ 0{,}3 & 0{,}6 & 0{,}1 \\ 0{,}2 & 0{,}02 & 0{,}78 \end{pmatrix} \begin{pmatrix} 200 & 150 & 100 \\ 50 & 40 & 10 \\ -10 & -20 & -50 \end{pmatrix} = \begin{pmatrix} 182{,}6 & 136{,}6 & 88{,}6 \\ 89 & 67 & 31 \\ 33{,}2 & 15{,}2 & -18{,}8 \end{pmatrix}.$$

Таким чином, буде:

$$\overline{g} = \begin{pmatrix} 182{,}6 \\ 67 \\ -18{,}8 \end{pmatrix}.$$

Початкове рівняння перепишіть у такому вигляді:

$$\begin{pmatrix} v_1(3) \\ v_2(3) \\ v_3(3) \end{pmatrix} = \left[\begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} + \begin{pmatrix} 0{,}9 & 0{,}06 & 0{,}04 \\ 0{,}3 & 0{,}6 & 0{,}1 \\ 0{,}2 & 0{,}02 & 0{,}78 \end{pmatrix} + \begin{pmatrix} 0{,}9 & 0{,}06 & 0{,}04 \\ 0{,}3 & 0{,}6 & 0{,}1 \\ 0{,}2 & 0{,}02 & 0{,}78 \end{pmatrix}^2 \right] \begin{pmatrix} 182{,}6 \\ 67 \\ -18{,}8 \end{pmatrix} +$$

$$+ \begin{pmatrix} 0{,}9 & 0{,}06 & 0{,}04 \\ 0{,}3 & 0{,}6 & 0{,}1 \\ 0{,}2 & 0{,}02 & 0{,}78 \end{pmatrix}^3 \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix} \to$$

$$\to \begin{pmatrix} v_1(3) \\ v_2(3) \\ v_3(3) \end{pmatrix} = \begin{pmatrix} 182{,}6 \\ 67 \\ -18{,}8 \end{pmatrix} + \begin{pmatrix} 167{,}61 \\ 93{,}1 \\ 23{,}2 \end{pmatrix} + \begin{pmatrix} 157{,}6 \\ 108{,}5 \\ 43{,}5 \end{pmatrix} = \begin{pmatrix} 507{,}81 \\ 268{,}6 \\ 47{,}9 \end{pmatrix}.$$

Таким чином, буде:

$$\begin{pmatrix} v_1(3) \\ v_2(3) \\ v_3(3) \end{pmatrix} = \begin{pmatrix} 507{,}81 \\ 268{,}6 \\ 47{,}9 \end{pmatrix}.$$

Отже, коли верстат за три місяці перейде у стан S_1 (справний), то прибуток виробника буде дорівнювати 507 грн 81 к.; для стану S_2 цей прибуток буде становити 268 грн 60 к. і для стану S_3 – лише 47 грн 90 к.

9.4. Вправи для самостійної роботи

9.1. Контроль за готовою продукцією фірми здійснюють 5 контролерів. Якщо виріб надходить на контроль, коли всі контролери зайняті перевіркою готових виробів, то він залишається неперевіреним. Середня кількість виробів, що випускаються фірмою, становить 25 виробів за годину. Середній час на перевірку одного виробу – 5 хв. Обчисліть основні характеристики СМО. Визначте ймовірність того, що виріб пройде перевірку; наскільки завантажені контролери та скільки їх необхідно постачити, щоб $P_{обсл.} \geq 0{,}96$.

9.2. Прибуткова каса міського району працює 8 год на день і приймає від населення плату за комунальні послуги та різні платежі в середньому від 280 осіб на день. У прибутковій касі працюють 4 оператори-касири. Середня тривалість обслуговування одного клієнта становить 4 хв. Визначте характеристики роботи прибуткової каси як об'єкта СМО.

9.3. На АЗС установлено 4 колонки для видавання бензину. Біля станції знаходиться майданчик на 2 машини для очікування заправлення. На станцію прибуває в середньому 20 машин на год. Середній час заправлення однієї машини – 3,5 хв. Обчисліть основні характеристики СМО. Визначте ймовірність відмови та середню довжину черги.

9.4. На станцію поточного технічного ремонту надходять машини, які утворюють найпростіший потік із параметром $\lambda = 0{,}8$ машин на год. У ремонті бере участь одна бригада робітників. Середній час ремонту однієї машини є випадковою величиною, що має експоненціальний закон розподілу з параметром $\mu = 1{,}8$ машини на год. На майданчику ремонтної станції одночасно може перебувати не більше як 4 машини. Визначте: імовірність того, що ремонтна станція буде простоювати без роботи; імовірність того, що майданчик станції буде повністю заповненим; середню кількість машин на ремонтній станції.

Відповідь: $P_0 = 0{,}6399$; $P_5 = 0{,}024$; $z = 0{,}7878$.

9.5. Досліджують роботу автозаправної станції (АЗС), яка має три заправні колонки. У середньому за кожну хвилину на АЗС надходять 2 машини. Заправлення 1 машини в середньому триває 4 хв. Кількість машин, що можуть перебувати в черзі для заправлення, практично не обмежена, і всі машини не залишають черги. Визначте: імовірність того, що на АЗС буде відсутня черга; середню кількість машин у черзі; середній час очікування для машин до початку їхнього обслуговування.

Відповідь: $P_0 = 0{,}606$; $z = 0{,}43$; $t = 0{,}11$ хв.

9.6. Знайдіть стаціонарні ймовірності для регулярних ланцюгів Маркова, що мають матриці однокрокового переходу системи:

а) $\pi = \begin{pmatrix} 0{,}9 & 0{,}1 \\ 0{,}2 & 0{,}8 \end{pmatrix}$; б) $\pi = \begin{pmatrix} 0{,}5 & 0{,}4 & 0{,}1 \\ 0{,}3 & 0{,}4 & 0{,}3 \\ 0{,}2 & 0{,}1 & 0{,}7 \end{pmatrix}$.

9.7. Дві фірми пропонують виробнику обладнання, яке умовно позначено як техніка фірми *А* і техніка фірми *В*. Ця техніка має такі імовірні матриці однокрокового переходу зі стану S_1 (працює добре) у стан S_2 (потребує ремонту).

Дані матриці мають такий вигляд:

$$\pi_A = \begin{pmatrix} 0{,}89 & 0{,}01 \\ 0{,}82 & 0{,}12 \end{pmatrix},\ \pi_B = \begin{pmatrix} 0{,}79 & 0{,}21 \\ 0{,}72 & 0{,}28 \end{pmatrix}.$$

Визначте, яке обладнання варто купити виробнику.

Відповідь: обладнання фірми *А*.

9.5. Тестові завдання

9.1. Одна робітниця обслуговує 20 ткацьких верстатів, забезпечуючи їхній запуск після обриву нитки. Модель такої системи масового обслуговування можна охарактеризувати як

А багатоканальну однофазову з обмеженою довжиною черги

Б одноканальну однофазову з необмеженою довжиною черги

В одноканальну багатофазову з обмеженою довжиною черги

Г одноканальну однофазову з обмеженою довжиною черги

Д багатоканальну однофазову з необмеженою довжиною черги

9.2. У теорії масового обслуговування для опису простого потоку заявок, що надходять на вхід системи, використовують такий розподіл імовірностей

А нормальний

Б експоненціальний

В пуассонівський

Г біноміальний

9.3. У теорії масового обслуговування передбачають, що кількість заявок у черзі є

А фіксованою або змінною

Б обмеженою або необмеженою

В відомою або невідомою

Г випадковою або детермінованою

9.4. Двома основними параметрами, які визначають конфігурацію системи масового обслуговування, є

А темп надходження і темп обслуговування

Б довжина черги та правило обслуговування

В розподіл часу між заявками й розподіл часу обслуговування

Г кількість каналів і кількість фаз обслуговування

9.5. У теорії масового обслуговування для опису часу, що витрачають на обслуговування заявок, зазвичай використовують такий розподіл імовірностей

А нормальний

Б експоненціальний

В пуассонівський

Г біноміальний

9.6. Ремонт комп'ютерів, що вийшли з ладу, здійснюють три фахівці, які працюють одночасно й незалежно один від одного. Модель такої системи масового обслуговування можна охарактеризувати як

А багатоканальну з обмеженою довжиною черги

Б одноканальну з необмеженою довжиною черги

В одноканальну з обмеженою довжиною черги

Г багатоканальну з необмеженою довжиною черги

9.7. Випадковий процес є процесом із неперервним часом, якщо

А його переріз у будь-який можливий момент часу t утворює неперервну випадкову величину

Б перехід системи з одного стану в інший може бути здійснений у будь-який можливий момент часу t за певний спостережуваний період

В у будь-який можливий момент часу *t* множина його станів є скінченною, або зліченною

Г відповідь відсутня

9.8. Ланцюг Маркова називають однорідним, якщо

А $p_{ij}(t)$ не залежить від часу

Б у кожний можливий стан система може перейти через певний інтервал часу

В якщо існує таке ціле число *k* (*k* > 0), що перехід системи з будь-якого можливого стану в інший може здійснитися за *k* кроків

Г відповідь відсутня

9.9. Ланцюг Маркова називають регулярним, якщо

А імовірність переходу зі стану S_i у стан S_j у момент часу *t* $p_{ij}(t)$ не залежить від часу

Б у кожний можливий стан система може перейти через певний інтервал часу

В якщо існує таке ціле число *k* (*k* > 0), що перехід системи з будь-якого можливого стану в інший може здійснитися за *k* кроків

Г відповідь відсутня

9.6. Запитання для самоперевірки

9.1. Що називають випадковим процесом?

9.2. Що називають реалізацією випадкового процесу?

9.3. Що називають перерізом випадкового процесу?

9.4. Що називають функцією розподілу перерізу?

9.5. Дайте визначення математичного сподівання та дисперсії перерізу випадкового процесу $X = x(t)$.

9.6. Що називають кореляційною функцією випадкового процесу? Які вона має властивості?

9.7. Дайте визначення найпростішого потоку подій. Наведіть властивості найпростішого потоку подій.

9.8. Який закон розподілу ймовірностей використовують у теорії масового обслуговування для характеристики найпростішого потоку заявок, що надходять на вхід системи?

9.9. Якими основними параметрами визначають конфігурацію системи масового обслуговування?

9.10. Що називають марковським процесом?

9.11. Що таке "ланцюг Маркова"? Наведіть класифікацію ланцюгів Маркова. Які ланцюги Маркова називають регулярними?

9.12. Який вигляд має матриця стаціонарних імовірностей? Як знаходять стаціонарні ймовірності для регулярних ланцюгів Маркова?

9.13. Що таке "матриця вартостей"?

9.7. Висновки за темою

Теорія випадкових процесів – це розділ математичної науки, який вивчає закономірності випадкових явищ у динаміці їхнього розвитку.

Предметом теорії масового обслуговування є побудова математичних моделей, що пов'язують задані умови роботи СМО з показниками її ефективності, які описують здатність цієї системи обробляти потоки заявок. Мета вивчення СМО полягає в тому, щоб узяти під контроль деякі характеристики системи обслуговування, установити залежність між кількістю обслуговуваних одиниць і якістю обслуговування.

У промисловості СМО застосовують під час постачання сировини, матеріалів, комплектних виробів на склад і видавання їх зі складу; оброблення широкої номенклатури деталей на одному тому самому устаткуванні; організації налагодження і ремонту устаткування; визначенні оптимальної кількості обслуговчих відділів і служб підприємств тощо. Розрізняють системи з відмовами, системи з необмеженими очікуваннями, системи змішаного типу з очікуванням та обмеженою довжиною черги.

Серед випадкових процесів особливе місце посідають марковські процеси. Марковський процес – це випадковий процес із дискретними станами, що ґрунтується на принципі Маркова. Ланцюг Маркова є випадковим процесом із дискретним часом.

Розділ 2
Математична статистика

10. Граничні теореми теорії ймовірностей

10.1 Мета та компетентності

Метою вивчення теми є ознайомлення зі змістом граничних теорем теорії ймовірностей та формування компетентностей щодо використання їх на практиці.

Компетентності, що формують під час вивчення теми:

знання особливостей застосування теорем закону великих чисел і центральної граничної теореми;

уміння застосовувати граничні теореми теорії ймовірностей для обчислення ймовірностей відхилення між числовими характеристиками випадкової величини та результатами їхнього статистичного оцінювання.

10.2. Основні теоретичні відомості

Граничні теореми теорії ймовірностей можна умовна розподілити на два класи: закон великих чисел і граничні теореми.

Закон великих чисел математично описує стійкість середніх значень масових випадкових явищ. Закон великих чисел складається з декількох теорем, у кожній із яких за певних умов затверджено факт збіжності середніх характеристик під час проведення великої кількості випробувань до певних невипадкових, постійних величин.

Нерівність Маркова. Якщо випадкова величина набуває лише невід'ємних значень та має математичне сподівання, то для будь-якого $\varepsilon > 0$ має місце нерівність:

$$P(x > \varepsilon) \leq \frac{M(X)}{\varepsilon}.$$

Нерівність Чебишова. Якщо випадкова величина має математичне сподівання та обмежену дисперсію, то для будь-якого $\varepsilon > 0$ має місце нерівність:

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

Нерівність Чебишова дає можливість оцінити помилку, яку допускають, якщо математичне сподівання замінюють середнім значенням обмеженої вибірки.

Оцінювання відхилення відносної частоти появи випадкової величини від імовірності її появи в окремому випробуванні здійснюють за такою формулою:

$$P\left(\left|\frac{m}{n}-p\right|<\varepsilon\right)\geq 1-\frac{pq}{n\varepsilon^2}.$$

Теорема Чебишова. Нехай $X_1, X_2, \ldots, X_n$ – попарно незалежні випадкові величини з обмеженими дисперсіями, тобто існує таке D, що $D(X_1)\leq D$, $D(X_2)\leq D$, ..., $D(X_n)\leq D$. Тоді для будь-якого $\varepsilon>0$ ймовірність відхилення середнього арифметичного значень цих величин від середнього арифметичного їхніх математичних сподівань на величину не більшу ніж ε скільки завгодно близька до одиниці, якщо кількість випадкових величин достатньо велика:

$$\lim_{n\to\infty} P\left(\left|\frac{X_1+X_2+\ldots+X_n}{n}-\frac{M(X_1)+M(X_2)+\ldots+M(X_n)}{n}\right|\leq\varepsilon\right)=1.$$

Теорему Чебишова можна розглядати як *необхідну і достатню умову* граничної збіжності середніх значень незалежних випадкових величин до середніх значень їхніх математичних сподівань. Сенс теореми полягає в тому, що середнє арифметичне достатньо великої кількості незалежних випадкових величин утрачає характер випадкової величини.

Теорема Бернуллі. Якщо ймовірність події в n випробуваннях постійна і дорівнює p, то ймовірність відхилення відносної частоти від цієї ймовірності на величину, що не перевершує $\varepsilon>0$, прагне до одиниці в разі достатньо великої кількості випробувань n:

$$\lim_{n\to\infty} P\left(\left|\frac{m}{n}-p\right|\leq\varepsilon\right)=1.$$

Теорема Бернуллі стверджує стійкість відносних частот за умови, що випробування є однорідними. У тому разі, якщо умови випробувань змінюються, тобто змінюється ймовірність появи події в одиничному випробуванні, теж має місце аналогічна стійкість результатів.

Теорема Хінчина. Якщо $X_1, X_2, ..., X_n$ – незалежні однаково розподілені випадкові величини зі скінченним математичним сподіванням a, то

$$\lim_{n\to\infty} P\left(\left|\frac{X_1 + X_2 + ... + X_n}{n} - a\right| < \varepsilon\right) = 1.$$

Установлені раніше зв'язки між послідовностями випадкових величин і їхніми математичними сподіваннями, насамперед, належать до *дискретних законів розподілу*.

Якщо розглянути випадкові величини з *неперервними законами розподілу,* то послідовності незалежних випадкових величин за певних умов також виявляють тенденцію до утворення граничних залежностей.

Граничні теореми, що встановлюють граничні закони розподілу випадкових величин, об'єднують загальною назвою **центральна гранична теорема**. Різні форми центральної граничної теореми відрізняються умовами, за яких виникає нормальний закон розподілу. Серед цих теорем важливе місце посідає теорема Ляпунова.

Теорема. Якщо випадкові величини $X_1, X_2, ..., X_n$ у послідовності незалежні, однаково розподілені та для них існують моменти другого порядку та

$$S_n = \sum_{i=1}^{n} X_i; \ \ D(X_i) = b_i^2; \ \ B_n^2 = \sum_{i=1}^{n} b_i^2; \ \ S_n^* = \frac{S_n - M(S_n)}{B_n},$$

то

$$\lim_{n\to\infty} P\left(S_n^* < x\right) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-\frac{t^2}{2}} dt,$$

тобто граничним розподілом для S_n^* є нормальний закон розподілу з нульовим математичним сподіванням та одиничною дисперсією.

Теорема Ляпунова. Якщо $X_1, X_2, ..., X_n$ незалежні випадкові величини, у кожної з яких математичне сподівання

$$M(X_i) = a_i,$$

дисперсія

$$D(X_i) = \sigma_i^2,$$

абсолютний центральний момент третього порядку

$$M\left(\left|X_i - a_i\right|^3\right) = m_i \text{ та } \lim_{n\to\infty} \frac{\sum_{i=1}^{n} m_i}{\left(\sum_{i=1}^{n} \sigma_i^2\right)^{3/2}} = 0,$$

то закон розподілу суми $Y = X_1 + X_2 + ... + X_n$ необмежено наближається до нормального з математичним сподіванням $\sum_{i=1}^{n} a_i$ та дисперсією $\sum_{i=1}^{n} \sigma_i^2$.

Сенс теореми Ляпунова для її використання на практиці полягає в такому: якщо випадкова величина є сумою дуже великої кількості взаємно незалежних випадкових величин, вплив кожної з яких на суму малий, то вона має розподіл, близький до нормального.

10.3. Приклади розв'язування задач

Приклад 10.1. Кількість палива, що витрачають на фірмі протягом доби, є випадковою величиною, математичне сподівання якої дорівнює 6 т. Оцініть імовірність того, що протягом найближчої доби витрати палива перевищать 10 т.

Розв'язання. За умовою $M(X) = 6$, $\varepsilon = 10$.

За нерівністю Маркова: $P(X > 10) \le \frac{6}{10}$.

Отже, з імовірністю, яка не перевищує 0,6, можна стверджувати, що витрати палива будуть більшими за 10 т.

Приклад 10.2. Електрична мережа має 18 000 ламп, імовірність увімкнення кожної з яких – 0,9. Оцініть імовірність того, що кількість увімкнених ламп відхиляється від свого математичного сподівання на величину не меншу ніж 200.

Розв'язання. За умовою $n = 18\,000$, $p = 0{,}9$.

Тоді

$$M(X) = np = 18\,000 \cdot 0{,}9 = 16\,200,$$
$$D(X) = npq = 18\,000 \cdot 0{,}9 \cdot 0{,}1 = 1\,620.$$

Згідно з нерівністю Чебишова:

$$P(|X - 16\,200| \le 200) \ge 1 - \frac{1\,620}{200^2} = 1 - 0{,}045 = 0{,}955;$$

$$P(|X - 16\,200| \ge 200) \ge 1 - 0{,}955 = 0{,}045.$$

Отже, з імовірністю понад 0,045, можна стверджувати, що кількість увімкнених ламп відхиляється від свого математичного сподівання на величину не меншу ніж 200.

Приклад 10.3. Оцініть імовірність того, що з 800 малих підприємств збанкрутує не менше ніж 50, якщо ймовірність банкрутства будь-якого навмання взятого малого підприємства постійна і дорівнює 0,05.

Розв'язання. Нехай випадкова величина X – це кількість збанкрутілих підприємств. Ця випадкова величина підпорядковується біноміальному розподілу з параметрами $n = 800,\ p = 0{,}05$. Тоді, згідно з формулою обчислення математичного сподівання випадкової величини, що має біноміальний розподіл, буде: $M(X) = np = 800 \cdot 0{,}05 = 40$.

Використайте нерівність Маркова для $\varepsilon = 50$:

$$P(X > \varepsilon) \le \frac{40}{50} = \frac{4}{5}.$$

Приклад 10.4. Середня температура у приміщенні в період опалювального сезону дорівнює 20 ^{o}C, а середнє квадратичне відхилення дорівнює 2 ^{o}C. За допомогою нерівності Чебишова оцініть імовірність того, що температура у квартирі відхилиться від середньої менше ніж на 4 ^{o}C.

Розв'язання. Нехай випадкова величина X – це температура у приміщенні. За умови $M(X) = 20,\ \sigma(X) = 2,\ \varepsilon = 4$.

Із нерівності Чебишова:

$$P(|X - 20| < 4) \ge 1 - \frac{4}{16} = \frac{3}{4}.$$

Приклад 10.5. Імовірність того, що телевізор витримає гарантійний термін роботи дорівнює 0,95 для всіх 200 телевізорів, які обслуговує гарантійна майстерня. Оцініть імовірність того, що кількість телевізорів, які витримають гарантійний термін роботи, буде в межах [185; 195].

Розв'язання. Дискретна випадкова величина X – це кількість телевізорів, які витримають гарантійний термін роботи, має біноміальний розподіл імовірності, тому

$$M(X) = np = 200 \cdot 0{,}95 = 190;\ D(X) = npq = 200 \cdot 0{,}05 \cdot 0{,}95 = 9{,}5.$$

Обчисліть $\varepsilon = 195 - 190 = |185 - 190| = 5$ і використайте нерівність Чебишова:

$$P(|X - 190| < 5) \geq 1 - \frac{9{,}5}{25} = 0{,}62.$$

Приклад 10.6. Середнє споживання електроенергії протягом травня в місті дорівнює 360 000 кВт · год. Оцініть імовірність того, що споживання електроенергії у травні поточного року перевищить 1 000 000 кВт · год. Оцініть ту саму ймовірність за умови, що середнє квадратичне відхилення споживання електроенергії за травень дорівнює 40 000 кВт · год.

Розв'язання. Випадкова величина X – це споживання електроенергії – набуває невід'ємних значень. Математичне сподівання її дорівнює 360 000.

Оцініть імовірність за допомогою нерівності Маркова:

$$P(X \geq 1\,000\,000) \leq \frac{360\,000}{1\,000\,000} = 0{,}36.$$

Оцініть цю саму нерівність, якщо відоме середнє квадратичне відхилення X.

Скористайтеся нерівністю Чебишова:

$$P(X \geq 1\,000\,000) = 1 - P(X < 1\,000\,000) = 1 - P(0 < X < 1\,000\,000) =$$
$$= 1 - P(-360\,000 < X < -MX < 640\,000) = 1 - P(|X - MX| < 640\,000) \leq$$
$$\leq 1 - 1 + \frac{(40\,000)^2}{(640\,000)^2} = \frac{1}{256}.$$

Приклад 10.7. Оцініть кількість деталей, необхідну для того, щоб з імовірністю не меншою ніж 0,98 можна було очікувати, що абсолютна величина відхилення відносної частоти стандартних деталей від імовірності, яка дорівнює $p = 0{,}95$, була меншою за 0,05.

Розв'язання. Застосуйте нерівність Чебишова.

У цьому прикладі $p = 0{,}95,\ P \geq 0{,}98,\ \varepsilon = 0{,}05$.

Знаходять n за нерівністю:

$$0{,}98 \leq 1 - \frac{0{,}95 \cdot 0{,}05}{n \cdot (0{,}05)^2} \Rightarrow n \geq \frac{0{,}95 \cdot 0{,}05}{0{,}02(0{,}05)^2}.$$

Отже, визначають, що $n \geq 950$, тобто необхідно не менше ніж 950 деталей, щоб з імовірністю не меншою ніж 0,98 можна було очікувати, що абсолютна величина відхилення відносної частоти стандартних деталей від $p = 0{,}95$ є меншою за 0,05.

Приклад 10.8. Контролер перевіряє деякі вироби. На першому етапі перевірки, який триває 10 с, він або відразу оцінює виріб, або вирішує, що перевірку треба повторити. Повторна перевірка триває 10 с, у результаті чого обов'язково ухвалюють рішення про якість продукції. Знайдіть імовірність того, що за семигодинний робочий день контролер перевірить понад 1 800 виробів; понад 1 640 виробів; не менш як 1 500 виробів. Передбачають, що кожний виріб, незалежно від інших, з імовірністю 0,5 проходить повторну перевірку.

Розв'язання. Нехай X_i – час (хв), потрібний для перевірки *i*-го виробу. Визначте послідовність незалежних однаково розподілених випадкових величин $X_1, X_2, ..., X_n, ...$.

Кожна з них має такий закон розподілу:

$$x_1 = \frac{1}{6};\ p(x_1) = \frac{1}{2};$$

$$x_2 = \frac{1}{3};\ p(x_2) = \frac{1}{2}.$$

Числові характеристики такі:

$$M(X_i) = \frac{1}{12} + \frac{1}{6} = \frac{1}{4};\ D(X_i) = \frac{1}{72} + \frac{1}{18} - \frac{1}{16} = \frac{1}{144}.$$

Загальний час S_n, що його витрачає робітник на перевірку *n* виробів, є сумою *n* незалежних однаково розподілених величин:

$$S_n = \sum_{i=1}^{n} X_i .$$

Згідно з теоремою, величина S_n має закон розподілу, близький до нормального.

Якщо $n = 1\,800,$ то $M(S_n) = 450,\ D(S_n) = 12{,}5.$ Тривалість зміни становить 420 хв.

Тоді за інтегральною теоремою Лапласа буде:

$$P(S_n < 420) = \Phi\left(\frac{420-450}{\sqrt{12{,}5}}\right) - \Phi\left(\frac{0-420}{\sqrt{12{,}5}}\right) = -\Phi\left(6\sqrt{2}\right) + \Phi\left(90\sqrt{2}\right) = 0.$$

Якщо $n = 1640,$ то $M(S_n) = 410,\ D(S_n) = \dfrac{205}{18} \approx 11{,}39.$

У такому разі:

$$P(S_n < 420) = \Phi\left(\frac{420-410}{\sqrt{11{,}39}}\right) - \Phi\left(\frac{0-410}{\sqrt{11{,}39}}\right) = \Phi(2{,}96) + \Phi(121{,}5) \approx 0{,}9985.$$

Без обчислень можна стверджувати, що за $n = 1\,500$

$$P(S_n < 420) = 1.$$

10.4. Вправи для самостійної роботи

10.1. Середнє значення ваги виробу – 50 г. За допомогою нерівності Маркова оцініть імовірність того, що навмання взятий виріб має вагу меншу за 90 г.

10.2. Використовуючи нерівність Чебишова, оцініть імовірність того, що $|X - M(X)| < 0{,}3$, якщо $D(X) = 0{,}0025$.

10.3. Відомо, що $P\{|X - M(X)| < \varepsilon\} \geq 0{,}9$ і $D(X) = 0{,}001$. Скориставшись нерівністю Чебишова, знайти ε.

10.4. Випадкову величину X задано законом розподілу (табл. 10.1):

Таблиця 10.1

Закон розподілу випадкової величини X

X	1	3	5	6
$P(X = x_i)$	0,1	0,3	0,4	0,2

Використовуючи нерівність Чебишова, оцініть імовірність того, що:

абсолютна різниця між значенням випадкової величини X і її математичним сподіванням менша за 5;

абсолютна різниця між значенням випадкової величини X і її математичним сподіванням не менша ніж 2.

10.5. Випадкову величину X задано законом розподілу (табл. 10.2):

Таблиця 10.2

Закон розподілу випадкової величини X

X	3	5
$P(X = x_i)$	0,4	0,6

Використовуючи нерівність Чебишова, оцініть імовірність того, що абсолютна різниця між значенням випадкової величини X і її математичним сподіванням менша за 2.

10.6. Імовірність того, що під час одиничного звернення банкомат спрацьовує правильно, дорівнює 0,95.

Оцініть імовірність того, що:

у разі 2 500 звернень відносна частота випадків правильної роботи банкомата відхилиться (за абсолютною величиною) від її ймовірності менше ніж на 0,02;

у разі 2 000 звернень кількість випадків правильної роботи банкомата перебуває в межах від 1 860 до 1 940.

10.7. Оцініть імовірність того, що за 1 500 незалежних підкидань грального кубика одне очко випаде менше ніж 1 000 разів.

10.8. Довжина виробів, що виробляють у цеху, – це випадкова величина, середнє значення якої дорівнює 90 см. Дисперсія цієї величини – 0,02. Оцініть імовірність того, що: а) відхилення довжини виготовленого виробу від його середнього значення за абсолютною величиною буде не більшим ніж 0,4; б) довжину виробу виражено числом, що міститься в межах [89,4; 90,6].

10.9. Імовірність народження дівчинки приблизно дорівнює 0,485. Оцініть імовірність того, що кількість дівчат серед 3 000 новонароджених буде відрізнятися від математичного сподівання цієї кількості за абсолютною величиною менше ніж на 55 дівчат.

10.10. Дисперсія кожної із 2 500 незалежних випадкових величин не перевищує 5. Оцініть імовірність того, що відхилення середнього арифметичного цих випадкових величин від середнього арифметичного їхніх математичних сподівань не перевищить 0,4.

10.11. Відомо, що цех виготовляє 80 % продукції вищого ґатунку. Оцініть імовірність того, що відносна частота виробів вищого ґатунку серед 20 000 виготовлених буде відрізнятися від імовірності виготовлення виробу вищого ґатунку не більше ніж на 0,02.

10.12. Імовірність настання події *A* в кожному випробуванні $p = \frac{1}{3}$. Яку найменшу кількість випробувань потрібно виконати, щоб з імовірністю не меншою за 0,99 можна було стверджувати, що частість настання події *A* відхилялася за абсолютною величиною від її ймовірності не більш ніж на 0,01? Для розв'язування скористайтеся: а) нерівністю Чебишова; б) інтегральною теоремою Лапласа.

10.5. Тестові завдання

10.1. Сенс закону великих чисел полягає

А у визначенні числових характеристик випадкових величин у разі великої кількості спостережуваних даних

Б у поведінці числових характеристик і законів розподілу спостережуваних значень випадкових величин

В у визначенні області застосовності нормального закону розподілу випадкових величин під час складання великої кількості випадкових величин

Г у поведінці числових характеристик і законів розподілу випадкових величин у разі збільшення кількості спостережень і дослідів

10.2. У запису теореми Чебишова вставити пропущений символ:

$$\lim_{n\to\infty} P\left(\left|\frac{X_1+X_2+\ldots+X_n}{n}-\frac{M(X_1)+M(X_2)+\ldots+M(X_n)}{n}\le\varepsilon\right|\right)=?$$

А 0 **Б** 1 **В** $\frac{1}{n}$ **Г** відповідь відсутня

10.3. Теорему Чебишова можна розглядати як

А достатню умову граничної збіжності середніх значень незалежних випадкових величин до середніх значень їхніх математичних сподівань

Б необхідну умову граничної збіжності середніх значень незалежних випадкових величин до середніх значень їхніх математичних сподівань

В необхідну і достатню умову граничної збіжності середніх значень незалежних випадкових величин до середніх значень їхніх математичних сподівань

Г відповідь відсутня

10.4. У запису теореми Хінчина вставити пропущений символ:

$$\lim_{n\to\infty} P\left(\left|\frac{X_1+X_2+\ldots+X_n}{n}-a\right|<\varepsilon\right)=?$$

А 0 **Б** 1 **В** $\frac{1}{n}$ **Г** відповідь відсутня

10.5. Уставити пропущений символ у нерівності Маркова:

$$P(x>\varepsilon)\le\frac{?}{\varepsilon}$$

А $D(X)$ **Б** 1 **В** $M(X)$ **Г** відповідь відсутня

10.6. Уставити пропущений символ у нерівності Чебишова:

$$P\left(\left|X - M(X)\right| \le \varepsilon\right) \ge 1 - \frac{?}{\varepsilon^2}$$

А $D(X)$ **Б** 1 **В** $M(X)$ **Г** відповідь відсутня

10.7. Для того щоб оцінити ймовірність відхилення випадкової величини від її математичного сподівання, необхідно використати

А теорему Бернуллі

Б теорему Чебишова

В центральну граничну теорему

Г нерівність Чебишова

10.8. Для того щоб оцінити за нерівністю Чебишова відхилення на ε випадкової величини від математичного сподівання, потрібно знати значення

А випадкової величини

Б математичного сподівання

В дисперсії

Г відповідь відсутня

10.6. Запитання для самоперевірки

10.1. Які теореми теорії ймовірностей називають граничними?

10.2. За яким принципом граничні теореми розподіляють на два класи?

10.3. Сформулюйте загальний сенс закону великих чисел.

10.4. Дайте означення закону великих чисел у формі Чебишова.

10.5. Наведіть теорему Бернуллі та розкрийте її зв'язок із статистичним означенням імовірності.

10.6. Поясніть сенс центральної граничної теореми та розкрийте її значення для математичної статистики.

10.7. Висновки за темою

Граничні теореми не тільки доводять сталість середніх результатів великої кількості випробувань, але й розкривають причини цієї сталості, саме тому вони мають виняткове значення для математичної статистики.

Граничні теореми розподіляють на два класи.

До одного класу належать теореми, які стверджують, що за ймовірністю певні характеристики випадкових величин збігаються до сталих значень, і ці сталі значення визначають за положеннями теорії ймовірностей. Закон великих чисел математично описує сталість середніх значень масових випадкових явищ. Закон великих чисел складається з декількох теорем, у кожній із яких за певних умов затверджено факт збіжності середніх характеристик під час проведення великої кількості випробувань до певних невипадкових, постійних величин.

До другого класу граничних теорем належать теореми, які розглядають граничні закони розподілу випадкової величини.

11. Первинне опрацювання статистичних даних

11.1. Мета та компетентності

Метою вивчення теми є ознайомлення із завданнями математичної статистики, особливостями вибіркового методу та формування компетентностей щодо можливостей його застосування до дослідження характеристик об'єктів у генеральній сукупності.

Компетентності, що формують під час вивчення теми:

знання основних методів відбору об'єктів генеральної сукупності у вибіркову сукупність;

розуміння впливу методів формування вибіркової сукупності на результати статистичних досліджень;

уміння здійснювати вибір методу формування вибіркової сукупності, відповідно до завдань дослідження та гіпотетичних властивостей досліджуваної випадкової величини.

11.2. Основні теоретичні відомості

Математична статистика розглядає такі групи завдань:

розроблення й обґрунтування способів відбору об'єктів для формування вибіркової сукупності та визначення обсягу вибіркової сукупності;

статистичне оцінювання числових характеристик випадкових величин та параметрів їхнього розподілу в генеральній сукупності за результатами дослідження вибірки;

перевірку статистичних гіпотез щодо значущості числових характеристик випадкової величини довільної вимірності, а також параметрів її розподілу та закону розподілу за результатами дослідження вибірки.

Словом "сукупність" у статистиці називають множину об'єктів, із яких здобувають вибірку. Усю сукупність, що вивчають, називають **генеральною сукупністю**. Генеральну сукупність можна вивчати шляхом суцільного вивчення всіх об'єктів або шляхом спостереження за частиною об'єктів.

Частину об'єктів, яку дістають із генеральної сукупності, називають **вибіркою** або **вибірковою сукупністю**.

Повну кількість об'єктів генеральної сукупності чи вибіркової сукупності називають їхнім **обсягом**. Обсяг генеральної сукупності позначають N, а обсяг вибіркової сукупності – n. Результати досліджень будь-якої ознаки генеральної сукупності будуть більш достовірними, якщо вибірку створити випадково.

Вибірку називають **випадковою**, якщо з генеральної сукупності елемент брати навмання і кожен із них може потрапити до неї з однаковою ймовірністю. На практиці випадкову вибірку можна дістати так: усі елементи нумерують від 1 до N, після чого вибирають послідовність n випадкових чисел ($1 \le n \le N$). Вибірку також можна вибрати і використавши датчик випадкових чисел.

Відбір об'єктів може бути повторним або безповторним. **Повторним** називають відбір, коли відібраний об'єкт повертають у генеральну сукупність (до відбору наступного об'єкта). **Безповторним** називають відбір, коли відібраний об'єкт не повертають до генеральної сукупності. Відбір об'єктів відбувається випадковим чином.

Якщо випадкова вибірка достатньо повно характеризує генеральну сукупність, то її називають **репрезентативною**.

Нехай є вибірка з генеральної сукупності $x_1, x_2, ..., x_n$. Якщо записати цю вибірку у вигляді зростаючої або спадної послідовностей, то буде дискретний **варіаційний ряд**.

Деякі значення дискретної випадкової величини можуть повторюватися, тоді варіаційний ряд записують у вигляді таблиці (табл. 11.1), де вказують можливі значення (x_i) випадкової величини X та їхні частоти (m_i), $\sum_{i=1}^{k} m_i = n$ – обсяг вибірки.

Таблиця 11.1

Дискретний варіаційний ряд

x_i	x_1	x_2	…	x_k
m_i	m_1	m_2	…	m_k

Числові значення, яких набуває досліджувана ознака, називають **варіантами**. **Частотою** називають кількість появи окремих значень випадкової величини (варіант).

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

Відносною частотою w_i називають відношення частоти появи ознаки до загального обсягу вибірки

$$w_i = \frac{m_i}{n}, \ \sum_{i=1}^{k} w_i = 1.$$

Відносна частота характеризує частку сукупності членів з однаковими значеннями ознаки.

Якщо обсяг вибірки великий $(n \geq 30)$, користуватися дискретним варіаційним рядом незручно. У цьому разі, а також у разі, якщо дані визначені в результаті вимірювання неперервної випадкової величини, за вибіркою складають **інтервальний варіаційний ряд**. Для його побудови весь інтервал розподілу ознаки розподіляють на k однакових частин довжиною h. Кількість k таких інтервалів, на які розподіляють вибіркову сукупність обсягом n за первинного групування, оцінюють за формулою Стерджеса:

$$k = 1 + 3{,}322 \cdot \lg n,$$

де кількість інтервалів округлюють до найближчого цілого числа. Зазвичай це число перебуває в межах від 8 до 12.

Для того щоб розподілити за інтервалами результати вимірювань неперервної випадкової величини, що становлять вибіркову сукупність, насамперед, необхідно визначити найменше x_{min} та найбільше x_{max} значення варіант.

Різницю між ними $R = x_{max} - x_{min}$ називають **розмахом** (її можна розглядати як додаткову числову характеристику емпіричного розподілу випадкової величини).

Поділивши розмах на кількість інтервалів, визначають довжину інтервалу, тобто **крок** $h = \frac{R}{k}$. Крок визначає відстань між верхньою і нижньою межами кожного інтервалу: $h = x_{i+1} - x_i$. Щоб оптимізувати обчислення, значення кроку доцільно округлити таким чином, щоб межі інтервалів і їхні середини були зручними для обчислення. Оскільки це супроводжується збільшенням розмаху, зміна значень x_1 та x_{k+1} не має перевищувати $0,5 \cdot h$. Отже, межі граничних інтервалів зсунуться: $x_1 < x_{min}$ та $x_{k+1} > x_{max}$. Для кожного інтервалу частоту m_i визначають як кількість значень випадкової величини, що потрапили до цього інтервалу. Кожний з інтервалів є напіввідкритим, тобто відкритим із боку верхньої межі.

Поставивши кожному з інтервалів у відповідність частоту m_i потраплянь у заданий інтервал або відносну частоту $w_i = \frac{m_i}{n}$, визначають статистичний розподіл неперервної випадкової величини, який задано у формі таблиці (табл. 11.2).

Таблиця 11.2

Інтервальний варіаційний ряд

$[x_i; x_{i+1})$	$[x_1; x_2)$	$[x_2; x_3)$	…	$[x_k; x_{k+1})$
m_i (w_i)	m_1 (w_1)	m_2 (w_2)	…	m_k (w_k)

Якщо варіаційний ряд задано у вигляді табл. 11.1 і 11.2, то його називають **статистичним рядом розподілу**.

Статистичний ряд розподілу випадкової величини X, визначений за емпіричними даними, називають також **емпіричним законом розподілу**. Такий ряд можна зобразити графічно. Для цього на осі абсцис наносять варіанти x_i, а на осі ординат – частоти m_i (або відносні частоти w_i), тобто є певні точки $M_i(x_i, m_i)$. З'єднавши точки M_i відповідних варіант і частот, визначають ламану лінію, яку називають **полігоном розподілу**.

Емпіричною функцією розподілу вибірки називають функцію $F^*(x)$, яка для будь-якого значення x визначає відносну частоту події, що задовольняє умові $X < x$, тобто випадкова величина набуде значення, меншого за x:

$$F^*(x) = \frac{m_x}{n}, \tag{11.1}$$

де m_x – сума частот варіант для значень аргументу, менших за x;

n – обсяг вибірки.

Емпіричну функцію розподілу визначають шляхом послідовного додавання відносних частот варіант, менших за x. Графіком емпіричної функції розподілу є **кумулята**, або графік накопичених відносних частот. Цей графік – східчаста фігура, яка має точки розриву, якщо значення абсцис дорівнюють числовим значенням, яких набуває випадкова величина.

Для генеральної сукупності функцію розподілу позначають $F(x)$. Різниця між емпіричною та теоретичною функціями полягає в тому, що *теоретична функція* $F(x)$ визначає ймовірність події $X < x$, а *емпірична функція* $F^*(x)$ визначає відносну частоту цієї події.

Згідно з теоремою Бернуллі, можна сказати, що в разі великої кількості n значення $F^*(x)$ і $F(x)$ мало відрізняються одне від одного в тому сенсі, що $\lim\limits_{n\to\infty} P\left(\left|F(x) - F^*(x)\right| < \varepsilon\right) = 1 \ (\varepsilon > 0)$.

Отже, емпіричну функцію розподілу вибірки може бути використано для оцінювання теоретичної функції розподілу генеральної сукупності.

Функція $F^*(x)$ має такі самі властивості, що й функція $F(x)$, тобто:

значення емпіричної функції належить відрізку $[0;1]$;

$F^*(x)$ – неспадна функція: $F^*(x_2) > F^*(x_1)$, якщо $x_2 > x_1$;

якщо x_1 – найменша варіанта, то $F^*(x) = 0$ за $x \le x_1$, якщо x_k – найбільша варіанта, то $F^*(x) = 1$ за $x > x_k$.

У разі інтервальних варіаційних рядів, визначених для неперервної випадкової величини, графічним зображенням є **гістограма**. Розрізняють гістограму частот і гістограму відносних частот.

Гістограмою частот називають східчасту фігуру, яка складається із прямокутників, площа кожного з яких дорівнює частоті потрапляння випадкової величини у відповідний інтервал. Основами прямокутників є інтервали, довжиною h, а висоти дорівнюють щільності частоти $\frac{m_i}{h}$. Для побудови гістограми на осі абсцис відкладають часткові інтервали, а над ними проводять відрізки, які паралельні осі абсцис на відстані $\frac{m_i}{h}$. Площа i-го часткового прямокутника дорівнює $\frac{m_i}{h} \cdot h = m_i$ – сумі частот варіант i-го інтервалу.

Таким чином, площа гістограми дорівнює сумі всіх частот, тобто обсягу вибірки.

Гістограмою відносних частот називають східчасту фігуру, яка складається із прямокутників, основою яких є інтервали довжиною h, а висоти дорівнюють відношенню $\frac{w_i}{h}$. Для побудови гістограм відносних частот на осі абсцис відкладають часткові інтервали, а над ними проводять відрізки, паралельні осі абсцис на відстані $\frac{w_i}{h}$ від неї. Площа гістограми відносних частот дорівнює сумі всіх відносних частот, тобто одиниці.

11.3. Приклади розв'язування задач

Приклад 11.1. Протягом місяця магазин із продажу канцелярських товарів вів облік щоденної кількості покупців, які користуються платіжними картками. Результати спостережень наведено в табл. 11.3.

Таблиця 11.3

Кількість користувачів платіжними картками

12	15	10	8	12	14	11	12	13	12
11	14	14	9	7	15	16	10	11	13
8	11	15	17	10	13	9	12	12	9

За вибірковою сукупністю побудуйте варіаційний ряд, а також полігон частот.

Розв'язання. Оскільки серед результатів 30 вимірювань, які містяться у вибірковій сукупності, є тільки 11 різних варіант, розташуйте їх у порядку зростання та підрахуйте, скільки разів кожну з них зустрічали у вибірковій сукупності.

Результати надайте у вигляді таблиці (табл. 11.4).

Таблиця 11.4

Дискретний ряд

$X = x_i$	7	8	9	10	11	12	13	14	15	16	17
m_i	1	2	3	3	4	6	3	3	3	1	1

Перевірте, що $\sum_{i=1}^{11} m_i = 30$, та побудуйте полігон частот (рис. 11.1).

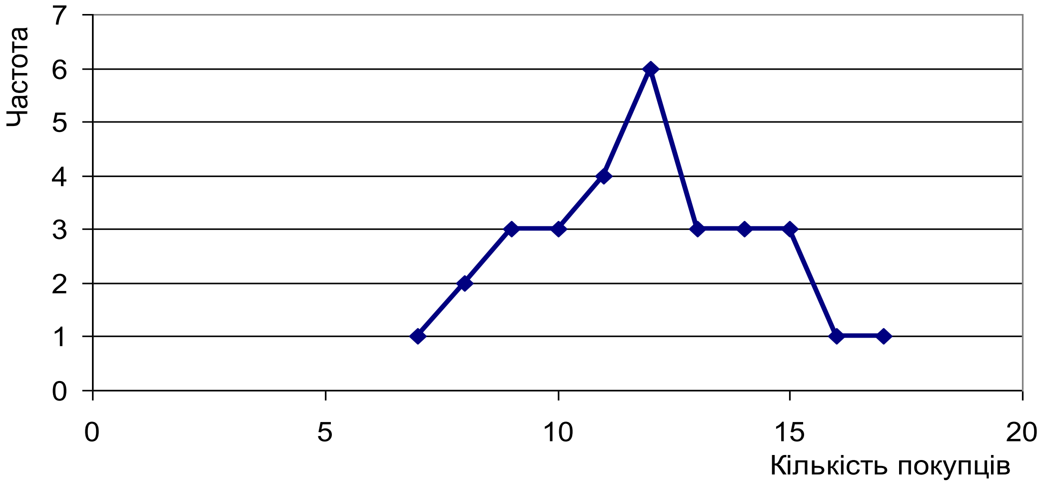


Рис. 11.1. **Полігон частот за даними прикладу 11.1**

Приклад 11.2. Задано вибірку рядом розподілу (табл. 11.5).

Таблиця 11.5

Ряд розподілу

x_i	1	6	11	16
m_i	20	10	40	30

Знайдіть та побудуйте емпіричну функцію розподілу $F^*(x)$.

Розв'язання. Знайдіть обсяг вибірки:

$$n=\sum_{i=1}^{4} m_i = 20+10+40+30=100.$$

Найменше значення величини X : $x_1=1$, тому $F^*(1)=0$, якщо $X\le 1$. Значення $X<6$ спостерігали 20 разів, отже, $F^*(x)=\frac{20}{100}=0{,}2$, якщо $1<x\le 6$. Значення $X<11$, тобто 1 і 6, спостерігали $20+10=30$ разів, отже, $F^*(x)=\frac{30}{100}=0{,}3$, якщо $6<x\le 11$. Значення $X<16$, тобто 1, 6 і 11, спостерігали $20+10+40=70$ разів, отже, $F^*(x)=\frac{70}{100}=0{,}7$, якщо $11<x\le 16$. Найбільшим значенням величини X є $x_4=16$, отже, $F^*(x)=1$, якщо $X>16$.

Таким чином, можна записати емпіричну функцію розподілу вибірки у вигляді:

$$F^*(x)=\begin{cases}0, & x\le 1;\\ 0{,}2, & 1<x\le 6;\\ 0{,}3, & 6<x\le 11;\\ 0{,}7, & 11<x\le 16;\\ 1, & x>16.\end{cases}$$

Графік цієї функції зображено на рис. 11.2.

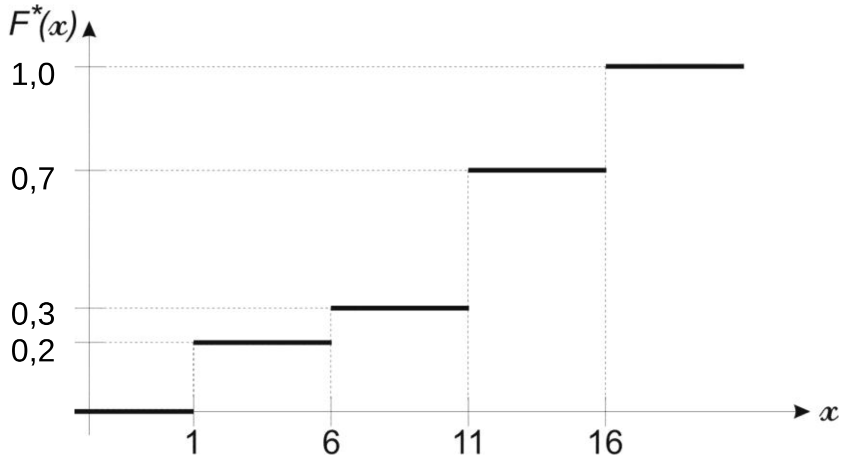


Рис. 11.2. **Емпірична функція розподілу**

Приклад 11.3. Проведено 40 спостережень швидкості автобусів на певній дільниці дороги (км/год), результати яких наведено в табл. 11.6.

Таблиця 11.6

Дані спостережень швидкості автобусів

41	41	29	25	41	43	42	34	41	30
33	48	50	36	35	46	28	46	50	41
50	27	43	53	48	47	34	35	29	42
30	35	38	41	36	38	45	59	44	43

Складіть інтервальний ряд статистичного розподілу.

Розв'язання. Оскільки обсяг вибіркової сукупності великий ($n = 40$), виконайте групування варіант заданої вибірки за інтервалами.

Отже, $x_{\min} = 25,\quad x_{\max} = 59,\quad R = x_{\max} - x_{\min} = 34$.

Для визначення кількості інтервалів застосуйте формулу (11.1):

$$k = 1 + 3{,}322 \cdot \lg 40 \approx 6.$$

Довжину інтервалу обчисліть за такою формулою:

$$h = \frac{x_{\max} - x_{\min}}{k}.$$

Тоді буде:

$$h = \frac{59 - 25}{1 + 3{,}322 \cdot \lg 40} = 5{,}52.$$

Оскільки початок першого інтервалу можна зсунути на півкроку в бік менших значень, то виберіть крок $h = 5$, а початком першого інтервалу вважайте $x_1 = x_{\min} - 0{,}5h$, тобто $x_1 = 25 - 2{,}5$. Межі останнього інтервалу повинні задовольняти умові: $x_k < x_{\max} < x_{k+1}$.

Підрахуйте частоти m_i улучення значень випадкової величини в кожний інтервал і результати запишіть у таблицю (табл. 11.7).

Таблиця 11.7

Інтервальний варіаційний ряд

$[x_i; x_{i+1})$	[22,5; 27,5)	[27,5; 32,5)	[32,5; 37,5)	[37,5; 42,5)	[42,5; 47,5)	[47,5; 52,5)	[52,5; 57,5)	[57,5; 62,5)	Сума
m_i	2	5	8	11	7	5	1	1	n = 40

Приклад 11.4. Для інтервального варіаційного ряду (табл. 11.8) з обсягом вибірки $n = 20$ побудуйте гістограму частот.

Таблиця 11.8

Вихідний інтервальний варіаційний ряд

$(x_{i-1} \div x_i]$	9 – 10	10 – 11	11 – 12	12 – 13	13 – 14	14 – 15	15 – 16
m_i	2	6	4	3	3	1	1

Розв'язання. Для побудови гістограми частот (рис. 11.3) на осі абсцис відкладіть часткові проміжки та на них побудуйте прямокутники, висоти яких дорівнюють $\frac{m_i}{h}$ ($h = 1$ – довжини інтервалів).

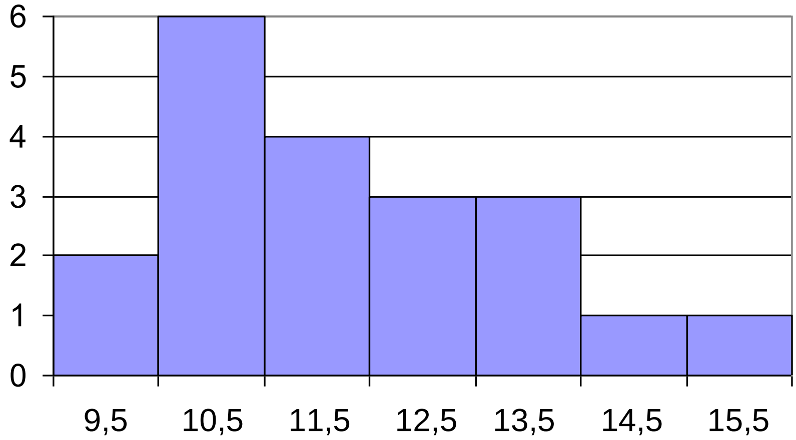


Рис. 11.3. **Гістограма частот**

Площа побудованих прямокутників дорівнює обсягу вибірки $n = 20$.

11.4. Вправи для самостійної роботи

11.1. Обстеження терміну, протягом якого здійснюють обслуговування одного клієнта на терміналі, дало такі значення (с):

34, 34, 34, 33, 35, 32, 34, 32, 37, 36, 34, 35, 35, 32, 33, 35, 34, 36, 38, 33, 35, 34, 36, 34, 33, 35, 32, 37, 32, 34, 36, 32, 34, 33, 35, 35, 33, 34, 35, 38, 36, 34, 35, 33, 37, 35, 34, 37, 33, 35.

Необхідно:

1) скласти дискретний варіаційний ряд, який характеризує розподіл терміну обслуговування клієнтів;

2) скласти інтервальний ряд, узявши за довжину інтервалу величину 2 с.

11.2. У цеху встановлено 5 верстатів. Протягом 25 днів реєстрували кількість верстатів, які не працювали. Здобуто такі значення:

0, 1, 2, 1, 1, 2, 3, 2, 1, 4, 2, 0, 0, 2, 2, 3, 3, 1, 0, 1, 2, 1, 3, 5, 0.

Побудуйте статистичну функцію розподілу.

11.3. Для дискретного варіаційного ряду (табл. 11.9):

визначте відносні частоти (частості) та накопичені частості;

запишіть функцію розподілу;

побудуйте полігон відносних частот та кумуляту.

Таблиця 11.9

Дискретний варіаційний ряд до вправи 11.3

x_i	6	7	8	9	10	11	12
m_i	2	3	4	5	4	1	1

11.4. Кількість балів (x_i), що набрали абітурієнти однієї зі шкіл на зовнішньому незалежному тестуванні з математики, наведено в табл. 11.10:

Таблиця 11.10

Дискретний варіаційний ряд до вправи 11.4

x_i	115	125	135	145	155	165	175	185	195
m_i	2	10	15	20	25	15	8	2	3

Побудуйте полігон частот і кумуляту.

11.5. За інтервальним розподілом (табл. 11.11) побудуйте гістограму частот та графік емпіричної функції розподілу. Сформулюйте гіпотезу про закон розподілу сукупності.

Таблиця 11.11

Інтервальний варіаційний ряд до вправи 11.5

$[x_i; x_{i+1})$	[2; 7)	[7; 12)	[12; 17)	[17; 22)	[22; 27)
m_i	5	10	25	6	4

11.6. Розподіл робітників підприємства за часом (хв), який вони витрачають на оброблення однієї деталі, наведено у вигляді інтервального статистичного ряду (табл. 11.12).

Таблиця 11.12

Інтервальний варіаційний ряд до вправи 11.6

$[x_i; x_{i+1})$	[2; 4)	[4; 6)	[6; 8)	[8; 10)	[10; 12)
Кількість робітників	42	73	154	205	26

Побудуйте гістограму відносних частот та кумуляту.

11.5. Тестові завдання

11.1. Задано вибірку: 2, 4, 5, 3, 5, 7, 5, 6, 8, 9, 3, 7, 6, 7. Її обсяг

А 7 **Б** 9 **В** 14 **Г** 15

11.2. Для того щоб задану в завданні 11.1 вибірку перетворити на варіаційний ряд, потрібно

А виписати один раз ті варіанти, які зустрічають у виборці

Б записати всі варіанти в порядку зростання

В нічого не змінювати

Г відповідь відсутня

11.3. У результаті випробувань випадкова величина X набула таких значень:

15, 4, 17, 11, 16, 17, 7, 5, 6, 23, 19, 22, 20, 7, 9, 13, 19, 18, 21, 25, 1, 15, 11, 5, 21.

Інтервальний ряд для 5 однакових інтервалів $(x_i; x_{i+1}]$.

А

Інтервали	(1; 4]	(4; 8]	(8; 13]	(13; 19]	(19; 25]
частоти	2	7	3	6	7

Б

Інтервали	(1; 4]	(4; 8]	(8; 13]	(13; 19]	(19; 25]
частоти	4	5	4	7	5

В

Інтервали	(0; 5]	(5; 10]	(10; 15]	(15; 20]	(20; 25]
частоти	4	4	5	7	5

Г

Інтервали	(1; 6]	(6; 10]	(10; 15]	(15; 19]	(19; 25]
частоти	3	8	4	5	5

Д

Інтервали	(0; 5]	(5; 10]	(10; 15]	(15; 20]	(20; 25]
частоти	2	5	6	7	3

11.4. Гістограму будують для

А дискретного розподілу

Б інтервального розподілу

11.5. Ламаною може бути графік

А кумуляти

Б полігону

В гістограми

11.6. Частотою називають

А кількість появи окремих значень випадкової величини

Б повну кількість об'єктів вибіркової сукупності

В різницю між найменшим x_{min} та найбільшим x_{max} значеннями варіант

Г відповідь відсутня

11.7. Сума частот усіх варіант вибірки дорівнює

А одиниці

Б обсягу вибірки

В кількості різних варіант вибірки

Г відповідь відсутня

11.8. Відносна частота – це

А відношення частоти до максимальної частоти

Б відношення частоти до обсягу вибірки

В відношення максимальної частоти до обсягу вибірки

Г відповідь відсутня

11.9. Сума відносних частот усіх варіант вибірки дорівнює

А одиниці

Б обсягу вибірки

В кількості різних варіант вибірки

Г відповідь відсутня

11.10. Сума площ усіх прямокутників, із яких складається гістограма частот статистичного розподілу, дорівнює

А одиниці

Б обсягу вибірки

В кількості різних варіант вибірки

Г відповідь відсутня

11.6. Запитання для самоперевірки

11.1. Сформулюйте, у чому полягають завдання математичної статистики.

11.2. Поясніть, у чому полягає сенс вибіркового методу та чому виникає потреба у його застосуванні.

11.3. Яким вимогам має відповідати вибіркова сукупність?

11.4. Які способи відбору застосовують для утворення вибірки?

11.5. Дайте означення генеральної та вибіркової сукупностей.

11.6. Що таке "варіанта"? Як утворюють варіаційний ряд?

11.7. Які існують способи задавання емпіричного закону розподілу дискретної випадкової величини?

11.8. Які властивості має емпірична функція розподілу?

11.9. Що таке "інтервальний варіаційний ряд"? Які принципи його формування?

11.10. Які існують способи задавання емпіричного закону розподілу неперервної випадкової величини?

11.11. Яким вимогам мають відповідати гістограма частот та гістограма відносних частот?

11.12. Як побудувати полігон для неперервної випадкової величини?

11.13. Які особливості розподілу випадкової величини можна спостерігати за гістограмою?

11.7. Висновки за темою

Завдання математичної статистики полягає у створенні методів збирання та оброблення статистичних даних для здобуття наукових і практичних висновків. Предметом математичної статистики є вивчення випадкових величин за результатами спостережень, дослідів, повторних випробувань.

Під час проведення статистичних досліджень застосовують **вибірковий метод**, який полягає в тому, що із загальної кількості об'єктів, поєднаних за певною ознакою, у сукупність, яка має назву генеральної сукупності, вибирають меншу за обсягом сукупність, для якої і здійснюють обстеження кількісних чи якісних ознак об'єктів. Остання має назву вибіркової сукупності.

Статистичний розподіл дискретної випадкової величини можна подати графічно за допомогою полігону частот або полігону відносних частот. Статистичний розподіл неперервної випадкової величини можна подати графічно за допомогою гістограми.

12. Статистичне оцінювання параметрів розподілу

12.1. Мета та компетентності

Метою вивчення теми є ознайомлення з методами визначення точкових та інтервальних оцінок основних числових характеристик одновимірної випадкової величини за вибірковими даними.

Компетентності, що формують під час вивчення теми:

знання методів визначення числових характеристик емпіричного закону розподілу;

розуміння точності вибіркових оцінок та особливостей обчислення довірчого інтервалу для основних числових характеристик;

уміння здійснювати ідентифікацію кількісних характеристик економічних процесів за допомогою вибіркового методу.

12.2. Основні теоретичні відомості

Одним із завдань математичної статистики є статистичне оцінювання числових характеристик випадкових величин та параметрів їхнього розподілу в генеральній сукупності за результатами дослідження вибірки. За результатами вибірки можна знайти не значення певного параметра θ , що характеризує випадкову величину в генеральній сукупності, а лише його **статистичну оцінку** θ^*, яка є випадковою величиною.

Для того щоб статистична оцінка достатньо точно відображала значення параметра, для якого здійснюють оцінювання, вона має відповідати таким вимогам, як: незсунутість, ефективність та ґрунтовність (спроможність).

Статистична оцінка θ^* є **незсунутою оцінкою** параметра θ, якщо математичне сподівання цієї оцінки за будь-якого обсягу вибіркової сукупності дорівнює параметру, що оцінюють: $M\left(\theta^*\right) = \theta$. Якщо математичне сподівання статистичної оцінки не дорівнює параметру, який оцінюють, то таку статистичну оцінку називають **зсунутою**.

Ефективною називають статистичну оцінку, якій для заданого обсягу вибірки відповідає найменша з можливих дисперсій. Цю вимогу можна зрозуміти таким чином: недостатньо дати точкову оцінку деякого параметра, необхідно також, щоб інтервал, до якого буде належати цей параметр генеральної сукупності, був якомога меншим, а надійність, із якою гарантується потрапляння параметра до цього інтервалу, якомога більшою.

Статистична оцінка є **ґрунтовною** (**спроможною**), якщо в разі необмеженого зростання обсягу вибірки вона за ймовірністю наближається до параметра, для якого це оцінювання здійснюють:

$$\lim_{n \to \infty} P\left(\left|\theta^* - \theta\right| < \varepsilon\right) = 1,$$

де ε – заздалегідь задане, як завгодно мале додатне число $\varepsilon > 0$.

Для будь-якої одновимірної випадкової величини, незалежно від її природи, основними числовими характеристиками статистичного оцінювання є її математичне сподівання і середнє квадратичне відхилення. У зв'язку із цим розглядають точкові та інтервальні статистичні оцінки. Якщо як статистичну оцінку розглядають тільки одне число, то таку оцінку називають **точковою**.

У деяких випадках доцільно можливі значення параметра, що підлягає оцінюванню, характеризувати двома числами – початком і кінцем інтервалу, до якого цей параметр θ буде належати з певною надійністю (довірчою ймовірністю). Таку статистичну оцінку, що визначають двома числами, називають **інтервальною**.

За даними статистичного розподілу визначають основні точкові числові характеристики випадкової величини.

Вибіркова середня $\bar{x}$ визначає центр вибіркової сукупності. За незгрупованими даними вибіркову середню визначають як середнє арифметичне вибіркових даних:

$$\bar{x} = \frac{1}{n} \cdot \sum_{i=1}^{n} x_i \,. \qquad (12.1)$$

За варіаційним рядом вибіркову середню обчислюють як середнє виважене варіант, кожну з яких беруть з вагою, що відповідає її відносній частоті:

$$\bar{x} = \frac{1}{n} \cdot \sum_{i=1}^{k} x_i m_i \,. \qquad (12.2)$$

Вибіркова середня є незсунутою оцінкою математичного сподівання випадкової величини, оскільки виконується таке співвідношення:

$$M\left(\bar{x}\right) = M(X).$$

Вибіркову дисперсію D^* за варіаційним рядом визначають як середнє виважене квадратів відхилення значень випадкової величини від вибіркової середньої:

$$D^* = \frac{1}{n} \cdot \sum_{i=1}^{k} (x_i - \bar{x})^2 m_i \text{, або } D^* = \frac{1}{n} \cdot \sum_{i=1}^{k} x_i^2 m_i - (\bar{x})^2 \,. \qquad (12.3.)$$

Вибіркова дисперсія є зсунутою оцінкою дисперсії випадкової величини генеральній сукупності.

Ураховуючи це, для вибіркової дисперсії вводять поправку на зсув, яка дорівнює $\frac{n}{n-1}$.

Для визначення **виправленої дисперсії**, що позначають S_x^2, користуються такою формулою:

$$S_x^2 = \frac{n}{n-1} D^*. \qquad (12.4)$$

Середнє квадратичне відхилення вибірки σ^* визначають як корінь квадратний із вибіркової дисперсії: $\sigma^* = \sqrt{D^*}$. Воно характеризує розсіювання випадкової величини навколо центра вибіркової сукупності. На відміну від дисперсії, його вимірність збігається з вимірністю випадкової величини.

Величина $S_x = \sqrt{S_x^2}$ є **виправленим середнім квадратичним відхиленням**. Саме ця величина і є оцінкою середнього квадратичного відхилення теоретичного розподілу, яке характеризує розсіювання значень випадкової величини навколо центра сукупності.

Разом із основними є додаткові числові характеристики статистичного розподілу:

мода – це варіанта, якій відповідає найбільша частота;

медіана – це варіанта, що за кількістю варіант розподіляє навпіл статистичний ряд розподілу;

коефіцієнт варіації: $V^* = \frac{\sigma_в}{\overline{x}} \cdot 100\,\%$;

початкові моменти k *-го порядку:* $\nu_k = \overline{X^k} = \frac{\sum_i x_i^k m_i}{n}$;

центральні моменти k *-го порядку:* $\mu_k = \overline{\left(X - \overline{X}\right)^k} = \frac{\sum_i \left(x_i - \overline{x}\right)^k m_i}{n}$.

Інтервальні оцінки використовують у тих випадках, якщо точкові оцінки не достатньо точно відображають характеристики ознаки. Тоді ознаку, яку вивчають, покривають довірчим інтервалом.

Довірчим інтервалом називають інтервал, що покриває всі значення випадкової величини із заданою ймовірністю або із заданим рівнем значущості. Загальноприйнятими є два рівні довірчої ймовірності: 0,95 та 0,99. Разом із надійністю оцінки γ розглядають **рівень значущості** $\alpha = 1 - \gamma$. Він визначає ймовірність того, що за заданим рівнем надійності значення параметра θ, який оцінюють за вибірковою сукупністю, може вийти за межі довірчого інтервалу.

Стандартна форма надання **довірчого інтервалу для оцінювання математичного сподівання**:

$$\bar{x} - t_\gamma \frac{S_x}{\sqrt{n}} < a < \bar{x} + t_\gamma \frac{S_x}{\sqrt{n}}. \qquad (12.5)$$

Із довірчою ймовірністю $P = \gamma$ похибка ε, із якою математичне сподівання a оцінюють за вибірковою середньою $\bar{x}$, не перевищує значення:

$$\varepsilon = t_\gamma \cdot \frac{S_x}{\sqrt{n}}. \qquad (12.6)$$

Для невеликих за обсягом вибіркових сукупностей значення t_γ знаходять за спеціальною таблицею критичних точок розподілу Стьюдента (додаток Е), відповідно до рівня довірчої ймовірності γ і кількості ступенів свободи $n-1$.

Зі збільшенням $n \to \infty$ розподіл Стьюдента наближається до нормального розподілу.

Тому для великих вибіркових сукупностей величину t_γ визначають як аргумент інтегральної функції Лапласа (додаток В).

За формулою (12.6) знаходять граничний обсяг вибірки:

$$n \geq \left(\frac{t_\gamma \cdot S_x}{\Delta} \right)^2, \quad n \in N \qquad (12.7)$$

де Δ – напівширина довірчого інтервалу, до якого з довірчою ймовірністю не меншою за γ буде належати математичне сподівання випадкової величини, розподіленої за нормальним законом.

Довірчий інтервал для оцінювання середнього квадратичного відхилення

$$S_x(1-q) \le \sigma \le S_x(1+q). \qquad (12.8)$$

Значення величини q, яка залежить від заданої надійності та обсягу вибірки, знаходять у спеціальній таблиці (додаток Е).

12.3. Приклади розв'язування задач

Приклад 12.1. Вибіркову сукупність задано таблицею розподілу (табл. 12.1).

Таблиця 12.1

Вихідна вибіркова сукупність

x_i	3	3,5	4	4,5	5
m_i	5	4	9	4	8

Знайдіть виправлену вибіркову дисперсію.

Розв'язання. За формулою (12.3) обчислюють виправлену вибіркову дисперсію.

Обсяг вибірки: $n = 5 + 4 + 9 + 4 + 8 = 30$.

Знаходять вибіркову середню за формулою (12.2):

$$\bar{x} = \frac{3 \cdot 5 + 3{,}5 \cdot 4 + 4 \cdot 9 + 4{,}5 \cdot 4 + 5 \cdot 8}{30} = 4{,}1$$

та вибіркову дисперсію за формулою (12.3):

$$D = \frac{(3-4{,}1)^2 \cdot 5 + (3{,}5-4{,}1)^2 \cdot 4 + (4-4{,}1)^2 \cdot 9 + (4{,}5-4{,}1)^2 \cdot 4 + (5-4{,}1)^2 \cdot 8}{30},$$

$$D = 0{,}49.$$

Отже, виправлена вибіркова дисперсія: $S_x^2 = \frac{30}{29} \cdot 0{,}49 \approx 0{,}505$.

Приклад 12.2. Із метою перевірки стану внесків у банк, аудитор відібрав 100 рахунків. Середнє значення внеску $\bar{x} = 12{,}57$ (тис. грн), середнє квадратичне відхилення для всієї генеральної сукупності (усіх можливих внесків) $\sigma = 5$ (тис. грн). Знайдіть 95-відсотковий довірчий інтервал середнього внеску в банку.

Розв'язання. $P = 2\Phi(t) = 0{,}95$, тоді $\Phi(t) = 0{,}475$.

За таблицею значень функції $\Phi(t)$ (додаток В) визначають $t = 1{,}96$.

Тоді за формулою (12.5) буде:

$$12{,}57 - \frac{5 \cdot 1{,}96}{\sqrt{100}} \le a \le 12{,}57 + \frac{5 \cdot 1{,}96}{\sqrt{100}},\ 11{,}59 \le a \le 13{,}56.$$

Таким чином, із надійністю $95\,\%$ довірчий інтервал середнього внеску для цього банку становить $(11{,}59;\ 13{,}56)$ тис. грн.

Приклад 12.3. Нехай $\bar{x} = 2$, $\varepsilon = 0{,}2$, $\sigma = 1$. Визначте, якого обсягу треба взяти вибірку, щоб побудувати довірчий інтервал для математичного сподівання довжини не більш ніж 2ε із надійністю $99\,\%$.

Розв'язання. $P = 2\Phi(t) = 0{,}99$, тоді $\Phi(t) = 0{,}495$.

За таблицею значень функції $\Phi(t)$ (додаток В) визначають $t = 2{,}58$.

За формулою $\varepsilon = \dfrac{\sigma t}{\sqrt{n}}$ знаходять $n = \dfrac{\sigma^2 t^2}{\varepsilon^2}$, звідки

$$n = \frac{1 \cdot (2{,}58)^2}{(0{,}2)^2} = 166{,}41.$$

Отже, $n = 167$.

Приклад 12.4. Нехай $n = 50$, $s = 2{,}82$. Із надійністю $95\,\%$ побудуйте довірчий інтервал для $\sigma_{\text{ген.}}$.

Розв'язання. За табл. Е.2 додатка Е $q\,(0{,}95;\ 50) = 0{,}21$.

Тоді за формулою (12.8) буде:

$$2{,}82(1 - 0{,}21) \le \sigma_{\text{ген.}} \le 2{,}82(1 + 0{,}21),$$
$$2{,}228 \le \sigma_{\text{ген.}} \le 3{,}412.$$

Таким чином, із надійністю 95 % будь-яке число із цього інтервалу можна взяти за генеральне середньоквадратичне відхилення.

12.4. Вправи для самостійної роботи

12.1. Задано інтервальний ряд розподілу заробітної плати (тис. грн) робітників цеху (табл. 12.2).

Таблиця 12.2

Розподіл заробітної плати

x_i	4,00 – 4,40	4,40 – 4,80	4,80 – 5,20	5,20 – 5,60	5,60 – 6,00
m_i	10	15	20	15	10

Знайдіть середній заробіток, дисперсію та середнє квадратичне відхилення.

12.2. Виконано 15 вимірювань деякої величини, у результаті яких було знайдено значення вибіркової середньої $\bar{x} = 30$ і виправленої дисперсії $S^2 = 36$. Знайдіть межі довірчого інтервалу, до якого з надійністю 95 % буде потрапляти істинне значення цієї величини.

12.3. Випадковим способом було відібрано 20 абітурієнтів і визначено їхні бали на зовнішньому незалежному оцінюванні з математики. Визначені результати дослідження вибіркової сукупності наведено в табл. 12.3 у вигляді інтервального статистичного розподілу.

Таблиця 12.3

Розподіл балів оцінювання з математики

$[x_i, x_{i+1})$	[165,5; 170,5)	[170,5; 175,5)	[175,5; 180,5)	[180,5; 185,5)
m_i	4	6	8	2

Побудуйте довірчий інтервал, до якого з надійністю $\gamma = 99\,\%$ буде належати середній бал усіх абітурієнтів цього віку.

12.4. Виробник запчастин бажає оцінити середню вагу деталі, що випускає. За результатами обстеження вибіркової сукупності із 40 деталей знайдено середню вагу $\bar{x} = 250$ мг та середнє квадратичне відхилення $\sigma^* = 2$ мг. Знайдіть довірчий інтервал, до якого середня вага деталі для всієї партії буде належати з надійністю 95 %. Знайдіть необхідний обсяг вибірки, який забезпечить із надійністю 95 % точність оцінювання $\varepsilon = 1$ мг.

12.5. Знайдіть мінімальний обсяг вибірки, за якого точність оцінювання математичного сподівання за вибірковою середньою дорівнює значенню $\varepsilon = 0{,}2$ з надійністю 94 %, якщо відомо середнє квадратичне відхилення $\sigma = 1{,}5$ нормально розподіленої генеральної сукупності.

12.6. За заданим статистичним розподілом вибірки (табл. 12.4) знайдіть із надійністю $\gamma = 0{,}99$ інтервальну оцінку для середнього квадратичного відхилення нормально розподіленої випадкової величини X.

Таблиця 12.4

Статистичний розподіл вибірки

x_i	10	15	20	25	30
m_i	5	10	20	5	10

12.7. Визначте з надійністю $\gamma = 95\,\%$ довірчий інтервал для генерального середнього квадратичного відхилення нормально розподіленої випадкової величини X, якщо для вибіркової сукупності обсягом $n = 20$ відомо виправлене середнє квадратичне відхилення $S = 2{,}37$.

12.5. Тестові завдання

12.1. До оцінок генеральної сукупності висувають такі вимоги

А оцінка має бути стаціонарною, ергодичною й ефективною

Б оцінка має бути спроможною й ефективною

В оцінка має бути спроможною, ефективною і незміщеною

Г оцінка має бути незміщеною, стаціонарною й ефективною

12.2. Статистична оцінка θ^* є незсунутою оцінкою параметра θ, якщо

А для заданого обсягу вибірки їй відповідає найменша з можливих дисперсій

Б $M\left(\theta^*\right) = \theta$

В $\lim\limits_{n \to \infty} P\left(\left|\theta^* - \theta\right| < \varepsilon\right) = 1,\ \varepsilon > 0$

Г її визначають двома числами

12.3. Статистична оцінка θ^* є спроможною оцінкою параметра θ, якщо

А для заданого обсягу вибірки їй відповідає найменша з можливих дисперсій

Б $M\left(\theta^*\right) = \theta$

В $\lim\limits_{n \to \infty} P\left(\left|\theta^* - \theta\right| < \varepsilon\right) = 1,\ \varepsilon > 0$

Г її визначають двома числами

12.4. Статистична оцінка θ^* є ефективною оцінкою параметра θ, якщо

А для заданого обсягу вибірки їй відповідає найменша з можливих дисперсій

Б $M\left(\theta^*\right) = \theta$

В $\lim\limits_{n \to \infty} P\left(\left|\theta^* - \theta\right| < \varepsilon\right) = 1,\ \varepsilon > 0$

Г її визначають двома числами

12.5. Вибіркову середню за незгрупованими даними визначають як

А $\bar{x} = \frac{1}{n} \cdot \sum\limits_{i=1}^{k} x_i m_i$

Б $\bar{x} = \frac{1}{n} \cdot \sum\limits_{i=1}^{n} x_i$

В $\bar{x} = \sum\limits_{i=1}^{k} x_i m_i$

Г $\bar{x} = \sum\limits_{i=1}^{k} (x_i - x_j)^2 m_i$

12.6. Вибіркова середня є

А незміщеною оцінкою генерального середнього

Б зміщеною оцінкою генерального середнього

В відповідь відсутня

12.7. Вибіркову дисперсію визначають як

А $D^* = \frac{1}{n} \cdot \sum_{i=1}^{k} x_i^2 m_i$

Б $D^* = \frac{1}{n} \cdot \sum_{i=1}^{k} x_i^2 m_i - (\bar{x})^2$

В $D^* = \sum_{i=1}^{k} x_i^2 m_i - (\bar{x})^2$

12.8. Вибіркова дисперсія є

А незміщеною оцінкою дисперсії генеральної сукупності

Б зміщеною оцінкою дисперсії генеральної сукупності

В відповідь відсутня

12.9. Середину варіаційного ряду визначає

А мода

Б медіана

В вибіркова середня

12.10. Між імовірністю P та рівнем значущості α існує співвідношення

А $P = 1 - \alpha^2$

Б $P = 1 - \alpha$

В $P = 1/\alpha$

12.11. Стандартна форма надання довірчого інтервалу для оцінювання математичного сподівання з довірчою ймовірністю $P = \gamma$ має такий вигляд

А $\bar{x} - t_\gamma \frac{S_x}{\sqrt{n}} < a < \bar{x} + t_\gamma \frac{S_x}{\sqrt{n}}$

Б $\overline{x}-t_{\gamma}\frac{\sigma_x}{\sqrt{n}}<a<\overline{x}+t_{\gamma}\frac{\sigma_x}{\sqrt{n}}$

В $\overline{x}-t_{\gamma}\frac{1}{\sqrt{n}}<a<\overline{x}+t_{\gamma}\frac{1}{\sqrt{n}}$

Г $\overline{x}-\frac{S_x}{\sqrt{n}}<a<\overline{x}+\frac{S_x}{\sqrt{n}}$

12.6. Запитання для самоперевірки

12.1. Які існують види статистичних оцінок?

12.2. Сформулюйте вимоги, яким мають відповідати статистичні оцінки, і дайте їхнє аналітичне означення.

12.3. Наведіть приклади числових характеристик випадкової величини, обчислені за вибірковою сукупністю та які є статистичними оцінками відповідних характеристик випадкової величини в генеральній сукупності.

12.4. Що таке "інтервальна оцінка" і якими характеристиками її визначають?

12.5. У чому полягає необхідність у побудові інтервальних статистичних оцінок параметрів розподілу?

12.6. Наведіть формулу для визначення довірчого інтервалу для математичного сподівання випадкової величини, розподіленої за нормальним законом, за результатами вибірки.

12.7. Наведіть формулу для визначення довірчого інтервалу для середнього квадратичного відхилення випадкової величини, розподіленої за нормальним законом, за результатами вибірки.

12.7. Висновки за темою

Точкові статистичні оцінки θ^* є випадковими величинами, а тому наближена заміна θ на θ^* часто призводить до істотних похибок, особливо коли обсяг вибірки малий. У цьому разі застосовують інтервальні статистичні оцінки.

Інтервал $[\theta^*-\delta; \theta^*+\delta]$, що покриває оцінюваний параметр θ генеральної сукупності із заданою надійністю γ, називають довірчим.

13. Перевірка статистичних гіпотез

13.1. Мета та компетентності

Метою вивчення теми є ознайомлення з методами перевірки статистичних гіпотез та формування компетентностей щодо застосування статистичних критеріїв для перевірки певних статистичних гіпотез.

Компетентності, що формують під час вивчення теми:

знання основних типів статистичних гіпотез;

розуміння принципів побудови статистичних критеріїв та визначення їхніх критичних точок;

знання методик перевірки статистичних гіпотез;

уміння перевіряти відповідність між числовими характеристиками й законом розподілу випадкової величини в генеральній сукупності та їхніми оцінками за результатами дослідження вибіркової сукупності.

13.2. Основні теоретичні відомості

Під час використання методів математичної статистики можна допустити певний відсоток помилкових рішень, оскільки вони ґрунтуються на випадкових величинах і їх приймають в умовах невизначеності. Частку помилкових рішень, якою можна знехтувати, називають **рівнем значущості**. Найчастіше вона становить 5 або $1\,\%$, тобто $\alpha = 0{,}05$ або $\alpha = 0{,}01$.

У процесі порівняння декількох статистичних характеристик, обчислених за результатами статистичного оцінювання вибірок, виникає потреба встановити, чи істотна між ними різниця. **Істотною** називають відмінність, що за величиною перевищує ту, яку можна було б пояснити випадковими коливаннями.

Статистичною гіпотезою називають будь-яке припущення на певному рівні статистичної значущості щодо виду або параметрів невідомого закону розподілу.

Твердження про відсутність істотної відмінності між емпіричною та теоретичною характеристиками називають **нульовою (основною) гіпотезою** (H_0). Уводять **альтернативну (конкурентну) гіпотезу** (H_1) – це логічне заперечення нульової гіпотези.

Статистичні гіпотези можуть бути **спрямованими** та **неспрямованими** (табл. 13.1).

Таблиця 13.1

Формулювання спрямованих і неспрямованих гіпотез

Спрямовані	Неспрямовані
$H_0: \theta_1 \geq \theta_2$, $H_1: \theta_1 < \theta_2$	$H_0: \theta_1 = \theta_2$, $H_1: \theta_1 \neq \theta_2$

Статистичні гіпотези розподіляють на **параметричні** (гіпотези щодо невідомого значення параметра розподілу, що входить до параметричної сукупності розподілів) та **непараметричні** (припущення, за яким вид розподілу невідомий).

Правило, за яким вирішують підтвердити або відхилити нульову гіпотезу H_0, називають **статистичним критерієм**.

Під час перевірки статистичних гіпотез можливі помилки двох родів. **Помилка першого роду** полягає в тому, що нульову гіпотезу H_0 відхиляють, тоді як вона насправді є правильною. Беруть рівень значущості 0,01, або надійність 99 %. **Помилка другого роду** полягає в тому, що нульову гіпотезу H_0 підтверджують, а в дійсності вона є неправильною (рівень значущості 0,05, або надійність 95 %).

Для визначення кращого критерію перевірки гіпотези H_0 необхідно серед усіх критеріїв, які мають ту саму ймовірність похибки першого роду, вибрати той, для якого ймовірність похибки другого роду найменша. Похибку першого роду, яку можна припустити, можна задати заздалегідь. Множину значень, для яких нульову гіпотезу відхиляють, називають **критичною областю**.

Завдання перевірки гіпотези зведено до знаходження критичної області заданого рівня значущості. Імовірність недопущення помилки другого роду називають **потужністю критерію**.

У разі перевірки спрямованих гіпотез використовують однобічний критерій (лівобічну або правобічну критичну область), неспрямованих – двобічний, критична область якого удвічі менша, ніж для однобічного.

Схему перевірки статистичних гіпотез можна подати в послідовності таких процедур:

формулювання нульової та альтернативної гіпотез H_0, H_1;

вибір рівня значущості α;
вибір типу статистичного критерію;
розрахунки емпіричного значення статистичного критерію;
визначення критичної області критерію (знаходження критичних точок);
порівняння емпіричного та критичного значень критерію;
формулювання статистичних висновків.

Порівняння дисперсій двох нормально розподілених генеральних сукупностей

Нехай є вибірки із двох нормально розподілених генеральних сукупностей. Обсяги вибірок, відповідно, n_1 та n_2. Знайдено виправлені вибіркові дисперсії $S_1^2 > S_2^2$. Треба перевірити чи суттєво вони відрізняються.

Гіпотеза H_0 полягає в тому, що різницю дисперсій уважають незначущою, тобто $H_0 : S_1^2 = S_2^2$.

Тоді, якщо гіпотези спрямовані, то $H_1 : S_1^2 > S_2^2$; якщо гіпотези неспрямовані, то $H_1 : S_1^2 \neq S_2^2$.

Як емпіричне значення критерію обчислюють статистику **критерію Фішера – Снедекора:**

$$F = \frac{S_1^2}{S_2^2}.$$

Із таблиці критичних точок розподілу Фішера – Снедекора (додатки Ж, И) знаходять:

$F^* = F(\alpha, k_1, k_2)$ у разі спрямованих гіпотез,

$F^* = F(\alpha / 2, k_1, k_2)$ у разі неспрямованих гіпотез,
де $k_1 = n_1 - 1$, $k_2 = n_2 - 1$, α – рівень значущості.

Якщо $F > F^*$, то за рівнем значущості α гіпотезу H_0 відхиляють; якщо $F < F^*$, то з надійністю $1 - \alpha$ гіпотеза H_0 підтверджують.

Порівняння середніх двох нормально розподілених генеральних сукупностей із відомими дисперсіями (незалежні вибірки)

Нехай є дві генеральні сукупності X_1 та X_2, розподілені нормально, їхні виправлені дисперсії S_1^2 та S_2^2 відомі. За незалежними вибірками, обсяги яких, відповідно, n_1 та n_2, вилучених із цих сукупностей, знайдено вибіркові середні $\overline{X}_1$ та $\overline{X}_2$. Треба перевірити, чи однакові генеральні середні досліджуваних сукупностей.

Нульова й альтернативні гіпотези такі:

$$H_0 : \overline{X}_1 = \overline{X}_2,\ H_1 : \overline{X}_1 \neq \overline{X}_2$$ (неспрямовані гіпотези),

або

$$H_0 : \overline{X}_1 = \overline{X}_2,\ H_1 : \overline{X}_1 > \overline{X}_2,$$

або

$$H_0 : \overline{X}_1 = \overline{X}_2,\ H_1 : \overline{X}_1 < \overline{X}_2$$ (спрямовані гіпотези).

Як статистику критерію обчислюють таку величину:

$$t = \frac{\overline{X_1} - \overline{X_2}}{\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}}. \tag{13.1}$$

Якщо $H_1 : \overline{X}_1 \neq \overline{X}_2$, то будують двобічну критичну область. За таблицею інтегральної функції Лапласа (див. додаток В) визначають критичну точку t_α ($2\Phi(\mathrm{t}) = 1 - \alpha$). Якщо $|t| < t_\alpha$, то за рівнем значущості α гіпотезу H_0 підтверджують, якщо $|t| \geq t_\alpha$, то гіпотезу H_0 відхиляють із надійністю $P = 1 - \alpha$.

Якщо $H_1 : \overline{X}_1 > \overline{X}_2$, то будують правобічну критичну область. За таблицею інтегральної функції Лапласа (див. додаток В) визначають критичну точку t_α ($2\Phi(\mathrm{t}) = 1 - 2\alpha$). Якщо $t < t_\alpha$, то за рівнем значущості α гіпотезу H_0 підтверджують, якщо $t \geq t_\alpha$, то гіпотезу H_0 відхиляють із надійністю $P = 1 - \alpha$.

Якщо $H_1 : \overline{X}_1 < \overline{X}_2$, то будують лівобічну критичну область. За таблицею інтегральної функції Лапласа (див. додаток В) визначають критичну точку t_α $(2\Phi(\mathrm{t}) = 1 - 2\alpha)$. Якщо $t < t_\alpha$, то за рівнем значущості α гіпотезу H_0 відхиляють, якщо $t \geq t_\alpha$, то гіпотезу H_0 підтверджують із надійністю $P = 1 - \alpha$.

Порівняння середніх двох довільно розподілених генеральних сукупностей (великі незалежні вибірки)

За незалежними великими вибірками, обсяги яких, відповідно, n_1 та n_2, вилучених із довільно розподілених генеральних сукупностей, знайдено вибіркові середні $\overline{X}_1$ та $\overline{X}_2$.

Треба перевірити, чи суттєво вони відрізняються.

Нульова й альтернативні гіпотези такі:

$$H_0 : \overline{X}_1 = \overline{X}_2,\ H_1 : \overline{X}_1 \neq \overline{X}_2,$$

або

$$H_0 : \overline{X}_1 = \overline{X}_2,\ H_1 : \overline{X}_1 > \overline{X}_2,$$

або

$$H_0 : \overline{X}_1 = \overline{X}_2,\ H_1 : \overline{X}_1 < \overline{X}_2.$$

Завдяки великим обсягам вибірок, вибіркові дисперсії є гарними оцінками генеральних дисперсій і тому їх можна вважати відомими приблизно.

Тоді як приблизну статистику критерію обчислюють таку величину:

$$t = \frac{\overline{X_1} - \overline{X_2}}{\sqrt{\dfrac{S_1^2}{n_1} + \dfrac{S_2^2}{n_2}}}. \tag{13.2}$$

Далі приблизні висновки роблять такі, як і в разі порівняння середніх двох нормально розподілених генеральних сукупностей із відомими дисперсіями.

Порівняння середніх двох нормально розподілених генеральних сукупностей із невідомими дисперсіями (малі незалежні вибірки)

Нехай є дві малі вибірки з генеральних сукупностей, нормально розподілені. Обсяг першої вибірки – n_1, другої – n_2. Треба перевірити, чи суттєво різняться їхні вибіркові середні $\overline{X}_1$ та $\overline{X}_2$.

Нульова й альтернативні гіпотези такі:

$$H_0 : \overline{X}_1 = \overline{X}_2,\ H_1 : \overline{X}_1 \neq \overline{X}_2,$$

або

$$H_0 : \overline{X}_1 = \overline{X}_2,\ H_1 : \overline{X}_1 > \overline{X}_2,$$

або

$$H_0 : \overline{X}_1 = \overline{X}_2,\ H_1 : \overline{X}_1 < \overline{X}_2.$$

Додатково слід припустити, що невідомі генеральні дисперсії однакові. Інакше необхідно перевірити гіпотезу про рівність генеральних дисперсій за критерієм Фішера – Снедекора.

Перевірку нульової гіпотези здійснюють за ***t*-критерієм Стьюдента**, статистика якого має такий вигляд:

$$t = \frac{\overline{X_1} - \overline{X_2}}{\sqrt{\frac{(n_1 - 1)S_1^2 + (n_2 - 1)S_2^2}{n_1 + n_2 - 2}\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}.$$

Визначене емпіричне значення порівнюють із критичною точкою розподілу Стьюдента для заданого рівня значущості α і кількості ступенів свободи $k = n_1 + n_2 - 2$ (за таблицею додатка Е визначають t_α).

Якщо $H_1 : \overline{X}_1 \neq \overline{X}_2$, то будують двобічну критичну область. Якщо $|t| < t_\alpha$, то за рівнем значущості α гіпотезу H_0 підтверджують; якщо $|t| \geq t_\alpha$, то гіпотезу H_0 відхиляють із надійністю $P = 1 - \alpha$.

Якщо $H_1 : \overline{X}_1 > \overline{X}_2$, то будують правобічну критичну область. Якщо $t < t_\alpha$, то за рівнем значущості α гіпотезу H_0 підтверджують; якщо $t \geq t_\alpha$, то гіпотезу H_0 відхиляють із надійністю $P = 1 - \alpha$.

Якщо $H_1 : \overline{X}_1 < \overline{X}_2$, то будують лівобічну критичну область. Якщо $t < t_\alpha$, то за рівнем значущості α гіпотезу H_0 відхиляють; якщо $t \geq t_\alpha$, то гіпотезу H_0 підтверджують із надійністю $P = 1 - \alpha$.

Перевірка гіпотези про значення генеральної середньої

Для перевірки нульової гіпотези $H_0 : \overline{X}_\Gamma = a, \ (M(x) = a)$, де a є певним числом, за заданого рівня значущості α припускають, що дисперсія генеральної сукупності відома або знайдена за вибіркою великого обсягу.

Як статистику критерію обчислюють таку величину:

$$t = \frac{\sqrt{n}\left(\bar{x}_в - a\right)}{\sigma_\Gamma}. \tag{13.4}$$

Якщо $H_1 : \overline{X}_\Gamma > a$, то будують правобічну критичну область. За таблицею інтегральної функції Лапласа (див. додаток В) визначають критичну точку t_α ($2\Phi(\mathrm{t}) = 1 - 2\alpha$). Якщо $t < t_\alpha$, то за рівнем значущості α гіпотезу H_0 підтверджують; якщо $t \geq t_\alpha$ гіпотезу H_0 відхиляють із надійністю $P = 1 - \alpha$.

Якщо $H_1 : \overline{X}_\Gamma < a$, то будують лівобічну критичну область. За таблицею інтегральної функції Лапласа (див. додаток В) визначають критичну точку t_α ($2\Phi(\mathrm{t}) = 1 - 2\alpha$). Якщо $t < t_\alpha$, то за рівнем значущості α гіпотезу H_0 відхиляють; якщо $t \geq t_\alpha$, то гіпотезу H_0 підтверджують із надійністю $P = 1 - \alpha$.

Якщо $H_1 : \overline{X}_\Gamma \neq a$, то будують двобічну критичну область. За таблицею інтегральної функції Лапласа (див. додаток В) визначають критичну точку t_α ($2\Phi(\mathrm{t}) = 1 - \alpha$). Якщо $|t| < t_\alpha$, то за рівнем значущості α гіпотезу H_0 підтверджують; якщо $|t| \geq t_\alpha$, то гіпотезу H_0 відхиляють із надійністю $P = 1 - \alpha$.

Розглянутий підхід до побудови критичних областей придатний лише за умови, якщо відоме значення середнього квадратичного відхилення σ_Γ ознаки генеральної сукупності.

У разі, якщо значення $\sigma_Г$ є невідомим, його замінюють виправленим середнім квадратичним відхиленням:

$$S = \sqrt{\frac{\sum (x_i - \bar{x})^2 n_i}{n-1}}.$$

Тоді статистикою критерію є випадкова величина:

$$t = \frac{\bar{x}_B - a}{\frac{S}{\sqrt{n}}}. \qquad (13.5)$$

Критичні точки в цьому разі визначають за таблицею критичних точок розподілу Стьюдента (див. додаток Е) за заданим рівнем значущості α та кількості ступенів свободи $k = n - 1$.

Критерії згоди щодо закону розподілу

Часто закон розподілу випадкової величини в генеральній сукупності є невідомим, але певні припущення щодо його характеру можна зробити, судячи з гістограми у вибірковій сукупності. У цьому разі перевіряють нульову гіпотезу H_0 – це генеральна сукупність, розподілена за теоретичним законом. Тоді H_1 – це розподіл генеральної сукупності, що відрізняється від теоретичного закону.

Критерієм згоди називають критерій перевірки гіпотези щодо закону розподілу.

Згідно із **критерієм згоди Пірсона**, спостережуваний емпіричний розподіл вибіркової сукупності, виражений емпіричними частотами m_i згрупованого варіаційного ряду, порівнюють із допустимим теоретичним розподілом генеральної сукупності, відображеним теоретичними частотами $\tilde{m}_i$. **Емпіричними** називають частоти, які спостерігають у процесі аналізу вибірки, а **теоретичними** – ті, які обчислюють за формулами.

Якщо кількість спостережень дуже велика $(n \to \infty)$, то закон розподілу випадкової величини, незалежно від того, якому закону розподілу підпорядковано генеральну сукупність, наближається до розподілу χ^2 з k ступенями свободи, а сам критерій називають *критерієм згоди "хі-квадрат"* або *критерієм Пірсона*.

Для перевірки нульової гіпотези H_0 треба обчислити таку статистику:

$$\chi^2 = \sum_{i=1}^{s} \frac{(m_i - \widetilde{m}_i)^2}{\widetilde{m}_i}, \tag{13.6}$$

де s – кількість інтервалів згрупованого ряду розподілу;

m_i – емпіричні частоти;

$\widetilde{m}_i$ – теоретичні частоти ($\widetilde{m}_i = n\widetilde{p}_i$);

$\widetilde{p}_i = hf_i$ – теоретичні ймовірності;

$f_i = \dfrac{\varphi(t_i)}{\sigma}$;

$t_i = \dfrac{x_i - \overline{x}}{\sigma}$ – стандартизована величина;

$x_i = \dfrac{x_{i-1}^* + x_i^*}{2}$ – середнє інтервальне;

$\varphi(t_i)$ – диференціальна функція Лапласа.

Слід зауважити, що спостережень m_i у кожному інтервалі має бути не менше від п'яти відсотків від загальної кількості спостережень: $m_i \geq 0{,}05n$. Якщо їх буде менше, то необхідно збільшити інтервали.

Знайдену за формулою (13.6) величину порівнюють із критичними значеннями $\chi_\alpha^2(k)$, які знаходять у спеціальній довідковій таблиці (додаток Г).

Величину k визначають за формулою:

$$k = s - r - 1, \tag{13.7}$$

де s – кількість збільшених інтервалів;

r – кількість параметрів теоретичного закону розподілу (для нормального закону $r = 2$, для показникового – $r = 1$, для рівномірного – $r = 2$).

Величина α визначає рівень значущості. Для критерію Пірсона розглядають два рівні значущості: $\alpha = 0{,}05$ і $\alpha = 0{,}01$.

Якщо $\chi^2 < \chi^2_{0,05}(k)$, то нульову гіпотезу H_0 підтверджують, тобто теоретичний закон розподілу відповідає емпіричним даним і помилка у п'яти випадках зі ста в разі підтвердження можливо хибної гіпотези (помилка другого роду).

Якщо $\chi^2 > \chi^2_{0,01}(k)$, то нульову гіпотезу слід відхилити, тобто теоретичний закон розподілу не відповідає емпіричним даним, помилка в одному випадку зі ста в разі відхилення можливо правильної гіпотези (помилка першого роду).

Якщо $\chi^2_{0,05}(k) < \chi^2 < \chi^2_{0,01}(k)$, то є невизначеність і слід використати інші критерії.

За **критерієм згоди Пірсона за Романовським (критерій Романовського)** обчислюють таку величину:

$$t = \left| \frac{\chi^2 - k}{\sqrt{2k}} \right|,$$

де $k = s - 3$.

Якщо $t \le 1{,}96$, то з надійністю $95\,\%$ гіпотезу H_0 підтверджують, тобто закон розподілу відповідає нормальному; якщо $t \ge 2{,}58$, то з рівнем значущості $\alpha = 0{,}01$ гіпотезу H_0 відхиляють; якщо $t \in (1{,}96;\ 2{,}58)$, то це область невизначеності.

За **критерієм згоди Колмогорова** обчислюють таку величин:

$$\lambda = D\sqrt{n},$$

де $D = max\left|F(x) - \tilde{F}(x)\right|$, що є найбільшою за абсолютною величиною різницею між значеннями емпіричної функції розподілу та відповідної теоретичної функції розподілу.

За спеціальною таблицею (табл. 13.2) знаходять відповідне критичне значення λ_α.

Таблиця 13.2

Критичні значення критерію Колмогорова

α	0,4	0,3	0,2	0,1	0,05	0,025	0,01	0,005	0,001	0,0005
λ_α	0,89	0,97	1,07	1,22	1,36	1,48	1,63	1,73	1,95	2,03

Якщо $\lambda > \lambda_\alpha$, то гіпотезу H_0 відхиляють, тобто розходження між розподілами вважають значущим і воно не може бути викликано випадковими причинами. Якщо $\lambda \le \lambda_\alpha$, то гіпотезу H_0 погоджують із даними, тобто закон розподілу відповідає нормальному.

Перевірка гіпотези про однорідність вибірок

Якщо є дві незалежні вибірки з невідомими теоретичними функціями розподілу $F_1(x)$ та $F_2(x)$, то нульова гіпотеза про однорідність даних вибірок має такий вигляд: $H_0 : F_1(x) = F_2(x)$ за альтернативної $H_1 : F_1(x) \ne F_2(x)$.

Для перевірки нульової гіпотези за **критерієм Колмогорова – Смирнова** обчислюють таку величину:

$$\lambda = \sqrt{\frac{n_1 n_2}{n_1 + n_2}} \cdot max\left|F_{n_1}(x) - F_{n_2}(x)\right|,$$

де $F_{n_1}(x)$ та $F_{n_2}(x)$ – емпіричні функції розподілу, побудовані за двома вибірками обсягів n_1 та n_2.

Для вибірок великих обсягів за спеціальною таблицею (див. табл. 13.2) знаходять відповідне критичне значення λ_α.

Якщо $\lambda > \lambda_\alpha$, то гіпотезу H_0 відхиляють; якщо $\lambda \le \lambda_\alpha$, то гіпотезу H_0 не відхиляють.

13.3. Приклади розв'язування задач

Приклад 13.1. Є дві вибірки X та Y з нормально розподілених генеральних сукупностей. Для вибірки X : $n_1 = 10$, $S_x = 1{,}23$. Для вибірки Y : $n_2 = 18$, $S_y = 0{,}41$. Із надійністю 95 % перевірте гіпотезу про значущість різниці двох вибіркових дисперсій.

Розв'язання. Гіпотеза H_0 полягає в тому, що різницю дисперсій вважають незначущою. Нехай $H_1 : S_x^2 > S_y^2$, тобто гіпотези спрямовані.

За формулою обчисліть статистику критерію Фішера – Снедекора:

$$F = \frac{S_{max}^2}{S_{min}^2} = \frac{1{,}23}{0{,}41} = 3.$$

Обчисліть кількість ступенів свободи $k_1 = 10 - 1 = 9$, $k_2 = 18 - 1 = 17$, якщо за умовою завдання рівень значущості $\alpha = 0{,}05$.

Тоді з таблиці критичних точок розподілу Фішера – Снедекора (див. додаток Ж) знайдіть критичну точку $F(0{,}05;\ 9{,}17) = 2{,}5$.

Оскільки $F > F(0{,}05;\ 9{,}17)$, то гіпотезу H_0 відхиляють, тобто різниця дисперсій є значущою й дисперсія вибірки X суттєво більша від дисперсії вибірки Y.

Приклад 13.2. Зроблено дві серії пострілів по 15 у кожній серії (табл.13.3 і 13.4), X – довжина відстані від центру мішені.

Таблиця 13.3

Розподіл довжини відстані від центра мішені в першій серії пострілів

x_{1i}	−3	−2	−1	0	1	2	3
m_{1i}	1	3	2	2	2	3	2

Таблиця 13.4

Розподіл довжини відстані від центру мішені у другій серії пострілів

x_{2i}	−2	−1	0	1	2
m_{2i}	2	5	3	4	1

Перевірте з надійністю 95 %, чи суттєво різняться вибіркові середні.

Розв'язання. Обчисліть вибіркові середні $\overline{x_1} = 0{,}2$, $\overline{x_2} = -0{,}2$, вибіркові дисперсії $D(X_1) = 3{,}63$, $D(X_2) = 1{,}36$; виправлені вибіркові дисперсії $S_1^2 = 3{,}89$, $S_2^2 = 1{,}46$.

Гіпотеза H_0 – розбіжність між вибірковими середніми неістотна. $H_0 : \overline{X}_1 = \overline{X}_2$, $H_1 : \overline{X}_1 \neq \overline{X}_2$.

Нехай невідомі генеральні дисперсії однакові. Оскільки вибіркові сукупності є малими за обсягом, застосуйте критерій Стьюдента.

За формулою обчисліть статистику критерію Стьюдента:

$$t = \frac{|0{,}2 + 0{,}2|}{\sqrt{\frac{(15-1)\cdot 3{,}89 + (15-1)\cdot 1{,}46}{15+15-2}\left(\frac{1}{15} + \frac{1}{15}\right)}} = 0{,}67.$$

Знайдене емпіричне значення порівняйте із критичною точкою розподілу Стьюдента для заданого рівня значущості α і кількості ступенів свободи $k = 15 + 15 - 2 = 28$ (за таблицею додатка Ж визначте $t_{0{,}05}(28) = 2{,}05$).

Оскільки $t = 0{,}67 < t_{0{,}05}(28) = 2{,}05$, то з рівнем значущості 0,05 нульову гіпотезу H_0 про неістотність різниці вибіркових середніх не відхиляють.

Приклад 13.3. Розбіжність вимірів діаметрів металевої деталі є випадковою величиною, що має нормальний закон розподілу з $\sigma_\Gamma = 4$ мм. За рівня значущості α = 0,01 перевірте правильність нульової гіпотези $H_0 : a = 240$ мм за альтернативної гіпотези $H_1 : a > 240$ мм, якщо відомо, що вибіркове середнє значення діаметрів, виміряних у 100 однотипних деталей, $\bar{x}_B$ = 225 мм.

Розв'язання. Оскільки $H_1 : a > 240$ мм, будують правобічну критичну область. Для цього необхідно знайти критичну точку та побудувати правобічну критичну область.

Для знаходження критичної точки застосуйте вираз:

$$\Phi(t_\alpha) = \frac{1-2\alpha}{2} = \frac{1-2\cdot 0{,}01}{2} = \frac{1-0{,}02}{2} = \frac{0{,}98}{2} = 0{,}49.$$

За значенням $\Phi(t_\alpha) = 0{,}49$, скориставшись таблицею інтегральної функції Лапласа (див. додаток В), знаходять $t_\alpha \approx 2{,}34$.

Обчисліть значення статистики критерію за формулою (13.4):

$$t = \frac{225-240}{\frac{4}{\sqrt{100}}} = \frac{-15}{\frac{4}{10}} = -\frac{15}{0{,}4} = -\frac{150}{4} = -37{,}5.$$

Оскільки $t < t_\alpha$, то немає підстав для відхилення нульової гіпотези $H_0 : a = 240$ мм. Отже, нульову гіпотезу не відхиляють.

Приклад 13.4. На підприємстві було здійснено аналіз стажу роботи 100 працівників (табл. 13.5). Перевірте гіпотезу про нормальний розподіл досліджуваної випадкової величини – стажу роботи працівників.

Таблиця 13.5

Розподіл стажу роботи працівників

Стаж роботи, роки	[1; 3)	[3; 5)	[5; 7)	[7; 9)	[9; 11)	[11; 13)	[13; 15)	[15; 17)	[17; 19)	[19; 21)	[21; 23)
m_i	2	4	6	10	18	20	16	11	7	5	1

Розв'язання. Нульова гіпотеза H_0: досліджувана випадкова величина розподілена нормально. Альтернативна гіпотеза H_1: розподіл досліджуваної випадкової величини не підпорядковано нормальному закону.

Обчислення виконайте в табл. 13.6.

Таблиця 13.6

Визначення вирівнювальних частот

<table>
<tr><th>№</th><th>$\overline{x_i}$</th><th>m_i</th><th>$t_i = \dfrac{\overline{x_i} - \overline{x}}{S_x}$</th><th>$\varphi(t_i)$</th><th>$\hat{m}_i$</th><th>$\dfrac{(m_i - \hat{m}_i)^2}{\hat{m}_i}$</th></tr>
<tr><td>1</td><td>2</td><td rowspan="3">$\left.\begin{matrix}2\\4\\5\end{matrix}\right\}12$</td><td>–2,37</td><td>0,0241</td><td rowspan="3">$\left.\begin{matrix}1\\3\\7\end{matrix}\right\}11$</td><td rowspan="3">0,09</td></tr>
<tr><td>2</td><td>4</td><td>–1,89</td><td>0,0669</td></tr>
<tr><td>3</td><td>6</td><td>–1,42</td><td>0,1456</td></tr>
<tr><td>4</td><td>8</td><td>10</td><td>–0,95</td><td>0,2541</td><td>12</td><td>0,33</td></tr>
<tr><td>5</td><td>10</td><td>18</td><td>–0,48</td><td>0,3555</td><td>17</td><td>0,06</td></tr>
<tr><td>6</td><td>12</td><td>20</td><td>–0,01</td><td>0,3989</td><td>19</td><td>0,05</td></tr>
<tr><td>7</td><td>14</td><td>16</td><td>0,46</td><td>0,3589</td><td>17</td><td>0,06</td></tr>
<tr><td>8</td><td>16</td><td>11</td><td>0,93</td><td>0,2589</td><td>12</td><td>0,33</td></tr>
<tr><td>9</td><td>18</td><td>7</td><td>1,40</td><td>0,1497</td><td>7</td><td>0</td></tr>
<tr><td>10</td><td>20</td><td rowspan="2">$\left.\begin{matrix}5\\1\end{matrix}\right\}6$</td><td>1,87</td><td>0,0694</td><td rowspan="2">$\left.\begin{matrix}4\\1\end{matrix}\right\}5$</td><td rowspan="2">0,20</td></tr>
<tr><td>11</td><td>22</td><td>2,34</td><td>0,0252</td></tr>
<tr><td colspan="2">Сума</td><td>100</td><td>–</td><td>–</td><td>100</td><td>1,12</td></tr>
</table>

Визначте основні числові характеристики емпіричного розподілу. Оскільки вихідні дані наведено у вигляді інтервального ряду, то для визначення вибіркової середньої застосуйте таку формулу:

$$\overline{x} = \frac{1}{\sum\limits_{i=1}^{s} m_i} \sum_{i=1}^{s} \overline{x}_i m_i .$$

Значеннями випадкової величини $\overline{x}_i$ для цього ряду є середини відповідних інтервалів інтервального ряду емпіричного розподілу.

Визначають, що $\overline{x} = 12{,}04$ (р.).

Визначте виправлену дисперсію за співвідношенням:

$$S_x^2 = \frac{n}{n-1}\sum_{i=1}^{s}\left(\overline{x_i} - \overline{x}\right)^2 \cdot m_i = 4{,}26.$$

Перевірку нульової гіпотези виконують за критерієм Пірсона.

Після об'єднання малонасичених інтервалів їхня кількість становить $s = 8$, а тоді кількість ступенів свободи $k = s - r - 1 = 8 - 2 - 1 = 5$ (для нормального розподілу).

Для обчислення вирівнювальних частот застосовують співвідношення:

$$\hat{m}_i = \frac{n \cdot h}{S_x}\varphi(t_i).$$

Отже, визначають $\chi^2_{емп.} = 1{,}12$.

За таблицею розподілу Пірсона (додаток Г) знайдіть $\chi^2_{0{,}05}(5) = 15{,}1$.

Оскільки $\chi^2_{емп.} < \chi^2_{0{,}05}(k)$, то немає підстав для відхилення нульової гіпотези. Отже, закон розподілу стажу роботи працівників досліджуваного підприємства можна вважати нормальним.

Приклад 13.5. За умовами прикладу 13.4 перевірте гіпотезу про нормальний закон розподілу, використовуючи критерій Романовського.

Розв'язання. Обчисліть емпіричне значення критерію Романовського за формулою:

$$\frac{\left|\chi^2 - k\right|}{\sqrt{2k}},$$

де k – кількість ступенів свободи.

Для цієї задачі $k = 8 - 3 = 5$. Тоді $\frac{\left|\chi^2 - k\right|}{\sqrt{2k}} = \frac{\left|1{,}12 - 5\right|}{\sqrt{10}} = 1{,}23 < 1{,}96$.

Отже, розбіжність між емпіричним і теоретичним розподілами можна вважати статистично незначущою, тобто гіпотезу про нормальний закон розподілу не відхиляють.

Приклад 13.6. За рівнем значущості $\alpha = 0{,}05$ перевірте нульову гіпотезу, що кількість верстатів, які не працюють, серед 5 верстатів, які є в цеху, розподілено за біноміальним законом із $p = \frac{1}{3}$, якщо $D_n = \max_x \left| F(x) - F_n^*(x) \right| = 0{,}0683$ і $n = 25$.

Розв'язання. Використайте для перевірки нульової гіпотези критерій згоди Колмогорова.

За таблицею розподілу Колмогорова (див. табл. 13.1) $\lambda_\alpha = 1{,}36$.

За формулою обчисліть статистику критерію та порівняйте її із критичним значенням:

$$\sqrt{n} D_n = \sqrt{25} \cdot 0{,}0683 = 0{,}3415 < \lambda_\alpha = 1{,}36.$$

Значення статистичної характеристики не належить критичній області, тому гіпотезу про біноміальний закон розподілу в сукупності слід підтвердити.

Приклад 13.7. Протягом місяця вибірково здійснювали перевірку торговельних майданчиків міста із продажу овочів. Результати двох перевірок щодо недоважування покупців за одним видом овочів наведено в табл.13.7.

Таблиця 13.7

Результати двох перевірок щодо недоважування покупців

Інтервали недоважування, г	Кількість під час першої перевірки	Кількість під час другої перевірки
0 – 10	3	5
10 – 20	10	12
20 – 30	15	8
30 – 40	20	25
40 – 50	12	10
50 – 60	5	8
60 – 70	25	20
70 – 80	15	7
80 – 90	5	5
Разом	110	100

За рівнем значущості $\alpha = 0{,}05$ перевірте нульову гіпотезу про однорідність досліджуваних вибірок.

Розв'язання. Позначте $n_1 = 110, n_2 = 100$. Обчисліть накопичені частоти n_{i1}^{+}, n_{i2}^{+} вибірок та значення їхніх емпіричних функцій

$$F_{n1}(x_i) = n_{i1}^{+} / n_1, F_{n2}(x_i) = n_{i2}^{+} / n_2.$$

Нульова гіпотеза про однорідність даних вибірок має такий вигляд: $H_0 : F_1(x) = F_2(x)$ за альтернативної $H_1 : F_1(x) \neq F_2(x)$.

Для перевірки нульової гіпотези за критерієм Колмогорова – Смирнова необхідно обчислити за формулою таку величину:

$$\lambda = \sqrt{\frac{n_1 n_2}{n_1 + n_2}} \cdot \max\left|F_{n_1}(x) - F_{n_2}(x)\right|,$$

де $F_{n_1}(x)$ та $F_{n_2}(x)$ – емпіричні функції розподілу, побудовані за двома вибірками обсягів n_1 та n_2.

Допоміжні обчислення наведено в табл.13.8.

Таблиця 13.8

Обчислення емпіричних функцій розподілу

x_i	n_{i1}^{+}	n_{i2}^{+}	$F_{n1}(x_i)$	$F_{n2}(x_i)$	$\left|F_{n_1}(x_i) - F_{n_2}(x_i)\right|$
10	3	5	0,027	0,05	0,023
20	13	17	0,118	0,17	0,052
30	28	25	0,254	0,25	0,004
40	48	50	0,436	0,50	0,064
50	60	60	0,545	0,60	0,055
60	65	68	0,591	0,68	0,089
70	90	88	0,818	0,88	0,072
80	105	95	0,955	0,95	0,050
90	110	100	1	1	0

Отже, буде: $\lambda = \sqrt{\frac{110 \cdot 100}{110 + 100}} \cdot 0{,}089 = 0{,}644$.

За таблицею розподілу Колмогорова (див. табл. 13.1) $\lambda_\alpha = 1{,}36$.

Оскільки $\lambda < \lambda_\alpha$, то гіпотезу H_0 не відхиляють, тобто обваження покупців описують однією функцією розподілу.

Тим самим підтверджують, що обважування покупців є стійким та закономірним процесом із продажу овочів у цьому місті.

13.4. Вправи для самостійної роботи

13.1. Із генеральних сукупностей, розподілених за нормальним законом, було сформовано дві вибіркові сукупності обсягами $n=12$ і $m=16$. Для цих вибірок знайдено виправлені дисперсії $S_x^2=1{,}52$ та $S_y^2=0{,}76$. Перевірте за рівнем значущості $\alpha=0{,}05$, чи виконується для генеральної сукупності нульова гіпотеза $H_0: D(X)=D(Y)$ про рівність генеральних дисперсій.

13.2. За рівнем значущості $\alpha=0{,}05$ перевірте нульову гіпотезу $H_0: M(X)=M(Y)$ про рівність генеральних середніх нормально розподілених сукупностей. Вибіркові сукупності надано у вигляді дискретних варіаційних рядів (табл. 13.9 і 13.10).

Таблиця 13.9

Вибіркова сукупність *X*

x_i	12,3	12,5	12,8	13,0	13,5
n_i	1	2	4	2	1

Таблиця 13.10

Вибіркова сукупність *Y*

y_i	12,2	12,3	13,0
m_i	6	8	2

13.3. Досліджували продуктивність робітників у гривнях двох виробничих філій одного заводу. Результати досліджень подано двома статистичними розподілами (табл. 13.11 і 13.12).

Таблиця 13.11

Продуктивність робітників першої філії

x_i	140,8	160,8	180,8	200,8	220,8
n_i	2	6	32	8	2

Таблиця 13.12

Продуктивність робітників другої філії

y_i	150,6	160,6	170,6	180,6	190,6
m_i	12	28	40	18	2

Ознаки X і Y є незалежними та розподілені за нормальним законом. За рівнем значущості $\alpha = 0,01$ перевірте правильність нульової гіпотези $H_0 : M(X) = M(Y)$, якщо $H_1 : M(X) < M(Y)$.

13.4. У двох партіях містяться однотипні підшипники, виготовлені двома заводами. Вимірювання їхніх діаметрів (мм) дали результати, наведені в табл. 13.13 і 13.14.

Таблиця 13.13

Розподіл діаметрів підшипників першого заводу

x_i	6,58	6,60	6,80	7,00	7,20
n_i	6	8	10	4	2

Таблиця 13.14

Розподіл діаметрів підшипників другого заводу

y_i	6,60	6,70	6,74	6,78	6,82
m_i	2	4	8	6	4

За рівнем значущості α = 0,01 перевірте правильність нульової гіпотези $H_0 : M(X) = M(Y)$ за альтернативної $H_1 : M(X) \neq M(Y)$, якщо відомі значення $D_x = 50$; $D_y = 60$.

13.5. Із генеральної сукупності, ознака якої X має нормальний закон розподілу з $\sigma_Г = 5$, реалізовано вибірку і побудовано статистичний розподіл (табл. 13.15).

Таблиця 13.15

Вибірковий розподіл

x_i	10,9	11,0	11,2	11,3	11,5	11,6	11,8	11,9
n_i	2	4	1	3	4	1	2	3

За рівнем значущості α = 0,01 перевірте правильність нульової гіпотези $H_0 : a = 11{,}44$, якщо альтернативна гіпотеза $H_1 : a \neq 11{,}44$.

13.6. Розподіл випадкової величини X за даними вибірки визначено інтервальним варіаційним рядом (табл. 13.16).

Таблиця 13.16

Вихідний інтервальний варіаційний ряд

Інтервали	[3,0; 3,6)	[3,6; 4,2)	[4,2; 4,8)	[4,8; 5,4)	[5,4; 6,0)	[6,0; 6,6)	[6,6; 7,2)
Частоти (m_i)	2	8	35	43	22	15	5

За рівнем значущості $\alpha = 0,01$ перевірте гіпотезу про нормальний розподіл випадкової величини X у генеральній сукупності.

13.7. Як результат оброблення статистичних даних було знайдено значення емпіричних (m_i) і вирівнювальних ($\hat{m}_i$) частот у припущенні, що ознаку розподілено за нормальним законом (табл. 13.17).

Таблиця 13.17

Емпіричні та вирівнювальні частоти за нормального закону

m_i	3	7	11	20	28	19	10	2
$\hat{m}_i$	2	6	14	23	25	18	9	3

За рівнем значущості $\alpha = 0,05$ перевірте гіпотезу про нормальний закон розподілу ознаки генеральної сукупності.

13.8. Як результат оброблення статистичних даних було знайдено значення емпіричних (m_i) і вирівнювальних ($\hat{m}_i$) частот у припущенні, що ознаку розподілено за показниковим законом (табл. 13.18).

Таблиця 13.18

Емпіричні та вирівнювальні частоти за показникового закону

m_i	40	30	20	6	4
$\hat{m}_i$	48	25	13	7	7

За рівнем значущості $\alpha = 0,01$ перевірте нульову гіпотезу про показниковий розподіл випадкової ознаки X у генеральній сукупності.

13.9. За допомогою критеріїв χ^2 і Колмогорова перевірте гіпотезу про нормальний закон розподілу в сукупності – розміри деталей після шліфування, на підставі даних, які наведено в табл.13.19.

Таблиця 13.19

Розподіл розміру деталей після шліфування

Межі інтервалу	Частота
1	2
3,6 – 3,7	1
3,7 – 3,8	22
3,8 – 3,9	40

Закінчення табл. 13.19

1	2
3,9 – 4,0	79
4,0 – 4,1	27
4,1 – 4,2	26
4,2 – 4,3	4
4,3 – 4,4	1

Рівень значущості $\alpha = 0{,}02$. Оцінки для параметрів узяти на підставі вибіркових даних.

13.10. За даними завдань 13.3 та 13.4 перевірте гіпотезу про однорідність вибірок.

13.5. Тестові завдання

13.1. Статистична гіпотеза – це

А припущення щодо параметрів і виду закону розподілу генеральної сукупності

Б припущення щодо обсягу генеральної сукупності

В припущення щодо параметрів і виду закону розподілу вибірки

Г припущення щодо статистичного критерію

13.2. Рівень значущості α – це

А імовірність того, що відмінність визнано випадковою, а вона, насправді, є істотною

Б імовірність того, що відмінність визнано істотною, а вона, насправді, є випадковою

В відповідь відсутня

13.3. Помилку відхилення нульової гіпотези, тоді як вона є правильною, називають помилкою

А першого роду

Б другого роду

В відповідь відсутня

13.4. Помилку підтвердження нульової гіпотези, тоді як вона є неправильною, називають помилкою

А першого роду

Б другого роду

В відповідь відсутня

13.5. Потужністю критерію називають здатність не припуститися помилки

А першого роду

Б другого роду

В відповідь відсутня

13.6. Наведені гіпотези $\mathrm{H}_0 : \theta_1 \geq \theta_2, \mathrm{H}_1 : \theta_1 < \theta_2$ є

А спрямованими

Б не спрямованими

В відповідь відсутня

13.7. Моделі статистичних висновків щодо різниці середніх значень сукупностей для інтервальних даних, які мають нормальний розподіл і невідомі дисперсії, використовують

А t-критерій Стьюдента

Б F-критерій Фішера – Снедекора

В χ^2-критерій

Г відповідь відсутня

13.8. Для оцінювання рівня дисперсій сукупності використовують

А t-критерій Стьюдента

Б F-критерій Фішера – Снедекора

В χ^2-критерій

Г відповідь відсутня

13.9. χ^2-критерій застосовують для зіставлення

А тільки емпіричного розподілу з теоретичним

Б тільки декількох емпіричних розподілів між собою

В і емпіричного розподілу з теоретичним, і декількох емпіричних розподілів між собою

Г відповідь відсутня

13.10. У критерії Колмогорова за міру якості погодження емпіричного та теоретичного розподілу беруть

А відносну різницю між теоретичною та емпіричною частотами потрапляння випадкової величини в інтервал

Б максимум різниці за модулем між теоретичною та емпіричною частотами потрапляння випадкової величини в інтервал

В середнє квадратичне відхилення між теоретичною та емпіричною частотами потрапляння випадкової величини в інтервал

Г максимум модуля різниці між емпіричною та теоретичною функціями розподілу

Д максимум модуля різниці між емпіричною та теоретичною функціями щільності ймовірності

13.6. Запитання для самоперевірки

13.1. Поясніть, чому висновки про генеральну сукупність, зроблені за результатами дослідження вибірки, уважають статистичними гіпотезами.

13.2. Дайте означення нульової та альтернативної гіпотез. Яка із цих гіпотез підлягає перевірці?

13.3. Наведіть приклади відомих вам статистичних критеріїв і назвіть типи задач, у яких їх застосовують.

13.4. Чи можна, завдяки застосуванню критеріїв згоди, підтвердити нульову гіпотезу?

13.5. Дайте означення критичної області. Які бувають критичні області?

13.6. Що таке "критичні точки" і як їх визначають?

13.7. Наведіть загальний алгоритм перевірки нульової гіпотези.

13.8. Що таке "рівень значущості α" і як він пов'язаний із надійністю γ?

13.9. Що таке "помилки першого та другого роду"?

13.10. Що таке "потужність критерію"?

13.11. Які статистичні критерії застосовують для перевірки нульової гіпотези $H_0 : M(X)=M(Y)$?

13.12. У чому полягає нульова гіпотеза в разі перевірки відповідності закону розподілу у вибірковій сукупності припущенням щодо теоретичного закону розподілу випадкової величини?

13.13. Поясніть, у чому полягають відмінності між теоретичними та емпіричними частотами.

13.14. Як обчислюють емпіричне значення критерію згоди Пірсона і за якими параметрами визначають його теоретичне значення?

13.15. У чому полягає нульова гіпотеза в разі перевірки однорідності двох вибірок?

13.7. Висновки за темою

Головна цінність ухвалення статистичних рішень полягає в тому, що в межах імовірнісних категорій можна об'єктивно виміряти ступінь ризику, що відповідає тому чи тому рішенню.

Будь-які статистичні висновки, зроблені на підставі оброблення вибірки, називають статистичними гіпотезами.

Гіпотезу, що підлягає перевірці, називають основною (нульовою). Кожній нульовій гіпотезі можна протиставити кілька альтернативних (конкурентних) гіпотез, які позначають символом H_1, що заперечують твердження нульової.

Для перевірки правильності висунутої статистичної гіпотези вибирають статистичний критерій, керуючись яким відхиляють або не відхиляють нульову гіпотезу. Статистичний критерій є випадковою величиною, закон розподілу ймовірностей якої заздалегідь відомий.

Схему перевірки статистичних гіпотез можна подати в послідовності таких процедур: формулювання нульової та альтернативної гіпотез H_0, H_1; вибір рівня значущості α; вибір типу статистичного критерію; розрахунки емпіричного значення статистичного критерію; визначення критичної області критерію (знаходження критичних точок); порівняння емпіричного та критичного значень критерію; формулювання статистичних висновків.

14. Елементи теорії кореляції

14.1. Мета та компетентності

Метою вивчення теми є ознайомлення з елементами теорії кореляції та формування компетентностей щодо застосування кореляційного аналізу для визначення впливу фактора-аргументу на функціональний фактор.

Компетентності, що формують під час вивчення теми:

знання завдань кореляційного аналізу, означення вибіркового коефіцієнта кореляції та кореляційного відношення;

уміння оцінювати вибірковий коефіцієнт кореляції, кореляційне відношення та коефіцієнт рангової кореляції Спірмена;

навички у визначенні точкових оцінок основних числових характеристик двовимірної випадкової величини за вибірковими даними.

14.2. Основні теоретичні відомості

Кореляційний аналіз – це метод багатовимірного статистичного аналізу, який полягає в дослідженні коефіцієнтів кореляції між випадковими величинами. Завданнями кореляційного аналізу є визначення наявності, форми та тісноти кореляційного зв'язку, а також перевірка значущості впливу зміни значень однієї випадкової величини на середні значення другої. Для визначення наявності кореляційного зв'язку емпіричні дані заносять до кореляційної таблиці. **Кореляційну таблицю** створюють для згрупованих спостережень. Вона становить інтервали зміни величин X і Y та частоти сумісної появи заданої пари значень x та y.

Слід позначити n – загальну кількість спостережень; m_{ki} – частоту сумісної появи двох випадкових величин x_i і y_k, причому $\sum_k \sum_i m_{ki} = n$.

По горизонталі кореляційної таблиці (табл. 14.1) слід розмістити інтервали фактора X, по вертикалі – інтервали функціональної ознаки Y. На перетині стовпця x_i рядка y_k слід занести частоту m_{ki}. У крайній стовпець та рядок слід записати суму по рядках і стовпцях $m_{x_i} = \sum_{k=1}^{s} m_{ki}$ і $m_{y_k} = \sum_{i=1}^{l} m_{ki}$, де $\sum_{i=1}^{l} m_{x_i} = \sum_{k=1}^{s} m_{y_k} = n$. Через x_i і y_k позначити середини інтервалів.

Таблиця 14.1

Кореляційна таблиця

	Інтервали X	$x_0^* - x_1^*$	$x_1^* - x_2^*$	$\ldots$	$x_{i-1}^* - x_i^*$	$\ldots$	$x_{l-1}^* - x_l^*$	m_{y_k}
Інтервали Y	x_i / y_k	x_1	x_2	$\ldots$	x_i	$\ldots$	x_l	

Закінчення табл. 14.1

1	2	3	4	5	6	7	8	9
$y_0^* - y_1^*$	y_1	m_{11}	m_{12}	$\ldots$	m_{1i}	$\ldots$	m_{1l}	m_{y_1}
$y_1^* - y_2^*$	y_2	m_{21}	m_{22}	$\ldots$	m_{2i}	$\ldots$	m_{2l}	m_{y_2}
$\ldots$	$\ldots$	$\ldots$	$\ldots$	$\ldots$	$\ldots$	$\ldots$	$\ldots$	$\ldots$
$y_{k-1}^* - y_k^*$	y_k	m_{k1}	m_{k2}	$\ldots$	m_{ki}	$\ldots$	m_{kl}	m_{y_k}
$\ldots$	$\ldots$	$\ldots$	$\ldots$	$\ldots$	$\ldots$	$\ldots$	$\ldots$	$\ldots$
$y_{s-1}^* - y_s^*$	y_s	m_{s1}	m_{s2}	$\ldots$	m_{si}	$\ldots$	m_{sl}	m_{y_s}
m_{x_i}		m_{x_1}	m_{x_2}	$\ldots$	m_{x_i}	$\ldots$	m_{x_l}	n

Кореляційною залежністю називають таку залежність між двома випадковими величинами, за якої зі зміною однієї з них змінюється середнє значення другої.

Умовним середнім $\overline{y}_{x_i}$ називають середнє арифметичне значень Y, які відповідають $x = x_i$:

$$\overline{y}_{x=x_i} = \frac{\sum\limits_{k=1}^{s} y_k m_{ki}}{m_{x_i}}. \qquad (14.1)$$

Аналогічно для $Y = y_k$ умовне середнє $\overline{x}_{y=y_k}$ обчислюють за такою формулою:

$$\overline{x}_{y=y_k} = \frac{\sum\limits_{i=1}^{l} x_i m_{ki}}{m_{y_k}}. \qquad (14.2)$$

Кореляційні таблиці дають можливість визначати наявність кореляційної залежності: якщо $\overline{y}_x$ $\left(\overline{x}_y\right)$ змінюються від стовпця до стовпця (від рядка до рядка), то між величинами X і Y є кореляційний зв'язок.

Кореляційне поле – це графічне зображення кореляційної таблиці. Кореляційним полем називають точкову діаграму, у якій кожна точка становить результат окремого спостереження над двома змінними величинами. Кореляційне поле можна побудувати за кореляційною таблицею.

Якщо є лише спостереження, то кореляційне поле створюють разом зі складанням кореляційної таблиці. Записують інтервали величин X та Y і до кожної комірки на перетині стовпця і рядка заносять точки, кількість яких відповідають величинам m_{ki} (рис. 14.1).

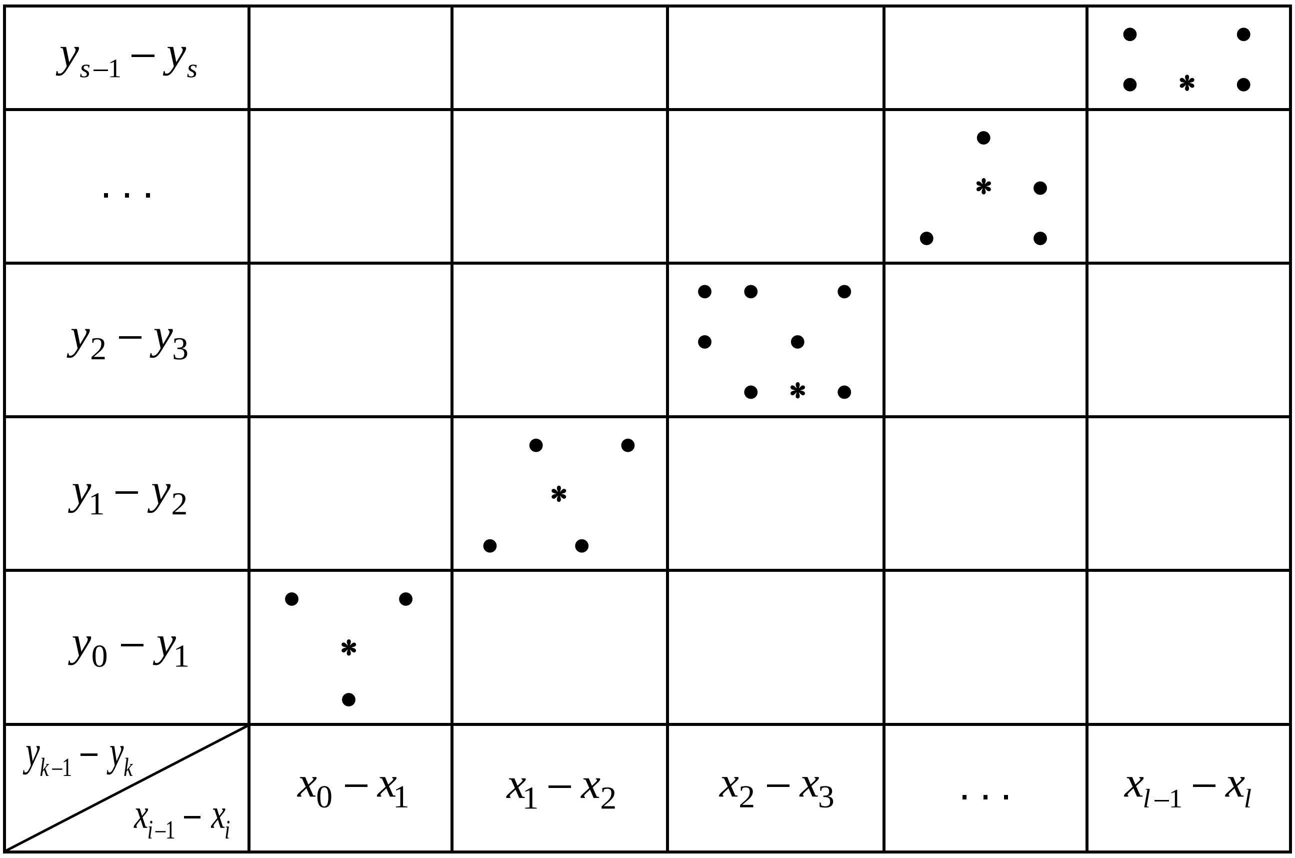


Рис. 14.1. **Кореляційне поле**

Якщо на кореляційному полі центри групування зміщуються, то можна судити про наявність кореляційного зв'язку між X та Y.

Статистичні висновки про тісноту лінійного взаємозв'язку між величинами X та Y, розподіленими за нормальним законом, ґрунтуються на значенні вибіркового коефіцієнта кореляції Пірсона.

Вибірковий коефіцієнт кореляції Пірсона r є статистичною оцінкою коефіцієнта кореляції генеральної сукупності, його обчислюють за такою формулою:

$$r = \frac{\sum_{i=1}^{n}(x_i - \overline{x})(y_i - \overline{y})}{\sqrt{\sum_{i=1}^{n}(x_i - \overline{x})^2 \cdot \sum_{i=1}^{n}(y_i - \overline{y})^2}},$$

де $\overline{x}$, $\overline{y}$ – вибіркові середні випадкових величин X і Y, відповідно.

Формулу застосовують у тому разі, якщо вибіркові дані є незгрупованими. Для даних, наданих у вигляді кореляційної таблиці, вибірковий коефіцієнт кореляції більш зручно оцінювати за такою формулою:

$$r = \frac{\overline{xy} - \overline{x} \cdot \overline{y}}{\sqrt{\left(\overline{x^2} - \overline{x}^2\right) \cdot \left(\overline{y^2} - \overline{y}^2\right)}}, \tag{14.3}$$

де $\overline{xy}$ – середня добутку випадкових величин;

$\overline{x^2}$, $\overline{y^2}$ – вибіркові середні квадратів випадкових величин X і Y.

Величина $\overline{xy} - \overline{x} \cdot \overline{y}$ має назву **коваріація** і є вибірковим **кореляційним моментом** μ_{xy}.

Властивості коефіцієнта кореляції r_{xy}:

1) якщо X і Y незалежні, то $r_{xy} = 0$ (оскільки $\mu_{xy} = 0$);

2) якщо випадкові величини лінійно залежні, то $r_{xy} = 1$;

3) $|r_{xy}| \le 1$, $-1 \le r \le 1$;

4) $r_{xy} = r_{yx}$ (за визначенням).

За величиною коефіцієнта кореляції отримують інформацію щодо статистичної значущості кореляційного зв'язку.

Якщо $|r| < 0{,}35$, то кореляційний зв'язок можна вважати статистично неістотним; якщо $|r| \ge 0{,}35$ – зв'язок уважають істотним, а якщо $|r| \ge 0{,}7$ – тісним.

У разі перевірки статистичної значущості вибіркового коефіцієнта кореляції (за заданим рівнем значущості) нульова гіпотеза полягає в тому, що коефіцієнт кореляції в генеральній сукупності дорівнює нулю, тобто $H_0: \ \rho = 0$. Альтернативною є гіпотеза $H_1: \ \rho \ne 0$, отже, їй відповідає двобічна критична область.

Якщо обсяг вибіркової сукупності невеликий, то як статистику критерію обчислюють таку величину:

$$t = \frac{r \cdot \sqrt{n-2}}{\sqrt{1 - r^2}}, \tag{14.4}$$

розподілену за статистикою Стьюдента з кількістю ступенів свободи $k = n - 2$.

Якщо $|t| \le t_\alpha(k)$, то нульову гіпотезу немає підстав відхилити, і кореляційний зв'язок між випадковими величинами X і Y є статистично незначущим; якщо $|t| > t_\alpha(k)$, то гіпотезу H_0 відхиляють, звідси $\rho \ne 0$.

За великого обсягу вибірки для перевірки нульової гіпотези як статистику критерію розглядають таку величину:

$$t = \frac{r \cdot \sqrt{n}}{\sqrt{1 - r^2}}, \tag{14.5}$$

розподілену за нормальним законом.

Критичне значення t_α обчислюють за умовою $\Phi(t_\alpha) = (1 - \alpha)/2$.

Якщо $|t| \le t_\alpha$, то нульову гіпотезу немає підстав відхилити, і кореляційний зв'язок між випадковими величинами X і Y є статистично незначущим; якщо $|t| > t_\alpha$, то гіпотезу H_0 відхиляють, звідси $\rho \ne 0$.

Також значущість лінійного кореляційного зв'язку для незгрупованих даних можна перевірити за допомогою критерію Фішера, статистика якого має такий вигляд:

$$F_r = \frac{r^2}{1 - r^2} \cdot \frac{n - 2}{1}.$$

Емпіричне значення критерію F_r порівнюють із критичними точками статистики Фішера – Снедекора (див. додатки Ж, И) за рівнем значущості 0,05 та 0,01 і кількостями ступенів свободи $k_1 = 1$ і $k_2 = n - 2$.

Якщо $F_r < F_{0,05}(1; n - 2)$, то нульову гіпотезу про те, що в генеральній сукупності $\rho = 0$, немає причин відхилити. Якщо визначають, що $F_r > F_{0,01}(1; n - 2)$, то з надійністю 99 % гіпотезу H_0 відхиляють.

Показником тісноти зв'язку між фактором X і ознакою Y також є величина $\eta_{y/x}$ – **кореляційне відношення**:

$$\eta_{y/x}=\sqrt{\frac{S^2_{\bar{y}_x}}{S^2_y}}, \qquad (14.6)$$

де $S^2_{\bar{y}_x}=\frac{n}{n-1}\cdot\frac{\sum_i\left(y_{x_i}-\bar{y}\right)^2 m_{x_i}}{n}=\frac{n}{n-1}\cdot\left(\frac{\sum_i \bar{y}^2_{x_i}m_{x_i}}{n}-\left(\bar{y}\right)^2\right)$.

Величина $\eta^2_{y/x}$ характеризує, яку частину повної мінливості функціональної ознаки Y пояснює вплив фактора-аргументу X.

Кореляційні відношення:

$$\eta_{y/x}=\frac{S_{\bar{y}_x}}{S_y},\ \eta_{x/y}=\frac{S_{\bar{x}_y}}{S_x} \qquad (14.7)$$

є значеннями, визначеними за емпіричними даними.

Величина η характеризує тісноту зв'язку в кореляційній залежності та має такі *властивості*:

1) $\eta\geq 0$ (за визначенням);

2) $0\leq\eta\leq 1$;

3) якщо $\eta=1$, то зв'язок між X і Y функціональний, тобто відсутнє розсіювання всередині групи для кожного x_i одне y_i);

4) якщо $\eta=0$, то між X і Y немає кореляційної залежності;

5) якщо величина η ближча до нуля, то зв'язок між X і Y слабкий; якщо величина η ближча до одиниці, то зв'язок між X і Y тісний;

6) якщо $\eta_{y/x}\approx|r|$, то має місце лінійна кореляційна залежність; якщо $\eta>>r$, то залежність між X і Y нелінійна.

Значущість кореляційного зв'язку між X і Y перевіряють за критерієм Фішера – Снедекора, статистика якого має такий вигляд:

$$F_\eta=\frac{\eta^2_{y/x}}{1-\eta^2_{y/x}}\cdot\frac{n-k}{k-1},$$

де k – кількість груп за аргументом X.

Якщо $F_\eta < F_{0,05}(k-1; n-k)$, то гіпотезу $H_0: \eta = 0$ немає підстав спростовувати, тобто кореляційний зв'язок між факторами X і Y відсутній. Якщо $F_\eta > F_{0,01}(k-1; n-k)$, то нульову гіпотезу спростовують на користь альтернативної $H_1: \eta \neq 0$, тобто кореляційна залежність є значущою.

Для лінійного кореляційного зв'язку коефіцієнт кореляції за модулем дорівнює кореляційному відношенню. Різниця між цими величинами є мірою відхилення кореляційного зв'язку від лінійної форми. Отже, нульовою є гіпотеза про відсутність такої розбіжності, тобто $H_0: r^2 = \eta^2$. Відповідно, альтернативною є гіпотеза $H_1: r < \eta$.

Для перевірки адекватності лінійної моделі розглядають F-статистику, тобто:

$$F_A = \frac{(\eta^2 - r^2)/(k-2)}{(1-\eta^2)/(n-k)}, \tag{14.8}$$

де k – кількість груп, у які за кореляційною таблицею об'єднано значення зовнішнього фактора X.

Емпіричне значення F-критерію порівнюють зі значеннями, які визначають за таблицею розподілу Фішера – Снедекора (див. додатки Ж, И) для рівня значущості α й кількості ступенів свободи $k_1 = k-2$ та $k_2 = n-k$.

Якщо $F_A < F_{0,05}(k-2; n-k)$, то з надійністю 95 % основну гіпотезу немає підстав відхилити. Тобто немає підстав уважати, що кореляційний зв'язок є нелінійним, отже, лінійна модель є адекватною. Навпаки, якщо $F_A > F_{0,01}(k-2; n-k)$, то нульову гіпотезу відхиляють, тобто лінійна модель є неадекватною, і треба шукати іншу форму кореляційного зв'язку.

Коефіцієнт кореляції Пірсона можна застосовувати тільки в разі, якщо величини X та Y розподілено за нормальним законом. В інших випадках застосовують коефіцієнт рангової кореляції Спірмена.

Вибірковий коефіцієнт кореляції Спірмена r_s, який є статистичною оцінкою коефіцієнта кореляції Спірмена ρ_s у генеральній сукупності, обчислюють за такою формулою:

$$r_s = 1 - \frac{6\sum(x_i' - y_i')^2}{n(n^2-1)},$$

де x_i', y_i' – ранги (номери) елементів, відповідно, x_i $(i=\overline{1,n})$ у варіаційному ряді компоненти X та y_i $(i=\overline{1,n})$ у варіаційному ряді компоненти Y.

Властивості вибіркового коефіцієнта кореляції Спірмена:

1) $|r_s| \le 1$;

2) якщо ранги компонентів X і Y збігаються для всіх $i=\overline{1,n}$, то вибірковий коефіцієнт рангової кореляції Спірмена дорівнює одиниці, тобто має місце повний прямий зв'язок;

3) якщо випадкові величини X і Y є незалежними, то математичне сподівання випадкової величини, якою є статистична оцінка коефіцієнта рангової кореляції Спірмена, дорівнює нулю $M(r_s)=0$, а її дисперсія становить $D(r_s)=\dfrac{1}{n-1}$.

Для перевірки гіпотези про значущість коефіцієнта рангової кореляції ($H_0:\rho_s=0$, $H_1:\rho_s\ne 0$) для вибірок, обсяг яких відповідає умові $4\le n\le 10$, потрібно обчислити величину:

$$t=\frac{r_s\cdot\sqrt{n-2}}{\sqrt{1-r_s^2}}, \tag{14.9}$$

розподілену за статистикою Стьюдента з кількістю ступенів свободи $k=n-2$. За рівнем значущості α значення цієї випадкової величини порівнюють із критичною точкою $t_\alpha(k)$ двобічної критичної області.

Якщо $|t|<t_\alpha(k)$, то нульову гіпотезу H_0 немає підстав відхилити, отже, на рівні значущості α зв'язок між рангами факторів X і Y є статистично неістотним. Якщо $|t|>t_\alpha(k)$, то нульову гіпотезу H_0 відхиляють на користь альтернативної, отже, $\rho_s\ne 0$.

За великих обсягів вибірки нульову гіпотезу щодо незалежності рангів спростовують, якщо виявиться, що

$$|r_s|>\frac{t_\alpha}{\sqrt{n-1}},$$

де t_α є аргументом функції Лапласа, згідно з умовою $\Phi(t_\alpha)=1-0{,}5\alpha$.

14.3. Приклади розв'язування задач

Приклад 14.1. За даними табл. 14.2, що містить інформацію про продуктивність праці (X) в ум. од. та преміальні виплати (Y) у %, складіть кореляційну таблицю.

Таблиця 14.2

Таблиця вихідних даних

x_i	y_i	x_i	y_i	x_i	y_i	x_i	y_i	x_i	y_i
10,2	3,4	21,8	5,1	24,4	5,7	24,5	5,1	27,2	5,3
9,4	3,7	22,9	4,6	21,5	3,9	21,3	4,4	27,1	6,1
13,8	4,1	17,7	5,8	20,9	4,7	29,9	6,1	16,1	5,2
19,1	3,9	23,3	5,6	27,9	6,4	12,8	4,0	20,6	3,6
19,7	4,6	11,3	4,2	25,3	4,9	26,5	5,9	21,7	4,7
7,6	2,6	17,8	5,3	25,5	5,8	17,9	4,2	22,8	6,2
17,1	3,6	26,7	4,7	22,8	5,0	26,6	5,2	22,4	4,9
14,7	4,6	14,7	3,3	7,9	3,9	19,4	6,1	32,4	6,3
20,6	5,0	22,2	4,9	16,8	4,4	22,5	4,4	20,8	5,3
20,6	4,1	13,1	5,1	19,6	4,7	18,1	4,0	25,3	5,4

Розв'язання. За умовою прикладу X є зовнішнім (незалежним) фактором, а Y – внутрішнім (функціональним) фактором.

Для побудови кореляційної таблиці визначають інтервали для кожного з них.

Так, для випадкової величини X буде $x_{max} = 32{,}4$, $x_{min} = 7{,}6$. Тоді розмах становить $R_X = 32{,}4 - 7{,}6 = 24{,}8$.

Візьміть 5 інтервалів, тоді крок буде:

$$h_X = \frac{R_X}{k_x} \approx \frac{25}{5} = 5.$$

Доцільно взяти довжину кроку $h_X = 5$.

Початок першого інтервалу виберіть таким чином, щоб за такої довжини кроку серединам інтервалів відповідали цілі числа (це зручно для подальших обчислень). Оскільки початок першого інтервалу в загальному випадку можна зсунути відносно x_{min} убік менших значень на півкроку, то виберіть нижньою межею першого інтервалу $x_1 = 7{,}5$. Тоді $x_2 = 12{,}5$, $x_3 = 17{,}5$, …, $x_6 = 32{,}5$.

Для випадкової величини Y буде $y_{max} = 6{,}4$, $y_{min} = 2{,}6$, відповідно, розмах становить $R_Y = 6{,}4 - 2{,}6 = 3{,}8$.

Поділіть цій розмах на 4 інтервали, тоді довжина кожного інтервалу дорівнює:

$$h_Y \approx \frac{4}{4} = 1.$$

Для того щоб мати цілі значення середини інтервалів, виберіть $y_1 = 2{,}5$. Тоді $y_2 = 3{,}5$, $y_3 = 4{,}5$, $y_4 = 5{,}5$, $y_5 = 6{,}5$.

Якщо розмістити інтервали X по горизонталі, а Y – по вертикалі, то буде складено таблицю, яку розподілено на комірки (табл. 14.3).

Таблиця 14.3

Кореляційна таблиця

Y \ X	[7,5; 12,5)	[12,5: 17,5)	[17,5; 22,5)	[22,5; 27,5)	[27,5; 32,5)	m_y
[2,5; 3,5)	2					2
[3,5; 4,5)	3	5	8			16
[4,5; 5,5)		3	10	8		21
[5,5; 6,5)			2	6	3	11
m_x	5	8	20	14	3	50

Тепер для кожної комірки цієї таблиці обчисліть кількість значень двовимірної випадкової величини $(X;Y)$, які одночасно відповідають умовам $x_i \le X < x_{i+1}$ та $y_j \le Y < y_{y+1}$, де $i = \overline{1{,}5}$, $j = \overline{1{,}4}$. Отже, для кожної з комірок таблиці визначте частоту m_{ij} потрапляння до неї значень двовимірної випадкової величини, що досліджують. Так, першій парі значень $(10{,}2;\ 3{,}4)$ із табл. 14.2 відповідає комірка табл. 14.3, що стоїть на перетині першого стовпця і першого рядка, парі $(9{,}4;\ 3{,}7)$ – комірка на перетині першого стовпця і другого рядка тощо.

Остаточно буде складено кореляційну таблицю (див. табл. 14.3), із якої видно, що частоти утворюють *хмару розсіювання* у формі еліпса, більшу вісь якого спрямовано вздовж головної діагоналі таблиці.

Доцільно припустити, що між факторами X і Y існує додатний кореляційний зв'язок.

Приклад 14.2. Задано кореляційну таблицю, у якій наведено результати іспитів із фізики (y) та математики (x) в одній з академічних груп I курсу (табл. 14.4).

Таблиця 14.4

Кореляційна таблиця результатів іспитів із фізики та математики

y \ x	2	3	4	5	l_k
2	1	2			3
3	1	6	3		10
4		5	5		10
5			1	1	2
h_i	2	13	9	1	25

Оцініть тісноту зв'язку між результатами іспитів із фізики та математики.

Розв'язання. Обчисліть:

$$\bar{x} = \frac{2\cdot 2 + 3\cdot 13 + 4\cdot 9 + 5\cdot 1}{25} = \frac{84}{25} = 3{,}36;$$

$$\bar{y} = \frac{2\cdot 3 + 3\cdot 10 + 4\cdot 10 + 5\cdot 2}{25} = \frac{86}{25} = 3{,}44;$$

$$\overline{x^2} = \frac{4\cdot 2 + 9\cdot 13 + 16\cdot 9 + 25\cdot 1}{25} = \frac{294}{25} = 11{,}76;$$

$$\overline{y^2} = \frac{4\cdot 3 + 9\cdot 10 + 16\cdot 10 + 25\cdot 2}{25} = \frac{312}{25} = 12{,}48;$$

$$\overline{xy} = \frac{\sum_i \sum_k m_{ki} x_i y_k}{n} = \frac{1}{25}\big(\big(1\cdot(2\cdot 2) + 2\cdot(3\cdot 2)\big) + \big(1\cdot(2\cdot 3) + 6\cdot(3\cdot 3) + 3\cdot(4\cdot 3)\big) +$$

$$+\big(5\cdot(3\cdot 4) + 5\cdot(4\cdot 4)\big) + \big(1\cdot(4\cdot 5) + 1\cdot(5\cdot 5)\big)\big) = \frac{1}{25}(16 + 96 + 140 + 45) = \frac{297}{25} = 11{,}88.$$

За формулою (14.3) буде:

$$r = \frac{11{,}88 - 3{,}36\cdot 3{,}44}{\sqrt{11{,}76 - 3{,}36^2}\cdot\sqrt{12{,}48 - 3{,}44^2}} \approx 0{,}58.$$

Оскільки $0{,}35 < r < 07$, то кореляційний зв'язок між результатами іспитів із фізики та математики є статистично значущим, але не дуже тісним.

Приклад 14.3. У припущенні лінійного кореляційного зв'язку обчисліть коефіцієнт кореляції, якщо за результатами дослідження визначено:

$$\sum x = 35; \quad \sum y = 136,0; \quad \sum xy = 2806,5; \quad \sum x^2 = 2038;$$

$$\sum y^2 = 36804,4; \qquad n = 50.$$

Розв'язання. Коефіцієнт кореляції обчисліть за формулою (14.3):

$$r = \frac{\overline{xy} - \bar{x}\cdot\bar{y}}{\sqrt{\overline{x^2} - \bar{x}^2}\cdot\sqrt{\overline{y^2} - \bar{y}^2}}.$$

Для цього знаходять усі величини, що входять до цієї формули:

$$\bar{x} = \frac{\sum x}{n} = \frac{35}{50} = 0,7; \quad \bar{y} = \frac{\sum y}{n} = \frac{136,0}{50} = 2,72; \quad \overline{xy} = \frac{\sum xy}{n} = \frac{2806,5}{50} = 56,13;$$

$$\overline{x^2} = \frac{\sum x^2}{n} = \frac{2038}{50} = 40,76; \qquad \overline{y^2} = \frac{\sum y^2}{n} = \frac{36804,4}{50} = 736,088.$$

Далі обчислюють такі числові характеристики:

вибірковий кореляційний момент:

$$\mu_{xy} = \overline{xy} - \bar{x} \cdot \bar{y} = 56{,}13 - 0{,}7 \cdot 2{,}72 = 54{,}226;$$

вибіркові середні квадратичні відхилення для обох факторів:

$$\sigma_x^* = \sqrt{\overline{x^2} - \bar{x}^2} = \sqrt{40{,}76 - 0{,}7^2} = \sqrt{40{,}27};$$

$$\sigma_y^* = \sqrt{\overline{y^2} - \bar{y}^2} = \sqrt{736{,}088 - 2{,}72^2} = \sqrt{728{,}69}.$$

Тоді

$$r = \frac{54{,}226}{\sqrt{40{,}27} \cdot \sqrt{728{,}69}} = 0{,}317.$$

Оскільки $|r| < 0{,}35$, то кореляційний зв'язок є статистично незначущим.

Приклад 14.4. Дані вибіркової сукупності наведено в табл. 14.5.

Таблиця 14.5

Вихідна вибіркова сукупність

$X = x_i$	1	2	4	6	7	9	10	12	16
$Y = y_i$	10,7	12,9	14,9	19,3	22,9	23,2	26,1	31,7	35,3

Обчисліть коефіцієнт кореляції та за критерієм Стьюдента перевірте його значущість із надійністю 95 %.

Розв'язання. Коефіцієнт кореляції обчислюють за формулою (14.3).

Допоміжні обчислення наведено в табл. 14.6.

Таблиця 14.6

Таблиця допоміжних обчислень

№ п/п	x_i	y_i	$x_i y_i$	x_i^2	y_i^2
1	2	3	4	5	6
1	1	10,7	10,7	1	114,49
2	2	12,9	25,8	4	166,41

Закінчення табл. 14.6

1	2	3	4	5	6
3	4	14,9	59,6	16	222,01
4	6	19,3	115,8	36	372,49
5	7	22,9	160,3	49	524,41
6	9	23,2	208,8	81	538,24
7	10	26,1	261,0	100	681,21
8	12	31,7	380,4	144	1004,89
9	16	35,3	564,8	256	1246,09
Сума	67	197	1787,2	687	4870,24

За даними рядка "Сума" табл. 14.6 обчислюють такі середні значення: вибіркову середня випадкової величини X:

$$\bar{x} = \frac{1}{9} \cdot 67 = 7,444.$$

вибіркову середня випадкової величини Y:

$$\bar{y} = \frac{197}{9} = 21,889;$$

вибіркову середня добутку випадкових величин:

$$\overline{xy} = \frac{1787,2}{9} = 198,578;$$

середні квадратів випадкових величин:

$$\overline{x^2} = \frac{687}{9} = 76,333;$$

$$\overline{y^2} = \frac{4870,24}{9} = 541,138.$$

Далі знаходять вибіркові середні квадратичні відхилення:

$$\sigma_x^2 = 76,333 - 7,444^2 = 20,920; \ \sigma_x^* = \sqrt{20,920} = 4,574$$

та

$$\sigma_y^2 = 541{,}138 - 21{,}889^2 = 62{,}010; \ \sigma_y^* = \sqrt{62{,}010} = 7{,}875.$$

За формулою (14.3) обчислюють значення коефіцієнта кореляції:

$$r = \frac{198{,}578 - 7{,}444 \cdot 21{,}889}{4{,}574 \cdot 7{,}875} = 0{,}989.$$

$r > 0{,}7$, що свідчить про тісний зв'язок між величинами X та Y.

Перевірте статистичну значущість коефіцієнта кореляції. Згідно з нульовою гіпотезою H_0: $r = 0$. Альтернативною є гіпотеза H_1: $r \neq 0$, тобто критична область є двобічною.

Оскільки обсяг вибірки невеликий, то для перевірки нульової гіпотези обчислюють емпіричне значення t-критерію за формулою (14.4):

$$t = \frac{0{,}989 \cdot \sqrt{9-2}}{\sqrt{1-0{,}989^2}} = \frac{0{,}989 \cdot 2{,}6458}{\sqrt{0{,}021879}} = 17{,}69.$$

За таблицею значень критичних точок розподілу Стьюдента (див. додаток Ж) за рівнем значущості $\alpha = 0{,}05$ та кількістю ступенів свободи $k = n - 2$ знаходять $t_{0{,}05}(7) = 2{,}36$.

Оскільки $|t| > t_{0{,}05}(7)$, то нульову гіпотезу слід відхилити.

Отже, із надійністю 95 % кореляційний зв'язок між факторами є статистично значущим.

Приклад 14.5. Знайдіть кореляційне відношення $\eta_{y/x}$ та поясніть результат, якщо $\bar{y} = 33{,}5$, $S_y^2 = 98{,}75$ і умовні середні фактора Y для сталого значення фактора X наведено в табл. 14.7.

Таблиця 14.7

Умовні середні фактора Y

$\bar{y}\vert(X = x_i)$	11,0	18,14	30,0	33,8	42,05	51,0
m_{x_i}	4	7	10	57	19	3

Розв'язання. Кореляційне відношення обчислюють за формулою (14.6):

$$\eta_{y/x} = \sqrt{\frac{S^2_{\bar{y}_x}}{S^2_y}},$$

де

$$S^2_{\bar{y}_x} = \frac{n}{n-1} \cdot \frac{\sum_i \left(y_{x_i} - \bar{y}\right)^2 m_{x_i}}{n} = \frac{n}{n-1} \cdot \left(\frac{\sum_i \bar{y}^2_{x_i} m_{x_i}}{n} - \left(\bar{y}\right)^2 \right).$$

Отже, знаходять:

$$S^2_{y_x} = \frac{1}{99}((11-33{,}5)^2 \cdot 4 + (18{,}14-33{,}5)^2 \cdot 7 + (30-33{,}5)^2 \cdot 10 +$$
$$+ (33{,}8-33{,}5)^2 \cdot 57 + (42{,}05-33{,}5)^2 \cdot 19 + (51-33{,}5)^2 \cdot 3) = 61{,}74.$$

Тоді

$$\eta_{y/x} = \sqrt{\frac{61{,}74}{98{,}75}} = \sqrt{0{,}6252} = 0{,}791.$$

Буде $\eta^2_{y/x} = 0{,}6252$.

Це означає, що 62,52 % мінливості фактора Y можна пояснити наявністю кореляційного зв'язку між Y та X, а решту 31,48 % пов'язано із впливом інших факторів, не врахованих цією моделлю.

Приклад 14.6. Перевірте гіпотезу щодо лінійної форми кореляційного зв'язку між факторами Y та X за рівнем значущості $\alpha = 0{,}05$, якщо для вибіркової сукупності обсягом $n = 100$ коефіцієнт кореляції – $r_{xy} = 0{,}526$, кореляційне відношення – $\eta_{y/x} = 0{,}625$, кількість груп за фактором X – $k = 15$.

Розв'язання. За нульовою гіпотезою H_0: $r^2 = \eta^2$, альтернативною є гіпотеза H_1: $r < \eta$.

Для перевірки припущення про лінійну форму зв'язку знайдіть емпіричне значення F-критерію за формулою (14.8):

$$F_A = \frac{(\eta^2 - r^2)/(k-2)}{(1-\eta^2)/(n-k)}, \text{ тобто } F_A = \frac{0,625^2 - 0,526^2}{1-0,625^2} \cdot \frac{100-15}{15-2} = 1,22$$

і порівняйте його із критичним.

Відповідне критичне значення за розподілом Фішера – Снедекора становить $F_{0,05}(13,85) = 1,86$ (див. додаток Ж).

Оскільки $F_A < F_{0,05}$, то з надійністю 95 % лінійна модель є адекватною, тобто відхилення зв'язку від лінійного є статистично несуттєвим. Отже, із надійністю 95 % кореляційний зв'язок між факторами можна вважати лінійним.

Приклад 14.7. Оцінки десяти студентів із вищої математики (A) та інформатики (B) за стобальною шкалою наведено в табл. 14.8.

Таблиця 14.8

Оцінки студентів із вищої математики та інформатики

A	95	90	86	84	75	70	62	60	57	50
B	92	93	83	80	55	60	45	72	62	70

Знайдіть вибірковий коефіцієнт рангової кореляції Спірмена між оцінками за двома тестами. За рівнем значущості 0,01 перевірте нульову гіпотезу про те, що генеральний коефіцієнт рангової кореляції Спірмена дорівнює нулю.

Розв'язання. Надайте ранги x_i оцінкам із вищої математики A.

Ці оцінки розташовано у спадному порядку, тому їх ранги x_i дорівнюють порядковим номерам:

Ранги x_i	1	2	3	4	5	6	7	8	9	10
Оцінки A	95	90	86	84	75	70	62	60	57	50

(*)

Надайте ранги y_i оцінкам з інформатики B, для чого спочатку розташуйте ці оцінки у спадному порядку та пронумеруйте їх:

1	2	3	4	5	6	7	8	9	10
93	92	83	80	72	70	62	60	55	45

(*)

Знайдіть ранг y_1.

Індекс i = 1 указує, що розглядають оцінку студента, який посідає з вищої математики в ряду (*) перше місце (ця оцінка дорівнює 95);

з умови видно, що з інформатики студент отримав оцінку 92, яка в ряду (**) розташована на другому місці. Таким чином ранг $y_1 = 2$.

Знайдіть ранг y_2.

Індекс i = 2 вказує, що розглядають оцінку студента, який посідає з вищої математики в ряду (*) друге місце;

з умови видно, що студент отримав з інформатики оцінку 93, яка в ряду (**) розташована на першому місці. Таким чином ранг $y_2 = 1$.

Аналогічно знайдіть інші ранги:

$y_3 = 3,\ y_4 = 4,\ y_5 = 9,\ y_6 = 8,\ \ y_7 = 10,\ y_8 = 5,\ y_9 = 7,\ y_{10} = 6$.

Записують послідовності рангів x_i та y_i:

x_i	1	2	3	4	5	6	7	8	9	10
y_i	2	1	3	4	9	8	10	5	7	6

(***)

Обчислюють різниці рангів: $d_1 = x_1 - y_1 = 1 - 2 = -1$.

Аналогічно, $d_2 = 1$, $d_3 = 0$, $d_4 = 0$, $d_5 = -4$, $d_6 = -2$, $d_7 = -3$, $d_8 = 3$, $d_9 = 2$, $d_{10} = 4$.

Сума квадратів різниць рангів:

$$\sum d_i^2 = 1 + 1 + 16 + 4 + 9 + 9 + 4 + 16 = 60.$$

Знайдіть коефіцієнт рангової кореляції Спірмена, ураховуючи, що $n = 10$:

$$r_s = 1 - \frac{6\sum d_i^2}{n^3 - n} = 1 - \frac{6 \cdot 60}{10^3 - 10} = 0,64.$$

Отже, $r_s = 0,64$.

Перевірте, чи є значущою рангова кореляційна залежність між оцінками із двох дисциплін.

Знайдіть критичну точку двобічної критичної області розподілу Стьюдента за рівнем значущості $\alpha = 0,01$ та кількістю ступенів свободи $k = n - 2 = 10 - 2 = 8$ (див. додаток Е):

$$t_{кр.}(0,01; 8) = 3,36.$$

Обчисліть значення статистики критерію:

$$t = \frac{r_s \cdot \sqrt{n-2}}{\sqrt{1 - r_s^2}} = \frac{0,64 \cdot \sqrt{8}}{\sqrt{1 - 0,64^2}} = 2,36.$$

Оскільки $t < t_{кр.}$, то немає підстав відхилити нульову гіпотезу про те, що генеральний коефіцієнт рангової кореляції Спірмена дорівнює нулю.

Отже, ранговий кореляційний зв'язок між оцінками із двох дисциплін незначущий, тобто оцінки незалежні між собою.

14.4. Вправи для самостійної роботи

14.1. Залежність швидкості руху людини Y (ум. од.) від віку X (роки) наведено статистичним розподілом вибірки (табл. 14.9).

Таблиця 14.9

Залежність швидкості руху від віку

Y	4	3	4	5	4	5	6	5	7	8	6	8	9	8	11	8	7	6	4	3
X	10	10	10	10	20	20	20	30	30	30	40	40	50	50	60	60	60	70	70	80

Побудуйте кореляційну таблицю та кореляційне поле залежності Y від X; обчисліть коефіцієнт кореляції та поясніть його значення.

14.2. Сумісний розподіл неперервних випадкових величин X, Y задано кореляційною таблицею (табл. 14.10).

Таблиця 14.10

Розподіл неперервних випадкових величин

Y \ X	0 – 2	2 – 4	4 – 6
8 – 10	1	5	
10 – 12	1	7	
12 –14		3	3

Обчисліть коефіцієнт кореляції та поясніть його значення.

14.3. Обчисліть коефіцієнт кореляції Пірсона, якщо $\sum x = 324$, $\sum y = 1975, \sum xy = 4\,197, \sum x^2 = 2\,324, \sum y^2 = 54\,834, n = 100.$

14.4. Залежність між собівартістю X та продуктивністю Y наведено в табл. 14.11.

Таблиця 14.11

Залежність між собівартістю та продуктивністю

Y = y_i, тис. грн	2,2	3,5	3,7	3,8	4,5	5,7
X = x_i, тис. шт.	1,5	1,4	1,2	1,1	0,9	0,8

Обчисліть коефіцієнт кореляції та за критерієм Стьюдента перевірте його значущість із надійністю 95 %.

14.5. За даними обчислень знайдіть кореляційне відношення, якщо $\overline{y} = 0{,}36; \sum \overline{y}_x^2 = 136{,}94; S_y^2 = 1{,}985; n = 100.$

14.6. Знайдіть значення кореляційного відношення $\eta_{y/x}$ та поясніть його, якщо $\overline{y} = 37{,}8,\ \ S_y^2 = 102{,}63$ і задано умовні середні (табл. 14.12).

Таблиця 14.12

Вихідні умовні середні

$\overline{y}\vert(X = x_i)$	17,9	22,9	35,1	41,1	46,9	59,0
m_{x_i}	3	15	19	47	13	3

14.7. Перевірте адекватність лінійної кореляційної моделі, якщо за вибірковими даними $\eta_{y/x} = 0{,}618$, $r = 0{,}503$, $n = 60$, а кількість груп за зовнішнім фактором X становить $k = 6$.

14.8. Перевірте адекватність лінійної кореляційної моделі, якщо $r = -0{,}386$, $S_y^2 = 270{,}1$ та задано умовні середні (табл. 14.13).

Таблиця 14.13

Вихідні умовні середні

$\overline{y}\vert(X = x_i)$	128,0	133,7	125,1	121,3	115,1	114,1	135,2
m_{x_i}	3	29	12	33	34	15	4

14.9. Вимірювання довжини руки X (см) і кількості влучень м'яча в кільце Y (разів) у 16 учасників спортивної секції з баскетболу дали результати, наведені в табл. 14.14.

Таблиця 14.14

Залежність між кількістю влучень та довжиною руки

X	66	61	67	73	51	59	48	47	58	44	41	54	52	47	51
Y	38	31	36	43	29	33	28	25	36	26	21	30	20	27	28

Знайдіть вибірковий коефіцієнт рангової кореляції Спірмена між цими показниками та перевірте його значущість за рівнем значущості $\alpha = 0{,}05$.

14.10. Два викладачі оцінили знання 12 учнів за стобальною шкалою та поставили їм оцінки, наведені в табл. 14.15. У першому рядку вказано кількість балів, поставлених першим викладачем, а у другому – другим).

Таблиця 14.15

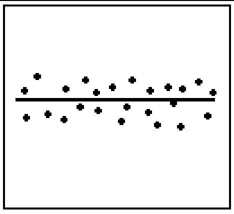
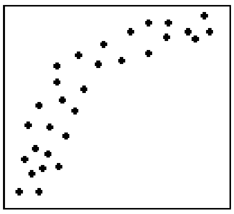
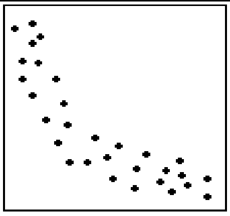
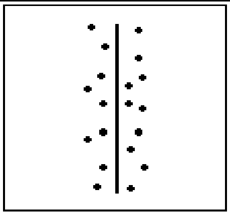
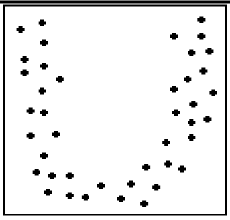
Оцінювання знань учнів першим і другим викладачами

98	94	88	80	76	70	63	61	60	58	56	51
99	91	93	74	78	65	64	66	58	53	48	62

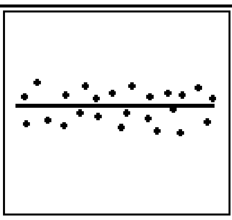
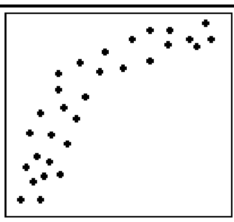
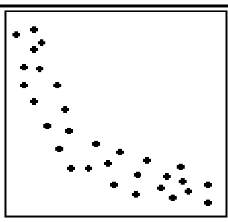
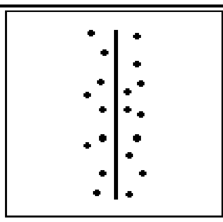
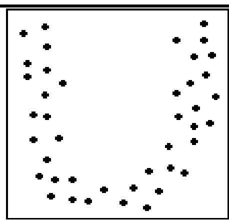
Знайдіть вибірковий коефіцієнт рангової кореляції Спірмена між оцінками двох викладачів. За рівнем значущості $\alpha = 0{,}05$ перевірте нульову гіпотезу про те, що генеральний коефіцієнт рангової кореляції Спірмена дорівнює нулю.

14.5. Тестові завдання

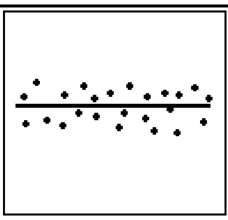
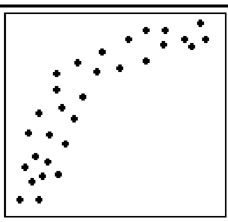
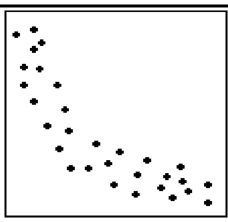
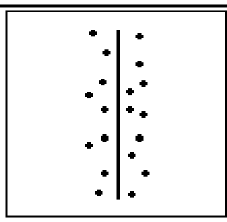
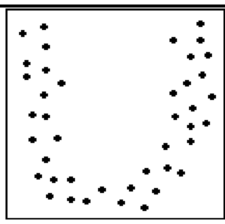
14.1. Форма зв'язку тільки спадна у випадку

А	Б	В	Г	Д
				

14.2. Кореляційний зв'язок $\overline{Y}$ від X відсутній у випадку

А	Б	В	Г	Д
				

14.3. Коефіцієнт кореляції набуває додатного значення у випадку

А	Б	В	Г	Д
				

14.4. Вибірковий коефіцієнт кореляції оцінюють за такою формулою

А $r = \dfrac{\overline{xy} - \overline{x} \cdot \overline{y}}{\sqrt{\left(\overline{x^2} - \overline{x}^2\right) \cdot \left(\overline{y^2} - \overline{y}^2\right)}}$

Б $r = \dfrac{\overline{xy} - \bar{x}\cdot\bar{y}}{\sqrt{\left(\bar{x} - x^2\right)^2 \cdot \left(\bar{y} - y^2\right)^2}}$

В $r = \dfrac{\bar{x}\cdot\bar{y} - \overline{xy}}{\sqrt{\left(\overline{x^2} - \bar{x}^2\right)\cdot\left(\overline{y^2} - \bar{y}^2\right)}}$

14.5. Якщо випадкові величини X і Y незалежні, то

А $r_{xy} = 0$

Б $r_{xy} = 1$

В $-1 \le r \le 1$

Г відповідь відсутня

14.6. Якщо випадкові величини X і Y лінійно залежні, то:

А $r_{xy} = 0$

Б $r_{xy} = 1$

В $-1 \le r \le 1$

Г відповідь відсутня

14.7. Якщо коефіцієнт кореляції дорівнює приблизно 0,68, то

А можна визнати факт існування прямого кореляційного зв'язку

Б необхідно перевірити значущість коефіцієнта кореляції та після цього формулювати статистичні висновки

В відповідь відсутня

14.8. Якщо $\eta = 1$, то

А зв'язок між X і Y функціональний

Б між X і Y немає кореляційної залежності

В відповідь відсутня

14.9. Якщо $\eta = 0$, то

А зв'язок між X і Y функціональний

Б між X і Y немає кореляційної залежності

В відповідь відсутня

14.6. Запитання для самоперевірки

14.1. Розкрийте зміст кореляційного аналізу та його завдання.

14.2. Що таке "статистична залежність", у чому полягають її особливості, порівняно з функціональною залежністю?

14.3. Як побудувати кореляційне поле? Що таке "хмара розсіювання"? Які висновки про вид кореляційного зв'язку можна зробити за формою цієї хмари?

14.4. Що таке "кореляційна залежність"? Який вигляд має кореляційна таблиця? Наведіть способи її побудови.

14.5. Як визначають вибірковий коефіцієнт кореляції Пірсона та які він має властивості?

14.6. Які існують критерії для визначення значущості кореляційного зв'язку за значенням вибіркового коефіцієнта кореляції Пірсона?

14.7. Що таке "кореляційне відношення" і як його визначають?

14.8. Як перевірити статистичну значущість кореляційного відношення? За яким критерієм здійснюють оцінювання адекватності лінійної моделі?

14.9. Наведіть формулу для обчислення рангового коефіцієнта кореляції Спірмена. Як перевірити статистичну значущість коефіцієнта Спірмена?

14.7. Висновки за темою

Кореляційний аналіз – це метод багатовимірного статистичного аналізу, який полягає в дослідженні коефіцієнтів кореляції між випадковими величинами. Завданнями кореляційного аналізу є визначення форми та тісноти кореляційного зв'язку, а також перевірка значущості впливу зміни значень однієї випадкової величини на середні значення іншої.

15. Елементи дисперсійного аналізу

15.1. Мета та компетентності

Метою вивчення теми є ознайомлення з основними принципами й особливостями дисперсійного аналізу та формування компетентностей щодо застосування дисперсійного аналізу до дослідження економічних процесів.

Компетентності, що формують під час вивчення теми:

розуміння змісту дисперсійного аналізу;

знання можливостей дисперсійного аналізу щодо визначення тісноти статистичного зв'язку;

уміння визначати тісноту статистичного зв'язку між компонентами двовимірної випадкової величини за допомогою дисперсійного аналізу.

15.2. Основні теоретичні відомості

Сутність **дисперсійного аналізу** полягає в тому, що загальну дисперсію досліджуваної ознаки розподіляють на окремі компоненти, обумовлені впливом певних конкретних факторів. Істотність їхнього впливу на цю ознаку визначають методом дисперсійного аналізу.

У разі здійснення дисперсійного аналізу досліджуваний масив даних, визначених під час експерименту, розподіляють на певні групи, які різняться дією на результати експерименту певних рівнів факторів.

Уважають, що досліджувана ознака має нормальний закон розподілу, а дисперсії в кожній окремій групі здобутих значень ознаки однакові. Ці припущення необхідно перевірити.

Дисперсійний аналіз, який досліджує наявність зв'язку між компонентами двовимірної випадкової величини, називають **однофакторним**, тобто він розглядає вплив на ознаку лише одного зовнішнього фактора.

У процесі однофакторного дисперсійного аналізу масив емпіричних даних складається зі значень однієї з компонент двовимірної випадкової величини, яку називають **ознакою**. Ці значення розподіляють на групи за рівнем фактора, що впливає на ознаку – це друга компонента двовимірної випадкової величини.

Результати експерименту зручно подавати у вигляді таблиці (табл. 15.1).

Таблиця 15.1

Таблиця вихідних даних однофакторного дисперсійного аналізу

Рівень фактора (номер групи)	Емпіричні значення якісної ознаки X за групами	Групові середні значення	Загальна (вибіркова) середня ознаки X
1	2	3	4
1	$x_{11}, x_{12}, x_{13}, ..., x_{1n_1}$	$\overline{x_1} = \frac{1}{n_1}\sum_{j=1}^{n_1} x_{1j}$	$\overline{x} = \frac{\sum_{i=1}^{p}\sum_{j=1}^{n} x_{ij}}{\sum_{i=1}^{p} n_i}$,

1	2	3	4
2	$x_{21}, x_{22}, x_{23}, ..., x_{2n_2}$	$\overline{x_2} = \frac{1}{n_2}\sum_{j=1}^{n_2} x_{2j}$	$\sum_{i=1}^{p} n_i = N$
...	...	...	
p	$x_{p1}, x_{p2}, x_{p3}, ..., x_{pn_p}$	$\overline{x_p} = \frac{1}{n_p}\sum_{j=1}^{n_p} x_{pj}$	

Згідно з положеннями дисперсійного аналізу, **загальну суму квадратів відхилень** значень випадкової величини X від її вибіркової середньої, що характеризує загальне розсіювання значень випадкової величини, надають у вигляді суми двох компонентів:

$$SST = SSR + SSE, \tag{15.1}$$

де SST – **загальна сума квадратів відхилень** значень випадкової величини від її вибіркової середньої (*sum of squares total*);

SSR – **сума квадратів відхилень, пов'язаних із регресією (факторна)**, тобто із впливом досліджуваного фактора (*sum of squares by regression*);

SSE – **сума квадратів помилок (залишкова)**, тобто відхилень, пов'язаних із факторами, які не розглядають у межах заданої моделі, і які вважають похибками (*sum of squares by errors*).

Загальну суму квадратів відхилень значень випадкової величини від її вибіркової середньої визначають за такою формулою:

$$SST = \sum_{i=1}^{p}\sum_{j=1}^{n_i}(x_{ij} - \bar{x})^2, \tag{15.2}$$

де $\bar{x}$ – загальна (вибіркова) середня, тобто середня вибіркової сукупності випадкових величин $X_1, X_2, ..., X_p$ обсягом $N = \sum_{i=1}^{p} n_i$.

Суму квадратів відхилень, пов'язаних із впливом зовнішнього фактора, визначають як виважену суму квадратів відхилень вибіркової середньої кожної з випадкових величин X_i $(i = \overline{1, p})$ від загальної середньої ознаки X:

$$SSR = \sum_{i=1}^{p} n_i \cdot \left(\overline{x_i} - \overline{x}\right)^2, \tag{15.3}$$

де $\overline{x_i}$ – вибіркова середня випадкової величини X_i $(i = \overline{1, p})$.

Суму квадратів помилок визначають за такою формулою:

$$SSE = \sum_{i=1}^{p} \sum_{j=1}^{n_i} \left(x_{ij} - \overline{x_i}\right)^2. \tag{15.4}$$

Для перевірки статистичної гіпотези щодо однорідності всієї вибіркової сукупності, тобто для перевірки нульової гіпотези H_0, розглядають випадкову величину:

$$F = \frac{SSR/(p-1)}{SSE/(N-p)}, \tag{15.5}$$

що має розподіл Фішера – Снедекора.

За значеннями α, $k_1 = p-1$, $k_2 = N-p$, знаходять критичну точку $F_\alpha(k_1; k_2)$ (див. додатки Ж, И).

Якщо емпіричне значення критерію, що обчислюють за формулою (15.5), не перевищує критичне значення $F_\alpha(k_1; k_2)$, то нульову гіпотезу про рівність групових середніх немає підстав відхилити, отже, вплив зовнішнього фактора на ознаку X слід уважати статистично незначущим.

Якщо емпіричне значення критерію перевищує критичне, то з надійністю $\gamma = 1-\alpha$ нульову гіпотезу відхиляють, тобто підтверджують наявність впливу певного зовнішнього фактора на ознаку X.

Результати однофакторного дисперсійного аналізу зручно подавати у вигляді таблиці (табл. 15.2).

Таблиця 15.2

Форма подання результатів однофакторного дисперсійного аналізу

Вид варіації ознаки	Сума квадратів відхилень	Кількість ступенів свободи	Статистична оцінка дисперсії	Значення критерію Фішера – Снедекора	
				емпіричне	критичне
Міжгрупова	$\sum_{i=1}^{p} n_i \cdot \left(\overline{x}_i - \overline{x}\right)^2$	$p-1$	$\frac{SSR}{p-1}$	F	$F_{\alpha}(k_1; k_2)$
Внутрішньо-групова	$\sum_{i=1}^{p} \sum_{j=1}^{n_i} \left(x_{ij} - \overline{x}_i\right)^2$	$N-p$	$\frac{SSE}{N-p}$		
Загальна	$\sum_{i=1}^{p} \sum_{j=1}^{n_i} \left(x_{ij} - \overline{x}\right)^2$	$N-1$	$\frac{SST}{N-1}$		

Нехай необхідно дослідити наявність впливу факторів A і B на ознаку X. Для цього застосовують **двофакторний дисперсійний аналіз**. Розв'язання цієї задачі передбачає виконання певного обсягу обчислень. Результати вимірювань та обчислень двофакторного аналізу зручно подавати в таблицях (табл. 15.3 і 15.4).

Спочатку для кожної пари значень факторів A і B обчислюють блокові середні, а також визначають середні значення ознаки за стовпцями та за рядками. Ці значення порівнюють із загальною вибірковою середньою ознаки. За всіма блоковими середніми обчислюють виправлені дисперсії. Відповідно визначають: виправлену дисперсію, зумовлену впливом на ознаку X лише фактора A; виправлену дисперсію, зумовлену впливом на ознаку X лише фактора B; виправлену дисперсією, зумовлену одночасним впливом на ознаку X і фактора A, і фактора B; виправлену дисперсію, обумовлену дією на ознаку X факторів, вплив яких не розглядають у межах заданої моделі.

Останній рядок табл. 15.4. містить формули для обчислення загальної суми квадратів відхилень, яка має дорівнювати сумі всіх квадратів відхилень, пов'язаних із впливом на ознаку кожного з факторів окремо, обох факторів у сукупності, а також розсіювання ознаки під впливом факторів, які не враховують у межах заданої моделі.

Таблиця 15.3

Середні значення двофакторного дисперсійного аналізу

Рівень фактора A	Рівень фактора B				Середня величина за рядками	Загальна середня величина
	B_1	B_2	...	B_q		
	блокова середня	блокова середня	...	блокова середня		
A_1	$\bar{x}_{11} = \frac{\sum_{i=1}^{n} x_{i11}}{n}$	$\bar{x}_{12} = \frac{\sum_{i=1}^{n} x_{i12}}{n}$	...	$\bar{x}_{1q} = \frac{\sum_{i=1}^{n} x_{i1q}}{n}$	$\bar{y}_1 = \frac{\sum_{i=1}^{n}\sum_{k=1}^{q} x_{i1k}}{nq}$	$\bar{x} = \frac{\sum_{i=1}^{n}\sum_{j=1}^{p}\sum_{k=1}^{q} x_{ikj}}{npq}$
A_2	$\bar{x}_{21} = \frac{\sum_{i=1}^{n} x_{i21}}{n}$	$\bar{x}_{22} = \frac{\sum_{i=1}^{n} x_{i22}}{n}$	...	$\bar{x}_{2q} = \frac{\sum_{i=1}^{n} x_{i2q}}{n}$	$\bar{y}_2 = \frac{\sum_{i=1}^{n}\sum_{k=1}^{q} x_{i2k}}{nq}$	
	...	...	...	...	...	
A_p	$\bar{x}_{p1} = \frac{\sum_{i=1}^{n} x_{ip1}}{n}$	$\bar{x}_{p2} = \frac{\sum_{i=1}^{n} x_{ip2}}{n}$	...	$\bar{x}_{pq} = \frac{\sum_{i=1}^{n} x_{ipq}}{n}$	$\bar{y}_p = \frac{\sum_{i=1}^{n}\sum_{k=1}^{q} x_{ipk}}{nq}$	
Середня величина за стовпцями	$\bar{z}_1 = \frac{\sum_{i=1}^{n}\sum_{j=1}^{p} x_{i1j}}{np}$	$\bar{z}_2 = \frac{\sum_{i=1}^{n}\sum_{j=1}^{p} x_{i2j}}{np}$	...	$\bar{z}_q = \frac{\sum_{i=1}^{n}\sum_{j=1}^{p} x_{iqj}}{np}$	x_{ikj} – конкретне значення ознаки X, якого вона набуває в разі i-го експерименту, за j-м рівнем фактора A і k-м рівнем фактора B	

Таблиця 15.4

Результати обчислень двофакторного дисперсійного аналізу

Джерело, що спонукає до розсіювання	Сума квадратів відхилень	Кількість ступенів свободи	Статистичні оцінки дисперсій (виправлені дисперсії)
Фактор A	$Q_1 = nq\sum\limits_{j=1}^{p}(\bar{y}_j - \bar{x})^2$	$k_1 = p - 1$	$S_1^2 = \dfrac{Q_1}{p-1}$
Фактор B	$Q_2 = np\sum\limits_{j=1}^{q}(\bar{z}_j - \bar{x})^2$	$k_2 = q - 1$	$S_2^2 = \dfrac{Q_2}{q-1}$
Одночасна дія факторів A і B	$Q_3 = \sum\limits_{j=1}^{q}\sum\limits_{k=1}^{p}(\bar{x}_{kj} - \bar{z}_j - \bar{y}_i + \bar{x})^2$	$k_3 = (p-1)(q-1)$	$S_3^2 = \dfrac{Q_3}{(p-1)(q-1)}$
Дія випадкової компоненти	$Q_4 = \sum\limits_{i=1}^{n}\sum\limits_{j=1}^{q}\sum\limits_{k=1}^{p}(x_{ikj} - \bar{x}_{kj})^2$	$k_4 = N - pq$	$S_4^2 = \dfrac{Q_4}{N - pq}$
Загальне відхилення	$Q = \sum\limits_{i=1}^{n}\sum\limits_{j=1}^{q}\sum\limits_{k=1}^{p}(x_{ikj} - \bar{x})^2$	$N - 1$	$S^2 = \dfrac{Q}{N-1}$

Має виконуватися тотожність:

$$S^2 = S_1^2 + S_2^2 + S_3^2 + S_4^2.$$

За даними табл. 15.4. визначають емпіричні значення критерію Фішера:

$$F_A = \frac{S_1^2}{S_4^2};\ \ F_B = \frac{S_2^2}{S_4^2};\ \ F_{AB} = \frac{S_3^2}{S_4^2}.$$

Емпіричні значення критерію Фішера для перевірки значущості впливу факторів A і B на ознаку X порівнюють із критичними точками статистики Фішера – Снедекора.

Критичні точки визначають за довідковою таблицею (див. додатки Ж, И) за рівнем значущості α і кількістю ступенів свободи, відповідно: $F_\alpha(k_1;k_4)$; $F_\alpha(k_2;k_4)$ та $F_\alpha(k_3;k_4)$.

Якщо $F_A > F_\alpha(k_1;\,k_4)$, то нульову гіпотезу про відсутність впливу на ознаку X фактора A відхиляють.

Якщо $F_B > F_\alpha(k_2;\,k_4)$, то нульову гіпотезу про відсутність впливу фактора B на ознаку X відхиляють.

Якщо $F_{AB} > F_\alpha(k_3;\,k_4)$, то нульову гіпотезу про відсутність спільного впливу факторів A і B відхиляють.

15.3. Приклади розв'язування задач

Приклад 15.1. Ступінь впливу мінеральних добрив на врожайність, наведено в табл. 15.5.

Таблиця 15.5

Аналіз впливу мінеральних добрив на врожайність

Ступінь впливу мінеральних добрив	Урожайність (т)
1	3,2; 3,1; 3,1; 2,8; 3,3; 3,0
2	2,6; 3,1; 2,7; 2,9; 2,7; 2,8
3	2,9; 2,6; 3,0; 3,1; 3,0; 2,8
4	3,7; 3,4; 3,2; 3,3; 3,5; 3,3
5	3,0; 3,4; 3,2; 3,5; 2,9; 3,1

З'ясуйте, чи істотно впливають мінеральні добрива на врожайність за α = 0,01.

Розв'язання. Використовуючи табл. 15.1 і 15.2, виконують відповідні обчислення. Результат наведено в табл. 15.6.

Таблиця 15.6

Дисперсійний аналіз впливу мінеральних добрив на врожайність

Ступінь впливу мінеральних добрив	Спостережуване значення (врожайність)	Групові середні	Загальна середня
1	3,2; 3,1; 3,1; 2,8; 3,3; 3,0	$\bar{x}_1$ = 3,083	
2	2,6; 3,1; 2,7; 2,9; 2,7; 2,8	$\bar{x}_2$ = 2,8	
3	2,9; 2,6; 3,0; 3,1; 3,0; 2,8	$\bar{x}_3$ = 2,9	$\bar{x}$ = 3,073
4	3,7; 3,4; 3,2; 3,3; 3,5; 3,3	$\bar{x}_4$ = 3,4	
5	3,0; 3,4; 3,2; 3,5; 2,9; 3,1	$\bar{x}_5$ = 3,18	
Вид варіації ознаки	Сума квадратів відхилень	Кількість ступенів свободи	Статистичні оцінки дисперсій
Міжгрупова	$\sum_{i=1}^{5} n_i \cdot \left(\bar{x}_i - \bar{x}\right)^2 = 1{,}338$	$p-1=$ $=5-1=4$	$\frac{SSR}{p-1}=0{,}335$
Внутрішньо-групова	$\sum_{i=1}^{5}\sum_{j=1}^{n_i}\left(x_{ij} - \overline{x_i}\right)^2 = 0{,}927$	$N-p=$ $=30-5=25$	$\frac{SSE}{N-p}=0{,}037$

За формулою (15.5) обчисліть:

$$F = \frac{0{,}335}{0{,}037} = 9{,}02.$$

Критичне значення: $F_{кр.}(\alpha = 0{,}01;\ k_1 = 4;\ k_2 = 25) = 6{,}6$.

Оскільки $F > F_{кр.}$, то вплив мінеральних добрив на врожайність є істотним.

Приклад 15.2. У табл. 15.7 наведено відсотки браку шліфування 36 деталей, залежно від швидкості (фактор A) та потужності (фактор B) шліфувального верстата.

За рівнем значущості $\alpha = 0{,}05$ перевірте, чи є статистично значущим вплив факторів A і B окремо та спільний вплив факторів A і B на результат шліфування деталей.

Таблиця 15.7

Брак шліфування деталей, %

Ступінь впливу фактора A	Ступінь впливу фактора B		
	B_1	B_2	B_3
A_1	3,6; 3,9; 4,1	2,9; 3,1; 3,0	2,7; 2,5; 2,9
A_2	4,2; 4,0; 4,1	3,3; 2,9; 3,2	3,7; 3,5; 3,6
A_3	3,8; 3,5; 3,6	3,6; 3,7; 3,5	3,2; 3,0; 3,4
A_4	3,4; 3,2; 3,2	3,4; 3,6; 3,5	3,6; 3,8; 3,7

Розв'язання. Використовуючи дані табл. 15.7, обчисліть і наведіть у вигляді таблиці значення блокових середніх $\overline{x_{ij}}$, умовних середніх $\overline{y_i}$ ($i = \overline{1,4}$), що відповідають сталому рівню фактора A, та умовних середніх $\overline{z_j}$ ($j = \overline{1,3}$), що відповідають сталому рівню фактора B, а також загальної вибіркової середньої $\overline{x}$ (табл. 15.8).

Таблиця 15.8

Результати обчислення середніх

Ступінь впливу фактора A	Ступінь впливу фактора B			$\overline{y_i}$
	B_1	B_2	B_3	
1	2	3	4	5
A_1	3,8(6)	3,0	2,7	3,189
A_2	4,1	3,1(3)	3,6	3,611
A_3	3,6(3)	3,6	3,2	3,478

Закінчення табл. 15.8

1	2	3	4	5
A_4	3,2(6)	3,5	3,7	3,442
$\overline{z_j}$	3,717	3,308	3,3	–
$\overline{x}$	3,442			

За табл. 15.8 визначте виправлені дисперсії, формули яких наведено в табл. 15.4:

$$S_1^2 = \frac{nq}{p-1}\sum_{i=1}^{p}\left(\overline{y_i} - \overline{x}\right)^2 = 0{,}2884;$$

$$S_2^2 = \frac{np}{q-1}\sum_{j=1}^{q}\left(\overline{z_j} - \overline{x}\right)^2 = 0{,}6808;$$

$$S_3^2 = \frac{n}{(p-1)(q-1)}\sum_{i=1}^{p}\sum_{j=1}^{q}\left(\overline{x_{ij}} - \overline{y_i} - \overline{z_j} + \overline{x}\right)^2 = 0{,}4790;$$

$$S_4^2 = \frac{1}{N-pq}\sum_{i=1}^{p}\sum_{j=1}^{q}\sum_{k=1}^{n}\left(x_{ijk} - \overline{x_{ij}}\right)^2 = 0{,}0236;$$

$$S^2 = \sum_{i=1}^{p}\sum_{j=1}^{q}\sum_{k=1}^{n}\left(x_{ijk} - \overline{x}\right)^2 = 5{,}6675.$$

Перевірте виконання тотожності:

$$S^2 = S_1^2 + S_2^2 + S_3^2 + S_4^2.$$

Дійсно: $S^2 = 5{,}6675 = 0{,}8653 + 1{,}3617 + 2{,}8739 + 0{,}5667$.

Отже, обчислення не містять помилок.

Тепер визначте емпіричні значення критерію Фішера – Снедекора за кожним із факторів $A,\ B$ та за їхньою сумісною дією:

$$F_A = \frac{S_1^2}{S_4^2} = \frac{0{,}2884}{0{,}0236} = 12{,}22;$$

$$F_B = \frac{S_2^2}{S_4^2} = \frac{0,6808}{0,0236} = 28,84;$$

$$F_{AB} = \frac{S_3^2}{S_4^2} = \frac{0,479}{0,0236} = 20,30.$$

За таблицею критичних точок розподілу Фішера – Снедекора (див. додаток Ж) визначте критичні точки статистичного критерію.

Буде: $F_{0,05}(2;\ 24) = 3,40$; $F_{0,05}(3;\ 24) = 3,01$; $F_{0,05}(6;\ 24) = 2,51$.

Отже, $F_A > F_{0,05}(3;\ 24)$, $F_B > F_{0,05}(2;\ 24)$, $F_{AB} > F_{0,05}(6;\ 24)$.

Тому всі нульові гіпотези відхиляють, тобто із надійністю 95 % вплив фактора A і фактора B окремо, а також вплив факторів A і B у сукупності слід уважати суттєвими.

15.4. Вправи для самостійної роботи

15.1. У результаті проведення досліду, із метою з'ясування впливу чорного пару на врожайність пшениці з ділянки 9 га (3 га були під чорним паром; 3 га – під картоплею; 3 га – під кормовими травами), дістали такі результати (табл. 15.9).

Таблиця 15.9

Вплив чорного пару на врожайність пшениці

Фактори	Урожайність, ц/га
Чорний пар	26,6; 26,6; 30,6
Площа під картоплею	24,3; 25,2; 25,2
Площа під кормовими травами	26,6; 28,0; 31,0

За рівнем значущості α = 0,01 перевірте істотність впливу чорного пару на врожайність пшениці.

Відповідь: вплив чорного пару на врожайність пшениці є істотним.

15.2. Експериментально досліджувався вплив на зносостійкість колінчастих валів технології їхнього виготовлення: вплив фактора *А*, який має чотири рівні, тобто застосовували чотири технології виготовлення валів.

Визначені результати наведено в табл. 15.10.

Таблиця 15.10

Вплив технології виготовлення на зносостійкість колінчастих валів

Ступінь впливу фактора A	Кількість відпрацьованих місяців
A_1	9; 8; 10; 12
A_2	10; 12; 11; 8
A_3	8; 16; 10; 18
A_4	9; 18; 10; 8

За рівнем значущості $\alpha = 0{,}01$ перевірте вплив технологій на зносостійкість валів.

Відповідь: вплив технологій на зносостійкість є неістотним.

15.3. Для перевірки впливу методики навчання виробничих навичок на якість підготовки із випускників виробничо-технічного училища навмання вибирають чотири групи учнів, які після закінчення навчання за різними методиками тестувалися на кількість виготовлених однотипних деталей протягом робочої зміни.

Результати тестування наведено в табл. 15.11.

Таблиця 15.11

Вплив методики навчання на якість підготовки

Ступінь впливу фактора A (методики)	Кількість виготовлених деталей за робочу зміну
A_1	60; 80; 75; 80; 85; 70
A_2	75; 66; 85; 80; 70; 80; 90
A_3	60; 80; 65; 60; 86; 75
A_4	95; 85; 100; 80

За рівнем значущості $\alpha = 0{,}05$ з'ясуйте вплив методики навчання на якість підготовки учнів.

Відповідь: вибір методики впливає на виробничі навички підготовлених фахівців.

15.4. Із кожної з 8 партій однотипних заготовок навмання бралися заготовки, які обробляли на трьох верстатах-автоматах різної модифікації. Кількість деталей, виготовлених верстатами, досліджували на стандартність.

Результати досліджень подано в табл.15.12.

Таблиця 15.12

Вплив модифікації верстатів-автоматів на якість деталей

Фактор *A* (тип верстатів-автоматів)	Кількість деталей, виготовлених верстатами-автоматами, що відповідають стандарту
A_1	100; 86; 90; 89; 95; 22; 80; 79
A_2	99; 82; 98; 88; 100; 96; 98; 100
A_3	100; 88; 86; 98; 98; 100; 99; 99

За рівнем значущості α = 0,01 перевірте вплив модифікації верстатів-автоматів на якість виготовлених деталей.

Відповідь: вплив неістотний.

15.5. Досліджують вплив на зносостійкість деталей таких факторів: *A* – матеріал для виготовлення деталей (застосовували три види сталі) і *B* – технологія виготовлення деталей (дві технології).

Результати досліджень наведено в табл. 15.13.

За рівнем значущості α = 0,05 визначте вплив факторів *A* і *B* та їхній спільний вплив на досліджувану ознаку.

Таблиця 15.13

Аналіз факторів впливу на зносостійкість деталей

Фактор *B*	Фактор *A*		
	A_1	A_2	A_3
1	2	3	4
B_1	10; 7; 8; 6; 12; 8; 11; 10; 14; 13	8; 14; 6; 10; 16; 14; 13; 12; 11; 15	15; 12; 11; 9; 8; 13; 11; 12; 16; 14

Закінчення табл. 15.13

1	2	3	4
B_2	12; 13; 6; 9; 8; 11; 10; 10; 13; 17	11; 12; 12; 16; 13; 8; 10; 9; 8; 15	13; 12; 14; 8; 6; 8; 16; 12; 14; 16

15.6. Досліджують вплив факторів *A* і *B* на кількість виготовлених втулок зі ста взятих, які відповідають нормам стандарту: *A* – використали дві технології виготовлення; *B* – заготовки надходили із трьох заводів. Результати досліджень наведено в табл. 15.14:

Таблиця 15.14

Аналіз факторів впливу на кількість виготовлення деталей

Фактор *B*	Фактор *A*	
	A_1	A_2
B_1	90; 88; 90; 96; 98; 76; 80; 95; 85; 80	100; 99; 82; 98; 95; 80; 96; 95; 99; 91; 89; 90
B_2	79; 88; 92; 76; 80; 83; 85; 90; 96; 75	81; 82; 100; 98; 89; 85; 96; 98; 75; 97
B_3	82; 78; 75; 79; 80; 81; 86; 89; 75; 90	80; 86; 90; 91; 78; 76; 75; 82; 73; 82

За рівнем значущості $\alpha = 0{,}05$ визначте вплив факторів *A* і *B* та їхній сумісний вплив на досліджувану ознаку.

15.5. Тестові завдання

15.1. У моделі однофакторного дисперсійного аналізу міжгрупову дисперсію зумовлено

А впливом досліджуваного фактора на ознаку *X*

Б впливом інших випадкових факторів

В відповідь відсутня

15.2. У моделі однофакторного дисперсійного аналізу внутрішньогрупову дисперсію зумовлено

А впливом досліджуваного фактора на ознаку *X*

Б впливом інших випадкових факторів

В відповідь відсутня

15.3. Якщо досліджуваний фактор не впливає на значення ознаки *X*, то в цьому разі

А $\frac{SSR}{p-1}$ і $\frac{SSE}{N-p}$ можна розглядати як незалежні оцінки загальної дисперсії *D*

Б вибірки слід уважати здійсненими з різних сукупностей

В відповідь відсутня

15.4. Які можливості двофакторного дисперсійного аналізу

А дозволяє оцінити лише вплив кожного з факторів

Б дозволяє оцінити лише взаємодію факторів

В дозволяє оцінити вплив кожного з факторів і їхню взаємодію

15.5. Якщо $F_A^* > F_{кр.}(\alpha; k_4, k_1)$, то

А нульову гіпотезу про відсутність впливу фактора *A* відхиляють

Б нульову гіпотезу про відсутність впливу фактора *A* не відхиляють

В нульову гіпотезу про відсутність спільного впливу факторів *A* і *B* відхиляють

15.6. Якщо $F_B^* > F_{кр.}(\alpha; k_3, k_1)$, то

А нульову гіпотезу про відсутність впливу фактора *B* відхиляють

Б нульову гіпотезу про відсутність впливу фактора *B* не відхиляють

В нульову гіпотезу про відсутність спільного впливу факторів *A* і *B* відхиляють

15.7. Якщо $F_{AB}^* > F_{кр.}(\alpha; k_2, k_1)$, то

А нульову гіпотезу про відсутність впливу фактора *B* відхиляють

Б нульову гіпотезу про відсутність спільного впливу факторів *A* і *B* не відхиляють

В нульову гіпотезу про відсутність спільного впливу факторів *A* і *B* відхиляють

15.6. Запитання для самоперевірки

15.1. Розкрийте можливості дисперсійного аналізу, його основні принципи.

15.2. Що таке "вибіркова", "групова" та "загальна" середні?

15.3. Що таке "групова", "внутрішньогрупова", "міжгрупова" та "загальна" дисперсії?

15.4. Сформулюйте загальний принцип обчислення статистичного критерію, за яким перевіряють нульову гіпотезу в дисперсійному аналізі.

15.5. Суми квадратів яких відхилень розглядають в однофакторному дисперсійному аналізі? Поясніть їхній зв'язок із питомими дисперсіями.

15.6. Яку математичну модель має двофакторний дисперсійний аналіз?

15.7. Наведіть формули для обчислення блокових середніх і середніх значень ознаки за рядками та стовпцями у двофакторному аналізі.

15.8. За якими формулами обчислюють спостережувані значення статистичних критеріїв за кожним фактором A і B та їхню спільну дію?

15.7. Висновки за темою

Дисперсійний аналіз є методом багатовимірної статистики, який дозволяє здійснювати дослідження впливу якісних факторів та аналізу значущості цього впливу.

Основною метою дисперсійного аналізу є дослідження значущості відмінності між середніми декількох груп даних або змінних.

Однофакторний дисперсійний аналіз використовують у дослідженнях зміни результативної ознаки під впливом зміни умов або градацій фактора. Двофакторний дисперсійний аналіз застосовують у дослідженнях одночасної дії двох факторів на різні вибірки об'єктів.

Сенс методу полягає в тому, що загальна дисперсія фактора, що характеризує певний процес, тобто внутрішнього фактора системи, є сумою двох компонентів: дисперсії, обумовленої впливом зовнішніх контрольованих факторів, і дисперсії, пов'язаної з похибками моделі, тобто обумовленої впливом факторів, які не розглядають у межах цієї моделі.

16. Елементи теорії регресії

16.1. Мета та компетентності

Метою вивчення теми є ознайомлення із принципами побудови моделі парної регресії та формування компетентностей щодо застосування регресійного аналізу для побудови моделей економічних явищ і прогнозування за цими моделями.

Компетентності, що формують під час вивчення теми:

розуміння змісту економічних величин, що входять до складу моделі парної регресії;

знання методів перевірки значущості параметрів моделі регресії й оцінювання адекватності моделі;

уміння будувати рівняння парної лінійної регресії, здійснювати прогноз за цим рівнянням і визначати його точність.

16.2. Основні теоретичні відомості

У разі кореляційної залежності Y від X будують рівняння умовного середнього $\overline{y}_x$ від x: $\overline{y}_x = f(x)$.

За результатами одного й того самого випробування можуть бути визначені дві залежності: $\overline{y}_x$ та $\overline{x}_y$, тобто треба вказати, яку зі змінних уважати фактором, а яку функціональною ознакою (де причина, а де наслідок).

Ламану лінію $\left(\overline{y}_x\right)$, яка з'єднує точки з координатами $\left(x_i, \overline{y}_{x_i}\right)$, називають **емпіричною лінією регресії** Y на X. Ламану лінію $\left(\overline{x}_y\right)$, яка з'єднує точки з координатами $\left(\overline{x}_{y_k}, y_k\right)$, називають **емпіричною лінією регресії** X на Y.

Граничне положення емпіричної лінії регресії, до якої вона наближається за необмеженої кількості спостережень, називають граничною **теоретичною лінією регресії**.

Рівняння $\overline{y}_x = f(x)$ називають **рівнянням регресії** Y на X, функцію $f(x)$ – **регресією** Y на X, а графік – **лінією регресії** Y на X. Можна ввести й іншу залежність: $\overline{x}_y = \varphi(y)$ – рівняння регресії X на Y.

Поняття регресії та кореляції безпосередньо пов'язані між собою. Тоді як у кореляційному аналізі оцінюють силу стохастичного зв'язку, у регресійному – досліджують його форму. Обидва види аналізу призначено для встановлення причинних співвідношень між явищами та означення наявності або відсутності зв'язку.

Розрізняють такі види кореляції та регресії:

1) за *характером:* **додатна** регресія (коли зі зростанням (зменшенням) аргументу x зростає (зменшується) значення y) і **від'ємна** регресія (коли зі зростанням (зменшенням) значення аргументу x значення y зменшуються (зростають));

2) за *кількістю змінних:* **парна** та **множинна** кореляція й регресія;

3) за *формою зв'язку:* **лінійна** та **нелінійна** кореляція й регресія.

Під час установлення форми кореляційного зв'язку між двома змінними перевагу віддають лінійній залежності:

$$\hat{y}_x = b_1 x + b_0 \text{ або } \hat{x}_y = a_1 y + a_0 .$$

Параметри рівняння регресії $\hat{y}_x = b_1 x + b_0$ визначають за методом найменших квадратів так, щоб сума квадратів відхилень $\hat{y}_i$ від $y_{i\ емп.}$ була найменшою.

Параметр $b_1 = \dfrac{\overline{xy} - \overline{x} \cdot \overline{y}}{\overline{x^2} - \left(\overline{x}\right)^2}$ називають **коефіцієнтом регресії** Y на X і позначають $\rho_{y/x}$. Коефіцієнт регресії за своїм знаком збігається зі знаком коваріації μ_{xy} = $\overline{xy} - \overline{x} \cdot \overline{y}$: якщо знак $\rho_{y/x}$ "+", то регресія додатна, якщо знак "–", то регресія від'ємна.

Таким чином:

$$\rho_{y/x} = \frac{\mu_{xy}}{\sigma_x^2},\ b_0 = \overline{y} - \rho_{y/x} \overline{x} . \qquad (16.1)$$

Коефіцієнт регресії $\rho_{y/x}$ – це величина розмірна. Він показує, на скільки одиниць свого вимірювання зменшиться середнє значення функціональної ознаки $\hat{y}_x$, якщо фактор X збільшиться на одиницю свого вимірювання.

Аналогічно можна визначити теоретичне рівняння регресії X на Y:

$$\hat{x} = \rho_{x/y} y + a_0, \tag{16.2}$$

де $\rho_{x/y} = \dfrac{\mu_{xy}}{\sigma_y^2}$; $\sigma_y^2 = \overline{y^2} - \left(\bar{y}\right)^2$; $a_0 = \bar{x} - \rho_{x/y}\bar{y}$.

Емпіричну та теоретичну лінії регресії зображують на кореляційному полі. Якщо параметри теоретичного рівняння регресії обчислено правильно, то помилки незначні.

Точка $M(\bar{x}, \bar{y})$ є центром регресії, крізь який проходять обидві теоретичні лінії регресії.

Оскільки

$$\rho_{y/x} = \frac{\mu_{xy}}{\sigma_x^2},\ \rho_{x/y} = \frac{\mu_{xy}}{\sigma_y^2}, \text{ то } \rho_{y/x} \cdot \rho_{x/y} = \frac{\mu_{xy}^2}{\sigma_x^2 \cdot \sigma_y^2} = r^2.$$

Звідки:

$$r_{xy} = \pm\sqrt{\rho_{y/x} \cdot \rho_{x/y}}, \tag{16.3}$$

де знак "+", якщо $\rho_{y/x}$, $\rho_{x/y}$ додатні;

знак "–", якщо $\rho_{y/x}$, $\rho_{x/y}$ від'ємні.

Кут між теоретичними лініями регресії $\hat{y}_x$ та $\hat{x}_y$ визначають за такою формулою:

$$tg\varphi = \frac{1 - r_{xy}^2}{r_{xy}} \cdot \frac{\sigma_x \sigma_y}{\sigma_x^2 + \sigma_y^2}. \tag{16.4}$$

Якщо $|r| = 1$, то $tg\varphi = 0$, $\varphi = 0$, і лінії регресії $\hat{y}_x$ та $\hat{x}_y$ збігаються (зв'язок лінійний);

якщо $r = 0$, то $tg\varphi = \infty$, $\varphi = \pi/2$, і лінії регресії взаємно перпендикулярні (немає кореляційного зв'язку);

чим більше значення коефіцієнта кореляції, тим ближче одна до одної розміщені спряжені теоретичні лінії регресії;

чим слабкіша кореляційна залежність, тим більший кут між спряженими теоретичними лініями регресії $\hat{y}_x$ та $\hat{x}_y$.

Для кожного з параметрів моделі $\hat{y}_x = b_1 x + b_0$ виконують перевірку, чи є параметр, значення якого обчислювали за вибірковою сукупністю, **статистично значущим**.

Нульову гіпотезу H_0: $b_1 = 0$ за альтернативної гіпотези H_1: $b_1 \neq 0$ перевіряють за допомогою статистики:

$$t_{b_1} = \frac{b_1}{\sigma_{b_1}}, \tag{16.5}$$

де $\sigma_{b_1} = \sqrt{\dfrac{\sigma_e^2}{\sum\limits_{i=1}^{n}(x_i - \bar{x})^2}}$ – стандартне відхилення для параметра b_1;

$\sigma_e = \sqrt{\dfrac{1}{n-2}\sum\limits_{i=1}^{n}(y_i - \hat{y}_i)^2}$ – незсунута оцінка середнього квадратичного відхилення випадкової величини e (помилок моделі).

Критичні значення випадкової величини t у разі невеликого обсягу вибірки визначають за таблицею розподілу Стьюдента (див. додаток Е), відповідно до рівня значущості α та кількості ступенів свободи $k = n - 2$. Будують двобічну критичну область.

Зі збільшенням обсягу вибірки критичне значення t_α є аргументом функції Лапласа $\Phi(t_\alpha)$ (див. додаток В), відповідно до умови $\Phi(t_\alpha) = \dfrac{1-\alpha}{2}$.

Якщо $|t_{b_1}| < t_{0,05}$, то за рівнем значущості $\alpha = 0{,}05$ статистичну гіпотезу H_0 немає підстав відхилити, тоді як за умови $|t_{b_1}| > t_{0,01}$ гіпотезу H_0 спростовують на користь альтернативної з надійністю 99 %.

Якщо виявиться, що гіпотезу H_0 немає підстав відхилити, то це означає, що вибіркове рівняння набуває вигляду: $y = \bar{y}$, тобто випадкова величина Y з урахуванням її розсіювання під впливом факторів, які не враховує регресійна модель, є сталою величиною на рівні свого середнього значення.

Аналогічно перевіряють статистичну значущість параметра b_0. За нульовою гіпотезою H_0: $b_0 = 0$, а за альтернативною гіпотезою H_1: $b_0 \neq 0$.

Для перевірки нульової гіпотези обчислюють емпіричне значення:

$$t_{b_0} = \frac{b_0}{\sigma_{b_0}}, \tag{16.6}$$

де $\sigma_{b_0} = \sqrt{\dfrac{\sigma_e^2 \cdot \sum\limits_{i=1}^{n} x_i^2}{n \cdot \sum\limits_{i=1}^{n} (x_i - \bar{x})^2}}$ – стандартне відхилення для параметра b_0.

Критичні значення t знаходять за таблицею розподілу Стьюдента (див. додаток Е), відповідно до заданого рівня значущості α та кількості ступенів свободи $k = n - 2$.

Якщо виявиться, що нульову гіпотезу відхиляють, то рівняння регресії буде містити вільний член: $b_0 \neq 0$.

Довірчі інтервали для параметрів регресійної моделі визначають, відповідно до вибраного рівня значущості за їхніми стандартними відхиленнями.

Довірчі інтервали для параметрів $b_i, (i = 0,1)$ за рівнем значущості α визначають за такою формулою:

$$[b_i - t_a \cdot \sigma_{b_i};\ b_i + t_a \cdot \sigma_{b_i}], \tag{16.7}$$

де для малих вибіркових сукупностей величину t_α знаходять за таблицею розподілу Стьюдента (див. додаток Е) для заданого рівня значущості α, відповідно до кількості ступенів свободи $k = n - 2$; для великих вибірок t_α визначають як аргумент функції Лапласа (див. додаток В), відповідно до умови $2\Phi(t_\alpha) = 1 - \alpha$.

Для лінії регресії, побудованої за відповідним вибірковим рівнянням, визначають **довірчу смугу**, до якої із заданою надійністю буде належати теоретична лінія регресії.

Для кожного значення зовнішнього фактора $X = x$ верхня та нижня межі довірчої смуги, до якої з надійністю $\gamma = 95\,\%$ буде належати теоретична лінія регресії, обчислюють, відповідно до такої формули:

$$\Delta y(x) = \pm t_{0,05} \cdot S_y \cdot \sqrt{\frac{1-r^2}{n-2}} \cdot \sqrt{1+\frac{(x-\bar{x})^2}{S_x^2}}\,. \qquad (16.8)$$

Для побудови **довірчого інтервалу прогнозу** за функцією регресії для кожного індивідуального значення x_{n+1} визначають межі інтервалу за таким співвідношенням:

$$\hat{y}\left(x_{n+1}\right) \pm t_\alpha(k) \cdot S_{прогн.}, \qquad (16.9)$$

де $t_\alpha(k)$ – критичне значення статистики Стьюдента (див. додаток Е) для рівня значущості α і кількості ступенів свободи $k = n-2$;

$$S^2_{прогн.} = \sigma_e^2 \left(1+\frac{1}{n}+\frac{\left(x_{n+1}-\bar{x}\right)^2}{\sum\limits_{i=1}^{n}(x_i-\bar{x})^2}\right)$$ – дисперсія прогнозу.

16.3. Приклади розв'язування задач

Приклад 16.1. Задано кореляційну таблицю (табл. 16.1), у якій наведено результати іспитів з інформатики (y) та математики (x) в одній з академічних груп I курсу.

Таблиця 16.1

Результати іспитів з інформатики та математики

y \ x	2	3	4	5	l_k
2	1	2			3
3	1	6	3		10
4		5	5		10
5			1	1	2
h_i	2	13	9	1	25

У припущенні лінійної залежності $\overline{y}_x = \rho_{y/x}x + b$ між X і Y знайдіть коефіцієнти $\rho_{y/x}$ і b.

Розв'язання. Обчисліть усі види середніх:

$$\bar{x} = \frac{2\cdot 2 + 3\cdot 13 + 4\cdot 9 + 5\cdot 1}{25} = \frac{84}{25} = 3{,}36;$$

$$\bar{y} = \frac{2\cdot 3 + 3\cdot 10 + 4\cdot 10 + 5\cdot 2}{25} = \frac{86}{25} = 3{,}44;$$

$$\overline{x^2} = \frac{4\cdot 2 + 9\cdot 13 + 16\cdot 9 + 25\cdot 1}{25} = \frac{294}{25} = 11{,}76;$$

$$\overline{y^2} = \frac{4\cdot 3 + 9\cdot 10 + 16\cdot 10 + 25\cdot 2}{25} = \frac{312}{25} = 12{,}48;$$

$$\overline{xy} = \frac{\sum\limits_i \sum\limits_k m_{ki} x_i y_k}{n} = \frac{1}{25}\big((1\cdot(2\cdot 2) + 2\cdot(3\cdot 2)) + (1\cdot(2\cdot 3) + 6\cdot(3\cdot 3) + 3\cdot(4\cdot 3)) +$$

$$+ (5\cdot(3\cdot 4) + 5\cdot(4\cdot 4)) + (1\cdot(4\cdot 5) + 1\cdot(5\cdot 5))\big) = \frac{1}{25}(16 + 96 + 140 + 45) = \frac{297}{25} = 11{,}88.$$

Тепер за формулою (16.1) буде:

$$\rho_{y/x} = \frac{11{,}88 - 3{,}36\cdot 3{,}44}{11{,}76 - 3{,}36^2} = \frac{11{,}88 - 11{,}56}{11{,}76 - 11{,}29} = \frac{0{,}32}{0{,}47} \approx 0{,}68;$$

$$\bar{b} = 3{,}44 - 0{,}68\cdot 3{,}36 = 3{,}44 - 2{,}28 = 1{,}16.$$

Таким чином, теоретичне рівняння регресії має такий вигляд:

$$\bar{y}_x = 0{,}68x + 1{,}16.$$

Зробіть перевірку:

$$x = 3:\ \bar{y}_x = 0{,}68\cdot 3 + 1{,}16 = 3{,}2;\ \bar{y}_x = \frac{2\cdot 2 + 3\cdot 6 + 4\cdot 5}{13} = 3{,}23;\ \varepsilon_1 = 0{,}03.$$

$$x = 4:\ \bar{y}_x = 0{,}68\cdot 4 + 1{,}16 = 3{,}88;\ \bar{y}_x = \frac{3\cdot 3 + 4\cdot 5 + 5\cdot 1}{9} = 3{,}78;\ \varepsilon_1 = 0{,}1.$$

Приклад 16.2. Залежність між змінними величинами X (кількість унесених добрив) і Y (урожайність) подано в табл. 16.2.

Таблиця 16.2

Залежність між кількістю внесених добрив і урожайністю

x_i	1,0	1,5	2,0	2,5	3,0
y_i	2,15	2,30	2,60	2,80	2,50

У припущенні лінійної залежності $\overline{y}_x = \rho_{y/x}\, x + b$ між X і Y знайдіть коефіцієнти $\rho_{y/x}$ і b.

Розв'язання. Вихідні дані й результати обчислень помістіть у табл. 16.3.

Таблиця 16.3

Результати обчислень

n	x_i	y_i	x_i^2	$x_i y_i$
1	1,0	2,15	1,00	2,15
2	1,5	2,30	2,25	3,45
3	2,0	2,60	4,00	5,20
4	2,5	2,80	6,25	7,00
5	3,0	2,50	9,00	7,50
$\sum$	10,0	12,35	22,50	25,30

Виконуючи обчислення, визначають:

$$\overline{x} = \frac{10}{5} = 2;\ \overline{y} = \frac{12{,}35}{5} = 2{,}47;\ \overline{x^2} = \frac{22{,}5}{5} = 4{,}5;\ \overline{xy} = \frac{25{,}3}{5} = 5{,}06.$$

Тоді:

$$\rho_{y/x} = 0{,}24,\ b = 1{,}99.$$

Теоретичне рівняння лінійної регресії має такий вигляд:

$$\overline{y}_x = 0{,}24x + 1{,}99.$$

На рис. 16.1 зображено емпіричні точки та знайдено теоретичну лінію регресії $\hat{y}_x$.

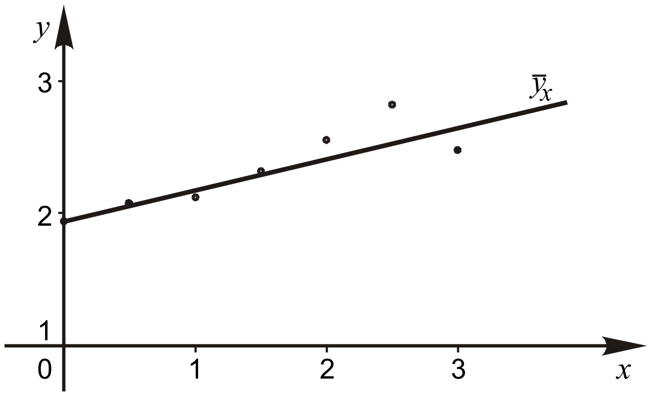


Рис. 16.1. **Теоретична лінія регресії $\hat{y}_x$ та емпіричні точки**

Перевірте рівняння:

$$\overline{y}_{x=1,5} = 0,24 \cdot 1,5 + 1,99 = 2,35; \; \overline{y}_{x=2,5} = 0,24 \cdot 2,5 + 1,99 = 2,6.$$

Відхилення:

$$\varepsilon_1 = |2,35 - 2,3| = 0,05; \; \varepsilon_2 = |2,8 - 2,6| = 0,2.$$

Приклад 16.3. Середню температуру окремого регіону вимірювали протягом 13 днів у січні (*A*) та лютому (*B*). Результати вимірювання наведено в табл.16.4.

Таблиця 16.4

Середня температура в січні (*A*) та лютому (*B*)

$A(X = x_i)$	–19,2	–14,8	–19,6	–11,1	–9,4	–16,9	–13,7
$B(Y = y_i)$	–21,8	–15,4	–20,8	–11,3	–11,6	–19,2	–13,0

Закінчення табл. 16.4

$A(X = x_i)$	–4,9	–13,9	–9,4	–8,3	–7,9	–5,3
$B(Y = y_i)$	–7,4	–15,1	–14,4	–11,1	–10,5	–7,2

Потрібно:

1) скласти рівняння парної лінійної регресії Y на X;

2) побудувати довірчі інтервали для параметрів теоретичного рівняння лінійної регресії з надійністю 95 %;

3) за рівнем значущості $\alpha = 0{,}05$ перевірити значущість параметрів теоретичного рівняння лінійної регресії;

4) із надійністю 99 % побудувати довірчий інтервал для прогнозованого індивідуального значення за умови, що $x = -22{,}5$.

Розв'язання:

1. Для обчислення значень параметрів теоретичного рівняння лінійної регресії використайте табл.16.5.

Обчислюють середні:

$$\bar{x} = \frac{\sum x_i}{n} = -\frac{154{,}4}{13} = -11{,}88;$$

$$\bar{y} = \frac{\sum y_i}{n} = -\frac{178{,}8}{13} = -13{,}75; \quad \overline{x^2} = \frac{\sum x_i^2}{n} = \frac{2121{,}68}{13} = 163{,}21;$$

$$\overline{y^2} = \frac{\sum y_i^2}{n} = \frac{2720{,}36}{13} = 209{,}26; \quad \overline{xy} = \frac{\sum x_i y_i}{n} = \frac{2385{,}96}{13} = 183{,}54.$$

Таблиця 16.5

Допоміжні обчислення

№ п/п	x_i	y_i	$x_i y_i$	x_i^2	y_i^2
1	2	3	4	5	6
1	–19,2	–21,8	418,56	368,64	475,24
2	–14,8	–15,4	227,92	219,04	237,16
3	–19,6	–20,8	407,68	384,16	432,64
4	–11,1	–11,3	125,43	123,21	127,69
5	–9,4	–11,6	109,04	88,36	134,56
6	–16,9	–19,2	324,48	285,61	368,64
7	–13,7	–13,0	178,10	187,69	169,00
8	–4,9	–7,4	36,26	24,01	54,76
9	–13,9	–15,1	209,89	193,21	228,01

Закінчення табл. 16.5

1	2	3	4	5	6
10	–9,4	–14,4	135,36	88,36	207,36
11	–8,3	–11,1	92,13	68,89	123,21
12	–7,9	–10,5	82,95	62,41	110,25
13	–5,3	–7,2	38,16	28,09	51,84
$\sum$	–154,4	–178,8	2 385,96	2 121,68	2 720,36

Знаходять параметри теоретичного рівняння лінійної регресії:

$$b_1 = \frac{\overline{xy} - \bar{x} \cdot \bar{y}}{\overline{x^2} - (\bar{x})^2} = \frac{183{,}54 - (-11{,}88)(-13{,}75)}{163{,}21 - (-11{,}88)^2} = \frac{20{,}19}{22{,}08} = 0{,}92;$$

$$b_0 = \bar{y} - b_1 \bar{x} = -13{,}75 - 0{,}92(-11{,}88) = -13{,}75 + 10{,}93 = -2{,}82.$$

Отже, рівняння регресії має такий вигляд:

$$\hat{y} \;= -2{,}82 + 0{,}92x.$$

2. Для визначення довірчих інтервалів параметрів теоретичного рівняння лінійної регресії необхідно обчислити значення

$$\sigma_e = \sqrt{\frac{1}{n-2} \sum_{i=1}^{n} (y_i - \hat{y}_i)^2}\,,$$

що є точковою незміщеною статистичною оцінкою середньоквадратичного відхилення випадкового фактора *e* (помилок моделі). Допоміжні обчислення наведено в табл. 16.6.

Таблиця 16.6

Обчислення довірчих інтервалів

x_i	y_i	$\hat{y}_i = b_0 + b_1 x_i$	$y_i - \hat{y}_i$	$(y_i - \hat{y}_i)^2$
1	2	3	4	5
–19,2	–21,8	–20,484	–1,316	1,732

Закінчення табл. 16.6

1	2	3	4	5
–14,8	–15,4	–16,436	1,036	1,073
–19,6	–20,8	–20,852	0,052	0,003
–11,1	–11,3	–13,032	1,732	2,999
–9,4	–11,6	–11,462	–1,138	0,003
–16,9	–19,2	–18,368	–0,832	0,692
–13,7	–13,0	–15,424	2,424	5,876
–4,9	–7,4	–7,328	–0,072	0,005
–13,9	–15,1	–15,608	0,508	0,258
–9,4	–14,4	–11,468	–2,932	8,597
–8,3	–11,1	–10,456	–0,644	0,415
–7,9	–10,5	–10,088	–0,412	0,169
–5,3	–7,2	–7,696	0,496	0,246
–154,4	–178,8	–	–	22,068

Отже, буде:

$$\sigma_e = \sqrt{\frac{1}{n-2}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2} = \sqrt{\frac{22,068}{13-2}} = \sqrt{2,006} = 1,416.$$

Далі обчислюють стандартні відхилення кожного параметра:

$$\sigma_{b_1} = \sqrt{\frac{\sigma_e^2}{\sum_{i=1}^{n}(x_i - \bar{x})^2}} = \sqrt{\frac{\sigma_e^2}{\sum_{i=1}^{n}x_i^2 - n(\bar{x})^2}} = \sqrt{\frac{2,006}{2\,121,68 - 13\cdot(-11,88)^2}} =$$

$$= \sqrt{\frac{2,006}{286,93}} = \sqrt{0,007} = 0,084;$$

$$\sigma_{b_0} = \sqrt{\frac{\sigma_e^2 \cdot \sum_{i=1}^{n}x_i^2}{n\cdot\sum_{i=1}^{n}(x_i - \bar{x})^2}} = \sqrt{\frac{\sigma_e^2 \cdot \sum_{i=1}^{n}x_i^2}{n\cdot\left(\sum_{i=1}^{n}x_i^2 - n(\bar{x})^2\right)}} = \sqrt{\frac{2,006\cdot 2121,68}{13\left(2\,121,68 - 13\cdot(-11,88)^2\right)}} =$$

$$= \sqrt{\frac{4256,1}{13\cdot 286,93}} = \sqrt{1,141} = 1,068.$$

За таблицею критичних точок розподілу Стьюдента (див. додаток Е) за заданим значенням $\gamma = 0{,}95$ і $k = n - 2 = 13 - 2 = 11$ знаходять:

$$t_\alpha = t(\gamma = 1 - \alpha = 0{,}95; k = 11) = 2{,}201.$$

Довірчі інтервали для параметрів $b_i, (i = 0{,}1)$ за рівнем значущості α визначають за такою формулою:

$$[b_i - t_a \cdot \sigma_{b_i};\ b_i + t_a \cdot \sigma_{b_i}].$$

Отже, із надійністю 95 % довірчий інтервал для параметра b_0 становить:

$$[-2{,}82 - 2{,}201 \cdot 1{,}068; -2{,}82 + 2{,}201 \cdot 1{,}068] = [-5{,}171; 0{,}469];$$

для параметра b_1:

$$[0{,}92 - 2{,}201 \cdot 0{,}084; 0{,}92 + 2{,}201 \cdot 0{,}084] = [0{,}735; 1{,}105].$$

3. За рівнем значущості $\alpha = 0{,}05$ перевірте значущість параметрів теоретичного рівняння лінійної регресії.

Нульову гіпотезу $H_0 : b_i = 0$ за альтернативної гіпотези $H_1 : b_i \neq 0$ перевіряють за допомогою статистики: $t_{b_i} = \dfrac{b_i}{\sigma_{b_i}}$.

Визначають:

$$t_{b_0} = \frac{b_0}{\sigma_{b_0}} = \frac{-2{,}82}{1{,}068} = -2{,}64;\ t_{b_1} = \frac{b_1}{\sigma_{b_1}} = \frac{0{,}92}{0{,}084} = 10{,}95.$$

Таким чином $|t_{b_i}| > t_{0,05}(11) = 2{,}201$, тобто нульову гіпотезу H_0 спростовують на користь альтернативної з надійністю 95 %. Це означає, що обидва параметри теоретичного рівняння лінійної регресії є значущими.

4. Із надійністю 99 % побудуйте довірчий інтервал для прогнозованого індивідуального значення за умови, що температура в січні буде $x = -22{,}5$ °С.

Визначте межі інтервалу за співвідношенням (16.9):

$$\hat{y}(x_{n+1}) \pm t_\alpha(k) \cdot S_{прогн.},$$

де прогнозоване індивідуальне значення –

$$\hat{y}(x_{n+1}) = \hat{y}(-22{,}5) = -2{,}82 + 0{,}92 \cdot (-22{,}5) = -23{,}52;$$

критичне значення статистики Стьюдента (див. додаток Е) –

$$t_{\alpha}(k) = t_{0,01}(11) = 3{,}11;$$

дисперсія прогнозу –

$$S^2_{прогн.} = \sigma_e^2 \left(1 + \frac{1}{n} + \frac{(x_{n+1} - \bar{x})^2}{\sum\limits_{i=1}^{n} (x_i - \bar{x})^2} \right) = 2{,}006 \left(1 + \frac{1}{13} + \frac{(-22{,}5 + 11{,}88)^2}{286{,}93} \right) =$$

$$= 2{,}949 \Rightarrow S_{прогн.} = \sqrt{2{,}949} = 1{,}717.$$

Отже, із надійністю 99 % прогнозоване індивідуальне значення температури в лютому буде належати довірчому інтервалу:

$$[-23{,}52 - 3{,}11 \cdot 1{,}717; -23{,}52 + 3{,}11 \cdot 1{,}717] = [-28{,}86; -18{,}18].$$

Приклад 16.4. Складіть рівняння парної лінійної регресії Y (середньодобова переробка сировини, тис. т) на X (основні виробничі фонди, млн грн), якщо за вибірковими даними буде: $\bar{x} = 1{,}381$; $\bar{y} = 2{,}905$; $\overline{xy} = 12{,}014$; $n = 100$; $\sigma_x^2 = 5{,}33$; $\sigma_y^2 = 10{,}24$.

Побудуйте вибіркову лінію регресії та довірчу смугу, до якої з надійністю 95 % буде належати лінія регресії генеральної сукупності.

Розв'язання. Рівняння регресії визначають у такому вигляді:

$$\hat{y}_x = b_1 x + b_0.$$

Обчислюють вибірковий коефіцієнт регресії – $b_1 = \dfrac{\overline{xy} - \bar{x} \cdot \bar{y}}{\sigma_x^2}$:

$$b_1 = \frac{12{,}014 - 2{,}381 \cdot 2{,}905}{5{,}33} = 0{,}956.$$

Тоді:

$$b_0 = 2{,}905 - 0{,}956 \cdot 1{,}381 = 0{,}629.$$

Отже, визначають вибіркове рівняння регресії: $\hat{y} = 0{,}629 + 0{,}956x$.

Оскільки графіком лінії регресії є пряма, то для його побудови достатньо обчислити за вибірковим рівнянням координати двох будь-яких точок, що належать цій прямій.

Якщо $x_1 = 0$, тоді $\hat{y}_1 = 0{,}629$; якщо $x_2 = 3$, тоді $\hat{y}_2 = 3{,}497$. Пряму за цими точками побудовано на рис.16.2.

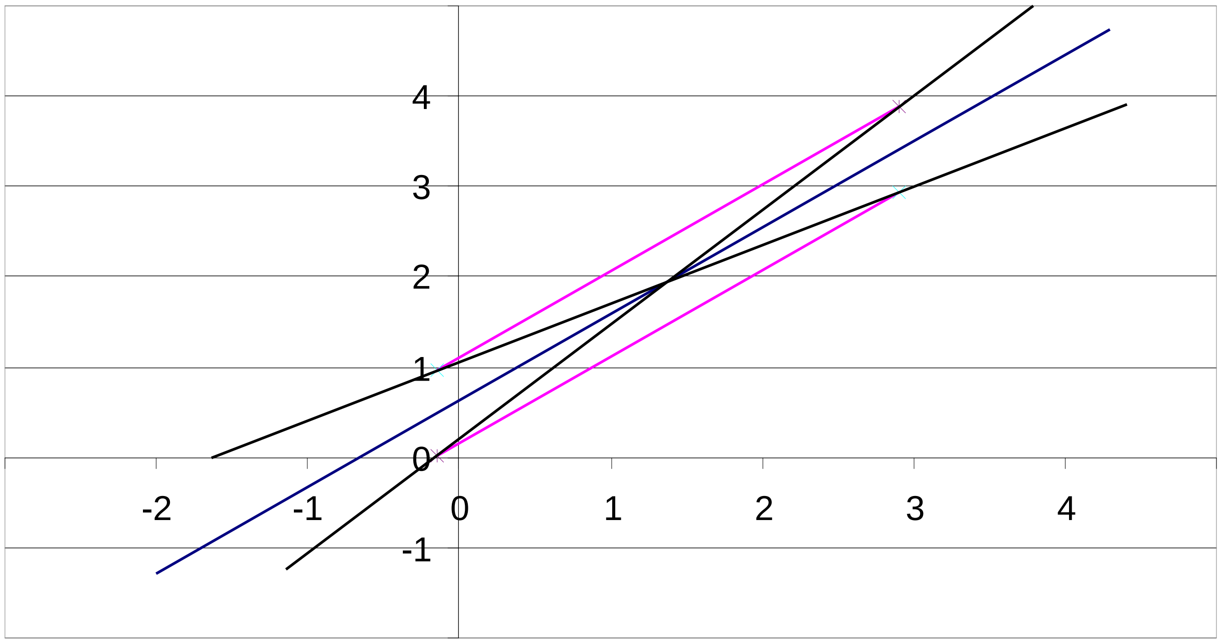


Рис. 16.2. **Схематичне подання довірчої смуги лінії регресії**

Тепер схематично побудуйте довірчу смугу для лінії регресії.

Для цього спочатку знаходять вибірковий коефіцієнт кореляції – $r = \dfrac{\overline{xy} - \overline{x} \cdot \overline{y}}{\sigma_x \sigma_y}$:

$$r = \frac{12{,}014 - 2{,}381 \cdot 2{,}905}{\sqrt{5{,}33} \cdot \sqrt{10{,}24}} = 0{,}69.$$

Далі обчислюють напівширину довірчої смуги для лінії регресії, яка відповідає проміжку, якщо відхилення значення фактора X від його вибіркової середньої не перевищує середнього квадратичного відхилення, тобто

$$\Delta y(x) \approx t_{0{,}05} \cdot S_y \cdot \sqrt{\frac{1 - r^2}{n - 2}},$$

де $S_y = \sqrt{\frac{n}{n-1}\cdot\sigma_y^2}$ – це виправлене середнє квадратичне відхилення фактора Y:

$$S_y = \sqrt{10{,}24\cdot\frac{100}{99}} = 3{,}216.$$

Для $\alpha = 0{,}05$ та $k = n-2$ за таблицею критичних точок розподілу Стьюдента (див. додаток Е) визначають, що $t_{0,05}(98) = 1{,}99$.

Отже, буде:

$$\Delta y(x) \approx 1{,}99\cdot 3{,}216\cdot\sqrt{\frac{1-0{,}69^2}{98}} = 0{,}468.$$

Для побудови проміжку $\left[\bar{x}-S_x;\, \bar{x}+S_x\right]$ визначають виправлене середнє квадратичне відхилення для фактора X:

$$S_x = \sqrt{\frac{n}{n-1}\cdot\sigma_x^2} \;\Rightarrow\; S_x = \sqrt{5{,}33\cdot\frac{100}{99}} = 1{,}523.$$

На проміжку $x \in [-0{,}142;\, 2{,}904]$ угору і вниз від лінії регресії відкладають $\Delta y(x) = 0{,}468$.

Визначають паралелограм, у якому проводять діагоналі та продовжують їх за цей паралелограм. Продовження діагоналей і є межами 95 % довірчої смуги для $|x-\bar{x}| > S_x$ (див. рис. 16.2).

16.4. Вправи для самостійної роботи

16.1. У припущенні, що кореляційний зв'язок між випадковими величинами X і Y є лінійним, знайдіть вибіркове рівняння регресії Y на X для даних табл. 16.7.

Таблиця 16.7

Вихідні дані залежності

X	0,9	2,5	3,4	5,6	6,9	9,0	9,8	11,2
Y	13,1	14,8	18,7	20,3	26,0	28,1	32,2	33,0

16.2. Складіть рівняння прямої регресії Y на X, якщо за вибірковими даними визначено такі результати: $\overline{x}=14$; $\overline{y}=1{,}8$ та $b_1=1{,}1$.

16.3. Знайдіть рівняння прямої регресії Y на X за результатами обчислення вибіркових даних: $\overline{x}=3{,}5$; $\overline{y}=21{,}8$; $\mu_{xy}=2{,}86$ та $\sigma_x^*=1{,}41$.

16.4. У припущенні, що кореляційний зв'язок між випадковими величинами X і Y є лінійним, складіть рівняння спряжених ліній регресії, якщо за даними обчислення 100 вимірювань визначено такі результати:

$$\sum x=421; \sum y=34{,}1; \sum x^2=68\,315; \sum y^2=576{,}4; \sum xy=9\,876{,}5.$$

16.5. Складіть рівняння спряжених прямих регресії за результатами вимірювання значень двовимірної випадкової величини (X; Y), наведених у вигляді кореляційної таблиці (табл. 16.8).

Таблиця 16.8

Вихідна кореляційна таблиця

Y \ X	1	2	3	4	5	m_y
1	2					2
2	1	7	2			10
3		2	5	1		8
4			1	2	2	5
m_x	3	9	8	3	2	25

16.6. У припущенні лінійного кореляційного зв'язку між факторами X та Y складіть спряжені вибіркові рівняння регресії, знайдіть кут між лініями регресії, а також обчисліть вибірковий коефіцієнт кореляції, якщо $n=100$; $\sum x=324$; $\sum y=1975$; $\sum xy=4197$; $\sum x^2=2\,324$; $\sum y^2=54\,834$.

16.7. Показники товарообігу *Y* та суми витрат *X*, які досліджували в 20 магазинах, наведено в табл. 16.9.

Таблиця 16.9

Залежність товарообігу від витрат

$Y=y_i$, грн	480	510	530	540	555	564	570	575	580	585
$X=x_i$, грн	30	25	31	32	38	41	40	46	49	54

Закінчення табл. 16.9

$Y = y_i$, грн	590	596	605	618	625	635	640	650	660
$X = x_i$, грн	58	60	64	75	78	82	83	85	90

Потрібно:

1) скласти рівняння парної лінійної регресії Y на X;

2) побудувати довірчі інтервали для параметрів теоретичного рівняння лінійної регресії з надійністю 95 %;

3) за рівнем значущості $\alpha = 0{,}05$ перевірити значущість параметрів теоретичного рівняння лінійної регресії;

4) із надійністю 99 % побудувати довірчий інтервал для прогнозованого індивідуального значення за умови $x = 92$;

5) побудувати вибіркову лінію регресії та довірчу смугу, до якої з надійністю 95 % буде належати лінія регресії генеральної сукупності.

16.5. Тестові завдання

16.1. Завданнями регресійного аналізу є

А виявлення зв'язку між випадковими величинами та оцінювання їхньої тісноти

Б виявлення зв'язку між випадковими величинами та їхніми числовими характеристиками

В визначення рівняння зв'язку між випадковими величинами

Г визначення рівняння зв'язку між випадковою величиною та невипадковими незалежними змінними й оцінювання невідомих значень залежної змінної

Д визначення рівняння зв'язку між невипадковою залежною величиною та випадковими незалежними змінними й оцінювання невідомих значень залежної змінної

16.2. Ламану лінію $\left(\overline{y}_x\right)$, яка з'єднує точки з координатами $\left(x_i;\ \overline{y}_{x_i}\right)$, називають

А емпіричною лінією регресії Y на X

Б емпіричною лінією регресії X на Y

В відповідь відсутня

16.3. За формою зв'язку розрізняють регресію

А додатну і від'ємну

Б парну та множинну

В лінійну й нелінійну

Г відповідь відсутня

16.4. За кількістю змінних розрізняють регресію

А додатну і від'ємну

Б парну та множинну

В лінійну й нелінійну

Г відповідь відсутня

16.5. Параметри теоретичного рівняння регресії $\hat{y}_x = b_1 x + b_0$ визначають за такими формулами

А $b_1 = \dfrac{\overline{xy} - \overline{x} \cdot \overline{y}}{\overline{x^2} - \left(\overline{x}\right)^2}$; $b_0 = \overline{y} - b_1 \overline{x}$

Б $b_0 = \dfrac{\overline{xy} - \overline{x} \cdot \overline{y}}{\overline{x^2} - \left(\overline{x}\right)^2}$; $b_1 = \overline{y} - b_0 \overline{x}$

В $b_1 = \dfrac{\overline{xy} - \overline{x} \cdot \overline{y}}{\overline{x^2} - \overline{x}}$; $b_0 = \overline{y} - b_1 \overline{x}$

Г відповідь відсутня

16.6. Чим більше значення коефіцієнта кореляції

А тим більший кут між спряженими теоретичними лініями регресії $\hat{y}_x$ та $\hat{x}_y$

Б тим ближче одна до одної розміщені спряжені теоретичні лінії регресії

В тим ближче лінії регресії до взаємної перпендикулярності

Г відповідь відсутня

16.7. Вільний член у рівнянні регресії $\hat{y}_x = b_1 x + b_0$ – це

А точка, у якій лінія регресії перетинає вісь *Y*

Б показник зв'язку між незалежною та залежною змінними

В завжди дорівнює одиниці

Г відповідь відсутня

16.8. Нульову гіпотезу H_0 : $b_i = 0$ за альтернативної гіпотези H_1: $b_i \neq 0$ перевіряють за допомогою такої статистики

А $t_{b_i} = \dfrac{\sigma_{b_i}}{b_i}$, де σ_{b_i} – стандартне відхилення для параметра b_i

Б $t_{b_i} = \dfrac{b_i}{\sigma_{b_i}}$, де σ_{b_i} – стандартне відхилення помилки

В $t_{b_i} = \dfrac{b_i}{\sigma_{b_i}}$, де σ_{b_i} – стандартне відхилення для параметра b_i

Г відповідь відсутня

16.6. Запитання для самоперевірки

16.1. Охарактеризуйте регресійний аналіз та наведіть його завдання.

16.2. Що таке "спряжені емпіричні лінії регресії"?

16.3. Які висновки про форму кореляційного зв'язку можна зробити за видом емпіричних ліній регресії?

16.4. Про що свідчить взаємне розташування емпіричних ліній регресії?

16.5. Розкрийте сутність методу найменших квадратів.

16.6. Наведіть формули, за якими здійснюють оцінювання параметрів рівняння парної лінійної регресії.

16.7. Як перевірити значущість кожного з параметрів рівняння парної регресії?

16.8. Як побудувати довірчі інтервали для оцінок параметрів лінії регресії за вибірковими даними?

16.9. Як визначити довірчу смугу, до якої теоретична лінія регресії буде належати із заданою надійністю?

16.10. Запишіть формулу для визначення меж довірчого інтервалу для прогнозного значення за вибірковим рівнянням регресії.

16.7. Висновки за темою

Одним із методів багатовимірної статистики є регресійний аналіз, головне завдання якого полягає в дослідженні взаємозв'язку між двома або більше випадковими величинами, визначенні форми цього зв'язку, обчисленні параметрів лінії регресії та перевірці статистичних гіпотез щодо значущості параметрів моделі й адекватності регресійної моделі, а також прогнозуванні за допомогою цієї моделі.

Регресійну модель слід розглядати як математичний вид реального закономірного зв'язку. В економічних дослідженнях становить інтерес не просто вивчення взаємозв'язків процесів і явищ, а кількісний вид цих взаємозв'язків. Тому до моделі, насамперед, висувають вимоги найбільшої відповідності характеру досліджуваного процесу, можливості економічної інтерпретації всіх його параметрів і наближення розрахункових результатів до досліджених даних.

Процес кореляційного і регресійного аналізу складається з таких послідовних етапів: попередніх угрупувань статистичних даних і виявлення форми зв'язку; складання рівнянь парної регресії за кожним фактором; оцінювання тісноти зв'язку, надійності й достовірності здобутої залежності; розроблення регресійної моделі явища, що вивчають, оцінювання її точності та визначення сили впливу врахованих факторів; аналізу досліджуваного явища за допомогою складеної моделі.

Регресійну модель здебільшого розглядають як інструмент аналізу, планування та управління. Звідси особливо жорсткі вимоги висувають до надійності, адекватності й точності кожного коефіцієнта моделі.

Рекомендована література

1. Збірник вправ з розділу "Теорія ймовірностей та математична статистика" навчальної дисципліни "Математика для економістів" для студентів галузі знань "Економіка і підприємництво" усіх форм навчання / уклад. Е. Ю. Железнякова, А. В. Ігначкова, З. Г. Попова та ін. – Харків : Вид. ХНЕУ, 2009. – 116 с.

2. Лабораторний практикум з навчальної дисципліни "Теорія ймовірностей та математична статистика" : навч. посіб. / Е. Ю. Железнякова, І. Л. Лебедєва, Л. О. Норік, К. В. Стєпанова. – Харків : ХНЕУ ім. С. Кузнеця, 2016. – 184 с.

3. Малярець Л. М. Математика для економістів. Теорія ймовірностей та математична статистика : навч. посіб. У 3-х ч. Ч. 3 / Л. М. Малярець, І. Л. Лебедєва, Л. Д. Широкорад. – Харків : Вид. ХНЕУ, 2011. – 568 с.

4. Малярець Л. М. Практикум з теорії ймовірностей та математичної статистики в Excel : навч.-практ. посіб. / Л. М. Малярець, І. Л. Лебедєва, Е. Ю. Железнякова. – Харків : Вид. ХНЕУ, 2007. – 160 с.

5. Малярец Л. М. Теория вероятностей и математическая статистика в примерах и задачах : учебное пособие для студентов-иностранцев отрасли знаний 0305 "Экономика и предпринимательство" / Л. М. Малярец, Э. Ю. Железнякова, А. В. Игначкова. – Харьков : ХНЭУ. – 2012. – 124 с.

6. Малярець Л. М. Теорія ймовірностей і математична статистика у вправах, прикладах та задачах : навч.-практ. посіб. / Л. М. Малярець, А. В. Ігначкова, Л. Д. Широкорад. – Харків : Вид. ХНЕУ, 2010. – 548 с.

7. Теорія ймовірностей та математична статистика : навч. посіб. / Л. М. Малярець, І. Л. Лебедєва, Е. Ю. Железнякова та ін. – Харків : Вид. ХНЕУ, 2010. – 404 с.

8. Барковський В. В. Теорія ймовірностей та математична статистика / В. В. Барковський, Н. В. Барковська, О. К. Лопатін. – 5-е вид. – Київ : Центр учбової літератури, 2010. – 424 с.

9. Гмурман В. Е. Руководство по решению задач по теории вероятностей и математической статистике / В. Е. Гурман. – Москва : Высшая школа, 2001. – 576 с.

10. Гмурман В. Е. Теория вероятностей и математическая статистика : учеб. пособ. для вузов / В. Е. Гурман. – 6-е изд. – Москва : Высшая школа, 1998. – 480 с.

11. Елисеева И. И. Теория статистики с основами теории вероятностей : учеб. пособ. / И. И. Елисеева, В. С. Князевский. – Москва : ЮНИТИ-ДАНА, 2001. – 446 с.

12. Жлуктенко В. І. Теорія ймовірностей і математична статистика : навч.-метод. посіб. У 2-х ч. Ч. І. Теорія ймовірностей / В. І. Жлуктенко, С. І. Наконечний. – Київ : КНЕУ, 2000. – 304 с.

13. Жлуктенко В. І. Теорія ймовірностей і математична статистика : навч.-метод. посіб. У 2-х ч. Ч. ІІ. Математична статистика / В. І. Жлуктенко, С. І. Наконечний, С. С. Савіна. – Київ : КНЕУ, 2001. – 336 с.

14. Кремер Н. Ш. Теория вероятностей и математическая статистика / Н. Ш. Кремер. – Москва : ЮНИТИ-ДАНА, 2000. – 544 с.

15. Теория вероятностей и математическая статистика. Базовый курс с примерами и задачами : учеб. пособ. / под ред. А. И. Кибзуна. – Москва : Физматлит, 2002. – 224 с.

16. Сайт персональних навчальних систем ХНЕУ ім. С. Кузнеця. – Режим доступу : http://www.ikt.hneu.edu.ua.

Додатки

Додаток А

Таблиця значень функції $\varphi(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$

x	0	1	2	3	4	5	6	7	8	9
0,0	0,3989	0,3989	0,3989	0,3988	0,3986	0,3984	0,3982	0,3980	0,3977	0,3973
0,1	0,3970	0,3965	0,3961	0,3956	0,3951	0,3945	0,3939	0,3932	0,3925	0,3918
0,2	0,3910	0,3902	0,3894	0,3885	0,3876	0,3867	0,3857	0,3847	0,3836	0,3825
0,3	0,3814	0,3802	0,3790	0,3778	0,3765	0,3752	0,3739	0,3726	0,3712	0,3698
0,4	0,3683	0,3668	0,3652	0,3637	0,3621	0,3605	0,3589	0,3572	0,3555	0,3538
0,5	0,3521	0,3503	0,3485	0,3467	0,3448	0,3429	0,3410	0,3391	0,3372	0,3352
0,6	0,3332	0,3312	0,3292	0,3271	0,3251	0,3230	0,3209	0,3187	0,3166	0,3144
0,7	0,3123	0,3101	0,3079	0,3056	0,3034	0,3011	0,2989	0,2966	0,2943	0,2920
0,8	0,2897	0,2874	0,2850	0,2827	0,2803	0,2780	0,2756	0,2732	0,2709	0,2685
0,9	0,2661	0,2637	0,2613	0,2589	0,2565	0,2541	0,2516	0,2492	0,2468	0,2444
1,0	0,2420	0,2396	0,2371	0,2347	0,2323	0,2299	0,2275	0,2251	0,2227	0,2203
1,1	0,2179	0,2155	0,2131	0,2107	0,2083	0,2059	0,2036	0,2012	0,1989	0,1965
1,2	0,1942	0,1919	0,1895	0,1872	0,1849	0,1826	0,1804	0,1781	0,1758	0,1736
1,3	0,1714	0,1691	0,1669	0,1647	0,1626	0,1604	0,1582	0,1561	0,1539	0,1518
1,4	0,1497	0,1476	0,1456	0,1435	0,1415	0,1394	0,1374	0,1354	0,1334	0,1315
1,5	0,1295	0,1276	0,1257	0,1238	0,1219	0,1200	0,1182	0,1163	0,1145	0,1127
1,6	0,1109	0,1092	0,1074	0,1057	0,1040	0,1023	0,1006	0,0989	0,0973	0,0957
1,7	0,0940	0,0925	0,0909	0,0893	0,0878	0,0863	0,0848	0,0833	0,0818	0,0804
1,8	0,0790	0,0775	0,0761	0,0748	0,0734	0,0721	0,0707	0,0694	0,0681	0,0669
1,9	0,0656	0,0644	0,0632	0,0620	0,0608	0,0596	0,0584	0,0573	0,0562	0,0551
2,0	0,0540	0,0529	0,0519	0,0508	0,0498	0,0488	0,0478	0,0468	0,0459	0,0449
2,1	0,0440	0,0431	0,0422	0,0413	0,0404	0,0395	0,0387	0,0379	0,0371	0,0363
2,2	0,0353	0,0347	0,0339	0,0332	0,0325	0,0317	0,0310	0,0303	0,0297	0,0290
2,3	0,0283	0,0277	0,0270	0,0264	0,0258	0,0252	0,0246	0,0241	0,0235	0,0229
2,4	0,0224	0,0219	0,0213	0,0208	0,0203	0,0198	0,0194	0,0189	0,0184	0,0180
2,5	0,0175	0,0171	0,0167	0,0163	0,0158	0,0154	0,0151	0,0147	0,0143	0,0139
2,6	0,0136	0,0132	0,0129	0,0126	0,0122	0,0119	0,0116	0,0113	0,0110	0,0107
2,7	0,0104	0,0101	0,0099	0,0096	0,0093	0,0091	0,0088	0,0086	0,0084	0,0081
2,8	0,0079	0,0077	0,0075	0,0073	0,0071	0,0069	0,0067	0,0065	0,0063	0,0061
2,9	0,0060	0,0058	0,0056	0,0055	0,0053	0,0051	0,0050	0,0048	0,0047	0,0046
3,0	0,0044	0,0043	0,0042	0,0040	0,0039	0,0038	0,0037	0,0036	0,0035	0,0034
3,1	0,0033	0,0032	0,0031	0,0030	0,0029	0,0028	0,0027	0,0026	0,0025	0,0025
3,2	0,0024	0,0023	0,0022	0,0022	0,0021	0,0020	0,0020	0,0019	0,0018	0,0018
3,3	0,0017	0,0017	0,0016	0,0016	0,0015	0,0015	0,0014	0,0014	0,0013	0,0013
3,4	0,0012	0,0012	0,0012	0,0011	0,0011	0,0010	0,0010	0,0010	0,0009	0,0009
3,5	0,0009	0,0008	0,0008	0,0008	0,0008	0,0007	0,0007	0,0007	0,0007	0,0006
3,6	0,0006	0,0006	0,0006	0,0005	0,0005	0,0005	0,0005	0,0005	0,0005	0,0004
3,7	0,0004	0,0004	0,0004	0,0004	0,0004	0,0004	0,0003	0,0003	0,0003	0,0003
3,8	0,0003	0,0003	0,0003	0,0003	0,0003	0,0002	0,0002	0,0002	0,0002	0,0002
3,9	0,0002	0,0002	0,0002	0,0002	0,0002	0,0002	0,0002	0,0002	0,0002	0,0001

Значення функції $P\left(k,\lambda\right)=\dfrac{\lambda^k}{k!}e^{-\lambda}$

k \ λ	**0,1**	**0,2**	**0,3**	**0,4**	**0,5**	**0,6**	**0,7**	**0,8**	**0,9**
0	0,9048	0,8187	0,7408	0,6703	0,6065	0,5458	0,4966	0,4493	0,4066
1	0,0905	0,1637	0,2222	0,2681	0,3033	0,3293	0,3476	0,3595	0,3659
2	0,0045	0,0164	0,0333	0,0536	0,0758	0,0988	0,1217	0,1438	0,1647
3	0,0002	0,0011	0,0033	0,0072	0,0126	0,0198	0,0284	0,0383	0,0494
4		0,0001	0,0002	0,0007	0,0016	0,003	0,0050	0,0077	0,0111
5				0,0001	0,0002	0,0004	0,0007	0,0012	0,0020
6							0,0001	0,0002	0,0003

k \ λ	**1,0**	**2,0**	**3,0**	**4,0**	**5,0**	**6,0**	**7,0**	**8,0**	**9,0**
0	0,3679	0,1353	0,0498	0,0183	0,0067	0,0025	0,0009	0,0003	0,0001
1	0,3679	0,2707	0,1494	0,0733	0,0337	0,0149	0,0064	0,0027	0,0011
2	0,1839	0,2707	0,2240	0,1465	0,0842	0,0446	0,0223	0,0107	0,0050
3	0,0613	0,1804	0,2240	0,1954	0,1404	0,0892	0,0521	0,0286	0,0150
4	0,0153	0,0902	0,1680	0,1954	0,1755	0,1339	0,0912	0,0572	0,0337
5	0,0031	0,0361	0,1008	0,1563	0,1755	0,1606	0,1277	0,0916	0,0607
6	0,0005	0,0120	0,0504	0,1042	0,1462	0,1606	0,1490	0,1221	0,0911
7	0,0001	0,0037	0,0216	0,0595	0,1044	0,1377	0,1490	0,1396	0,1171
8		0,0009	0,0081	0,0298	0,0653	0,1033	0,1304	0,1396	0,1318
9		0,0002	0,0027	0,0132	0,0363	0,0688	0,1014	0,1241	0,1318
10			0,0008	0,0053	0,0181	0,0413	0,0710	0,0993	0,1186
11			0,0002	0,0019	0,0082	0,0225	0,0452	0,0722	0,0970
12			0,0001	0,0006	0,0034	0,0126	0,0263	0,0481	0,0728
13				0,0002	0,0013	0,0052	0,0142	0,0296	0,0504
14				0,0001	0,0005	0,0022	0,0071	0,0169	0,0324
15					0,0002	0,0009	0,0033	0,0090	0,0194
16						0,0003	0,0014	0,0045	0,0109
17						0,0001	0,0006	0,0021	0,0058
18							0,0002	0,0009	0,0029
19							0,0001	0,0004	0,0014
20								0,0002	0,0006
21								0,0001	0,0003
22									0,0001

Додаток В

Таблиця значень функції $\Phi(x) = 1/\sqrt{2\pi}\int_0^x e^{-\frac{x^2}{2}}dx$

x	Φ(x)	x	Φ(x)	x	Φ(x)	x	Φ(x)
0,00	0,0000	0,39	0,1517	0,78	0,2823	1,17	0,3790
0,01	0,0040	0,40	0,1554	0,79	0,2852	1,18	0,3810
0,02	0,0080	0,41	0,1591	0,80	0,2881	1,19	0,3830
0,03	0,0120	0,42	0,1628	0,81	0,2910	1,20	0,3849
0,04	0,0160	0,43	0,1664	0,82	0,2939	1,21	0,3869
0,05	0,0199	0,44	0,1700	0,83	0,2967	1,22	0,3883
0,06	0,0239	0,45	0,1736	0,84	0,2995	1,23	0,3907
0,07	0,0279	0,46	0,1772	0,85	0,3023	1,24	0,3925
0,08	0,0319	0,47	0,1808	0,86	0,3051	1,25	0,3944
0,09	0,0359	0,48	0,1844	0,87	0,3078	1,26	0,3962
0,10	0,0398	0,49	0,1879	0,88	0,3106	1,27	0,3980
0,11	0,0438	0,50	0,1915	0,89	0,3133	1,28	0,3997
0,12	0,0478	0,51	0,1950	0,90	0,3159	1,29	0,4015
0,13	0,0517	0,52	0,1985	0,91	0,3186	1,30	0,4032
0,14	0,0557	0,53	0,2019	0,92	0,3212	1,31	0,4049
0,15	0,0596	0,54	0,2054	0,93	0,3238	1,32	0,4066
0,16	0,0636	0,55	0,2088	0,94	0,3264	1,33	0,4082
0,17	0,0675	0,56	0,2123	0,95	0,3289	1,34	0,4099
0,18	0,0714	0,57	0,2157	0,96	0,3315	1,35	0,4115
0,19	0,0753	0,58	0,2190	0,97	0,3340	1,36	0,4131
0,20	0,0793	0,59	0,2224	0,98	0,3365	1,37	0,4147
0,21	0,0832	0,60	0,2257	0,99	0,3389	1,38	0,4162
0,22	0,0871	0,61	0,2291	1,00	0,3413	1,39	0,4177
0,23	0,0910	0,62	0,2324	1,01	0,3438	1,40	0,4192
0,24	0,0948	0,63	0,2357	1,02	0,3401	1,41	0,4207
0,25	0,0987	0,64	0,2389	1,03	0,3485	1,42	0,4222
0,26	0,1026	0,65	0,2422	1,04	0,3508	1,43	0,4236
0,27	0,1064	0,66	0,2454	1,05	0,3531	1,44	0,4251
0,28	0,1103	0,67	0,2486	1,06	0,3554	1,45	0,4265
0,29	0,1141	0,68	0,2517	1,07	0,3577	1,46	0,4279
0,30	0,1179	0,69	0,2549	1,08	0,3599	1,47	0,4292
0,31	0,1217	0,70	0,2580	1,09	0,3621	1,48	0,4306
0,32	0,1255	0,71	0,2611	1,10	0,3643	1,49	0,4319
0,33	0,1293	0,72	0,2642	1,11	0,3665	1,50	0,4332
0,34	0,1331	0,73	0,2673	1,12	0,3686	1,51	0,4345
0,35	0,1368	0,74	0,2703	1,13	0,3708	1,52	0,4357
0,36	0,1406	0,75	0,2734	1,14	0,3729	1,53	0,4370
0,37	0,1443	0,76	0,2764	1,15	0,3749	1,54	0,4382
0,38	0,1480	0,77	0,2794	1,16	0,3770	1,55	0,4394

x	Ф(x)	x	Ф(x)	x	Ф(x)	x	Ф(x)
1,56	0,4406	1,81	0,4649	2,12	0,4830	2,62	0,4956
1,57	0,4418	1,82	0,4656	2,14	0,4838	2,64	0,4959
1,58	0,4429	1,83	0,4664	2,16	0,4846	2,66	0,4961
1,59	0,4441	1,84	0,4671	2,18	0,4854	2,68	0,4963
1,60	0,4452	1,85	0,4678	2,20	0,4861	2,70	0,4965
1,61	0,4463	1,86	0,4686	2,22	0,4868	2,72	0,4967
1,62	0,4474	1,87	0,4693	2,24	0,4875	2,74	0,4969
1,63	0,4484	1,88	0,4699	2,26	0,4881	2,76	0,4971
1,64	0,4495	1,89	0,4706	2,28	0,4887	2,78	0,4973
1,65	0,4505	1,90	0,4713	2,30	0,4893	2,80	0,4974
1,66	0,4515	1,91	0,4719	2,32	0,4898	2,82	0,4976
1,67	0,4525	1,92	0,4726	2,34	0,4904	2,84	0,4977
1,68	0,4535	1,93	0,4732	2,36	0,4909	2,86	0,4979
1,69	0,4545	1,94	0,4738	2,38	0,4913	2,88	0,4980
1,70	0,4554	1,95	0,4744	2,40	0,4918	2,90	0,4981
1,71	0,4564	1,96	0,4750	2,42	0,4922	2,92	0,4982
1,72	0,4573	1,97	0,4756	2,44	0,4927	2,94	0,4984
1,73	0,4582	1,98	0,4761	2,46	0,4931	2,96	0,4985
1,74	0,4591	1,99	0,4767	2,48	0,4934	2,98	0,4985
1,75	0,4599	2,00	0,4772	2,50	0,4938	3,00	0,49865
1,76	0,4608	2,02	0,4783	2,52	0,4941	3,20	0,49931
1,77	0,4616	2,04	0,4793	2,54	0,4945	3,34	0,49966
1,78	0,4625	2,06	0,4803	2,56	0,4948	3,60	0,49984
1,79	0,4633	2,08	0,4812	2,58	0,4951	3,80	0,49993
1,80	0,4641	2,10	0,4821	2,60	0,4953	4,00	0,49997

Критичні точки розподілу χ^2

Кількість ступенів свободи k	Рівень значущості α					
	0,01	0,025	0,05	0,95	0,975	0,99
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	0,362	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	2,7	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	8,76	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	9,59	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

Таблиця Д.1

Таблиця значень $t = t(\gamma, n)$

n	γ			n	γ		
	0,95	0,99	0,999		0,95	0,99	0,999
5	2,78	4,60	8,61	20	2,093	2,861	3,883
6	2,57	4,03	6,86	25	2,064	2,797	3,745
7	2,45	3,71	5,96	30	2,045	2,756	3,659
8	2,37	3,50	5,41	35	2,032	2,720	3,600
9	2,31	3,36	5,04	40	2.023	2,708	3,558
10	2,26	3,25	4,78	45	2,016	2,692	3,527
11	2,23	3,17	4,59	50	2,009	2,679	3,502
12	2,20	3,11	4,44	60	2,001	2,662	3,464
13	2,18	3,06	4,32	70	1,996	2,649	3,439
14	2,16	3,01	4,22	80	1,991	2,640	3,418
15	2,15	2,98	4,14	90	1,987	2,633	3,403
16	2,13	2,95	4,07	100	1,984	2,627	3,392
17	2,12	2,92	4,02	120	1,980	2,617	3,374
18	2,11	2,90	3,97	∞	1,960	2,576	3,291
19	2,10	2,88	3,92				

Таблиця Д.2

Таблиця значень $q = q(\gamma, n)$

n	γ			n	γ		
	0,95	0,99	0,999		0,95	0,99	0,999
5	1,37	2,67	5,64	20	0,37	0,58	0,88
6	1,09	2,01	3,88	25	0,32	0,49	0,73
7	0,92	1,62	2,98	30	0,28	0,43	0,63
8	0,80	1,38	2,42	35	0,26	0,38	0,56
9	0,71	1,20	2,06	40	0,24	0,35	0,50
10	0,65	1,08	1,80	45	0,22	0,32	0,46
11	0,59	0,98	1,60	50	0,21	0,30	0,43
12	0,55	0,90	1,45	60	0,188	0,269	0,38
13	0,52	0,83	1,33	70	0,174	0,245	0,34
14	0,48	0,78	1,23	80	0,161	0,226	0,31
15	0,46	0,73	1,15	90	0,151	0,211	0,29
16	0,44	0,70	1,07	100	0,143	0,198	0,27
17	0,42	0,66	1,01	150	0,115	0,160	0,211
18	0,40	0,63	0,96	200	0,099	0,136	0,185
19	0,39	0,60	0,92	250	0,089	0,120	0,162

Критичні точки розподілу Стьюдента

Рівень значущості α (двобічна критична область)				
Кількість ступенів свободи k	$\alpha = 0{,}01$	$\alpha = 0{,}02$	$\alpha = 0{,}05$	$\alpha = 0{,}1$
1	63,7	31,82	12,7	6,31
2	9,92	6,97	4,30	2,92
3	5,84	4,54	3,18	2,35
4	4,60	3,75	2,78	2,13
5	4,03	3,37	2,57	2,01
6	3,71	3,14	2,45	1,94
7	3,50	3,00	2,36	1,89
8	3,36	2,90	2,31	1,86
9	3,25	2,82	2,26	1,83
10	3,17	2,76	2,23	1,81
11	3,11	2,72	2,20	1,80
12	3,05	2,68	2,18	1,78
13	3,01	2,65	2,16	1,77
14	2,98	2,62	2,14	1,76
15	2,95	2,60	2,13	1,75
16	2,92	2,58	2,12	1,75
17	2,90	2,57	2,11	1,74
18	2,88	2,55	2,10	1,73
19	2,86	2,54	2,09	1,73
20	2,85	2,53	2,09	1,73
21	2,83	2,52	2,08	1,72
22	2,82	2,51	2,07	1,72
23	2,81	2,50	2,07	1,71
24	2,80	2,49	2,06	1,71
25	2,79	2,49	2,06	1,71
26	2,78	2,48	2,06	1,71
27	2,77	2,47	2,05	1,71
28	2,76	2,46	2,05	1,70
30	2,75	2,46	2,04	1,70
40	2,70	2,42	2,02	1,68
60	2,66	2,39	2,00	1,67
120	2,62	2,36	1,98	1,66
∞	2,58	2,33	1,96	1,64
k	$\alpha = 0{,}005$	$\alpha = 0{,}01$	$\alpha = 0{,}025$	$\alpha = 0{,}05$
Рівень значущості α (однобічна критична область)				

Додаток Ж

Критичні точки розподілу *F* Фішера – Снедекора за рівнем значущості $\alpha = 0{,}05$

k_2	k_1											
	1	2	3	4	5	6	7	8	9	10	11	12
1	161	200	216	225	230	234	237	239	241	242	243	244
2	18,5	19,0	19,2	19,2	19,3	19,3	19,3	19,4	19,4	19,4	19,4	19,4
3	10,1	9,55	9,28	9,13	9,01	8,94	8,88	8,84	8,81	8,78	8,76	8,74
4	7,71	6,94	6,59	6,39	6,26	6,16	6,09	6,04	6,00	5,96	5,93	5,91
5	6,61	5,79	5,41	5,19	5,05	4,95	4,88	4,82	4,78	4,74	4,70	4,68
6	5,99	5,14	4,76	4,53	4,39	4,28	4,21	4,15	4,10	4,06	4,03	4,00
7	5,59	4,74	4,35	4,12	3,97	3,87	3,79	3,73	3,68	3,63	3,60	3,57
8	5,32	4,46	4,07	3,84	3,69	3,58	3,50	3,44	3,39	3,34	3,31	3,28
9	5,12	4,16	3,86	3,63	3,48	3,37	3,29	3,23	3,18	3,13	3,10	3,07
10	4,96	4,10	3,71	3,48	3,33	3,22	3,14	3,07	3,02	2,97	2,94	2,91
11	4,84	3,98	3,59	3,36	3,20	3,09	3,01	2,95	2,90	2,86	2,82	2,79
12	4,75	3,88	3,49	3,26	3,11	3,00	2,92	2,85	2,80	2,76	2,72	2,69
13	4,67	3,80	3,41	3,18	3,02	2,92	2,84	2,77	2,72	2,67	2,63	2,60
14	4,60	3,74	3,34	3,11	2,96	2,85	2,77	2,70	2,65	2,60	2,56	2,53
15	4,54	3,68	3,29	3,06	2,90	2,79	2,70	2,64	2,59	2,55	2,51	2,48
16	4,49	3,63	3,24	3,01	2,85	2,74	2,66	2,59	2,54	2,49	2,45	2,42
17	4,45	3,59	3,20	2,96	2,81	2,70	2,62	2,55	2,50	2,45	2,41	2,38
20	4,35	3,49	3,10	2,87	2,71	2,60	2,52	2,45	2,40	2,35	2,31	2,28
24	4,26	3,40	3,01	2,78	2,62	2,51	2,43	2,36	2,30	2,26	2,22	2,18
30	4,17	3,32	2,92	2,69	2,53	2,42	2,34	2,27	2,21	2,16	2,12	2,09
40	4,08	3,23	2,84	2,61	2,45	2,34	2,25	2,18	2,12	2,07	2,04	2,00
50	4,03	3,18	2,79	2,56	2,40	2,29	2,20	2,13	2,07	2,02	1,98	1,95
70	3,98	3,13	2,74	2,50	2,35	2,23	2,14	2,07	2,01	1,97	1,93	1,89
100	3,94	3,09	2,70	2,46	2,30	2,19	2,10	2,03	1,97	1,92	1,88	1,85
200	3,89	3,04	2,65	2,41	2,26	2,14	2,05	1,98	1,92	1,87	1,83	1,80
400	3,86	3,02	2,62	2,39	2,23	2,12	2,03	1,96	1,90	1,85	1,81	1,78
∞	3,84	2,99	2,60	2,37	2,21	1,09	2,01	1,94	1,88	1,83	1,79	1,75

Критичні точки розподілу *F* Фішера – Снедекора за рівнем значущості $\alpha = 0{,}01$

k_2	k_1									
	1	2	3	4	5	6	8	12	24	∞
1	4 052	4 999	5 403	5 625	5 764	5 859	5 981	6 106	6 234	6 366
2	98,5	99,0	99,2	99,3	99,3	99,4	99,3	99,4	99,5	99,5
3	34,1	30,8	29,5	28,7	28,2	27,9	27,5	27,1	26,6	26,1
4	21,2	18,0	16,7	16,0	15,5	15,2	14,8	14,4	13,9	13,5
5	16,3	13,3	12,1	11,4	11,0	10,7	10,3	9,9	9,5	9,6
6	13,7	10,9	9,8	9,2	8,8	8,5	8,1	7,7	7,3	6,9
7	12,3	9,6	8,5	7,9	7,5	7,2	6,8	6,5	6,1	5,7
8	11,3	8,7	7,6	7,0	6,6	6,5	6,0	5,7	5,3	4,9
9	10,6	8,0	7,0	6,4	6,1	5,8	5,5	5,1	4,7	4,3
10	10,0	7,6	6,6	6,0	5,6	5,4	5,1	4,7	4,3	3,9
11	9,7	7,2	6,2	5,7	5,3	5,1	4,7	4,4	4,0	3,6
12	9,3	6,9	6,0	5,4	5,1	4,8	4,5	4,2	3,8	3,4
13	9,1	6,7	5,7	5,2	4,9	4,6	4,3	4,0	3,6	3,2
14	8,9	6,5	5,6	5,0	4,7	4,5	4,1	3,8	3,4	3,0
15	8,7	6,4	5,4	4,9	4,6	4,3	4,0	3,7	3,3	2,9
16	8,5	6,2	5,3	4,8	4,4	4,2	3,9	3,6	3,2	2,8
17	8,4	6,1	5,2	4,7	4,3	4,1	3,8	3,5	3,1	2,7
18	8,3	6,0	5,1	4,6	4,3	4,0	3,7	3,4	3,0	2,6
19	8,2	5,9	5,0	4,5	4,2	3,9	3,6	3,3	2,9	2,4
20	8,1	5,9	4,9	4,4	4,1	3,9	3,6	3,2	2,9	2,4
22	7,9	5,7	4,8	4,3	4,0	3,8	3,5	3,1	2,8	2,3
24	7,8	5,6	4,7	4,2	3,9	3,7	3,3	3,0	2,7	2,2
26	7,7	5,5	4,6	4,1	3,8	3,6	3,3	3,0	2,6	2,1
28	7,6	5,5	4,6	4,1	3,8	3,5	3,2	2,9	2,5	2,1
30	7,6	5,4	4,5	4,0	3,7	3,5	3,2	2,8	2,5	2,0
40	7,3	5,2	4,3	3,8	3.5	3,3	3,0	2,7	2,3	1,8
60	7,1	5,0	4,1	3,7	3,3	3,1	2,8	2,5	2,1	1,6
120	6,9	4,8	4,0	3,5	3,2	3,0	2,7	2,3	2,0	1,4
∞	6,6	4,6	3,8	3,3	3,0	2,8	2,5	2,2	1,8	1,0

Предметний покажчик

НАВЧАЛЬНЕ ВИДАННЯ

Железнякова Еліна Юріївна
Норік Лариса Олексіївна

ТЕОРІЯ ЙМОВІРНОСТЕЙ ТА МАТЕМАТИЧНА СТАТИСТИКА

Практикум

Самостійне електронне текстове мережеве видання

Відповідальний за видання *Л. М. Малярець*

Відповідальний редактор *М. М. Оленич*

Редактор *О. Г. Доценко*

Коректор *О. Г. Доценко*

План 2019 р. Поз. № 34-ЕНП. Обсяг 321 с.

Видавець і виготовлювач – ХНЕУ ім. С. Кузнеця, 61166, м. Харків, просп. Науки, 9-А

Свідоцтво про внесення суб'єкта видавничої справи до Державного реєстру
ДК № 4853 від 20.02.2015 р.