**А. М. Алілуйко, Н. В. Дзюбановська,
В. О. Єрьоменко, О. М. Мартинюк,
М. І. Шинкарик**

ПРАКТИКУМ
З ТЕОРІЇ ІМОВІРНОСТЕЙ ТА МАТЕМАТИЧНОЇ СТАТИСТИКИ

*НАВЧАЛЬНИЙ ПОСІБНИК
для студентів
економічних спеціальностей*

™
Тернопіль
Видавництво «Підручники і посібники»
2018

УДК 519.21..519.22
П 69

Рецензенти: **І. М. Черевко** — доктор фізико-математичних наук, професор, декан факультету математики та інформатики Чернівецького національного університету імені Юрія Федьковича;
В. А. Кривень — доктор фізико-математичних наук, професор, завідувач кафедри математичних методів в інженерії Тернопільського національного технічного університету імені Івана Пулюя;
В. З. Чорний — кандидат фізико-математичних наук, доцент, завідувач кафедри математики та методики її навчання Тернопільського національного педагогічного університету імені Володимира Гнатюка.

Алілуйко А.

П 69 **Практикум з теорії імовірностей та математичної статистики**. Навчальний посібник для студентів економічних спеціальностей / Алілуйко А. М., Дзюбановська Н. В., Єрьоменко В. О., Мартинюк О. М., Шинкарик М. І. — Тернопіль : Підручники і посібники, 2018. — 352 с.

Посібник містить в конспективній формі теоретичний матеріал з основних розділів теорії імовірностей та математичної статистики, який супроводжується великою кількістю задач (в тому числі економічного змісту), наведених із розв'язками та для самостійної роботи. При цьому основний акцент робиться на алгоритмізацію розв'язування задач, а також на створення читачем цілісної уяви про зв'язки між темами і розділами та їх подальше використання.

Для студентів економічних спеціальностей ВНЗ, для економістів та осіб, котрі займаються самоосвітою.

УДК 519.21..519.22

ПЕРЕДМОВА

Досвід економічно розвинутих країн показав, що в умовах ринкової економіки переваги при працевлаштуванні і наступному посадовому рості мають економісти з високим рівнем аналітичних якостей. Одним із найпотужніших засобів покращення цих якостей є грунтовне засвоєння циклу математичних дисциплін, серед яких особливе місце займає “Теорія імовірностей і математична статистика” (ТІМС). Така особливість зумовлена, з одного боку, тим, що за останні роки в економічних ВНЗ зростає кількість навчальних дисциплін, які або грунтуються на ТІМС, або використовують її методи. З другого боку, багаторічний досвід викладання ТІМС авторами показує, що основним джерелом труднощів для студентів при вивченні цієї дисципліни і особливо при виконанні індивідуальних розрахункових робіт є слабкі навики аналізу різних ситуацій та їх найпростішого моделювання.

Пропонований посібник є другим виданням практикуму з ТІМС [9]. Він доповнений параграфами “Системи випадкових величин (багатовимірні випадкові величини)” та “Функція випадкових величин”. Оновлено задачі для самостійного розв’язування та проміжного контролю знань студентів.

Кожний параграф містить теоретичний матеріал, розв’язування великої кількості різнотипових задач, а також значну кількість задач економічного спрямування для самостійної роботи.

Відбір теоретичного матеріалу здійснювався з метою “автономності” посібника (відсутності потреби, як правило, користуватися додатковими джерелами), подаючи теорію так, щоб: 1) детально та ілюстровано висвітлити означення, які стають інструментарієм аналізу та розв’язування задач; 2) читач створив цілісну уяву про зв’язки між темами і розділами дисципліни.

Особливу увагу автори приділили **алгоритмізації** розв’язування задач, а також генеруванню ідей (в процесі розв’язування задач), які стають ключовими при доведенні більш складних тверджень.

Посібник складається з двох частин: I “Теорія імовірностей” і II “Математична статистика”. Нумерація формул та задач здійснюється по наростанню в межах параграфа і має дві позиції, перша з яких вказує на номер параграфа. При цьому нумерація параграфів для кожної із частин автономна. У випадку використання в другій частині посилання на формулу з першої частини до її нумерації добавляється ЧІ, що усуває можливу двозначність. Символи ○ та ● означають відповідно початок і завершення розв’язування задачі.

ЧАСТИНА ПЕРША
ТЕОРІЯ ІМОВІРНОСТЕЙ

§ 1. ВИЗНАЧЕННЯ ІМОВІРНОСТІ

Події та їх види. Класичне означення імовірності випадкової події. Властивості імовірностей. Елементи комбінаторики в теорії імовірностей. Відносна частота випадкової події. Статистична імовірність. Геометрична імовірність.

КОРОТКІ ТЕОРЕТИЧНІ ВІДОМОСТІ

Під **випробуванням** будемо розуміти здійснення намічених дій і отримання результату при виконанні певного комплексу умов S. При цьому припускається, що ці умови є фіксованими; вони або об'єктивно існують, або створюються штучно і можуть бути відтворені необмежене число разів.

Результатом випробування є подія. Розрізняють події **достовірні**, **неможливі** та **випадкові**.

Достовірною називають подію, яка при випробуванні обов'язково відбувається. **Неможливою** називають подію, яка при випробуванні обов'язково не відбувається. **Випадкова** – це та подія, яка при випробуванні може як відбутися, так і не відбутися.

Достовірну подію позначимо літерою Ω, а неможливу – $\varnothing$.

Розглянемо деякі **властивості випадкових подій**.

Дві події називаються **несумісними** (**сумісними**), якщо при випробуванні відбуття однієї **виключає (не виключає)** відбуття іншої.

Сукупність випадкових подій утворює **повну групу**, якщо одна з них при випробуванні обов'язково відбувається, а ***будь-які*** дві події є несумісними.

Елементарними будемо називати найпростіші випадкові події, які можуть відбутися при випробуванні.

Події, «породжені» одним випробуванням, назвемо **рівноможливими**, якщо є підстави вважати, що жодна з них не є більш можливою, ніж інші.

Чисельну міру можливості відбуття випадкової події дає ймовірність цієї події.

Означення. Класичною імовірністю події A називається відношення числа елементарних рівноможливих подій, що сприяють появі події A, до загального числа елементарних рівноможливих подій, що утворюють повну групу.

Кожна з елементарних випадкових подій, по суті, є одним із наслідків випробування. Такий підхід є корисним при аналізі задач.

З урахуванням цього зауваження аналітичний вираз класичного означення набере такого виду:

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

де $P(A)$ – класична імовірність події A;

m – число елементарних рівноможливих подій, що сприяють появі події A (число наслідків випробування, в яких відбувається подія А);

n – число елементарних рівноможливих подій, що утворюють повну групу (загальне число рівноможливих наслідків випробування).

Дане означення дозволяє сформулювати **основні властивості класичної імовірності**:

1) $P(\Omega) = 1$ (імовірність достовірної події дорівнює 1);

2) $P(\varnothing) = 0$ (імовірність неможливої події дорівнює 0);

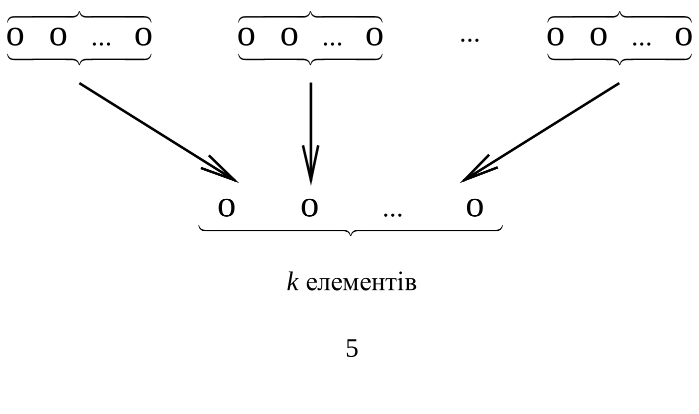
3) якщо A – випадкова подія, тоді:

$$0 < P(A) < 1. \tag{1.2}$$

Для того, щоб мати деякі стандартні методи при розрахунках по схемі класичної імовірності, наведемо **основну формулу комбінаторики**, а також розглянемо поняття **комбінацій**, **розміщень** та **перестановок**.

Нехай є k груп елементів, чисельність кожної з яких відповідно дорівнює n_1, n_2, ..., n_k. Виберемо довільним чином по **одному** елементу з кожної групи:

1-а група (n_1 елементів) 2-а група (n_2 елементів) k-а група (n_k елементів)



k елементів

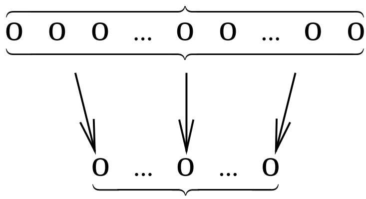
Тоді загальне число N способів, якими можна здійснити такий відбір, визначається співвідношенням

$$N = n_1 \cdot n_2 \cdot \ldots \cdot n_k, \tag{1.3}$$

яке називається **основною формулою комбінаторики**.

Розглянемо сукупність **різних** елементів довільної природи, чисельністю n. Будемо утворювати групи по m $(m \le n)$ **різних** елементів із цієї сукупності:

n елементів



m елементів

Такі групи в теорії імовірностей часто називаються **вибірками**.

Нехай $m < n$. **Комбінаціями** називаються такі групи, які відрізняються одна від іншої хоча б одним елементом. Загальне число комбінацій C_n^m (читається: це з n по m) знаходиться за формулою

$$C_n^m = \frac{n(n-1)\cdot(n-2)\cdot\ldots\cdot(n-m+1)}{m!}, \tag{1.4}$$

де $m! = m(m-1)\cdot\ldots\cdot 2\cdot 1$ (читається: m факторіал).

Зауваження. В чисельнику (1.4) є m співмножників. Якщо чисельник і знаменник помножити на $(n-m)!$, тоді отримається така рівність:

$$C_n^m = \frac{n!}{(n-m)!\,m!}. \tag{1.4*}$$

Нехай $m \le n$. **Розміщеннями** називаються такі групи, які відрізняються одна від іншої або хоча б одним елементом, або порядком розташування цих елементів в групі. Число розміщень A_n^m (читається: а з n по m) знаходиться за формулою:

$$A_n^m = \underbrace{n(n-1)\cdot(n-2)\cdot\ldots\cdot(n-m+1)}_{m\ \text{співмножників}}. \tag{1.5}$$

Якщо в формулі (1.5) $m = n$, то A_n^m – число таких розміщень, які відрізняються тільки порядком розташування елементів, а не самими елементами. Такі розміщення називаються **перестановками**. Їх число P_n за формулою (1.5)

$$A_n^n = n(n-1)(n-2)...2\cdot 1 = n! = P_n,$$

тобто

$$P_n = A_n^n = n!. \tag{1.6}$$

Число n може набирати не тільки натуральні значення, воно може також дорівнювати нулю. Порожна множина (вибірка) є підмножиною довільної множини і природно вважати, що вона може бути впорядкована тільки одним способом. Тому вважається, що **0! = 1**.

Число комбінацій володіє такими властивостями:

1) $C_n^0 = C_n^n = 1$; 2) $C_n^1 = C_n^{n-1} = n$; 3) $C_n^m = C_n^{n-m}$;

4) $C_n^0 + C_n^1 + C_n^2 + \ldots + C_n^n = 2^n$.

Відмітимо, що числа розміщень, перестановок і комбінацій пов'язані рівністю

$$A_n^m = P_m C_n^m.$$

При розв'язуванні задач комбінаторики використовуються такі правила:

Правило суми. Якщо деякий об'єкт α може бути відібраний із сукупності об'єктів k способами, а другий об'єкт β може бути відібраний s способами, то відібрати або α, або β можна $k + s$ способами.

Правило добутку. Якщо об'єкт α можна вибрати із сукупності об'єктів k способами і після кожного такого відбору об'єкт β можна вибрати s способами, то пара об'єктів (α, β) у вказаному порядку може бути вибрана $k \cdot s$ способами.

Зауваження. Ще раз нагадаємо, що в означеннях комбінацій, розміщень і перестановок суттєвим є те, що **всі елементи** в групах **різні**. Якщо ж у комбінаціях, розміщеннях і перестановках деякі із елементів (або всі) можуть виявитися однаковими, то такі групи називаються **комбінаціями з повтореннями**, **розміщеннями з повтореннями** і **перестановками з повтореннями** відповідно. Формули для знаходження числа такого виду груп ми не наводимо, проте в §2 вкажемо методи знаходження імовірностей випадкових подій, в яких фігурують групи (комбінації, розміщення, перестановки) з повтореннями.

Разом із імовірністю до основних понять теорії імовірностей належить відносна частота.

Відносною частотою випадкової події називається відношення числа випробовувань, в яких подія відбулася, до загального числа фактично проведених випробовувань. Тобто, відносна частота події A визначається формулою:

$$W(A) = \frac{M}{N}, \tag{1.7}$$

де M – число появ події A, N – загальне число випробувань.

Співставлення означень імовірності і відносної частоти дозволяє зробити висновок: імовірність обчислюють до випробування (тобто вона є апріорною величиною), а відносну частоту – після випробування (апостеріорна величина).

Із означення (1.7) для **випадкової** події A випливає така подвійна нерівність (порівняйте з (1.2)!):

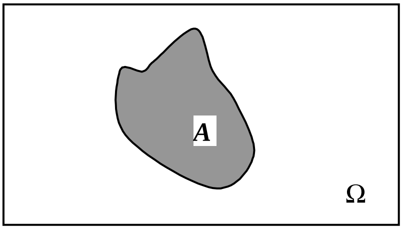
$$0 \le W(A) \le 1.$$

При невеликій кількості випробовувань відносна частота випадкової події може помітно змінюватися від однієї серії випробовувань до іншої. Проте тривалі спостереження показали, що коли в однакових умовах проводиться достатньо велике число випробовувань, то відносна частота виявляє властивість **стійкості**. Ця властивість полягає в тому, що в різних серіях випробовувань відносна частота **змінюється мало** (тим менше, чим більше число випробовувань), коливаючись навколо деякого постійного числа. Виявилось, що це стале число є імовірністю випадкової події. Математичне формулювання цієї властивості стійкості буде дано в темі «Закон великих чисел» (теорема Я. Бернуллі).

Властивість стійкості відносної частоти, а також можливість, хоча б принципово, проводити необмежене число випробовувань відносно випадкової події дозволяють сформулювати **статистичне означення імовірності**: в якості статистичної імовірності випадкової події береться відносна частота цієї події.

Один із недоліків класичного означення імовірності, пов'язаний із неможливістю використання такого означення для випадку випробувань із нескінченним числом наслідків випробування, може бути усунутий з допомогою **геометричної імовірності** – імовірності попадання точки в область (відрізок, частину площини, частину тіла тощо).

Припустимо, що здійснюється випробування – на прямокутник Ω, в якому міститься довільна фігура $\boldsymbol{A}$, навмання кидається точка (мал. 1.1).



Мал.1.1.

При цьому вважається, що виконуються такі припущення: вона може опинитися в будь-якій точці прямокутника Ω, імовірність (можливість) попадання точки на фігуру $\boldsymbol{A}$ пропорційна площі цієї фігури і не залежить ні від розташування $\boldsymbol{A}$ відносно Ω, ні від форми $\boldsymbol{A}$. Хай випадкова подія A – точка попала в фігуру $\boldsymbol{A}$. Тоді **геометричне означення імовірності** події A дається рівністю:

$$P(A) = \frac{S(\boldsymbol{A})}{S(\Omega)}, \qquad (1.8)$$

де $S(\mathscr{A})$ – площа фігури $\boldsymbol{A}$, $S(\Omega)$ – площа прямокутника Ω.

Означення (1.8) є частинним випадком загального означення геометричної імовірності. Якщо позначити міру (довжину, площу, об’єм) області через *mes*, то імовірність попадання точки, кинутої навмання (із збереженням вище наведених припущень) в область d – частину області D, визначається рівністю:

$$P = \frac{mes\ d}{mes\ D}. \qquad (1.9)$$

РОЗВ’ЯЗУВАННЯ ТИПОВИХ ЗАДАЧ

Задача 1.1. Перед змаганнями проводиться жеребкування серед спортсменів. В урні є 300 жетонів, пронумерованих від 1 до 300. Знайти імовірність того, що навмання витягнутий жетон першим спортсменом матиме хоча б одну цифру: а) 8; б) 2.

○ а) Подія A полягає в тому, що навмання взятий жетон матиме хоча б одну цифру 8. Випробування – взяття жетона. Число всіх наслідків випробування (елементарних подій) $n = 300$, оскільки будь-який із жетонів може бути витягнутим. Всі вони рівноможливі. Подія A відбувається кожний раз тоді, коли жетон матиме або одну, або дві цифри 8. Для того, щоб підрахувати число всіх таких наслідків випробування, достатньо знайти число жетонів з хоча б однією цифрою 8 в одній сотні (8 не може бути на місці сотень) і результат помножити на три. Зокрема, в першій сотні 8 може зустрітися на місці одиниць в десяти жетонах (8, 18, ..., 88, 98), а на місці десятків також в десяти (80, 81, ..., 88, 89). Проте 88 пораховано в першій серії, тому всього наслідків випробування, в яких подія A відбувається, для першої сотні дорівнює 19, а $m = 3 \cdot 19 = 57$.

Згідно з (1.1) $P(A) = 57/300 = 0{,}19$.

б) B – витягнутий жетон має хоча б одну цифру 2.

Очевидно, що $n = 300$. В першій сотні є 19 жетонів з хоча б однією цифрою 2, в другій – 20 (вона завершується числом 200), в третій (201, 202, ..., 299, 300) – 99. Отже, $m = 19 + 20 + 99 = 138$ і $P(B) = 138/300 = 0{,}46$. ●

Зауваження. Отримані результати вказують на те, що подія B в 2,42 рази більш імовірніша в порівнянні із подією A.

Наступну задачу пов'язують із помилкою відомого французького математика і філософа Д'Аламбера (1717–1783), яка полягала в ігноруванні рівноможливості елементарних подій – однієї із основних вимог класичного означення імовірності.

Задача 1.2. Монету кидають двічі на горизонтальну тверду поверхню. Знайти імовірність того, що хоча б раз випаде герб.

○ Для спрощення міркувань припустимо, що кидаються дві монети (до початкової постановки задачі повернемося в §3 «Повторні незалежні випробування»).

Подія A – герб випадає принаймні один раз. Випробування – кидання двох монет. Всі наслідки випробування: (*г*; *г*), (*г*; *н*), (*н*; *г*), (*н*; *н*), де *г* – герб, *н* – напис, перша позиція відповідає результату кидання першої монети, друга – другої. Вони є рівноможливими і утворюють повну групу. Отже, $n = 4$, $m = 3$, $P(A) = 3/4$. ●

Д'Аламбер аналізував: герб з'явиться або при першому киданні, або при другому, або зовсім не появиться. Всіх наслідків три, з яких сприяють події A два, а тому шукана імовірність дорівнює 2/3.

Розглянемо ще одну задачу, яка була однією із історично перших, для розв'язування яких було використане класичне означення імовірності.

Задача 1.3. Знайти імовірність всіх можливих значень суми очок на верхніх гранях двох кинутих гральних кубиків.

○ Очевидно, що мінімальним значенням суми є число 2, а максимальним – 12. Тому потрібно буде знайти імовірності випадкових подій ($\Sigma = 2$), ($\Sigma = 3$), ..., ($\Sigma = 12$), де Σ – сума очок на верхніх гранях кубиків. Проаналізуйте, чому кожна із цих подій є випадковою. Випробуванням є кидання двох кубиків, наслідками є сукупність пар очків на верхніх гранях. Підрахуємо число цих пар. «Одиниця» першого кубика може зустрітися з 1, 2, ..., 6 другого, тобто вона «породжує» 6 наслідків випробування. Таку ж кількість «породжує» і «двійка». При цьому слід мати на увазі, що наслідки випробування ($1 \div 2$) і ($2 \div 1$) – різні (перша цифра відповідає числу очок на верхній грані першого кубика, друга – другого). Аналогічна ситуація з рештою чисел першого кубика. Отже, $n = 6 \cdot 6 = 36$. Кожен із цих наслідків відповідає вимогам означення імовірності: вони рівноможливі і утворюють повну групу. Справді, вони попарно несумісні, оскільки поява одного із наслідків при випробуванні виключає появу будь-якого іншого в цьому ж випробуванні. При випробуванні обов'язково відбудеться один із наслідків. Нарешті, рівноможливість досягається за рахунок симетричності кубиків та однорідності матеріалу, з яких вони виготовлені.

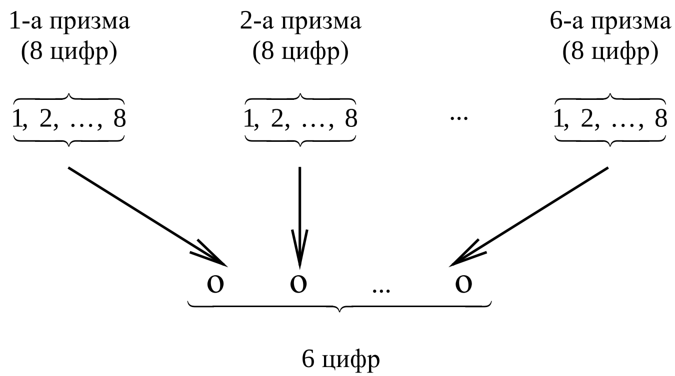
Події ($\Sigma = 2$) сприяє тільки один наслідок ($1 \div 1$). Тому $P(\Sigma = 2) = 1/36$. Події ($\Sigma = 3$) сприяють вже два наслідки випробування: ($1 \div 2$), ($2 \div 1$). Тому $P(\Sigma = 3) = 2/36$. Продовжуючи обчислення ймовірностей решти випадкових подій, отримаємо таку таблицю:

k	2	3	4	5	6	7	8	9	10	11	12
$P(\Sigma = k)$	1/36	2/36	3/36	4/36	5/36	6/36	5/36	4/36	3/36	2/36	1/36

Аналіз таблиці дозволяє зробити висновок, зокрема, що імовірність події ($\Sigma = 7$) в шість разів більша від імовірностей подій ($\Sigma = 2$) та ($\Sigma = 12$). ●

Задача 1.4. Навісний замок із «секретом» має шість восьмикутних призм, кожна з яких повертається навколо своєї осі незалежно від інших. Бічні грані кожної із призм пронумеровані цифрами від 1 до 8. Замок відкривається, якщо обертанням призм на чільній стороні замка буде набрано певне шестизначне число. Знайти імовірність того, що після довільного набору шестизначного числа на чільній стороні замок відкриється. Зробити висновок про надійність цього замка.

○ Випробування – обертання кожної із шести призм. Подія A – набір секретного шестизначного числа. Наслідок випробування – шестизначне число. Для знаходження всіх наслідків випробування потрібно порахувати число всіх різних шестизначних чисел. Ці числа відрізняються одне від іншого або цифрами, або порядком їх розташування. На перший погляд групи із шести елементів (цифр) є розміщеннями. Проте необхідною вимогою як розміщень, так і комбінацій є те, що **всі** елементи групи є **різними**. В той час як в даній задачі групи (числа) можуть містити деякі або всі однакові елементи (цифри). Аналіз схеми утворення груп елементів, для якої виведена основна формула комбінаторики (1.3), дає підстави стверджувати про можливість застосування її в даному випадку:



Згідно з даною формулою комбінаторики (1.3), де $k = 6$,
$n_1 = n_2 = ... = n_6 = 8$, отримуємо
$n = N = 8 \cdot 8 \cdot ... \cdot 8 = 8^6 = 262\,144$.

З усіх наслідків випробування тільки в одному замок відкривається (коли буде набрано секретне число), тому $m = 1$. Остаточно $P(A) = 1/262\,144 = 0{,}0000038$.

Характеристика надійності замка визначається числом n: не знаючи секретного числа, потрібно здійснити всього 262 144 набори різних шестизначних чисел. ●

Задача 1.5. Банк протягом місяця може видати в кредит позику п'ятьом своїм клієнтам, в той час як поступили замовлення на кредит від 15 клієнтів першого району і 10 клієнтів другого району. Для збереження клієнтів банк розглядає як тимчасову вимушену міру – розігрування випадковим чином п'яти позик серед тих, від кого поступило замовлення. Знайти імовірність того, що число клієнтів першого району, яким дістанеться позика, дорівнює: а) 5; б) 0; в) 3.

○ Випробування – проведення розігрування серед клієнтів банку. Наслідок випробування – п'ятірка клієнтів. Число наслідків випробування дорівнює числу всеможливих п'ятірок, які можна утворити із сукупності чисельністю 25 елементів (клієнтів банку).

Для таких груп елементів характерним є те, що вони відрізняються одна від іншої хоча б одним елементом. Крім того, як основна сукупність елементів (з 25 клієнтів), так і утворені групи по 5 елементів, складаються з **різних** елементів (відсутність повторів). Це дає підстави зробити висновок, що такі групи є комбінаціями. І їх число для всіх трьох підзадач дорівнює:

$$n = C_{25}^{5} = \frac{25\cdot 24\cdot 23\cdot 22\cdot 21}{5\cdot 4\cdot 3\cdot 2\cdot 1} = 53130.$$

а) A – власниками кредиту є п'ять клієнтів першого району. Число наслідків випробування, для яких відбувається подія A, дорівнює числу всеможливих п'ятірок клієнтів, котрі можна утворити із загального числа 15 клієнтів першого району. Ці п'ятірки знову є комбінаціями і їх загальне число

$$m = C_{15}^{5} = \frac{15\cdot 14\cdot 13\cdot 12\cdot 11}{5\cdot 4\cdot 3\cdot 2\cdot 1} = 3003.$$

За класичним означенням імовірності

$$P(A) = \frac{m}{n} = \frac{3003}{53130} \approx 0,057.$$

б) B – власниками кредиту є п'ять клієнтів другого району. Число наслідків випробування m, для яких відбувається подія B, дорівнює числу всеможливих п'ятірок клієнтів, що можна утворити із 10 клієнтів другого району. Знову ж ці п'ятірки є комбінаціями і

$$m = C_{10}^{5} = \frac{10\cdot 9\cdot 8\cdot 7\cdot 6}{5\cdot 4\cdot 3\cdot 2\cdot 1} = 252,$$

а

$$P(B) = \frac{m}{n} = \frac{252}{53130} \approx 0,005.$$

в) C – власниками кредиту є три клієнти першого району і два клієнти другого району. Всеможливі трійки клієнтів першого району можна утворити C_{15}^{3} способами, а всеможливі двійки клієнтів другого району – $C_{10.}^{2}$ Використавши **правило добутків**, отримаємо:

$$m = C_{15}^{3} \cdot C_{10}^{2} = \frac{15 \cdot 14 \cdot 13}{3 \cdot 2 \cdot 1} \cdot \frac{10 \cdot 9}{2 \cdot 1} = 455 \cdot 45 = 20475,$$

звідки $$P(C) = \frac{20475}{53130} \approx 0{,}385. \;\bullet$$

Задача 1.6. Керівництво приватизованого підприємства оголосило конкурс з виплатою трьох перших премій на кращий проект модернізації виробництва. На конкурс поступило 9 проектів. Скільки можна скласти всіх можливих трійок власників премій?

○ З дев'яти **різних** елементів утворюються групи по три **різних** елементи. Порядок «розташування» елементів суттєвий для них. Справді, нехай проекти під номерами 5, 6, 8 вибороли премії. Але, зокрема, трійки: 1) № 5 – І-а премія, № 6 – ІІ-а, № 8 – ІІІ-я; 2) № 5 – ІІІ-я, № 6 – І-а, № 8 – ІІ-а – різні. Тому такі групи є розміщеннями і їх число дорівнює $A_{9}^{3} = 9 \cdot 8 \cdot 7 = 504$. ●

Зауваження. Нехай задача 1.6 доповнена таким чином: «Знайти імовірність того, що після підведення підсумків конкурсу власниками трьох перших премій стануть автори проектів під номерами: № 7 – І-а премія, № 5 – ІІ-а премія, № 9 – ІІІ-я премія». Тоді цю імовірність, взагалі кажучи, не можна знайти за класичним означенням імовірності, оскільки наслідки випробувань не будуть рівноможливими випадковими подіями (проекти суттєво можуть відрізнятися один від одного, а тому можуть мати різні шанси бути відзначеними).

Задача 1.7. На чотирьох картках написані літери *Е*, *У*, *Н*, *Т*. Картки перемішуються і розкладаються в ряд. Знайти імовірність того, що в результаті буде отримане слово «*ТНЕУ*».

○ Випробування – розкладання карток в ряд після перемішування. Подія A – отримання слова «ТНЕУ». Тільки один наслідок випробування сприяє появі цієї події, тобто $m = 1$. Наслідки випробування – це всеможливі групи по 4 елементи (літери), для яких суттєвий порядок розташування елементів, і їх число $n = P_4 = 4! = 4 \cdot 3 \cdot 2 \cdot 1 = 24$. Отже, $P(A) = 1/24$. ●

Задача 1.8. На книжковій полиці розташовано 26 книг, вартості яких відповідно рівні: 8 по 5 грн., 12 по 4 грн. і 6 по 3 грн. Навмання беруться дві книги. Яка імовірність того, що їх сумарна вартість складає 8 грн.?

○ Подія *A* – сумарна вартість відібраних двох книг дорівнює 8 грн. Випробування – відбір двох книг. Число наслідків дорівнює числу всеможливих способів відбору двох книг. Такі групи по дві книги є комбінаціями, бо для них несуттєвий порядок розташування в групі елементів (книг). Тому $n = C_{26}^{2} = \dfrac{26 \cdot 25}{2 \cdot 1} = 325.$

Подія *A* відбувається або тоді, коли обидві книги коштують по 4 грн. (таких наслідків випробування є C_{12}^{2}), або одна книга коштує 5 грн., а друга – 3 грн. (таких наслідків є $8 \cdot 6 = C_8^1 \cdot C_6^1$). Згідно з правилом суми комбінаторики $m = C_{12}^{2} + C_8^1 \cdot C_6^1 = 66 + 48 = 114.$ Остаточно P(A) = 114/325 = 0,351. ●

В деяких випадках при знаходженні числа комбінацій доцільніше користуватися формулою (1.4*), а не (1.4). При цьому корисним є використання основних властивостей числа комбінацій.

Задача 1.9. Партія із 1000 плиток шоколаду, серед яких 4 із призами, довільним чином розподілена на дві рівні частини, кожна з яких відправлена двом споживачам. Знайти ймовірність того, що: а) всі призи дістануться одному споживачу; б) обом споживачам порівну.

○ а) Випробування – відбір 500 плиток шоколаду із 1000. Подія *A* – всі призи дістануться одному споживачу. Загальне число наслідків випробування $n = C_{1000}^{500}$. Подія *A* відбувається , коли із 500 плиток, відправлених одному споживачу, буде або 496 без призів із 996 і всі 4 із призами, або 500 без призів і 0 із призами. Число таких наслідків $m = C_{996}^{496} \cdot C_4^4 + C_{996}^{500} \cdot C_4^0$. За першою властивістю числа комбінацій $C_4^4 = C_4^0 = 1$, а за третьою $C_{996}^{500} = C_{996}^{996-500} = C_{996}^{496}$. Тому

$$P(A) = \frac{m}{n} = \frac{C_{996}^{496} \cdot C_4^4 + C_{996}^{500} \cdot C_4^0}{C_{1000}^{500}} = \frac{2 \cdot C_{996}^{496}}{C_{1000}^{500}} = 2 \cdot \frac{\dfrac{996!}{500! \cdot 496!}}{\dfrac{1000!}{500! \cdot 500!}} =$$

$$= 2 \cdot \frac{996! \cdot 500! \cdot 500!}{1000! \cdot 500! \cdot 496!} = 2 \cdot \frac{996! \cdot 496! \cdot 497 \cdot 498 \cdot 499 \cdot 500}{996! \cdot 997 \cdot 998 \cdot 999 \cdot 1000 \cdot 496!} =$$

$$= \frac{2 \cdot 497 \cdot 498 \cdot 499 \cdot 500}{997 \cdot 998 \cdot 999 \cdot 1000} \approx 0{,}124 ,$$

де $1000! = 996! \cdot 997 \cdot 998 \cdot 999 \cdot 1000,\ 500! = 496! \cdot 497 \cdot 498 \cdot 499 \cdot 500.$

б) Подія *B* – в одній частині продукції є 2 призи. Тоді такою ж властивістю буде володіти і друга частина. Подія *B* відбуватиметься, якщо із 500 плиток, відправлених одному споживачу, буде 498 без призів і дві з призами з чотирьох. Число таких наслідків $m = C_{996}^{498} \cdot C_4^2$. Тому

$$P(B) = \frac{C_{996}^{498} \cdot C_4^2}{C_{1000}^{500}} = \frac{\dfrac{996!}{498! \cdot 498!} \cdot \dfrac{4!}{2! \cdot 2!}}{\dfrac{1000!}{500! \cdot 500!}} = \frac{996! \cdot 500! \cdot 500! \cdot 4!}{1000! \cdot 498! \cdot 498! \cdot 2! \cdot 2!} =$$

$$= \frac{996! \cdot (498! \cdot 499 \cdot 500)^2 \cdot 4 \cdot 3 \cdot 2 \cdot 1}{996! \cdot 997 \cdot 998 \cdot 999 \cdot 1000 \cdot (498!)^2 \cdot (2 \cdot 1)^2} =$$

$$= \frac{(499 \cdot 500)^2 \cdot 6}{997 \cdot 998 \cdot 999 \cdot 1000} \approx 0{,}376 . \ \bullet$$

Розв’язування наступної задачі ілюструє можливість отримання додаткової інформації на підставі знання імовірності випадкової події.

Задача 1.10. В контейнері знаходяться стандартні і браковані деталі. Імовірність того, що навмання витягнуті три деталі будуть стандартними, дорівнює 2/3. Яка мінімальна кількість деталей в контейнері?

○ Позначимо: *k* – число всіх деталей в контейнері, *s* – кількість стандартних серед них. Випробування – відбір трьох деталей. Подія *A* – три деталі стандартні. Згідно з умовою *P*(*A*)=2/3. За класичною формулою $P(A) = \dfrac{m}{n} = \dfrac{C_s^3}{C_k^3} = \dfrac{s \cdot (s-1) \cdot (s-2)}{k \cdot (k-1) \cdot (k-2)}$.

Отже, $\dfrac{s \cdot (s-1) \cdot (s-2)}{k \cdot (k-1) \cdot (k-2)} = \dfrac{2}{3}$.

Ціле число *s* задовольняє подвійну нерівність $3 \le s \le k-1$, оскільки для *s*=0, 1, 2 подія *A* стає неможливою, а для *s*=*k* – достовірною. Замінивши *s* в попередньому рівнянні числом *k*-1, отримаємо нерівність $\dfrac{(k-1) \cdot (k-2) \cdot (k-3)}{k \cdot (k-1) \cdot (k-2)} \ge \dfrac{2}{3}$, рівносильну нерівності $3\cdot(k-3)\ge 2\cdot k$, звідки $k\ge 9$. Залишається тільки з'ясувати існування цілого числа *s*, яке задовольняє умову задачі. Проста перевірка показує, що значенням *s* є число 8. ●

Задача 1.11. В ліфт на першому поверсі дванадцятиповерхового будинку зайшло 5 людей, кожна із яких може вийти незалежно одна від іншої на довільному поверсі з другого по дванадцятий включно. Знайти імовірність того, що всі пасажири вийдуть: а) на восьмому поверсі; б) на одному поверсі; в) на різних поверхах.

○ Випробування – спроба виходу із ліфту п'яти пасажирів на потрібному поверсі. а) Подія *A* – всі пасажири вийдуть на восьмому поверсі. Кожний із пасажирів може вийти із 2-го по 12-й поверх 11 способами. За основною формулою комбінаторики (1.3) число таких наслідків випробування $n=11^5$. І тільки в одному наслідку випробування відбудеться подія *A*, тобто $m=1$. Отже,

$$P(A)=\frac{m}{n}=\frac{1}{11^5}\approx 0{,}0000062.$$

б) Подія *B* – всі пасажири вийдуть на одному поверсі. Події *B* сприяють 11 наслідків (всі пасажири можуть вийти або на 2-му поверсі, або на 3-му, ..., або на 12-му). Тому

$$P(B)=\frac{m}{n}=\frac{11}{11^5}=\frac{1}{11^4}\approx 0{,}0000683.$$

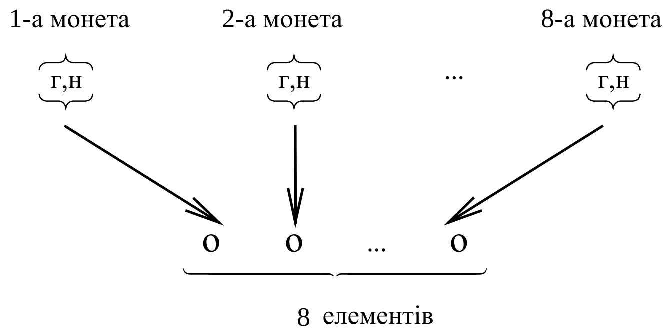
в) Подія *C* – всі пасажири вийдуть на різних поверхах. Наслідки випробування, в яких відбувається подія *C*, утворюють групи різних елементів (різних номерів поверху) чисельністю 5, які утворюються із різних 11 елементів. Ці групи відрізняються одна від іншої хоча б одним елементом і порядок розташування є суттєвим, бо він врахований в числі *n*. Отже, ці наслідки є розміщеннями і $m=A_{11}^5$.

Таким чином,

$$P(C)=\frac{m}{n}=\frac{A_{11}^5}{11^5}=\frac{11\cdot 10\cdot 9\cdot 8\cdot 7}{11^5}\approx 0{,}344239.$$ ●

Задача 1.12. Знайти імовірність того, що при киданні восьми монет герб випаде на шести з них.

○ Випробування – кидання восьми монет. Подія *A* – герб випадає на шести монетах (і при цьому не випадає на двох). Наслідки випробування – це всеможливі групи з восьми елементів, які утворюються за такою схемою:



Тут г – герб, н – напис. Приклад наслідка випробування: “нггнгннн”. Загальне число різних наслідків випробування знаходиться за основною формулою комбінаторики (1.3): $n = 2^8$. При цьому всі вони є **рівноможливими** випадковими подіями, оскільки для кожної із монет поява герба чи напису має **однакову імовірність**, рівну 0,5. Знайдемо число наслідків випробування, для яких подія А відбувається. Приклади таких наслідків: “ггггггнн”, “гггггнгн”, ... Всіх їх є стільки, скільки можна утворити груп різних елементів чисельністю 6 із загальної сукупності з восьми різних елементів (зверніть увагу на те, що герб на 1-й монеті і герб на 2-й монеті – це два різні елементи). Ці групи відрізняються одна від іншої тільки складом елементів (хоча б одним елементом). Тому вони є комбінаціями і, отже, $m = C_8^6$.

Остаточно

$$P(A) = \frac{m}{n} = \frac{C_8^6}{2^8} = \frac{C_8^2}{2^8} = \frac{8 \cdot 7}{1 \cdot 2 \cdot 2^8} = \frac{7}{2^6} \approx 0{,}1094,$$

де за третьою властивістю числа комбінацій $C_8^6 = C_8^{8-6} = C_8^2$. ●

Зауваження 1. В умові попередньої задачі вважалося, що для всіх монет імовірність випадання герба дорівнює 0,5. Якщо ж ця умова порушується (наприклад, якась із монет деформована), тоді наслідки випробування перестають бути рівноможливими, а тому класичну формулу вже не можна використовувати.

Зауваження 2. Результат попередньої задачі отримується і у випадку, коли 8 разів кидається **одна** монета (проаналізуйте).

Задача 1.13. Цифри 1,2,3,4,5,6,7 написані на однакових картках, які перемішані. Тричі навмання беруть по одній картці і кладуть їх зліва направо. Знайти імовірності того, що утворене тризначне число виявиться: а) парним; б) кратним трьом; в) кратним 5.

○ Випробування – відбір трьох карток. Загальне число наслідків випробування дорівнює числу всеможливих тризначних чисел, які можна утворити із семи цифр. Вони відрізняються або складом цифр, або їх розташуванням. Отже $n = A_7^3$.

а) Подія A – отримане тризначне число є парним. Цій події сприяють всі ті наслідки випробування, в яких остання цифра є парною, тобто або 2 , або 4, або 6. Зокрема, якщо остання цифра 2, то перші дві можуть бути відібрані з решти шести (включаючи 4 та 6). Ці двійки цифр є розміщеннями з шести по два і їх число дорівнює $A_6^2 = 6 \cdot 5 = 30$. Це ж саме число отримується і для останніх цифр 4 та 6. Отже, $m = 3 \cdot 30 = 90$ і

$$P(A) = \frac{m}{n} = \frac{90}{A_7^3} = \frac{90}{7 \cdot 6 \cdot 5} = \frac{3}{7} \approx 0,4286 .$$

б) Подія B – утворене тризначне число ділиться на 3. Число ділиться на 3 тоді і тільки тоді, коли сума його цифр ділиться на три. Тому із даних семи цифр потрібно виділити всеможливі трійки таким чином, щоб їхня сума була кратною трьом. Всі вони вичерпуються такими трійками: (1,2,3), (1,2,6), (1,3,5), (1,4,7), (1,5,6), (2,3,4), (2,3,7), (2,4,6), (2,6,7), (3,4,5), (3,5,7), (4,5,6), (5,6,7). Кожна з цих тринадцяти трійок “породжує” $P_3 = 3 \cdot 2 \cdot 1 = 6$ різних тризначних чисел, а всього їх буде $m = 13 \cdot 6 = 78$. Отже,

$$P(B) = \frac{m}{n} = \frac{78}{210} \approx 0,3714 .$$

в) Подія C – отримане тризначне число є кратним 5. Оскільки число ділиться на 5 тоді і тільки тоді, коли його останньою цифрою є 0 або 5, то сприятливими для події C є наслідки випробування, в яких останньою є цифра 5, попередні дві – будь-які з решти шести цифр. Тобто, $m = A_6^2 = 6 \cdot 5 = 30$. Остаточно

$$P(C) = \frac{m}{n} = \frac{30}{210} \approx 0,1429 . \bullet$$

Задача 1.14. Представниками кожної із п’яти фірм є директор і референт. Після підписання угоди планується провести товариську вечерю. Згідно із домовленістю місця за столом розігруються випадковим чином. Знайти імовірність того, що представники фірми *K* займуть сусідні місця.

○ Випробування – спроба посадовити 10 осіб. Подія *A* - представники фірми *K* сидітимуть поруч. Всіх наслідків випробування є стільки, скільки є різних десяток, для яких суттєвий порядок розташування осіб. Тобто $n = P_{10} = 10!$

Знайдемо число наслідків випробування, сприятливих для настання події *A*. Директор фірми *K* може зайняти кожне із десяти місць, а його референт може сісти двома способами поруч біля нього (зліва або справа). Тобто, вони можуть зайняти 20 різних позицій, перебуваючи поруч. В той же час вісімка їх партнерів може сісти за стіл 8! різними способами, утворивши перестановки із восьми осіб. Тому $m = 20 \cdot 8!$,

$$P(A) = \frac{m}{n} = \frac{20 \cdot 8!}{10!} = \frac{20 \cdot 8!}{8! \cdot 9 \cdot 10} = \frac{2}{9} \approx 0{,}2222 \text{ . } \bullet$$

Задача 1.15. В партії з 300 деталей відділ технічного контролю виявив шість нестандартних деталей. Знайти відносну частоту нестандартних деталей.

○ *A* – деталь нестандартна, випробування – перевірка деталі відділом технічного контролю. За умовою *N* = 300, *M* = 6. Тому згідно з (1.7) *W*(*A*) = 6/300 = 0,02. ●

Задача 1.16. Із 1000 навмання взятих деталей 4 браковані. Скільки бракованих виявиться серед 2400 деталей (приблизно)?

○ *A* – деталь бракована, випробування – перевірка деталі. За умовою *W*(*A*) = 0,004. Якщо серед 2400 деталей *x* бракованих, то $P(A) = \frac{x}{2400}$. Оскільки $P(A) \approx W(A)$, то $\frac{x}{2400} \approx 0{,}004$, звідки $x \approx 10$. ●

Задача 1.17. Після буревію з’ясувалося, що телефонна лінія пошкоджена на ділянці між 20-м і 40-м кілометрами. Яка імовірність того, що пошкодження сталося між 25-м і 30-м кілометрами лінії?

○ Припустимо, що імовірність знаходження точки розриву лінії на відрізку l пропорційна довжині цього відрізка і не залежить від його розташування відносно відрізка L. Тоді згідно із формулою (1.9)

$$P(A)=\frac{довжина \quad l}{довжина \quad L},$$

де випадкова подія $A=\{x\in l\}$, x – координата точки розриву на числовій осі.

В даному випадку

$$P(A)=P(25\le x\le 30)=\frac{30-25}{40-20}=\frac{5}{20}=0{,}25\,. \ \bullet$$

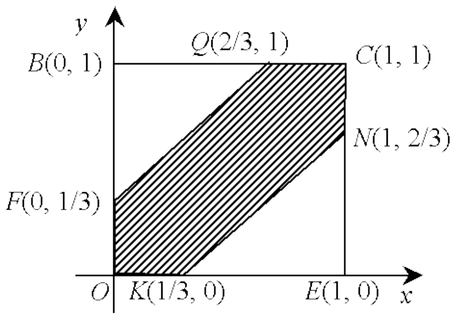
Задача 1.18. Два студенти домовилися зустрітися в певному місці між 14 і 15 годинами. Студент, який прийде першим, чекає другого протягом 20 хвилин, після чого залишає місце зустрічі. Знайти імовірність того, що зустріч відбудеться.

○ Розглянемо спочатку більш простий випадок, коли зустріч повинна відбутися між 0 і 1 годинами. Нехай x та y – моменти приходу першого та другого студентів відповідно. Тоді за умови спрощеної задачі виконуються подвійні нерівності $0\le x\le 1$, $0\le y\le 1$. Символічний запис випадкової події A – зустріч відбудеться, має такий вид:

$$A=\{|x-y|\le 1/3\},$$

де 1/3 відповідає 20 хвилинам.

Введемо прямокутну декартову систему координат xOy, в якій всі елементарні рівноможливі події (момент зустрічі $(x;\ y)$) заповнюють квадрат $\Omega = OBCE$ (мал. 1.2).



Мал.1.2.

Зобразимо графічно сукупність елементарних подій, для яких подія A відбувається. Нерівність $|x-y|\le 1/3$ рівносильна подвійній нерівності $-1/3 \le x - y \le 1/3$.

Перша з них визначає півплощину, яка знаходиться нижче граничної прямої l_1: $-1/3 = x - y$. Ця пряма проходить через точки F (0; 1/3) і Q (2/3; 1). Друга визначає півплощину, що розташована над граничною прямою l_2: $x - y = 1/3$. l_2 проходить через точки K (1/3; 0) і N (1; 2/3). Перетин отриманої «смуги» з квадратом $0BCE$ і визначає (в множинному розумінні) подію A: $A = OFQCNK$. З побудови отримуємо: $S(A) = S_{OBCE} - 2S_{\Delta FBQ} = 1 - \frac{4}{9} = \frac{5}{9}$. Згідно з рівністю (1.9)

$$P(A) = \frac{S(A)}{S(\Omega)} = \frac{5}{9}. \ \bullet$$

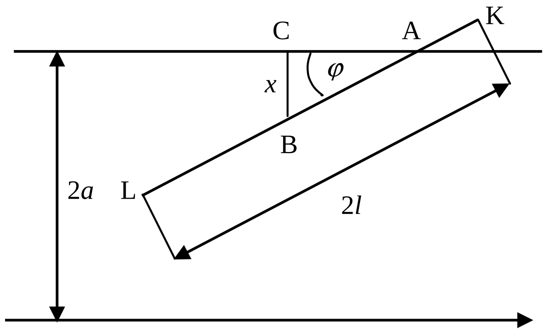
Задача 1.19. (задача Бюффона*). Площина розграфлена паралельними прямими, віддаль між якими складає $2a$. На площину навмання кидається тонка голка довжиною $2l$ ($l<a$). Знайти імовірність того, що голка перетне яку-небудь із цих прямих.

○ Подія A – голка перетне одну із прямих. На мал.1.3, зображений один із можливих наслідків відбуття події A.

Для того, щоб врахувати всі можливі наслідки випробування, а також ті, в яких відбувається подія A, введемо в рогляд позначення: x – віддаль від середини голки до найближчої прямої, φ - кут, утворений голкою із цією прямою (мал.1.3). Відмітимо, що положення голки повністю визначається заданням конкретних значень x і φ, при цьому x може змінюватися від 0 до a включно, а φ - від 0 до π. Отже, середина голки може попасти в будь-яку із точок прямокутника із довжинами сторін a і π (мал.1.4).

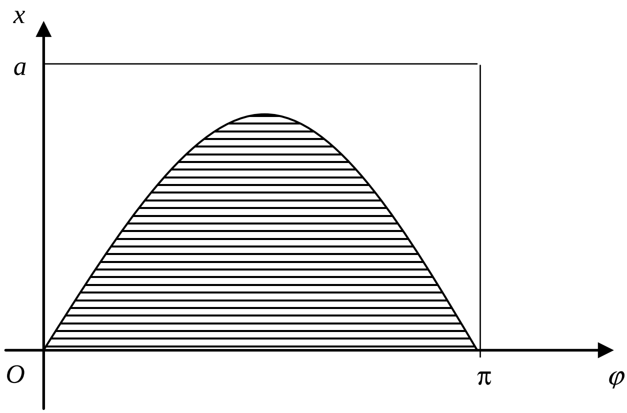
Знайдемо тепер фігуру d, кожна точка якої сприяє появі події A. Для цього розглянемо прямокутний трикутник ABC (мал.1.3), для якого $x = BC = AB\sin\varphi$. Але оскільки $AB<BK=l$, то для випадку, розглянутому на мал.1.3, отримаємо нерівність $x < l\sin\varphi$. Якщо ж точка K знаходиться на прямій, то $x = l\sin\varphi$.

* Жорж Луї Леклерк – французький природничник XVIIIст.



Мал.1.3.

Отже, голка перетне найближчу до неї пряму (матиме хоча б одну спільну точку із нею) тоді і тільки тоді, коли для x виконується подвійна нерівність $0 \le x \le l\sin\varphi$, тобто якщо середина голки попаде в будь-яку із точок фігури d, заштрихованої на мал.1.4.



Мал.1.4.

Її площа

$$\text{пл. } d = \int_0^{\pi} l\sin\varphi d\varphi = -l\cos\varphi\Big|_0^{\pi} = 2l\,.$$

Остаточно за формулою (1.9)

$$P(A) = \frac{пл.d}{пл.D} = \frac{2l}{\pi a}\,. \bullet$$

ЗАДАЧІ ДЛЯ САМОСТІЙНОГО РОЗВ'ЯЗУВАННЯ.

Завдання №1.

1. В семиповерховому готелі функціонує ліфт. На першому поверсі в ліфт зайшли 6 пасажирів, кожний із яких може вийти незалежно один від інших на довільному поверсі з другого по сьомий включно. Знайти імовірність того, що всі пасажири вийдуть: а) на різних поверхах; б) на сьомому поверсі; в) на одному поверсі.

2. Студент підготував на залік 30 питань із 40. Знайти імовірність того, що він складе залік за першим разом, якщо для цього достатньо правильно відповісти на три навмання витягнуті питання (кожне із 40 питань надруковане на окремій картці).

3. Знайти імовірність того, що навмання кинута в круг точка попаде у вписаний в нього квадрат.

4. В партії 200 виробів, з яких 4 бракованих. Партія довільним чином розподілена на дві рівні частини, які відправлені двом споживачам. Знайти імовірність того, що: а) браковані вироби потраплять до споживачів в однаковій кількості; б) всі браковані вироби дістануться одному споживачу.

5. За підсумком року акції десяти фірм мали прибуток, чотирьох фірм знецінились, а акції шести фірм зберегли свою номінальну вартість. Яка імовірність того, що випадково куплені шість акцій різних фірм матимуть прибуток?

6. Банк протягом місяця мав видати в кредит позику дванадцяти клієнтам першого району і восьми клієнтам другого району. Ця операція здійснюється поетапно. Знайти імовірність того, що за перший тиждень кредити отримають два клієнти першого району і три клієнти другого, якщо всі клієнти мають однакові можливості отримати позику.

7. В кейсі містяться 20 акцій різних видів, 5 із яких високоприбуткові. Випадковим чином всі акції розподіляються на дві рівні частини. а) Яка ймовірність попадання двох високоприбуткових акцій в одну частину і трьох в другу? б) Яка ймовірність того, що всі високоприбуткові акції потраплять в одну частину?

8. Серед 20 телевізорів, що продаються, 6 вимагають додаткового регулювання. Знайти імовірність того, що з п'яти куплених телевізорів три потребують додаткового регулювання.

9. На кожній із шести однакових карток надрукована одна із літер Е, Н, А, І, Т, У. Картки витягують навмання послідовно і складають зліва направо. Яка імовірність того, що в результаті отримається слово «ТНЕУ»?

10. Замок містить на спільній осі 4 диски, кожний із яких розподілений на 6 секторів, відмічених цифрами. Замок відкривається тільки в тому випадку, якщо всі диски займають визначене положення відносно корпусу замка (їх цифри утворюють певне число, яке складає «секрет» замка). Яка імовірність відкрити замок, набравши довільний набір цифр?

11. До контролера поступила партія однотипних виробів кількістю 16 шт. Серед них є п’ять бракованих, але про це йому невідомо. Контролер навмання бере чотири вироби для перевірки. Якщо всі відібрані вироби виявляться доброякісними, то партія пропускається. Знайти імовірність того, що партія буде пропущена контролером.

12. Цифри 1, 2, 3, 4, 5, 6, 7 написані на однакових картках, які ретельно перемішані. Тричі навмання беруть по одній картці і кладуть їх зліва направо. Знайти імовірність того, що утворене тризначне число виявиться: а) парним; б) кратним трьом; в) кратним 5.

13. Першість області по баскетболу виборюють 18 команд, які жеребкуванням розподіляються на дві групи по 9 команд в кожній. 5 команд зазвичай займають перші місця. Яка імовірність попадання всіх лідируючих команд в одну групу? Яка імовірність попадання двох лідируючих команд в одну групу і трьох – в іншу?

14. Продавець електродеталей має в коробці 18 мікросхем, серед яких 11 – аналогових і 7 – цифрових. Знайти імовірність того, що з п’яти навмання відібраних мікросхем число цифрових виявилось рівним: а) 3; б) 5; в) 0.

15. Пасажир забув дві останні цифри коду комірки автоматичної камери схову, де він залишив речі. Знайти імовірність того, що після першого набору коду із двома останніми навмання набраними цифрами комірка відкриється, а також імовірність цієї ж події у випадку, коли пасажир пам’ятає, що ці цифри різні.

16. В обласних змаганнях з футболу бере участь 16 команд, чотири з яких є найбільш сильними за підсумками минулорічних змагань. Ці команди випадковим чином розбиваються на дві підгрупи по вісім

команд в кожній. Знайти імовірність того, що: а) чотири найсильніші команди потраплять в одну підгрупу; б) в кожній із підгруп буде по дві найсильніші команди.

17. Цифри 2,3,4,5,6,7 написані на однакових картках. Випадковим чином ці картки розставляють у рядок. Обчислити імовірності випадкових подій: 1) А – цифра 3 стоятиме на першому місці, а 6 – на останньому; 2) В – цифри в числі утворюють спадну послідовність; 3) С – цифри утворюють парне число.

18. На пульті обленерго з'явилась інформація про пошкодження після буревію лінії електропередач між 30-м і 50-м кілометрами. Знайти імовірність того, що це сталося на ділянці між 40-м і 45-м кілометрами.

19. При складанні заліку студент навмання витягнув 5 питань із 80, з яких він знає 60. Якщо він знатиме хоча б 4 із п'яти, то отримає залік. Знайти імовірність успішної здачі заліку.

20. Навісний замок із «секретом» має чотири восьмикутні призми, кожна з яких повертається навколо своєї осі незалежно від інших. Бічні грані кожної призми пронумеровані цифрами від 1 до 8. Замок відкривається, якщо обертанням призм на чільній стороні замка буде набрано певне чотиризначне число. Знайти імовірність того, що особа, яка не знає «секрету», зможе відкрити за першою спробою замок, набравши на призмах довільне чотиризначне число.

21. Президент фірми хоче створити команду дизайнерів для розробки нової моделі виробу у складі двох інженерів і трьох маркетологів. Яка імовірність того, що команда такого складу буде створена, якщо з групи 9 інженерів і шести маркетологів вибрати навмання 5 осіб?

22. Із шести літер розрізної абетки складено слово «книжка». Неграмотний хлопчик змішав літери, а потім навмання їх зібрав. Яка імовірність того, що він знову отримає те ж саме слово?

23. Серед 20 видів акцій будівельних організацій 9 стали прибутковими, 5 – збитковими, а 6 залишилися без змін. Яка імовірність того, що серед п'яти навмання придбаних акцій різних видів прибутковими виявляться три?

24. Робітник отримав 16 деталей, виготовлених I заводом та 4 деталі – II заводом. Він навмання бере дві деталі. Знайти ймовірність того,

що хоча б одна з них виготовлена I заводом.

25. Із кошика з шістьма зеленими грушами і п'ятьма жовтими беруться (без повернення) випадковим чином три груші. Знайти імовірність того, що серед цих трьох груш: а) дві жовті, б) хоча б одна зелена?

26. На восьми сторінках газети розміщені рекламні оголошення, 5 сторінок присвячені соціально-політичним проблемам, 3 – спортивним новинам. Використали 4 сторінки з цієї газети. Яка імовірність того, що серед них немає сторінок із спортивними новинами?

27. Знайти імовірність того, що власник однієї картки спортлото «5 з 36» закреслить чотири виграшні числа.

28. У папці є 4 відомості, сформовані одним бухгалтером, і 5 відомостей – другим. Навмання вибирається три відомості. Знайти імовірність того, що: 1) всі три відомості сформовані другим бухгалтером; 2) відомостей, сформованих другим бухгалтером, виявиться більше, ніж першим.

29. У конверті 20 акцій, серед яких три фірми А. Навмання відібрано 4 акції. Яка імовірність того, що серед них буде одна акція фірми А?

30. У шухляді є вісім однотипних деталей, три з яких браковані (решта – стандартні). Навмання із шухляди беруть три деталі. Знайти імовірність подій A_i, де i – число бракованих деталей серед взятих (i = 0, 1, 2, 3).

31. З двадцяти пісень, трансльованих на радіо-FM, 12 є україномовними. Яка імовірність того, що слухач передачі з перших п'яти прослуханих пісень мав нагоду чути тільки українську мову?

32. В кіоску на початок зміни було 5 упаковок кави «Jacobs», 6 – «Nescafe», 8 – «Галка». Попит на кожний із цих видів кави був однаковий. За зміну було продано п'ять упаковок. Яка імовірність того, що вся кава «Jacobs» залишилася в кіоску?

33. Куб, всі грані якого зафарбовані, розпиляли на 64 кубики однакового розміру, які потім змішали. Знайти імовірність того, що навмання взятий кубик матиме зафарбованих граней: а) одну; б) дві; в) три; г) чотири; д) не матиме жодної.

34. Для деякого міста телефонний номер складається із шести цифр,

перша з яких може бути 2, 3, 4, 5, 7, 9. Знайти імовірність того, що навмання набраний номер телефону матиме: а) всі різні цифри; б) непарні цифри; в) всі однакові цифри.

35. Монета кидається 6 разів. Яка ймовірність того, що герб випаде 5 разів?

36. На п'яти картках написано по одній цифрі із набору 1, 2, 3, 4 і 5. Навмання вибираються одна за одною дві картки. Яка імовірність того, що цифра на другій картці виявиться більшою, ніж на першій?

37. В контейнері є 20 деталей, серед яких 6 нестандартних. Знайти імовірність того, що число нестандартних деталей серед п'яти навмання взятих деталей виявиться рівним: а) 0; б) 2; в) 5.

38. Знайти імовірність того, що власник однієї картки спортлото «6 із 40» закреслить числа виграшних номерів: а) три; б) п'ять; в) шість.

39. У скрині є 5 жовтих, 7 синіх і 4 білих куль. Випадковим чином виймають 6 куль. Знайти ймовірність того, що серед них будуть 1 жовта, 2 синіх і 3 білих кулі.

40. Серед 20 телевізорів фірми «Sony» 12 апаратів мають модуль для підключення телевізора до домашньої безпровідної мережі. Яка імовірність того, що серед п'яти випадково відібраних апаратів три телевізори будуть мати цей модуль.

41. Навмання обрано два додатні числа x та y, кожне з яких не перевищує 10. Знайти ймовірність того, що $x + y < 9$, а $y < 2x$.

42. До станції метро можна дістатися автобусом за 10 хвилин або пішки за 13 хвилин. Інтервал руху автобусів становить 7 хвилин. Будемо вважати, що варто їхати автобусом, якщо ймовірність випадкової події A = {швидше дістатися до метро автобусом, аніж пішки} перевищує 50 %. Чи варто чекати автобус?

43. В авіакасі заброньовано 8 квитків до Варшави, 5 квитків — до Львова і 3 — до Харкова. Однак тільки 10 квитків були викуплені. Знайти ймовірність того, що викуплено 6 квитків до Варшави і 4 — до Львова.

44. Задано множину цілих чисел {1, 2, 3, 4, 5}. Її елементи навмання розставляють у рядок. Обчислити ймовірність того, що розставлені в ряд числа утворюють зростаючу послідовність.

45. Задана множина цілих чисел {0, 1, 2, 3, 4, 5, 6, 7, 8, 9}. Яка

ймовірність того, що навмання взяті чотири числа, що розміщені в рядок, утворять число 2019?

46. 12 виробів, серед яких 4 нестандартних, випадковим чином розбивають на дві рівні партії. Знайти ймовірність того, що усі нестандартні вироби будуть в одній партії.

§ 2. ТЕОРЕМИ МНОЖЕННЯ І ДОДАВАННЯ ІМОВІРНОСТЕЙ ТА ЇХ НАСЛІДКИ

Дії над подіями (алгебра подій). Діаграми В'єнна. Умовна імовірність. Теорема множення імовірностей. Теореми додавання імовірностей. Основна властивість подій, які утворюють повну групу. Алгоритм розв'язування задач з використанням теорем додавання та множення імовірностей. Імовірність появи **хоча б однієї** події. Імовірність відбуття **тільки однієї** події. Формула повної імовірності. Формули Байєса. Алгоритм розв'язування задач з використанням формул повної імовірності та Байєса.

КОРОТКІ ТЕОРЕТИЧНІ ВІДОМОСТІ

Введемо в розгляд дії **над подіями**.

Дві події, які утворюють повну групу, називаються протилежними.

Означення. Під подією $\overline{A}$ (читається: не A) розуміється подія, яка полягає в тому, що при випробуванні подія A не відбувається і відбувається подія, протилежна до події A. Дія над подією, яка визначається рисочкою над цією подією, називається **запереченням** (цієї події).

Сумою подій A_1, A_2, ..., A_k називається подія A, яка полягає в тому, що при випробуванні відбувається хоча б одна з подій A_1, A_2, ..., A_k. Символічний запис: $A = A_1 + A_2 + ... + A_k$. Зокрема, якщо $k = 2$, тоді подія $A = A_1 + A_2$ полягає в тому, що при випробуванні відбувається **або** подія A_1, **або** подія A_2, **або** A_1 та A_2 (якщо A_1 і A_2 є сумісними). В зв'язку з цим отримуємо таке **мнемонічне правило**: знак суми асоціюється із словом «або» («+» $\leftrightarrow$ «або»). Якщо A_1 та A_2 несумісні, то $A_1 + A_2$ – це подія, яка полягає в тому, що при випробуванні відбувається або подія A_1, або A_2.

Добутком подій A_1, A_2, ..., A_k називається подія A, яка полягає в тому, що при випробуванні відбуваються всі події A_1, A_2, ..., A_k.

Символічно: А = $A_1 \cdot A_2 \cdot ... \cdot A_k$ (або $A = A_1 A_2 ... \ A_k$).
Відбуття всіх k подій передбачає відбуття **і** події A_1, **і** події A_2, ..., **і** події A_k. Отже, мнемонічно дія множення асоціюється із «і» («×» $\leftrightarrow$ «і»).

Нехай випробування, в результаті якого може відбутися випадкова подія В, доповнюється умовою про відбуття випадкової події А. Тоді **імовірність події В, знайдена при умові, що подія А відбулася, називається умовною імовірністю події** B і позначається $P_A(B)$ або $P(B|A)$.

Дві події називаються **незалежними**, якщо імовірність однієї з них не змінюється від того, відбулася чи ні інша. Аналітичним критерієм незалежності подій A та B є рівність

$$P_A(B) = P_{\overline{A}}(B). \qquad (2.1)$$

У випадку виконання (2.1) можна вважати, що

$$P_A(B) = P_{\overline{A}}(B) = P(B), \qquad (2.2)$$

де остання імовірність називається **безумовною**.

Дві події називаються **залежними**, якщо імовірність однієї з них змінюється внаслідок відбуття іншої. Тобто, якщо події A та B залежні, то рівність (2.2) порушується:

$$P_A(B) \neq P_{\overline{A}}(B). \qquad (2.3)$$

Теорема множення імовірностей

Імовірність добутку двох сумісних подій дорівнює добутку імовірностей однієї з них на умовну імовірність іншої події, обчислену в припущенні, що перша подія відбулася:

$$P(AB) = P(A)P_A(B)\big(= P(B)P_B(A)\big). \qquad (2.4)$$

Зауваження. Продовження рівності в дужках вказує на «рівноправність» подій A та B з врахуванням комутативності дії множення ($AB = BA$).

Наслідок 1. Імовірність добутку *k* подій дорівнює добутку імовірності однієї з них на умовні імовірності всіх решти, причому імовірність кожної наступної події знаходиться в припущенні, що всі попередні події вже відбулися:

$$P(A_1A_2A_3...A_k) = P(A_1)P_{A_1}(A_2)P_{A_1A_2}(A_3)...P_{A_1A_2...A_{k-1}}(A_k). \qquad (2.5)$$

Наслідок 2. Якщо події A та B незалежні, тоді

$$\boldsymbol{P(AB) = P(A)\,P(B).} \qquad (2.6)$$

Для узагальнення наслідка 2 на випадок довільного числа подій розглянемо такі означення.

Декілька подій називаються попарно незалежними, якщо **кожні** дві з них незалежні. Наприклад, події A_1, A_2, A_3 попарно незалежні, якщо незалежні події A_1 і A_2, A_1 і A_3, A_2 і A_3.

Декілька подій називаються незалежними в сукупності (або просто **незалежними**), якщо вони попарно незалежні, а також є незалежними кожна з них і всі можливі добутки інших. Наприклад, якщо події A, B, C незалежні в сукупності, то незалежні події A і B, A і C, B і C, A і BC, B і AC, C і AB.

Виявляється, що **попарна незалежність декількох подій ще не**

гарантує їх незалежність в сукупності.

Наслідок 3. Імовірність добутку k подій, незалежних в сукупності, дорівнює добутку імовірностей цих подій:

$$\boldsymbol{P(A_1 A_2 \ldots A_k) = P(A_1)\, P(A_2) \ldots P(A_k).} \qquad (2.7)$$

Теорема 1. Імовірність суми двох несумісних подій дорівнює сумі імовірностей цих подій:

$$P(A + B) = P(A) + P(B). \qquad (2.8)$$

Наслідок. **Імовірність суми k попарно несумісних подій дорівнює сумі імовірностей цих подій:**

$$\boldsymbol{P(A_1 + A_2 + \ldots + A_k) = P(A_1) + P(A_2) + \ldots + P(A_k).} \qquad (2.9)$$

Теорема 2. Імовірність суми двох сумісних подій дорівнює сумі імовірностей цих подій без імовірності їх сумісної появи:

$$\boldsymbol{P(A + B) = P(A) + P(B) - P(AB).} \qquad (2.10)$$

Теорема. Сума імовірностей подій, які утворюють повну групу, дорівнює одиниці:

$$\boldsymbol{P(A_1) + P(A_2) + \ldots P(A_n) = 1.} \qquad (2.11)$$

Для $n = 2$ події A_1, A_2 є протилежними, тому рівність (2.11) набуває такого виду:

$$P(A) + P(\overline{A}) = 1 \qquad \text{або} \qquad p + q = 1, \qquad (2.11^*)$$

де $p = P(A),\ q = P(\overline{A})$.

Алгоритм розв’язування задач з використанням теорем додавання та множення імовірностей

1) Вводяться в розгляд подія, імовірність якої треба знайти, а також більш простіші події, імовірності яких відомі або можуть бути знайдені за класичним означенням.

2) «Шукана» випадкова подія (імовірність якої потрібно знайти) виражається через простіші події за допомогою алгебри подій, тобто операцій суми, добутку, заперечення (протилежної події). При цьому потрібно керуватися мнемонічними правилами: «+» $\leftrightarrow$ **або**, «×» $\leftrightarrow$ **і**.

3) В залежності від виду отриманого виразу використовуються теореми додавання імовірностей або(і) теорема множення імовірностей та її наслідки. При реалізації цього пункту необхідно з’ясовувати властивості подій (сумісність, несумісність, залежність, незалежність, протилежність або повноту пари чи групи подій).

Зауваження. Слід мати на увазі те, що в багатьох задачах реалізація пункту 2) неєдина. В таких випадках бажано вибрати найкомпактнішу, переконавшись у співпаданні остаточних результатів після виконання пункту 3). Якщо ж результати не співпадають, то необхідно перевірити правильність побудови в п. 2) або коректність виконання п. 3).

РОЗВ'ЯЗУВАННЯ ТИПОВИХ ЗАДАЧ

Задача 2.1. Довести, що

а) $\overline{A_1 + A_2 + \ldots + A_k} = \overline{A_1} \cdot \overline{A_2} \cdot \ldots \cdot \overline{A_k}$;

б) $\overline{A_1 \cdot A_2 \cdot \ldots \cdot A_k} = \overline{A_1} + \overline{A_2} + \ldots + \overline{A_k}$.

○ а) подія $A_1 + A_2 + \ldots + A_k$ полягає в тому, що при випробуванні відбудеться хоча б одна із подій $A_1, A_2, \ldots, A_k$. Тоді $\overline{A_1 + A_2 + \ldots + A_k}$ - протилежна до неї, тобто при випробуванні не відбудеться жодна із подій $A_1, A_2, \ldots, A_k$ (і не відбувається подія A_1, і не відбувається A_2, ..., і не відбувається A_k - компактний запис: $\overline{A_1} \cdot \overline{A_2} \cdot \ldots \cdot \overline{A_k}$).

б) Оскільки подія $A_1 \cdot A_2 \cdot \ldots \cdot A_k$ полягає в тому, що при випробуванні відбудуться **всі** події ($A_1, A_2, \ldots, A_k$), то протилежна до неї подія ($\overline{A_1 \cdot A_2 \cdot \ldots \cdot A_k}$) означає невідбуття хоча б однієї події із $\overline{A_1}, \overline{A_2}, \ldots, \overline{A_k}$, тобто суму $\overline{A_1} + \overline{A_2} + \ldots + \overline{A_k}$.

Задача 2.2. В кейсі є 5 акцій першого виду, 6 – другого і 3 – третього. Знайти імовірність того, що три навмання взяті акції виявляться одного і того ж виду.

○ Позначимо: A – три навмання взяті акції є одного виду, B_i – три взяті акції i-го виду ($i = 1, 2, 3$). Подія A відбувається тоді, коли відбуваються або подія B_1, або B_2, або B_3. Інших можливостей для появи A немає. Тому мають місце такі рівності: $A = B_1 + B_2 + B_3$, $P(A) = P(B_1 + B_2 + B_3)$. Події B_1, B_2, B_3 попарно несумісні і згідно з наслідком теореми 1 (рівність (2.9))

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

За класичним означенням імовірності

$$P(B_1) = \frac{C_5^3}{C_{14}^3} = \frac{5}{182}, \quad P(B_2) = \frac{C_6^3}{C_{14}^3} = \frac{5}{91}, \quad P(B_3) = \frac{C_3^3}{C_{14}^3} = \frac{1}{364}.$$

Остаточно

$$P(A)=\frac{5}{182}+\frac{5}{91}+\frac{1}{364}=\frac{31}{364}.$$

Покажемо, як можна знайти імовірності подій B_1, B_2 та B_3, використовуючи теорему множення імовірностей. З цією метою позначимо: $B_j^{(i)}$ – j-та відібрана акція (j = 1, 2, 3) є акцією i-го виду (i = 1, 2, 3). Тоді, зокрема, B_1 відбудеться, якщо всі акції (**і** перша, **і** друга, **і** третя) є акціями першого виду, тобто коли відбуваються всі події $B_1^{(1)}, B_2^{(1)}, B_3^{(1)}$.

Отже, $B_1 = B_1^{(1)}B_2^{(1)}B_3^{(1)}$, $P(B_1)=P\left(B_1^{(1)}B_2^{(1)}B_3^{(1)}\right)$ і згідно з рівністю (2.5) для випадку k = 3

$$P(B_1)=P\left(B_1^{(1)}\right)\,P_{B_1^{(1)}}\left(B_2^{(1)}\right)\,P_{B_1^{(1)}B_2^{(1)}}\left(B_3^{(1)}\right)=\frac{5}{14}\cdot\frac{4}{13}\cdot\frac{3}{12}=\frac{5}{182}.$$

Відмітимо, що вибір рівності (2.5) на противагу рівності (2.7) зумовлений залежністю $B_1^{(1)}, B_2^{(1)}, B_3^{(1)}$ (доведіть!).

Аналогічно

$$P(B_2)=P\left(B_1^{(2)}\right)\,P_{B_1^{(2)}}\left(B_2^{(2)}\right)\,P_{B_1^{(2)}B_2^{(2)}}\left(B_3^{(2)}\right)=\frac{6}{14}\cdot\frac{5}{13}\cdot\frac{4}{12}=\frac{5}{91},$$

$$P(B_3)=P\left(B_1^{(3)}\right)\,P_{B_1^{(3)}}\left(B_2^{(3)}\right)\,P_{B_1^{(3)}B_2^{(3)}}\left(B_3^{(3)}\right)=\frac{3}{14}\cdot\frac{2}{13}\cdot\frac{1}{12}=\frac{1}{364}.$$

Отже, другий метод знаходження імовірностей подій B_1, B_2 та B_3 дає ті ж самі результати. ●

Задача 2.3. Підприємство планує здійснювати поставки двох видів виробів. Імовірність зриву поставок для першого виду виробів складає 0,05, а для другого 0,08. Згідно із проектом контракту при порушенні термінів поставок хоча б одного виду продукції до виробника застосовуються штрафні санкції, які приводять до нерентабельності виробництва обох видів виробів. Знайти імовірність нерентабельності виробництва цих виробів.

○ Позначимо: C – нерентабельність виробництва обох видів продукції, A та B – зрив поставки виробів першого та другого видів відповідно. Згідно з умовою задачі подія C відбувається, якщо відбувається **або** подія A, **або** подія B, **або** події A та B разом, тобто хоча б одна з них. Тому $C = A + B$, де A та B сумісні події. Згідно з теоремою додавання для сумісних подій

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

За умовою $P(A) = 0,05$, $P(B) = 0,08$ і можна вважати, що події A та B незалежні, тобто $P(AB) = P(A)\,P(B)$.

Остаточно $P(C) = 0,05 + 0,08 - 0,05\cdot 0,08 = 0,126$.

Другий метод розв'язування. Подію C можна представити через більш прості події A та B ще й таким чином:

$$C = A\overline{B} + \overline{A}B + AB.$$

Доданки-події справа є попарно несумісними випадковими подіями. Справді, припустимо, що події $A\overline{B}$ та $\overline{A}B$ є сумісними, тобто можуть відбутися у випробуванні. Тоді отримується суперечність: подія A, зокрема, в одному випробуванні і відбувається, і не відбувається. Це протиріччя вказує на хибність припущення. Переконайтеся в попарній несумісності цих подій, використавши діаграми В'єнна. Використавши рівності (2.9) та (2.6) і незалежність подій A та $\overline{B}$, $\overline{A}$ та B, отримаємо:

$$P(C) = P(A\overline{B} + \overline{A}B + AB) = P(A\overline{B}) + P(\overline{A}B) + P(AB) =$$

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

$$= 0,05\cdot 0,92 + 0,95\cdot 0,08 + 0,05\cdot 0,08 = 0,046 + 0,076 + 0,004 = 0,126,$$

де згідно із рівністю (2.11*)

$$P(\overline{A}) = 1 - P(A) = 1 - 0,05 = 0,95,$$

$$P(\overline{B}) = 1 - P(B) = 1 - 0,08 = 0,92.$$

Третій метод. Протилежною до події C є подія $\overline{C}$, яка полягає в тому, що внаслідок здійснення поставок продукції виробництво обох видів виробів є рентабельним. Подія $\overline{C}$ відбувається тоді, коли **і** першого виду вироби вчасно поставляються, **і** другого, тобто відбуваються події і $\overline{A}$, і $\overline{B}$. Тому $\overline{C} = \overline{A}\overline{B}$, $P(\overline{C}) = P(\overline{A})P(\overline{B}) = 0,95\cdot 0,92 = 0,874$. Але згідно з (2.11*) $P(C) + P(\overline{C}) = 1$, звідки $P(C) = 1 - P(\overline{C}) = 1 - 0,874 = 0,126$.

Висновки. Перевага третього методу, зокрема, полягає в тому, що він дозволяє розв'язати задачу для довільного скінченного числа видів продукції. Що ж до отриманої відповіді, то знайдену імовірність слід інтерпретувати таким чином: в середньому в 126 випадках з кожної тисячі (12,6%) очікується нерентабельність виробництва обох видів продукції. Оскільки така імовірність достатньо велика, то керівництву підприємства потрібно подбати про зменшення імовірностей зривів поставок або(і) зменшення штрафних санкцій. ●

Задача 2.4. У зв'язці шість різних ключів, з яких тільки одним можна відкрити замок. Навмання вибирається ключ і робиться спроба відкрити ним замок. Ключ, що не підійшов, більше не використовується. Знайти імовірність того, що для відкривання буде використано не більше трьох ключів.

○ Позначимо: A_k ($k = 1, 2, 3$) – замок буде відкрито k-тим за порядку відбору ключем, B – замок відкривається після використання не більше трьох ключів. Подія B відбудеться, якщо до замка підійде **або** перший ключ (відбувається A_1), **або** другий (при цьому перший ключ не підійшов – відбувається подія $\overline{A}_1 A_2$), **або** третій (перший і другий ключі не підійшли – відбувається подія $\overline{A}_1 \overline{A}_2 A_3$). Тобто вираз B через простіші події A_1, A_2, A_3 має такий вид:

$$B = A_1 + \overline{A}_1 A_2 + \overline{A}_1 \overline{A}_2 A_3.$$

Для знаходження $P(B)$ потрібно використати рівність (2.9), оскільки доданки є попарно несумісними подіями, а потім теорему множення імовірностей для залежних подій (обчисливши $P_{\overline{A}_1}(A_2)$ і $P_{A_1}(A_2)$, переконайтеся у тому, що події $\overline{A}_1$ і A_2 є залежними):

$$P(B) = P(A_1 + \overline{A}_1 A_2 + \overline{A}_1 \overline{A}_2 A_3) = P(A_1) + P(\overline{A}_1 A_2) + P(\overline{A}_1 \overline{A}_2 A_3) =$$
$$= P(A_1) + P(\overline{A}_1) P_{\overline{A}_1}(A_2) + P(\overline{A}_1) P_{\overline{A}_1}(\overline{A}_2) P_{\overline{A}_1 \overline{A}_2}(A_3) =$$
$$= \frac{1}{6} + \frac{5}{6} \cdot \frac{1}{5} + \frac{5}{6} \cdot \frac{4}{5} \cdot \frac{1}{4} = 0{,}5. \; \bullet$$

Задача 2.5. В шухляді є 11 карток розрізної абетки: 4 картки з літерою «и», 5 – з літерою «р» і 2 – «м». Навмання витягуються три картки і розкладаються на стіл зліва направо. Знайти імовірність того, що в результаті отримається слово «мир».

○ Нехай A – отримання слова «мир», B_1 – перша картка має літеру «м», B_2 – друга картка має літеру «и», B_3 – третя – «р». Подія A відбудеться, якщо відбудеться **і** подія B_1, **і** подія B_2, **і** подія B_3, тобто $A = B_1 B_2 B_3$. Тоді за теоремою множення імовірностей для залежних подій (рівність (2.5) для $k = 3$)

$$P(A) = P(B_1 B_2 B_3) = P(B_1) P_{B_1}(B_2) P_{B_1 B_2}(B_3) = \frac{2}{11} \cdot \frac{4}{10} \cdot \frac{5}{9} = \frac{4}{99}. \; \bullet$$

Зауваження. Цю задачу можна було б розв’язати і за класичним означенням імовірності. Але наслідки випробування (витягування трьох карток) вже не є розміщеннями, оскільки серед 11 вихідних елементів є однакові (пригадайте означення розміщення!). Ці наслідки не можна також назвати і комбінаціями по цій самій причині (крім того, розташування літер суттєве). Тому для обчислення m та n треба використовувати більш складні формули комбінаторики.

Розглянемо ще одну задачу, до розв'язування якої стосується попереднє зауваження з деякими доповненнями.

Задача 2.6. На чотирнадцяти окремих картках написано по одній літері: 5 карток з літерою «А», 4 – з літерою «І», 3 – з літерою «Р» і 2 – «Т». Навмання витягуються шість карток і розкладаються зліва направо. Яка імовірність того, що отримається слово «АРАРАТ»?

○ Позначимо випадкові події: B – отримання слова «АРАРАТ»,

A_1 – перша картка має літеру «А»,

A_2 – друга картка має літеру «Р»,

A_3 – третя картка має літеру «А»,

A_4 – четверта картка має літеру «Р»,

A_5 – п'ята картка має літеру «А»,

A_6 – шоста картка має літеру «Т»,

Подія B відбудеться, якщо відбудуться всі події $A_1, A_2, ..., A_6$, тобто $B = A_1A_2A_3A_4A_5A_6$. Використавши теорему множення імовірностей для залежних подій, отримаємо:

$$P(B) = P(A_1A_2A_3A_4A_5A_6) = P(A_1)\cdot P_{A_1}(A_2)\cdot P_{A_1A_2}(A_3)\cdot P_{A_1A_2A_3}(A_4)\cdot$$

$$\cdot P_{A_1A_2A_3A_4}(A_5)\cdot P_{A_1A_2A_3A_4A_5}(A_6) = \frac{5}{14}\cdot\frac{3}{13}\cdot\frac{4}{12}\cdot\frac{2}{11}\cdot\frac{3}{10}\cdot\frac{2}{9} = \frac{1}{3003} \approx 0,00033\,. \bullet$$

Задача 2.7. Однотипні деталі виготовляються трьома автоматами, продуктивності яких відносяться як 3:2:5. Із нерозсортованої партії деталей навмання беруться дві. Знайти імовірність того, що: а) одна із них виготовлена першим автоматом; б) обидві виготовлені одним автоматом; в) виготовлені різними автоматами.

○ Випробування – відбір двох деталей.

а) Позначимо:

B – одна із двох відібраних деталей виготовлена першим автоматом;

A_i – деталь виготовлена i-тим автоматом (i=1, 2, 3).

Подія B відбудеться, якщо одна із двох деталей виготовлена першим автоматом, а друга – іншими (другим або третім). Проте, зокрема, події A_1A_2 і A_2A_1 – різні, хоча добуток володіє властивістю комутативності A_1A_2=A_2A_1, оскільки A_1A_2 – **перша** взята виготовлена I-м і **друга** взята деталь виготовлена II-м автоматом, а A_2A_1 **– перша** деталь з II-го автомата і **друга** – з I-го. Отже,

$B = A_1A_2 + A_2A_1 + A_1A_3 + A_3A_1$.

Згідно з умовою A_1, A_2, A_3 є незалежними в сукупності і за класичним означенням імовірності

$$P(A_1)=\frac{3}{3+2+5}=0{,}3;\ P(A_2)=\frac{2}{3+2+5}=0{,}2;\ P(A_3)=\frac{5}{3+2+5}=0{,}5.$$

Використавши теорему суми для попарно несумісних подій, а потім добутку (для незалежних подій), отримаємо

$$P(B)=P(A_1A_2+A_2A_1+A_1A_3+A_3A_1)=$$
$$=P(A_1A_2)+P(A_2A_1)+P(A_1A_3)+P(A_3A_1)=$$
$$=P(A_1)\cdot P(A_2)+P(A_2)\cdot P(A_1)+P(A_1)\cdot P(A_3)+P(A_3)\cdot P(A_1)=$$
$$=0{,}3\cdot 0{,}2+0{,}2\cdot 0{,}3+0{,}3\cdot 0{,}5+0{,}5\cdot 0{,}3=0{,}42.$$

б) Подія C – обидві відібрані деталі виготовлені одним автоматом. Тоді $C = A_1A_1 + A_2A_2 + A_3A_3$ і

$$P(C)=0{,}3\cdot 0{,}3+0{,}2\cdot 0{,}2+0{,}5\cdot 0{,}5=0{,}38.$$

в) Подія D – обидві відібрані деталі виготовлені різними автоматами. Протилежною до події D є подія C, тобто $\overline{D}=C$, звідси $P(\overline{D})=P(C)=0{,}38$.

За формулою (2.11*)

$$P(D)=1-P(\overline{D})=1-0{,}38=0{,}62.$$ ●

Зауваження. Згідно із означенням добутку $A\cdot A=A$, але за означенням випробування $A_1\cdot A_1\neq A_1$, оскільки $A_1\cdot A_1$ - подія, яка полягає в тому, що і I-а, і II-а деталі виготовлені першим автоматом. По суті, слід було б вводити в розгляд 6 найпростіших подій $A_j^{(i)}$ - i-та відібрана деталь (i=1,2) виготовлена j-м автоматом (j=1,2,3) (див. другу частину розв'язування задачі 2.2).

Задача 2.8. Кожний із двох гравців почергово кидає по два гральних кубики. Виграє той із них, у кого першим випаде 12 очок (сума очок на гранях, що випали). Яка імовірність виграшу для гравця, який кидає гральні кубики: а) першим; б) другим?

○ Випробування – кидання двох гральних кубиків.

а) Позначимо:

B – виграш гравця, який почав гру першим,

A_i – випадання 12 очок в i-му випробовуванні (i=1, 2,...). Для довільного i (див. розв'язання задачі 1.3)

$$P(A_i)=\frac{1}{36}, P(\overline{A}_i)=\frac{35}{36} \quad .$$

Подія B зображається через простіші події таким чином:

$$B=A_1+\overline{A}_1\overline{A}_2A_3+\overline{A}_1\overline{A}_2\overline{A}_3\overline{A}_4A_5+\overline{A}_1\overline{A}_2\overline{A}_3\overline{A}_4\overline{A}_5\overline{A}_6A_7+\ldots \ .$$

Відмітимо, що непарним індексам відповідають спроби першого гравця, а парним – другого. Тому, зокрема, A_3 може відбутися тільки при невдалій першій своїй спробі $\overline{A}_1$ і невдалій першій спробі суперника ($\overline{A}_2$ – в протилежному випадку виграв би суперник). Аналогічна ситуація із іншими сприятливими для першого гравця наслідками (доданками у наведеній вище сумі).

Використавши теорему суми для попарно несумісних випадкових подій, а потім теорему добутку для незалежних в сукупності випадкових подій отримаємо

$$P(B)=P(A_1)+P(\overline{A}_1\cdot\overline{A}_2\cdot A_3)+P(\overline{A}_1\cdot\overline{A}_2\cdot\overline{A}_3\cdot\overline{A}_4\cdot A_5)+$$
$$+P(\overline{A}_1\cdot\overline{A}_2\cdot\overline{A}_3\cdot\overline{A}_4\cdot\overline{A}_5\cdot\overline{A}_6\cdot A_7)+\ldots=$$
$$=\frac{1}{36}+\left(\frac{35}{36}\right)^2\cdot\frac{1}{36}+\left(\frac{35}{36}\right)^4\cdot\frac{1}{36}+\left(\frac{35}{36}\right)^6\cdot\frac{1}{36}+\ldots .$$

Отриманий числовий ряд є геометричним з першим членом $a_1={}^1\!/_{36}$ і знаменником $q=\left({}^{35}\!/_{36}\right)^2<1$, його сума дорівнює $a_1/(1-q)$.

Отже,

$$P(B)=\frac{a_1}{1-q}=\frac{1/36}{1-(35/36)^2}=\frac{36}{71}\approx 0{,}507 .$$

б) Подія $\overline{B}$ – виграш гравця, який вступає у гру другим, має меншу ймовірність:

$$P(\overline{B})=1-P(B)=1-\frac{36}{71}=\frac{35}{71}\approx 0{,}493 . \ \bullet$$

Для випадку б) перевірити отриманий результат, зобразивши подію $\overline{B}$ через простіші події, а потім використавши теореми додавання і множення імовірностей.

Задача 2.9. Відомо, що випадкові події A та B незалежні, причому $P(A\overline{B})=0{,}46$, $P(\overline{A}B)=0{,}34$, $P(A+B)=0{,}8$. Знайти $P(A)+P(B)$ і з'ясувати питання: події A та B сумісні чи ні.

○ Використовуючи означення суми двох подій *A* та *B* (див. другий метод розв'язування задачі 2.3), отримаємо

$A+B = A\overline{B} + \overline{A}B + AB$.

Тоді з урахуванням умов задачі

$$P(A+B) = P(A\overline{B} + \overline{A}B + AB) = P(A\overline{B}) + P(\overline{A}B) + P(AB) =$$

$$= 0{,}46 + 0{,}34 + P(AB) = 0{,}8 + P(AB).$$

З другого боку, за умовою $P(A+B) = 0{,}8$.

Отже, $0{,}8 + P(AB) = 0{,}8$, $P(AB) = 0$, що означає несумісність подій *A* та *B*.

Таким чином, $P(A) + P(B) = 0{,}8$. ●

Імовірність появи хоча б однієї події

Нехай в результаті випробування можуть відбутися події $A_1, A_2, ..., A_k$, імовірності появи кожної з яких відомі і які є **незалежними в сукупності**. Позначимо *A* – поява хоча б однієї із цих подій у випробуванні. Тоді згідно з означенням суми подій $A = A_1 + A_2 + ... + A_k$. Спроба узагальнити теорему 2 на більше число подій-доданків наштовхується на труднощі, які ілюструє рівність для трьох **сумісних** подій:

$$P(A_1 + A_2 + A_3) = P(A_1) + P(A_2) + P(A_3) - P(A_1A_2) -$$
$$- P(A_1A_3) - P(A_2A_3) + P(A_1A_2A_3).$$

Доведіть цю рівність, використавши теорему 2 спочатку до суми подій $B = A_1 + A_2$ і A_3, а потім до суми A_1A_3 і A_2A_3. На останньому кроці слід скористатися рівністю $A_1A_3A_2A_3 = A_1A_2A_3$ (переконайтеся в її правильності).

Разом з тим третій метод розв'язування задачі 2.2 вказує шлях до знаходження $P(A)$, підсумком якого є така рівність:

$$P(A) = 1 - q_1 q_2 ... q_k, \qquad (2.12)$$

де $q_1 = P(\overline{A}_1), q_2 = P(\overline{A}_2), \ldots, q_k = P(\overline{A}_k)$.

В частинному випадку, коли $P(A_1) = P(A_2) = ... = P(A_k) = p$, отримується така формула

$$P(A) = 1 - q^k, \qquad (2.12^*)$$

де $q = 1 - p$.

Нарешті, якщо події $A_1, A_2, ..., A_k$ є **залежними** (не володіють властивістю незалежності в сукупності), тоді

$$P(A) = 1 - P(\overline{A}_1) P_{\overline{A}_1}(\overline{A}_2) \ldots P_{\overline{A}_1\overline{A}_2 ... \overline{A}_{k-1}}(\overline{A}_k). \qquad (2.13)$$

Задача 2.10. Підприємство отримує від суміжників продукцію

$П_1, П_2, П_3, П_4$ і виконає план, якщо вчасно отримає якісну продукцію в потрібній кількості. Ймовірність зриву поставок (як по кількості, так і по якості) по кожній із цих компонент технологічного процесу відповідно дорівнює 0,02; 0,01; 0,1; 0,05. Знайти імовірність невиконання заводом плану.

○ Позначимо: A – невиконання заводом плану, A_i – зрив поставки компоненти $П_i$ (i = 1, 2, 3, 4). Якщо відбувається хоча одна із подій A_1, A_2, A_3, A_4, тоді відбувається подія A. За умовою задачі $P(A_1) = 0{,}02$; $P(A_2) = 0{,}01$; $P(A_3) = 0{,}1$; $P(A_4) = 0{,}05$ і ці події можна вважати незалежними в сукупності. Використовуючи формулу (2.12), отримаємо:

$$q_1 = 1 - 0{,}02 = 0{,}98; \quad q_2 = 1 - 0{,}01 = 0{,}99;$$

$$q_3 = 1 - 0{,}1 = 0{,}9; \qquad q_4 = 1 - 0{,}05 = 0{,}95,$$

$$\mathrm{P(A)} = 1 - 0{,}98 \cdot 0{,}99 \cdot 0{,}9 \cdot 0{,}95 = 1 - 0{,}829521 \approx 0{,}171. \; \bullet$$

Задача 2.11. На залік виноситься 60 питань, кожне з яких надруковане на окремій картці. За правилами здачі цього предмету, студент навмання витягує 4 картки. Якщо він знатиме хоча б одне питання, тоді йому не потрібно повторно захищати розрахункові індивідуальні завдання (навіть якщо він і не отримує залік). Знайти імовірність того, що перший студент не буде повторно захищати свої індивідуальні завдання, якщо він підготував 40 питань.

○ Нехай A – студент не буде повторно захищати свої завдання, A_i – студент знає i-е витягнуте питання $(i = \overline{1,4})$. Подія A відбудеться при появі хоча б однієї із подій A_1, A_2, A_3, A_4. Але ці події не володіють властивістю незалежності у сукупності. Для доведення цього достатньо показати, наприклад, залежність подій A_1 та A_2:

$$P_{A_1}(A_2) = \frac{39}{59}; \; P_{\overline{A}_1}(A_2) = \frac{40}{59} \quad \Rightarrow \quad P_{A_1}(A_2) \neq P_{\overline{A}_1}(A_2).$$

За формулою (2.13) для випадку $k = 4$

$$P(A) = 1 - P(\overline{A}_1) P_{\overline{A}_1}(\overline{A}_2) P_{\overline{A}_1 \overline{A}_2}(\overline{A}_3) P_{\overline{A}_1 \overline{A}_2 \overline{A}_3}(\overline{A}_4) =$$

$$= 1 - \frac{20}{60} \cdot \frac{19}{59} \cdot \frac{18}{58} \cdot \frac{17}{57} = 1 - \frac{17}{1711} \approx 0{,}99.$$

Другий спосіб розв'язування. Протилежною до події A є $\overline{A}$ – всіх чотирьох питань студент не знає. За класичним означенням

$$P(\overline{A})=\frac{C_{20}^{4}}{C_{60}^{4}}=\frac{20\cdot19\cdot18\cdot17}{60\cdot59\cdot58\cdot57}=\frac{17}{1711}.$$

Згідно із властивістю протилежності подій

$$P(A)=1-P(\overline{A})=1-\frac{17}{1711}\approx0{,}99.$$ ●

Задача 2.12. Взимку при включенні запалення двигун почне працювати з імовірністю 0,7. Знайти імовірність того, що для запуску двигуна доведеться включати запалення менше чотирьох разів.

○ Випробування – включення запалення.

Позначимо: B – двигун запрацює до четвертого включення запалення, A_i – двигун запускається при i-му включенні запалення, де i=1, 2, 3, оскільки 3<4. На перший погляд подія B відбудеться тоді, коли появиться хоча б одна із подій A_1, A_2, A_3. Це веде до використання формули (2.12). Проте, якщо відбудеться, наприклад, подія A_2 , то немає сенсу виключати запалення, а потім знову його включати (зокрема, не достовірно, що двигун запрацює при третьому включенні запалення, якщо він запрацював при другому включенні).

Отже, подія B відбудеться або при першій вдалій спробі (A_1), або при другій ($\overline{A}_1A_2$) або при третій ($\overline{A}_1\overline{A}_2A_3$):

$$B=A_1+\overline{A}_1A_2+\overline{A}_1\overline{A}_2A_3\,.$$

Використання теорем додавання та множення ймовірностей дає:

$$P(B)=P(A_1+\overline{A}_1A_2+\overline{A}_1\overline{A}_2A_3)=$$

$$=P(A_1)+P(\overline{A}_1)\cdot P(A_2)+P(\overline{A}_1)\cdot P(\overline{A}_2)\cdot P(A_3)=$$

$$=0{,}7+0{,}3\cdot0{,}7+0{,}3\cdot0{,}3\cdot0{,}7=0{,}7+0{,}21+0{,}063=0{,}973\,.$$ ●

Висновки: 1. Використанню тієї чи іншої формули повинен передувати вдумливий аналіз умов задачі.

2. Імовірності вдалої спроби різко зменшуються із збільшенням номера спроби. Тому слід подбати про збільшення ефективності запалювальної системи і якість роботи акумулятора.

Задача 2.13. Скільки карток спортлото «5 із 36» треба придбати, щоб з імовірністю, не менше від α, виграла хоча б одна із них?

○ Нехай придбано k (поки що невідоме число) карток. Введемо в розгляд події: A_j – виграла j-та картка $(j=\overline{1;k})$, A – виграшною виявиться хоча б одна із k карток (відбудеться хоча б одна із подій $A_1, A_2, ..., A_k$). Оскільки картки заповнюються довільним (випадковим чином) і загальне число різних їх заповнень дуже велике $\left(C_{36}^{5} = 376992\right)$, то можна вважати події $A_1, A_2, ..., A_k$ незалежними у сукупності і виконаними рівності $P(A_1) = P(A_2) = ... = P(A_k) = p$. Навмання заповнена картка буде виграшною, якщо в результаті проведення тиражу буде вгадано 3 числа (B_3), або 4 (B_4), або 5 (B_5). З урахуванням несумісності цих подій $p = P(B_3) + P(B_4) + P(B_5)$. За класичним означенням

$$P(B_3) = \frac{C_5^3 \cdot C_{31}^2}{C_{36}^5} = \frac{4650}{376992}, \quad P(B_4) = \frac{C_5^4 \cdot C_{31}^1}{C_{36}^5} = \frac{155}{376992},$$

$$P(B_5) = \frac{C_5^5}{C_{36}^5} = \frac{1}{376992}.$$

Тому $p = \dfrac{4650}{376992} + \dfrac{115}{376992} + \dfrac{1}{376992} = \dfrac{4766}{376992}$, $\quad q \approx 0{,}987.$

За формулою (2.12*) $P(A) = 1 - q^k$. Але за умовою k повинно бути таким, щоб виконувалась нерівність $P(A) \geq \alpha$, звідки $q^k \leq 1 - \alpha$. Прологарифмувавши ліву і праву частини і врахувавши те, що $\ln q < 0$, остаточно отримаємо:

$$k \geq \frac{\ln(1-\alpha)}{\ln q}, \text{ де } q \approx 0{,}987. \ \bullet$$

Задача 2.14. Імовірності появи кожної з трьох незалежних у сукупності подій A_1, A_2, A_3 відповідно рівні p_1, p_2, p_3. Знайти імовірність появи у випробуванні тільки однієї з цих подій.

○ Нехай B – відбуття тільки однієї із подій A_1, A_2, A_3. Поява події B передбачає **або** відбуття події A_1 (і не відбуття A_2 та A_3), **або** події A_2 (і не відбуття A_1 та A_3), **або** події A_3 (і не відбуття A_1 та A_2), тобто

$$B = A_1\overline{A}_2\overline{A}_3 + \overline{A}_1 A_2 \overline{A}_3 + \overline{A}_1\overline{A}_2 A_3.$$

Події-доданки є попарно несумісними (доведіть від протилежного), а співмножники кожного доданку – незалежні у сукупності. Використавши теореми суми, а потім добутку імовірностей, отримаємо:

$$P(B) = P(A_1\overline{A}_2\overline{A}_3 + \overline{A}_1 A_2 \overline{A}_3 + \overline{A}_1\overline{A}_2 A_3) = P(A_1\overline{A}_2\overline{A}_3) + P(\overline{A}_1 A_2 \overline{A}_3) +$$
$$+P(\overline{A}_1\overline{A}_2 A_3) = P(A_1)P(\overline{A}_2)P(\overline{A}_3) + P(\overline{A}_1)P(A_2)P(\overline{A}_3) +$$
$$+P(\overline{A}_1)P(\overline{A}_2)P(A_3) = p_1 q_2 q_3 + q_1 p_2 q_3 + q_1 q_2 p_3,$$

де $q_i = 1 - p_i$, $i = 1, 2, 3$. ●

Задача 2.15. Виходячи з умови задачі 2.14, знайти імовірність того, що у випробуванні відбудуться: а) всі; б) жодна; в) тільки дві; г) хоча б дві; д) не більше однієї; е) не більше двох.

○ Обмежимося виконанням пункту 2) алгоритму, залишаючи завершення для самостійної роботи.

а) $A_1A_2A_3$; б) $\overline{A}_1\overline{A}_2\overline{A}_3$; в) $A_1A_2\overline{A}_3 + A_1\overline{A}_2A_3 + \overline{A}_1A_2A_3$;

г) $A_1A_2\overline{A}_3 + A_1\overline{A}_2A_3 + \overline{A}_1A_2A_3 + A_1A_2A_3$;

д) $\overline{A}_1\overline{A}_2\overline{A}_3 + A_1\overline{A}_2\overline{A}_3 + \overline{A}_1A_2\overline{A}_3 + \overline{A}_1\overline{A}_2A_3$;

е) протилежною подією до події, імовірність якої треба знайти, є $A_1A_2A_3$; далі використати властивість протилежної події. ●

Задача 2.16. Для вчасного виконання термінового замовлення протягом зміни достатньо, щоб безперебійно працювали два станки-автомати. В цеху є три такі станки, проте імовірності їх поломок складають відповідно 0,1; 0,05; 0,2. Знайти імовірність вчасного виконання замовлення.

○ Нехай B – замовлення виконується вчасно, A_i –протягом зміни безвідмовно працює i-ий станок (i=1,2,3).

За умовою $P(\overline{A}_1) = 0{,}1$, $P(\overline{A}_2) = 0{,}05$, $P(\overline{A}_3) = 0{,}2$. Тоді $P(A_1) = 0{,}9$, $P(A_2) = 0{,}95$, $P(A_3) = 0{,}8$.

Оскільки події A_1, A_2, A_3 – випадкові, то в якості підстраховки (виконання завдання) слід враховувати можливість сумісної роботи всіх трьох станків, тобто помилково вважати, що подія B полягає у відбутті **тільки двох подій** із трьох у випробуванні. Отже, “структура” події B така: $B = A_1 \cdot A_2 \cdot \overline{A}_3 + A_1 \cdot \overline{A}_2 \cdot A_3 + \overline{A}_1 \cdot A_2 \cdot A_3 + A_1 \cdot A_2 \cdot A_3$.

Використання теорем суми і добутку імовірностей дає:

$$P(B) = P(A_1)\cdot P(A_2)\cdot P(\overline{A}_3) + P(A_1)\cdot P(\overline{A}_2)\cdot P(A_3) + P(\overline{A}_1)\cdot P(A_2)\cdot P(A_3) +$$
$$+ P(A_1)\cdot P(A_2)\cdot P(A_3) = 0{,}9\cdot 0{,}95\cdot 0{,}2 + 0{,}9\cdot 0{,}05\cdot 0{,}8 + 0{,}1\cdot 0{,}95\cdot 0{,}8 +$$
$$+0{,}9\cdot 0{,}95\cdot 0{,}8 = 0{,}967\,.$$ ●

Задача 2.17. Імовірність (статистична) влучання в кошик при кожному виконанні штрафного кидка для першого баскетболіста дорівнює 0,8, а для другого – 0,9. Обидва вони, починаючи з І-го, почергово здійснюють кидки, проте виконують їх не більше, ніж по два. При цьому кожний баскетболіст виконує другий кидок тільки при умові, що перший зроблений ним кидок буде невдалим. Знайти ймовірність того, що в підсумку виявиться два влучних кидки.

○ Позначимо: C – баскетболісти в сумі набирають 2 (два влучних кидки), A_i та B_i – в i-му кидку (i=1,2) влучає перший та другий баскетболіст відповідно. За умовою $P(A_1) = P(A_2) = 0{,}8$, $P(B_1) = P(B_2) = 0{,}9$, звідки $P(\overline{A}_1) = P(\overline{A}_2) = 0{,}2$, $P(\overline{B}_1) = P(\overline{B}_2) = 0{,}1$.

Знову же таки, формально, відбуття події C передбачає появу тільки двох подій (двох влучних кидків). Переберемо всі варіанти відбуття події C:

1) кожний перший кидок обох баскетболістів влучний;
2) у І-го баскетболіста перший кидок вдалий (він не має права на наступний кидок), у ІІ-го – перший промах, другий влучний;
3) у І-го баскетболіста перший кидок невдалий, другий влучний і у ІІ-го перший кидок влучний;
4) перші кидки у обох невдалі, а другі – влучні.

Тому $C = A_1B_1 + A_1\overline{B}_1B_2 + \overline{A}_1B_1A_2 + \overline{A}_1\overline{B}_1A_2B_2$.

Відмітимо, що доданки $A_1\overline{B}_1A_2$ та $\overline{A}_1B_1\overline{A}_2B_2$ не можуть фігурувати в сумі, оскільки за умовою задачі кожний із баскетболістів виконує другий кидок при умові, що перший був невдалим.

Використання теорем суми і добутку імовірностей дає:

$$P(C) = P(A_1B_1 + A_1\overline{B}_1B_2 + \overline{A}_1B_1A_2 + \overline{A}_1\overline{B}_1A_2B_2) =$$
$$= 0{,}8\cdot 0{,}9 + 0{,}8\cdot 0{,}1\cdot 0{,}9 + 0{,}2\cdot 0{,}9\cdot 0{,}8 + 0{,}2\cdot 0{,}1\cdot 0{,}8\cdot 0{,}9 = 0{,}9504 .$$ ●

Формула повної імовірності

Нехай подія A може відбутися тільки при умові появи однієї із подій $B_1, B_2, ..., B_n$, які утворюють повну групу. Нехай відомі імовірності $P(B_k)$, $P_{B_k}(A)$, $k = (\overline{1,n})$. Відповідь на питання: як знайти $P(A)$? – дає

Теорема. Імовірність події A, яка може відбутися тільки після появи однієї із подій $B_1, B_2, ..., B_n$, які утворюють повну групу, знаходиться за так званою формулою повної імовірності:

$$P(A) = P(B_1)P_{B_1}(A) + P(B_2)P_{B_2}(A) + \ldots + P(B_n)P_{B_n}(A). \qquad (2.14)$$

Зауваження. Оскільки до випробування невідомо, після якої із подій $B_1, B_2, ..., B_n$ відбудеться подія $\boldsymbol{A}$, то ці події називаються

гіпотезами (припущеннями).

Формули Байєса

Нехай виконуються умови теореми відносно подій $B_1, B_2, ..., B_n$ та A. Припустимо, що проведено випробування, в результаті якого відбулася подія A. Виникає питання: як «переоцінити» імовірності гіпотез $B_1, B_2, ..., B_n$ з урахуванням відбуття події A, тобто знайти умовні імовірності $P_A(B_k)$, $(k=\overline{1,n})$? Відповідь дають так звані формули Байєса:

$$P_A(B_k)=\frac{P(B_k)P_{B_k}(A)}{\sum_{i=1}^{n}P(B_i)P_{B_i}(A)},k=1,2,\ldots,n. \qquad (2.15)$$

Алгоритм розв'язування задач з використанням формул повної імовірності та Байєса

Рекомендується така послідовність розв'язування задач.

1) Формулюють гіпотези $B_1, B_2, ..., B_n$ і подію A. При цьому слід перевірити **повноту групи гіпотез**, а також те, що **подія *A* може відбутися тільки після появи однієї із гіпотез**.

2) Знаходяться імовірності гіпотез. Правильність розрахунків контролюється виконанням рівності $P(B_1) + P(B_2) + ... + P(B_n) = 1$. Обчислюються умовні імовірності $P_{B_1}(A), P_{B_2}(A), \ldots, P_{B_n}(A)$.

3) Вибирається формула повної імовірності або формули Байєса. Останні використовуються тоді, коли є інформація про відбуття випадкової події.

Задача 2.18. Три верстати-автомати штампують однотипні деталі, що потрапляють на спільний конвейєр. Продуктивність другого автомату на 40% вища від продуктивності першого і вдвічі – від третього. Відсоток браку для кожного з автоматів дорівнює відповідно 3, 6, 2. а). Яка імовірність того, що навмання взята з конвейєра деталь виявиться бракованою? б). Навмання взята деталь виявилася бракованою. Що імовірніше: ця деталь виготовлена першим чи третім автоматом?

○ а) Першу частину задачі розв'язуємо за формулою повної імовірності, оскільки відсутня інформація про відбуття випадкової події.

При відборі довільним чином деталі з конвейєра невідомо, яким автоматом вона виготовлена. Цю **невизначеність не можна трактувати як відбуття випадкової події** (типова помилка студентів). Можна зробити припущення: будь-яка деталь (стандартна чи бракована) виготовлена або першим автоматом (подія B_1), або другим (B_2), або третім (B_3). Неважко переконатися в тому, що події B_1, B_2, B_3 утворюють повну групу (доведіть!).

Позначимо: A – навмання взята деталь бракована. Ясно, що подія A може відбутися після появи однієї із гіпотез, адже щоб говорити про бракованість деталі, потрібно її спочатку виготовити.

Знайдемо імовірності гіпотез. Для цього позначимо через a кількість деталей, що виготовить перший автомат за деякий проміжок часу. Тоді за умовою задачі за цей же проміжок часу другий автомат виготовить $1{,}4a$ деталей, а третій – $0{,}7a$. За класичним означенням імовірності

$$P(B_1)=\frac{a}{3{,}1a}=\frac{10}{31};\;\; P(B_2)=\frac{1{,}4a}{3{,}1a}=\frac{14}{31};\;\; P(B_3)=\frac{0{,}7a}{3{,}1a}=\frac{7}{31}.$$

Перевіримо обчислення. Оскільки гіпотези утворюють повну групу, то сума їх імовірностей повинна дорівнювати одиниці:

$$P(B_1)+P(B_2)+P(B_3)=\frac{10}{31}+\frac{14}{31}+\frac{7}{31}=1.$$

Згідно з умовою задачі і за класичним означенням

$$P_{B_1}(A)=\frac{3}{100}=0{,}03;\;\; P_{B_2}(A)=\frac{6}{100}=0{,}06;\;\; P_{B_3}(A)=\frac{2}{100}=0{,}02.$$

Прочитайте ліві частини цих рівностей з урахуванням змісту подій A, B_1, B_2, B_3, а також переконайтеся у правильності цих рівностей.

Формула повної імовірності (2.14) у цьому випадку має такий вид:

$$P(A)=P(B_1)P_{B_1}(A)+P(B_2)P_{B_2}(A)+P(B_3)P_{B_3}(A).$$

Підставивши обчислені дані в праву частину, отримаємо

$$P(A)=\frac{10}{31}\cdot 0{,}03+\frac{14}{31}\cdot 0{,}06+\frac{7}{31}\cdot 0{,}02\approx 0{,}0401.$$

б) У другій частині задачі використаємо формули Байєса, оскільки є інформація про відбуття випадкової події (A – деталь виявилася бракованою). При цьому ми повинні знайти $P_A(B_1)$, $P_A(B_3)$ і порівняти їх. Але більшою з них буде та імовірність, чисельник в правій частині формули Байєса для якої буде більшим. Тобто, достатньо порівняти $P(B_1)P_{B_1}(A)$ і $P(B_3)P_{B_3}(A)$:

$$P(B_1)P_{B_1}(A)=\frac{0,3}{31},\quad P(B_3)P_{B_3}(A)=\frac{0,14}{31}.$$ Отже, $P_A(B_1) > P_A(B_3)$, і можна зробити висновок: більш імовірніше, що бракована деталь виготовлена першим автоматом. ●

Задача 2.19. Студент знає відповіді на 20 білетів із 30. Коли йому краще взяти екзаменаційний білет: першим чи другим?

○ Потрібно знайти імовірності події A – студент знає витягнутий білет для двох випадків: 1) він іде першим; 2) він бере білет другим. В залежності від отриманих результатів треба буде зробити висновок.

В першому випадку за класичним означенням імовірності

$$P(A)=\frac{m}{n}=\frac{20}{30}=\frac{2}{3}.$$

В другому випадку перед студентом знаходиться 29 білетів і тому необхідно розглянути припущення: попередній студент взяв білет, який знає другий (подія B_1), попередній студент взяв білет, якого другий студент не підготував (подія B_2). Переконайтеся в тому, що події B_1, B_2 утворюють повну групу (є протилежними), а подія A може відбутися після появи однієї із подій B_1 та B_2. В задачі відсутня інформація про відбуття випадкової події, тому треба використати формулу повної імовірності:

$$P(A)=P(B_1)P_{B_1}(A)+P(B_2)P_{B_2}(A).$$

За класичним означенням знайдемо:

$$P(B_1)=\frac{20}{30}=\frac{2}{3},\ P(B_2)=\frac{10}{30}=\frac{1}{3},\ P(B_1)+P(B_2)=1,$$

$$P_{B_1}(A)=\frac{19}{29},\ P_{B_2}(A)=\frac{20}{29}.$$

Тоді $P(A)=\frac{2}{3}\cdot\frac{19}{29}+\frac{1}{3}\cdot\frac{20}{29}=\frac{38+20}{3\cdot 29}=\frac{58}{3\cdot 29}=\frac{2}{3}.$

Отже, отримані імовірності для обох випадків рівні, і для студента байдуже, яким чином брати білет: першим чи другим. ●

Розроблення «стратегії» здачі екзамену – актуальне питання для студента. Тому природно поставити питання, чи зміниться ситуація, коли він братиме білет третім по рахунку. Число гіпотез в цьому випадку збільшиться: до нього взяли два білети, з яких він не знає жодного (B_1), один (B_2), обидва (B_3). Пропонується здійснити обчислення за формулою повної імовірності в цьому випадку і переконатися в тому, що імовірність знання взятого білета знову не зміниться.

Розглянемо ще одну із задач, розв'язування яких (у більшості випадків) викликає труднощі у студентів в плані перевірки виконання умов, при яких мають місце формули повної імовірності або Байєса.

Задача 2.20. Два студенти незалежно один від другого здійснили по одному пострілу в одну мішень. Імовірність влучання в мішень для першого студента дорівнює 0,8, для другого – 0,6. Після залпу в мішені виявлена одна пробоїна. Знайти імовірність того, що у мішень влучив перший студент.

○ Сформулюємо гіпотези, не звертаючи уваги на результат випробування. Позначимо: B_1 – в мішень не попаде ні один студент, B_2 – попаде перший і не влучить другий, B_3 – не влучить перший і влучить другий, B_4 – обидва влучать, A – в мішені одна пробоїна. Обчислимо імовірності гіпотез, попередньо ввівши більш прості випадкові події: C – влучання в мішень першим студентом, D – влучання другим студентом. Тоді згідно із змістом подій B_k :

$$B_1 = \overline{C} \cdot \overline{D}, \quad B_2 = C \cdot \overline{D}, \quad B_3 = \overline{C} \cdot D, \quad B_4 = C \cdot D,$$

і оскільки C та D незалежні випадкові події, то за теоремою множення імовірностей

$$P(B_1) = P(\overline{C})P(\overline{D}) = 0{,}2 \cdot 0{,}4 = 0{,}08, \ P(B_2) = P(C)P(\overline{D}) = 0{,}8 \cdot 0{,}4 = 0{,}32,$$

$$P(B_3) = P(\overline{C})P(D) = 0{,}2 \cdot 0{,}6 = 0{,}12, \ P(B_4) = P(C)P(D) = 0{,}8 \cdot 0{,}6 = 0{,}48,$$

$$\sum_{i=1}^{4} P(B_i) = 0{,}08 + 0{,}32 + 0{,}12 + 0{,}48 = 1.$$

При знаходженні умовних імовірностей $P_{B_k}(A), k = \overline{1{,}4},$ уважно прочитайте їх з урахуванням змісту подій $B_k, k = \overline{1{,}4},$ та A і переконайтеся в правильності таких рівностей:

$$P_{B_1}(A) = 0, \ P_{B_2}(A) = P_{B_3}(A) = 1, \ P_{B_4}(A) = 0.$$

За формулою повної імовірності для чотирьох гіпотез отримуємо

$$P(A) = \sum_{k=1}^{4} P(B_k)P_{B_k}(A) = 0{,}08 \cdot 0 + 0{,}32 \cdot 1 + 0{,}12 \cdot 1 + 0{,}48 \cdot 0 = 0{,}44.$$

В даній задачі потрібно використати формулу Байєса, бо за умовою задачі є інформація про відбуття випадкової події (в мішені виявлена пробоїна після залпу – подія A). Тому шукана імовірність

$P_A(B_2)$ знаходиться за цією формулою:

$$P_A(B_2) = \frac{P(B_2) \cdot P_{B_2}(A)}{P(A)} = \frac{0{,}32 \cdot 1}{0{,}44} = \frac{8}{11} \text{ . } \bullet$$

Задача 2.21. Відомо, що в період між переналадками обладнання в середньому 96% випущеної підприємством продукції задовільняє умовам стандарту. На підприємстві в якості експерименту почали використовувати спрощену систему контролю якості продукції. Результати показали, що ця система визнає придатним виріб з ймовірністю 0,99, якщо він справді придатний, і з ймовірністю 0,07, якщо він бракований.

1) Знайти імовірність того, що навмання відібраний виріб буде визнано стандартним згідно цієї спрощеної системи контролю.
2) Яка ймовірність того, що виріб справді стандартний, якщо він: а) визнаний стандартним за спрощеним контролем; б) двічі пройшов (визнаний стандартним) за спрощеним контролем?

○ 1) Навмання взятий виріб може бути або справді стандартним (гіпотеза B_1), або бракованим (гіпотеза B_2). Тоді A – виріб визнано стандартним згідно спрощеної схеми – може відбутися тільки після появи однієї із гіпотез.

За умовою

$$P(B_1) = 0{,}96, \; P(B_2) = 0{,}04, \; P_{B_1}(A) = 0{,}99, \; P_{B_2}(A) = 0{,}07 \text{ .}$$

Відсутня також інформація про відбуття випадкової події. Тому за формулою повної імовірності (2.14) при n=2 отримаємо

$$P(A) = 0{,}96 \cdot 0{,}99 + 0{,}04 \cdot 0{,}07 = 0{,}9532 \text{ .}$$

2. а) Згідно з умовою подія A відбулася. Тому за формулою Байєса

$$P_A(B_1) = \frac{P(B_1) \cdot P_{B_1}(A)}{P(A)} = \frac{0.96 \cdot 0{,}99}{0{,}9532} \approx 0{,}9971.$$

б) Позначимо A^* – виріб двічі визнаний стандартним за спрощеним контролем, A_i – виріб визнано стандартним при i-тій перевірці за спрощеним контролем (i=1,2). Тоді $A^* = A_1 A_2$, де події A_1 і A_2 – незалежні, бо відбуття A_1 – не змінює імовірності A_2 (проаналізуйте!). Знайдемо спочатку $P(A^*)$ за формулою повної імовірності:

$$P(A^*) = P(B_1)\cdot P_{B_1}(A^*) + P(B_2)\cdot P_{B_2}(A^*),$$

де згідно із теоремою множення імовірностей для незалежних подій

$$P_{B_1}(A^*) = P_{B_1}(A_1A_2) = P_{B_1}(A_1)\cdot P_{B_1}(A_2) = 0{,}99\cdot 0{,}99 = 0{,}9801,$$

$$P_{B_2}(A^*) = P_{B_2}(A_1)\cdot P_{B_2}(A_2) = 0{,}07\cdot 0{,}07 = 0{,}0049.$$

Тоді

$$P(A^*) = 0{,}96\cdot 0{,}9801 + 0{,}04\cdot 0{,}0049 = 0{,}941092.$$

Оскільки подія A^* відбулася, то за формулою Байєса

$$P_{A^*}(B_1) = \frac{P(B_1)\cdot P_{B_1}(A^*)}{P(A^*)} = \frac{0.96\cdot 0{,}9801}{0{,}941092} = 0{,}999792. \ \bullet$$

Зауваження. Організація контролю якості продукції є актуальною задачею кожного виробництва. З одного боку – контроль вимагає економічних витрат, що відображаються на собівартості продукції. З другого боку, ігнорування його або зменшення уваги до нього може призвести до штрафних санкцій і втрати становища на ринку. Тому природно здійснювати здешевлення контролю якості продукції при збереженні певного рівня достовірності висновків. В зв'язку з цим стосовно попередньої задачі (п.2б) можна зробити такий **висновок**: оскільки $P_{A^*}(B_2) = 1 - P_{A^*}(B_1) = 1 - 0{,}999792 = 0{,}000208$ дуже мала, то гіпотезу B_2 про те, що що виріб, який двічі пройшов спрощений контроль (як стандартний), є бракованим, можна відкинути як практично неможливу подію.

Задача 2.22. Імовірність того, що деякий товар знаходиться на складі, дорівнює $p \in (0;1)$, причому він може знаходитися в довільній із дев'яти секцій складу з однаковою імовірністю. Перевірка восьми секцій показала, що там він відсутній. Знайти ймовірність того, що товар знаходиться у дев'ятій секції складу.

○ Оскільки невідомо, чи є в наявності на складі товар, то можна зробити припущення: B_1 – товар є на складі, B_2 – він відсутній на складі. Подія A – товар відсутній у восьми секціях складу.

За умовою $P(B_1)=p$, тоді $P(B_2) = P(\overline{B_1}) = 1 - p$. Подія A відбулася, тому використаємо формулу Байєса.

Знайдемо умовні імовірності, врахувавши те, що товар може знаходитися (якщо він взагалі є на складі) в кожній із секцій з однаковою імовірністю. Зокрема, $P_{B_1}(A)$ – це імовірність того, що товар відсутній у восьми секціях, якщо він є в наявності на складі. Всіх наслідків випробування (рівноможливих і які утворюють повну групу) є 9, а сприятливих для настання події A – 1, тому $P_{B_1}(A) = 1/9$. Якщо ж товар відсутній на складі, тобто відбувається подія B_2, то достовірною є подія A (відсутність товару у восьми секціях). Отже, $P_{B_2}(A) = 1$.

Якщо товар є на складі і він відсутній у восьми секціях, то це означає, що він може бути тільки в дев'ятій секції, тобто $P_A(B_1)$ справді шукана імовірність, яку обчислюємо за формулою Байєса:

$$P_A(B_1) = \frac{P(B_1)\cdot P_{B_1}(A)}{P(B_1)\cdot P_{B_1}(A) + P(B_2)\cdot P_{B_2}(A)} = \frac{p\cdot \frac{1}{9}}{p\cdot \frac{1}{9} + (1-p)\cdot 1} =$$

$$= \frac{p}{9-8p}\ . \quad \bullet$$

Задача 2.23. В кінці потокової лінії по виготовленню приладів встановлено два автомати-контролери, які визначають: має прилад вищу категорію якості, чи ні. Статистично встановлено, що 20% продукції задовольняють вимогам вищої категорії якості, а контролери роблять помилкові висновки стосовно якості приладу відповідно у 5% і 1% випадків. Випадково один і той самий прилад був перевірений обома автоматами: перший визначив, що прилад не задовольняє вимогам вищої категорії, а другий, що задовольняє. Якому із висновків вірити?

○ Стосовно двічі перевіреного приладу можна висунути два припущення: або він справді задовольняє вимогам вищої категорії якості (подія B_1) або не задовольняє (подія B_2). За умовою задачі $P(B_1)=0{,}2$, $P(B_2)=0{,}8$.

Розглянемо спочатку ситуацію відносно висновку першого контролера. Подія A – визнання ним відсутності у перевіреного приладу якостей вищої категорії. Ця випадкова подія відбулася, тому використаємо формулу Байєса.

За умовою у 5% випадків перший контролер помиляється відносно якості приладу, а тому у 95% робить правильні висновки. В зв'язку з цим

$P_{B_1}(A) = 0{,}05$, $P_{B_2}(A) = 0{,}95$

і за формулою Байєса

$$P_A(B_2) = \frac{P(B_2)\cdot P_{B_2}(A)}{P(B_1)\cdot P_{B_1}(A) + P(B_2)\cdot P_{B_2}(A)} = \frac{0{,}8\cdot 0{,}95}{0{,}2\cdot 0{,}05 + 0{,}8\cdot 0{,}95} = 0{,}987013$$

Отже, імовірність того, що перший контролер зробив правильний висновок, дорівнює 0,987013.

Нехай A^* - визнання вищої категорії якості перевіреного приладу другим контролером. З використанням умови задачі отримаємо:

$P_{B_1}(A^*) = 0{,}99$, $P_{B_2}(A^*) = 0{,}01$,

$$P_{A^*}(B_1) = \frac{P(B_1)\cdot P_{B_1}(A^*)}{P(B_1)\cdot P_{B_1}(A^*) + P(B_2)\cdot P_{B_2}(A^*)} = \frac{0{,}2\cdot 0{,}99}{0{,}2\cdot 0{,}99 + 0{,}8\cdot 0{,}01} = 0{,}961165$$

Тобто, імовірність того, що другий контролер зробив правильний висновок, складає 0,961165.

Порівняння знайдених імовірностей дозволяє зробити висновок: потрібно вірити висновку першого автомата-контролера (точніше, більш імовірніше, що перший контролер зробив правильний висновок). ●

Задача 2.24. У корзині є 20 куль (12 білих і 8 чорних). Знайти імовірність того, що навмання взята з корзини куля виявиться чорною, якщо перед цим випадковим чином було взято: 1) дві кулі; 2) три кулі.

○ 1) В умові задачі відсутня інформація про те, які саме дві кулі (за кольором) були відібрані перед відбором третьої. Тому можна висунути такі припущення:

B_1 – обидві кулі білі;

B_2 – одна біла, друга чорна;

B_3 – обидві чорні.

Подія A – відібрана куля (після взяття двох) виявиться чорною.

За класичним означенням імовірності

$P(B_1) = C_{12}^2 / C_{20}^2 = 33/95$,

$P(B_2) = C_{12}^1 \cdot C_8^1 / C_{20}^2 = 48/95$,

$P(B_3) = C_8^2 / C_{20}^2 = 14/95$.

Обчислимо умовні імовірності:

$$P_{B_1}(A)=\frac{8}{18}=\frac{4}{9},\ P_{B_2}(A)=\frac{7}{18},\ P_{B_3}(A)=\frac{6}{18}=\frac{1}{3}.$$

За формулою повної імовірності

$$P(A)=P(B_1)\cdot P_{B_1}(A)+P(B_2)\cdot P_{B_2}(A)+P(B_3)\cdot P_{B_3}(A)=$$

$$=\frac{33}{95}\cdot\frac{4}{9}+\frac{48}{95}\cdot\frac{7}{18}+\frac{14}{95}\cdot\frac{1}{3}=0{,}4.$$

2) Аналогічно попередньому розглянемо гіпотези стосовно трьох відібраних куль:

B_1 – три кулі білі;

B_2 – дві кулі білі, одна чорна;

B_3 – одна куля біла, дві чорні;

B_4 – три кулі чорні.

Подія A – відібрана куля (після взяття трьох) виявиться чорною.

Обчислимо безумовні та умовні імовірності, а потім знову за формулою повної імовірності знайдемо шукану ймовірність:

$$P(B_1)=C_{12}^3\big/C_{20}^3=11/57,\ P(B_2)=C_{12}^2\cdot C_8^1\big/C_{20}^3=44/95,$$

$$P(B_3)=C_{12}^1\cdot C_8^2\big/C_{20}^3=28/95,\ P(B_4)=C_8^3\big/C_{20}^3=14/285,$$

$$P_{B_1}(A)=\frac{8}{17},\ P_{B_2}(A)=\frac{7}{17},\ P_{B_3}(A)=\frac{6}{17},\ P_{B_4}(A)=\frac{5}{17},$$

$$P(A)=P(B_1)\cdot P_{B_1}(A)+P(B_2)\cdot P_{B_2}(A)+P(B_3)\cdot P_{B_3}(A)+P(B_4)\cdot P_{B_4}(A)=$$

$$=\frac{11}{57}\cdot\frac{8}{17}+\frac{44}{95}\cdot\frac{7}{17}+\frac{28}{95}\cdot\frac{6}{17}+\frac{14}{285}\cdot\frac{5}{17}=0{,}4.$$ ●

ЗАДАЧІ ДЛЯ САМОСТІЙНОГО РОЗВ'ЯЗУВАННЯ

Завдання №2

1. В магазині продаються 20 телевізорів одного класу, 4 з яких вимагають тривалого додаткового регулювання. Яка імовірність того, що покупець придбає телевізор, якщо для вибору апарату без дефектів він здійснить не більше трьох спроб?
2. У папці 10 акцій I-го виду і 8 – II-го. Навмання беруть дві акції. Знайти імовірність того, що вони будуть одного виду.
3. Вищу кваліфікацію мають 3 аудитори із дев'яти і 4 програмісти із шести. У відрядження потрібно відправити групу із трьох аудиторів і двох програмістів. Кожний із п'ятнадцяти працівників фірми має однакові можливості поїхати у відрядження. Яка імовірність того,

що у складі групи виявиться хоча б один аудитор вищої кваліфікації і щонайменше один програміст вищої кваліфікації?

4. В трьох урах міститься відповідно: 10 куль (7 червоних і 3 білі), 8 (2 червоні і 6 білих), 6 (4 червоні і 2 білі). З кожної із них навмання береться по одній кулі. Знайти імовірність того, що вони матимуть однаковий колір.

5. Абонент забув останню цифру номера телефона і тому набирає її навмання. Знайти ймовірність того, що йому доведеться подзвонити не більше ніж до трьох абонентів.

6. Магазин отримав товар від трьох постачальників: 40% від першого постачальника, 25% - від другого, 35% - від третього. Знайти імовірність того, що три навмання відібрані одиниці товару виготовлені: 1) одним і тим же постачальником; 2) різними постачальниками.

7. Імовірність своєчасної сплати податків для першого підприємства дорівнює 0,8, для другого 0,7, а для третього – корінь квадратного рівняння $2p^2 - 3p + 1 = 0$. Знайти імовірність вчасної сплати податків не більш ніж одним підприємством.

8. У лотереї, присвяченій презентації нової продукції фірми, розігрується 1000 білетів, з яких виграшними є 4 речових вартістю 40, 60, 70 і 100 грн. і 5 грошових по 30 грн. кожний. Знайти імовірність того, що учасник лотереї, маючи три білети, виграє на суму, не меншу 40 грн.

9. Статистична імовірність попадання в мішень при кожному пострілі для I-го лучника дорівнює 0,6, а для II-го – 0,8. Обидва вони, починаючи з I-го, почергово стріляють по мішені, але виконують не більше двох пострілів. Знайти імовірність того, що в мішені виявиться: а) дві стріли; б) три.

10. Імовірність безвідмовної роботи блока протягом певного часу дорівнює 0,8. Цей блок входить у систему, решта елементів якої вважаються безвідмовними. Для збільшення надійності роботи системи встановлюються ще два блоки, які будуть знаходитися у резерві і вмикатимуться почергово у випадку виходу з ладу своїх "попередників". Знайти імовірність безвідмовної роботи системи протягом вказаного проміжку часу із врахуванням резервних блоків

11. Для вчасного збирання врожаю пшениці достатньо, щоб у полі працювало два комбайни. Знайти імовірність того, що пшениця буде

вчасно зібрана, якщо господарство має три комбайни, імовірності справної роботи яких відповідно рівні 0,4; 0,9; 0,8.

12. У скрині знаходиться 15 білих і 25 чорних куль, однакових за розмірами і на дотик. Знайти імовірність того, що з трьох навмання витягнутих куль хоча б одна буде білою.

13. Однотипні деталі, виготовлені трьома станками, скидаються на спільний конвейєр. Продукція І-го станка складає 40%, а ІІ-го і ІІІ-го – 50% і 10% відповідно. Із конвейєра навмання взято дві деталі. Яка імовірність того, що: а) одна із них виготовлена ІІ-м станком; б) обидві виготовлені одним станком?

14. Імовірність одного попадання в ціль при одному залпі з двох рушниць дорівнює 0,38. Знайти імовірність попадання в ціль при одному пострілі з першої гвинтівки, якщо відомо, що для другої ця імовірність дорівнює 0,7.

15. Серед 30 механічних годинників, які надійшли в ремонт, 6 вимагають загальної чистки механізму. Знайти імовірність того, що серед навмання взятих п’яти годинників принаймні два вимагатимуть загальної чистки механізму.

16. В пакеті є 30 акцій, серед яких 3 шукані. Навмання беруться 3 акції. Знайти імовірність того, що серед них виявиться хоча б дві шукані.

17. Робітник при складанні механізму встановлює дві однакові деталі. Бере він їх випадковим чином із дванадцяти штук, серед яких три деталі меншого розміру. Механізм не буде працювати, якщо обидві встановлені деталі мають менший розмір. Знайти імовірність того, що механізм буде працювати.

18. Три лучники випустили по одній стрілі у спільну мішень. Імовірності влучання для кожного із них відповідно рівні 0,8; 0,6; 0,7. Знайти імовірності того, що в мішені виявиться: а) дві стріли; б) хоча б дві стріли.

19. Групі студентів для проходження виробничої практики виділено 30 місць: 15 – у Хмельницькому, 8 – у Львові, 7 – у Луцьку. Ці місця розподіляються між студентами випадковим чином. Знайти імовірність того, що студент і студентка, які незабаром збираються одружитися, будуть направлені для проходження практики в одне і те ж місто.

20. В пачці 20 фотокарток, серед яких три шукані. Навмання відібрано

5 карток. Яка імовірність того, що серед них виявиться хоча б дві шукані.

21. Імовірність банкрутства для першої фірми – розв'язок рівняння $4p^2 - 3p = 0$, а для другої фірми ця імовірність на 20% більша. Знайти імовірність того, що хоча б одна із цих фірм збанкрутує.

22. Підприємство отримує сировину від трьох постачальників і не виконує контракт по виготовленню продукції, якщо хоча б один із постачальників зриває поставку сировини. Імовірності вчасної поставки сировини для постачальників відповідно рівні 0,97; 0,95; 0,99. Знайти імовірність виконання контракту підприємством-виробником.

23. Знайти імовірність невиграшу для однієї картки спортлото «5 із 36», якщо вона виявляється такою, коли число вгаданих чисел буде меншим трьох.

24. До контролера поступила партія однотипних виробів в кількості 20 шт. Серед них є 5 бракованих, але про це йому невідомо. Контролер навмання бере три вироби для перевірки. Якщо хоча б один із них виявиться бракованим, тоді вся партія бракується. Знайти імовірність того, що партія забракується.

25. Три спортсмени одночасно вистрілили з далекої відстані по повітряній кулі. Імовірності влучання для кожного із них відповідно рівні 0,6; 0,7; 0,5. Знайти імовірність знищення кулі.

26. Гральний кубик кидається доти, поки двічі підряд на верхній грані не випаде 5 очок. Знайти імовірність того, що дослід закінчиться до шостого кидання.

27. Студент знає 50 із 60 питань програми. Знайти імовірність того, що із трьох навмання витягнутих питань він знатиме: а) хоча б одне; б) тільки одне; в) не більше одного.

28. В аудиторії серед 15 комп'ютерів 12 справних. Знайти імовірність того, що з двох навмання вибраних комп'ютерів хоча б один виявиться несправним.

29. У скрині є 8 білих і 2 чорні кулі. Два гравці по черзі випадковим чином витягують кулі із скрині, повертаючи кожного разу взяту кулю в скриню і перемішуючи її вмістиме. Виграє той, хто першим дістане чорну кулю. Знайти імовірність виграшу для кожного гравця.

30. Імовірність покращення спортсменом особистого досягнення по стрибках у висоту дорівнює 0,1. Чому дорівнює імовірність того, що він покращить свій результат, якщо йому надана можливість зробити три спроби.

31. Для виготовлення деталі робітнику потрібно виконати чотири незалежні технологічні операції. Імовірність допустити брак при виготовленні кожної з них відповідно дорівнює 0,004; 0,005; 0,008; 0,001. Знайти імовірність того, що виготовлена робітником деталь виявиться бракованою.

31. Відомо, що випадкові події A та B незалежні, причому $P(A\overline{B}) = 0{,}52$, $P(\overline{A}B) = 0{,}37$, $P(A + B) = 0{,}89$. Знайти $P(A)$ + $P(B)$ і з'ясувати питання: події A та B сумісні чи ні.

32. Імовірності вчасної сплати податків для кожного із трьох підприємств відповідно рівні 0,4; 0,8; 0,6. Знайти імовірність вчасної сплати податків не більш, ніж двома підприємствами.

33. У контейнері є пряжа в мотках, серед якої 40% блакитної, решта – білої. Знайти імовірність того, що два навмання взяті мотки матимуть однаковий колір.

34. Імовірність промаху при полюванні на лисицю дорівнює 0,6 і зростає з кожним пострілом на 0,1. Знайти імовірність того, що після трьох пострілів лисиця все-таки втече.

35. Знайти імовірність виграшу для картки спортлото «6 із 40», якщо для цього потрібно закреслити не менше трьох виграшних номерів.

36. Двері відкриваються одним із 4-х ключів, які знаходяться у зв'язці. В темряві господар навмання вибирає ключ і, якщо двері не відкриваються, бере наступний. Знайти імовірність того, що двері будуть відкриті за три спроби.

37. Протипожежний пристрій складається із трьох незалежно працюючих сигналізаторів, які спрацьовують у випадку пожежі з імовірностями, що відповідно дорівнюють 0,95; 0,9; 0,98. Знайти імовірність того, що при пожежі спрацюють: а) тільки один сигналізатор; б) принаймні один; в) тільки два; г) хоча б два.

38. В конверті знаходиться 5 акцій, останні цифри номерів яких відповідно 1, 2, 3, 4, 5. Навмання витягують дві акції. Знайти імовірність того, що сума останніх цифр номерів витягнутих акцій буде не менша трьох.

39. У скрині є 4 червоних, 6 синіх і 5 зелених куль. Тричі підряд навмання витягується по одній кулі, не повертаючи в скриню. Знайти імовірність того, що всі вони виявляться: а) різних кольорів; б) одного кольору.

40. Імовірність виявлення захворювання туберкульозом при одній рентгеноскопії дорівнює 3/4. Знайти імовірність виявлення цього захворювання при трьох рентгеноскопіях.

41. У середньому з 9 вантажівок 6 мають повне комерційне завантаження. Знайти ймовірність того, що з 4 вантажівок хоча б три з повним комерційним завантаженням.

42. Обчислити ймовірність того,що дні народження 4 осіб припадають на одну дату.

Завдання №3

1. Виробник комп'ютерів отримує комплектуючі деталі від чотирьох постачальників, частки яких складають відповідно 15%, 25%, 50%, 10%. Деталі першого постачальника мають 0,8% браку, другого – 1%, третього – 1,5%, четвертого – 0,9%. 1) Яка імовірність того, що навмання відібрана деталь виявилася бракованою? 2) Випадковим чином взята деталь виявилася стандартною. Імовірніше всього яким постачальником вона виготовлена?

2. Два мисливці одночасно стріляють однаковими кулями у дика. В результаті дик був убитий однією кулею. Яким чином мисливці повинні поділити м'ясо вбитого дика, якщо відомо, що імовірність влучання для першого мисливця дорівнює 0,3, для другого – 0,6.

3. Два станки виготовляють однотипні деталі, які потрапляють на спільний конвеєр. В середньому з кожних 100 деталей першого станка одна нестандартна, а з кожної тисячі другого – 8 нестандартних. Продуктивність другого станка на 20% більша від першого. Знайти імовірність того, що навмання взята з конвеєра деталь виявиться стандартною.

4. Виготовлена продукція перевіряється двома автоматами-контролерами, які визначають, має прилад вищу категорію якості, чи ні. Статистично встановлено, що 25% продукції задовольняють вимогам вищої категорії якості, а контролери роблять помилкові висновки стосовно якості приладу відповідно у 1% і 4% випадків.

Випадковим чином один і той самий прилад був перевірений обома автоматами, при цьому перший визначив вищу категорію якості, а другий – ні. Якому із цих висновків вірити?

5. В офісі чотири ноутбуки, виготовлених різними фірмами: Acer, Asus, Lenovo та HP. Імовірність того, що кожний з них витримає подвійний гарантійний термін, дорівнює відповідно 0,7; 0,9; 0,85; 0,8. Знайти імовірність того, що навмання взятий ноутбук витримає подвійний гарантійний термін.

6. Магазин отримує продукцію від двох виробників у співвідношенні 2:3. Імовірність продажу виробів першого постачальника дорівнює 0,95, а другого – 0,85. Знайти імовірність того, що навмання вибраний виріб не буде реалізовано.

7. Із 16 баскетболістів чотири влучають у кошик із штрафного кидка з імовірністю 0,9, сім – з імовірністю 0,8, три – з імовірністю 0,7 і два – з імовірністю 0,6. 1). Яка імовірність того, що навмання відібраний спортсмен влучить у кошик із штрафного? 2). Довільно відібраний баскетболіст виконав один штрафний кидок і не влучив у кошик. До якої групи імовірніше всього він належить?

8. В першій шухляді є чотири стандартні і дві браковані деталі, в другій – п’ять стандартних і три браковані, третя – порожня. З першої шухляди навмання взято дві деталі, з другої – одну, і все це перекладають у третю. Знайти імовірність того, що навмання взята з третьої шухляди деталь виявиться стандартною.

9. При заповненні певного документу перший бухгалтер помиляється з імовірністю 0,05, а другий – 0,1. За певний час перший бухгалтер заповнив 80 таких документів, а другий – 120. Всі ці документи в порядку їх заповнення складалися в одну папку. Навмання витягнутий із цієї папки документ виявився з помилкою. Що більш імовірніше: помилку допустив перший чи другий бухгалтер?

10. Деталь може надійти для обробки на перший автомат з імовірністю 0,3, на другий – з імовірністю 0,2, а на третій – з імовірністю 0,5. При обробці на першому верстаті імовірність браку складає 0,01, на другому – 0,03, а на третьому – 0,08. Вибрана навмання деталь виявилася бракованою. Яка імовірність того, що її виготовлено на другому автоматі?

11. Страхова компанія розподіляє застрахованих по чотирьом класам ризику: I клас – малий ризик, II клас – середній, III клас – великий

ризик, IV – дуже великий. Серед клієнтів компаній 40% – I-го класу, 30% – II-го, 25% – III-го і 5% – IV-го. Імовірність необхідності виплати страхового відшкодування для кожного із цих класів ризику становить відповідно 0,009; 0,02; 0,04; 0,12. Знайти імовірність того, що: а) застрахований отримає страхове відшкодування за період страхування; б) клієнт, який отримав страхове відшкодування, належить до групи малого ризику.

12. Дві стріли залишилися в мішені після пострілу трьох лучників по ній. Імовірності влучання для кожного із лучників відповідно дорівнюють 0,5; 0,6; 0,4. Знайти імовірність того, що у мішені була стріла: а) третього лучника; б) першого і третього лучників.

13. На гуртову базу надходять телевізори від трьох постачальників у співвідношенні a:b:c – 2:5:3. Статистичні дані показали, що телевізори від I-го постачальника не вимагають ремонту протягом гарантійного терміну у 97%, а для II-го та III-го – у 92% та 95% відповідно. 1) Знайти імовірність того, що навмання відібраний телевізор не вимагатиме ремонту протягом гарантійного терміну. 2) Проданий телевізор зламався до завершення гарантійного терміну. Від якого постачальника імовірніше всього надійшов цей телевізор?

14. В магазині три морозильники, в яких закінчується морозиво. У першому – 9 пломбірів і 3 шоколадних, у другому – 4 пломбіри і 5 шоколадних, у третьому – 5 пломбірів і 8 шоколадних. Навмання вибирають морозильник і виймають з нього морозиво. Визначити ймовірність того, що воно шоколадне.

15. На складі майстерні побутової техніки знаходяться три комплекти однотипних деталей: в першому – 100 деталей, з яких дві браковані, в другому – 200, відсоток браку складає 2; в третьому – 1500, всі стандартні. Знайти імовірність того, що навмання взята деталь із випадково вибраного комплекту виявиться стандартною.

16. Статистично встановлено, що 90% продукції підприємства є стандартною. Використання спрощеної системи контролю якості показано, що вона визнає виріб стандартним з імовірністю 0,95, якщо цей виріб справді стандартний і з імовірністю 0,06, якщо він бракований. 1) Знайти імовірність того, що навмання відібраний виріб буде визнано стандартним згідно цієї спрощеної системи контролю. 2) Навмання взятий виріб пройшов за спрощеною системою. Яка імовірність того, що цей виріб стандартний?

17. Два автомати штампують однорідні деталі, які потрапляють на спільний конвеєр. Продуктивність першого автомату втричі більша, ніж продуктивність другого. Відсоток браку для кожного із них відповідно дорівнює 0,4 та 0,5. Яка імовірність того, що навмання взята з конвеєра деталь буде стандартною?

18. Клапани, виготовлені цехом заводу, перевіряють три контролери. З кожної сотні клапанів, що поступають на перевірку, 20 потрапляють до першого контролера, 50 – до другого і 30 – до третього. Імовірність того, що бракована деталь не буде виявлена першим контролером, дорівнює 0,01, другим – 0,09 і третім – 0,02. Під час контрольної перевірки незабракованих контролерами клапанів один виявився бракованим. Яка імовірність того, що цей клапан перевіряв другий контролер?

19. Для формування інститутської команди з I-го курсу виділено 5 студентів, з II-го – 7, з III-го – 8, з IV-го – 6. Імовірність того, що будь-який студент кожного з курсів буде включений до складу збірної інституту, відповідно дорівнює 0,6; 0,4; 0,8; 0,45. Навмання відібраний учасник змагань потрапив до складу збірної. На якому курсі імовірніше всього він навчається?

20. Три заводи виготовляють однакові вироби, причому перший завод випускає 50%, другий – 20%, третій – 30% всієї продукції. Відсотки браку для кожного із них складають відповідно 1, 6, 3. Навмання відібраний виріб виявляється бракованим. Знайти імовірність того, що він був виготовлений на другому заводі.

21. На підприємстві виготовляються однотипні вироби на трьох поточних лініях. На першій лінії виготовляється 20% виробів від усього обсягу їх виробництва, на другій – 30%, на третій – 50%. Кожна із ліній характеризується відповідно такими відсотками стандартних виробів: 97%, 98% і 95%. Знайти імовірність того, що навмання взятий виріб, виготовлений на підприємстві, виявиться бракованим, а також імовірності того, що цей бракований виріб виготовлений на: а) першій лінії; б) другій; в) третій.

22. Імовірність того, що деякий товар знаходиться на складі, дорівнює р, причому він може знаходитися в довільній із восьми секцій складу з однаковою імовірністю. Перевірка семи секцій показала, що там він відсутній. Знайти імовірність того, що товар знаходиться у восьмій секції складу.

23. В двох контейнерах є по 20 деталей, причому в першому – 5 бракованих, а в другому – 3 браковані деталі. З першого контейнера навмання береться одна деталь і перекладається в другий. Знайти імовірність того, що навмання взята після цього з другого контейнера деталь виявиться стандартною.

24. У товарному поїзді 50 вагонів, завантажених вугіллям двох сортів: 25 вагонів містять 70% вугілля першого сорту, 15 вагонів – 60% і 10 вагонів – 85% вугілля другого сорту. Випадково взятий для аналізу шматок вугілля виявився другого сорту. Знайти імовірність того, що він взятий із вагону другої групи.

25. При збиранні телевізорів використовуються мікросхеми двох постачальників, відсотковий склад яких становить відповідно 70% та 30%. Бракована продукція складає для кожного постачальника відповідно 2% та 3%. Знайти імовірність того, що взята навмання мікросхема виявиться стандартною.

26. У першому комплекті міститься 20 деталей, 6 з яких нестандартні; в другому – 10, 3 з яких нестандартні. З кожного комплекту навмання виймають по одній деталі, а потім із цих двох деталей навмання вибирають одну. Знайти імовірність того, що ця деталь виявиться стандартною.

27. У першій скрині є 4 білих і 6 чорних куль, у другій – 7 білих і 3 чорні куль. Із першої скрині навмання витягнута куля перекладається у другу, і вмістиме її перемішується. Знайти імовірність того, що взята після цього із другої скрині куля виявиться білою.

28. Виріб перевіряється на стандартність одним із товарознавців. Імовірність того, що виріб попаде до першого товарознавця, дорівнює 0,65, а до другого – 0,35. Імовірність того, що стандартний виріб буде підтверджений стандартним першим товарознавцем, дорівнює 0,9, а другим – 0,98. Стандартний виріб при перевірці було підтверджено стандартним. Знайти імовірність того, що цей виріб перевірив другий товарознавець.

29. На ткацьку фабрику надходить пряжа, виготовлена двома цехами прядильної фабрики, причому 30% пряжі – це продукція цеху № 1, а решта – цеху № 2. Продукція цеху № 1 містить 80% пряжі вищої якості, а цеху № 2 – 60%. Знайти імовірність того, що навмання взята шпуля матиме пряжу вищої якості.

30. Два із трьох незалежно працюючих елементів обчислювального

пристрою вийшли з ладу. Знайти імовірність того, що з ладу вийшли перший і другий елементи, якщо імовірності виходу з ладу для кожного з них відповідно рівні 0,2; 0,4; 0,1.

31. На конвеєр надходять деталі, які виготовляються двома автоматами. Імовірність одержання нестандартної деталі на першому автоматі дорівнює 0,05, на другому – 0,07. Продуктивність другого автомата на 60% вища, ніж першого. Знайти імовірність того, що навмання взята з конвеєра деталь виявиться нестандартною.

32. Монітор до комп'ютера може належати одній із чотирьох партій з імовірностями 0,4; 0,1; 0,2 і 0,3 відповідно. Імовірність того, що монітор відпрацює подвійний гарантійний термін, дорівнює відповідно для кожної партії 0,7; 0,8; 0,6; 0,9. Знайти імовірність того, що навмання вибраний монітор буде працювати подвійний гарантійний термін.

33. У першій корзині є 80 куль, з яких 30 червоного кольору, в другій 120, 60% яких червоного кольору. Навмання взята куля із навмання вибраної корзини виявилась червоною. Яка імовірність того, що вона була взята із другої корзини?

34. У першому контейнері є 30 деталей, з яких 4 браковані, у другому відповідно 20 і 3. Навмання взята деталь із випадковим чином вибраного контейнера виявилась стандартною. Яка імовірність того, що деталь була взята із першого контейнера?

35. Імовірність того, що двокамерний холодильник «ВЕКО» не зіпсується протягом гарантійного терміну, дорівнює 0,8, а для однокамерного ця імовірність на 10% більша. Знайти імовірність того, що навмання куплений холодильник із шести двокамерних і десяти однокамерних не зіпсується протягом гарантійного терміну.

36. На фінансовому ринку продаються акції чотирьох фірм. Їх кількість відносно загальної кількості всіх чотирьох становить відповідно 25, 30, 15 і 30 відсотків. Але серед них є фальшиві і відсотковий склад таких відповідно рівний 10, 4, 1 і 3. Знайти імовірність того, що навмання придбана акція є фальшивою.

37. Два студенти незалежно один від другого здійснили постріл по спільній мішені. Імовірність влучання в мішень для першого студента дорівнює 0,8, а для другого – 0,6. Після залпу в мішені виявлена одна пробоїна. Знайти імовірність того, що у мішень влучив другий студент.

38. Три верстати-автомати штампують однотипні деталі, що потрапляють на спільний конвейєр. Продуктивності автоматів визначаються відношенням 2 : 1 : 3. Відсотки браку для кожного автомату дорівнюють відповідно 2; 0,5; 4. 1). Яка імовірність того, що навмання взята з конвейєра деталь виявиться бракованою? 2). Навмання взята деталь виявилась бракованою. Знайти імовірність того, що вона була виготовлена на 2-му верстаті.

39. У скрині знаходиться 13 куль, з яких п'ять – білі. Знайти імовірність того, що навмання витягнута із скрині куля виявиться білою, якщо перед цим навмання було взято: а) дві кулі; б) три кулі.

40. У піраміді знаходиться 20 гвинтівок, 6 з яких обладнані оптичним прицілом. Імовірність влучення із гвинтівки з оптичним прицілом дорівнює 0,9, без оптичного прицілу – 0,6 (для певного стрільця). Цей стрілець із навмання взятої гвинтівки виконав постріл і влучив у ціль. Що імовірніше: стрілець стріляв із гвинтівки з оптичним прицілом чи із гвинтівки без оптичного прицілу?

41. Екіпажу для безпечного проходження грозового фронту з однаковою ймовірністю може бути задано три напрями: ліворуч, праворуч або над центром грозової активності. Ймовірність успішного проходження літаком грозового фронту ліворуч дорівнює 0,8; праворуч — 0,9; над центром — 0,5. Літак благополучно перетнув грозовий фронт. Знайти ймовірність того, що він обходив фронт над його центром.

42. Вершкове масло фасується на двох технологічних лініях молокозаводу. Ймовірність виходу кондиційної продукції з першої лінії дорівнює 0,88, а з другої — 0,95. Навмання взятий пакет масла виявився кондиційним. З якої лінії найімовірніше фасовано цей пакет?

43. Уздовж траси з бензоколонкою проїжджає вдвічі більше вантажних автомашин, аніж легкових. Імовірність того, що заправлятися буде вантажівка, дорівнює 0,1, а для легкової автомашини вона становить 0,2. На заправку під'їхала машина. Знайти ймовірність того, що вона легкова.

44. На заводі «Росинка» встановлено вітчизняний та імпортний автомати розливу води, продуктивність яких відноситься як 4 : 6. Пляшки з водою надходять на загальний конвеєр. Вітчизняний автомат дає в середньому 5 % браку, а імпортний — 2%. Знайти ймовірність того, що навмання узята з конвеєра пляшка води виявиться бракованою.

45. У рекламному агентстві працює три групи дизайнерів: перша обслуговує 35 фірм, друга — 25, третя — 40. Протягом одного місяця кошти, витрачені на рекламу дизайнерами першої групи, повертаються до 50 % фірм, другої — до 55 %, третьої -— до 35 %. Яка ймовірність того, що фірма, котра окупила впродовж місяця витрачені на рекламу кошти, обслуговувалась третьою групою дизайнерів?

46. Прилад складається із двох незалежно працюючих елементів, кожний з яких необхідний для роботи приладу в цілому. Імовірність безвідмовної роботи протягом часу Т для першого елементу дорівнює 0,9, а для другого – 0,7. За вказаний проміжок часу прилад вийшов з ладу через поломку одного із елементів. Знайти імовірність того, що з ладу вийшов перший елемент.

§ 3. ПОВТОРНІ НЕЗАЛЕЖНІ ВИПРОБУВАННЯ

Формула Бернуллі. Найімовірніше число появи події. Локальна формула Лапласа. Формула Пуассона. Інтегральна формула Лапласа. Імовірність відхилення відносної частоти події від її постійної імовірності. Алгоритм розв'язування задач для повторних незалежних випробувань.

КОРОТКІ ТЕОРЕТИЧНІ ВІДОМОСТІ

На практиці часто зустрічаються випадки, коли проводиться не одне випробування, а декілька (можливо, дуже велике число). Такі випробування називаються **повторними**, а їх сукупність – **схемою повторних випробувань**, або **схемою Бернуллі**. В кожному із таких випробувань може відбутися одна і та ж випадкова подія A. Якщо $P(A)$ залишається незмінною для **кожного** випробування, то такі випробування будемо називати **незалежними** (**відносно події** A).

Формула Бернуллі

Теорема. Імовірність того, що в n повторних незалежних випробуваннях випадкова подія A відбудеться рівно m разів, знаходиться за формулою Бернуллі:

$$P_n(m) = C_n^m p^m q^{n-m}, \tag{3.1}$$

де C_n^m – число комбінацій, визначене формулою (1.4) або (1.4*),

p – імовірність появи події A в **одному** випробуванні ($p = P(A)$),

q – імовірність **непояви** події A в одному випробуванні ($q = P(\overline{A})$).

Найімовірніше число появи події

В n повторних незалежних випробуваннях подія A може відбутися число разів

$$0,\ 1,\ \ldots,\ m_0-1,\ m_0,\ m_0+1,\ \ldots,\ n \tag{3.2}$$

з відповідними імовірностями (які можна знайти за формулою Бернуллі):

$$P_n(0),\ P_n(1),\ \ldots,\ P_n(m_0-1),\ P_n(m_0),\ P_n(m_0+1),\ \ldots,\ P_n(n). \tag{3.3}$$

Означення. Число m_0 в послідовності (3.2) називається найімовірнішим числом появи події в n незалежних випробуваннях (або модою), якщо йому відповідає найбільша імовірність в послідовності (3.3).

Найімовірніше число можна знайти з такої подвійної нерівності:

$$np - q \le m_0 \le np + p. \tag{3.4}$$

Оскільки різниця між правою і лівою частинами цієї нерівності дорівнює 1, то можна зробити висновок, що максимальне число найімовірніших подій рівне двом.

Локальна формула Лапласа

Користуватися формулою Бернуллі для великих значень n досить важко. Наприклад, при $p = 0{,}2$, $q = 0{,}8$ $P_{50}(30) = C_{50}^{30}(0{,}2)^{30}(0{,}8)^{20}$, де $C_{50}^{30} = 4712921 \cdot 10^7$.

Одну із реалізацій наближеного знаходження правої частини формули (3.1) дає

Локальна теорема Лапласа (Муавра–Лапласа).

Якщо імовірність *p* появи випадкової події *A* в кожному із *n* повторних випробувань залишається незмінною (причому $0 < p < 1$), а число випробувань достатньо велике, то імовірність того, що в *n* випробуваннях подія відбудеться *m* разів, знаходиться за наближеною формулою

$$P_n(m) \approx \frac{1}{\sqrt{npq}}\varphi(x), \tag{3.5}$$

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

Зауваження. Формула (3.5) називається **локальною формулою Лапласа**. Її точність зростає при збільшенні n.

Функція Гаусса протабульована і її значення наведені в табл. 1 додатків. Для правильного користування цією таблицею слід враховувати такі властивості функції Гаусса: 1) $\varphi(x)$ визначена для всіх $x \in R$; 2) $\varphi(x)$ – парна функція $(\varphi(-x) = \varphi(x))$; 3) для додатних x $\varphi(x)$ дуже швидко прямує до 0 при збільшенні x, зокрема $\varphi(3{,}99) = 0{,}0001$. Згідно із цими властивостями в таблиці наведені значення функції Гаусса для x з проміжку [0; 5].

Практично можна вважати, що локальна формула Лапласа дає добре наближення, якщо $npq \geq 9$. Якщо ж вимоги до точності значення вищі, то слід вимагати виконання нерівності $npq \geq 25$.

Формула Пуассона

Нерівність $npq \geq 9$ навіть для великих n може не виконуватися (а отже, похибка при використанні локальної формули Лапласа буде дуже великою) у випадку рідкісних (малоімовірних) подій A, тобто таких подій, для яких p значно менше 0,1 (p << 0,1). В таких випадках слід користуватися іншим наближенням правої частини формули Бернуллі. Одне із них дається таким твердженням.

Теорема Пуассона. Якщо в кожному із *n* повторних випробувань імовірність *p* появи події *A* стала і мала (*p* << 0,1), а число випробувань *n* досить велике, то імовірність того, що подія *A* настане в цих випробуваннях рівно *m* разів, знаходиться за формулою

$$P_n(m) \approx \frac{\lambda^m e^{-\lambda}}{m!}, \tag{3.6}$$

де $\lambda = np$.

Зауваження. Похибка в наближеній рівності (3.6), яка називається **формулою Пуассона**, тим менша, чим більше число випробувань n.

Значення функції $P(m,\lambda) = \dfrac{\lambda^m e^{-\lambda}}{m!}$ двох змінних m та λ для деяких m та λ наведені в табл. 2 додатків.

Відмітимо, що формула Пуассона використовується також до числа непояви події A, якщо q << 0,1, а $nq < 9$.

Інтегральна формула Лапласа

Інтегральна теорема Лапласа (Муавра–Лапласа). Якщо імовірність *p* появи події *A* в кожному із *n* повторних випробувань є сталою (0 < *p* < 1), а число випробувань досить велике, то імовірність того, що подія *A* в цих випробуваннях відбудеться не менше m_1 разів і не більше m_2 разів, знаходиться за такою наближеною рівністю (інтегральною формулою Лапласа):

$$P_n(m_1 \le m \le m_2) \approx \Phi(x_2) - \Phi(x_1), \tag{3.7}$$

де $\Phi(x) = \dfrac{1}{\sqrt{2\pi}}\int\limits_0^x e^{-t^2/2}dt$ – функція Лапласа,

$$x_1 = \frac{m_1 - np}{\sqrt{npq}}, x_2 = \frac{m_2 - np}{\sqrt{npq}}.$$

Функція Лапласа протабульована і її значення наведені в табл. 3 додатків. При користуванні цією таблицею потрібно враховувати такі властивості функції Лапласа:

1) $\Phi(x)$ визначена для всіх $x \in R$;

2) $\Phi(x)$ непарна функція ($\Phi(-x) = -\Phi(x)$);

3) $\Phi(x)$ монотонно зростає для всіх $x \in R$, при цьому $y = -0,5$ – лівостороння асимптота, а $y = 0,5$ – правостороння;

4) швидкість зростання $\Phi(x)$ на проміжку [0; 5] дуже висока (зокрема $\Phi(5) = 0,499997$), тому для всіх $x > 5$ з мізерною похибкою $\Phi(x) \approx 0,5$.

Точність інтегральної формули Лапласа тим більша, чим більше число випробувань n. Як і у випадку локальної, **інтегральна формула дає добрі наближення, якщо $npq \geq 9$**. Якщо ж вимоги до точності значно вищі, то потрібно вимагати виконання нерівності $npq \geq 25$.

Імовірність відхилення відносної частоти події від її постійної імовірності

Теорема. Якщо імовірність p появи випадкової події A в кожному з n повторних випробувань стала, а число випробувань досить велике, то імовірність того, що відхилення відносної частоти m/n події A від її імовірності p по абсолютній величині не перевищить заданого числа ε > 0, знаходиться за формулою

$$P\left(\left|m/n - p\right| \leq \varepsilon\right) \approx 2\Phi\left(\varepsilon\sqrt{n/pq}\right). \qquad (3.8)$$

Алгоритм розв'язування задач для повторних незалежних випробувань

В цьому параграфі, як і для попередніх параграфів, залишається актуальним питання вибору тієї чи іншої формули при розв'язуванні конкретних задач. Це зумовлено, по-перше, тим, що у всіх трьох формулах (Бернуллі, локальній формулі Лапласа та Пуассона) ліві частини однакові. З другого боку, при знаходженні імовірності $P_n(m_1 \leq m \leq m_2)$ зовсім не обов'язково (а деколи й помилково) використовувати інтегральну формулу Лапласа.

1. Обчислення $P_n(m)$.

а) Якщо n мале ($n \leq 15$), то використовується формула Бернуллі для будь-яких значень p та q.

б) Якщо n велике, а p та q немалі, тобто при виконанні нерівності $npq \geq 9$, тоді використовується локальна формула Лапласа.

в) Якщо ж n велике, а p дуже мале (значно менше 0,1) і $\lambda = np < 9$, то застосовується формула Пуассона. При великому n, дуже малому q ($q << 0,1$) і при виконанні нерівності $\lambda^* = nq < 9$ слід перейти до числа невиконання події A.

2. Знаходження $P_n(m_1 \le m \le m_2)$.

а) Якщо n мале ($n \le 15$), тоді потрібно використати спочатку теорему додавання імовірностей, а потім формулу Бернуллі.

б) Для великих n і немалих p та q, тобто при виконанні нерівності $npq \ge 9$ використовується інтегральна формула Лапласа.

в) Для великих n і малих p використовується або теорема додавання імовірностей з наступним застосуванням формули Пуассона, або здійснюється перехід до протилежної події з наступним використанням теореми додавання імовірностей і формули Пуассона. При виборі однієї із альтернатив слід керуватися мінімізацією числа доданків в теоремі додавання імовірностей. Якщо n велике, а q мале і $\lambda^* = nq < 9$, тоді потрібно перейти до числа невиконання події A, а потім виконати рекомендації початку цього підпункту.

Зауваження. Якщо в n повторних незалежних випробуваннях потрібно знайти імовірності $P(m \le k)$ та $P(m \ge k)$, де k ціле число, що не перевищує n, тоді потрібно скористатися рівностями $P(m \le k) = P_n(0 \le m \le k)$, $P(m \ge k) = P_n(k \le m \le n)$ і перейти до п. 2 алгоритму.

Розглянемо реалізацію наведених вище рекомендацій.

РОЗВ'ЯЗУВАННЯ ТИПОВИХ ЗАДАЧ

Задача 3.1. Компанія володіє мережею дилерів на біржі. Імовірність того, що будь-який дилер буде грати вдало, становить 0,7. 1) Знайти імовірність того, що з п'яти дилерів будуть у збитках: а) два; б) хоча б два (вважається, що дії ділерів на біржі є незалежними). 2) Знайти найімовірніше число дилерів, які будуть грати вдало, а також імовірність такої кількості.

○ Випробування – гра дилера. Оскільки ділерів є 5, то $n = 5$. 1) Подія A – збиткова гра дилера. За умовою $P(\overline{A}) = q = 0{,}7$, тоді $p = 1 - q = 0{,}3$. Для випадку а) число появи події A $m = 2$, і потрібно знайти $P_5(2)$. Так як $n = 5$ – мале, то використаємо формулу Бернуллі:

$$P_5(2) = C_5^2 (0{,}3)^2 \cdot (0{,}7)^3 = 10 \cdot 0{,}09 \cdot 0{,}343 = 0{,}3087.$$

б) $P_5(m \ge 2) = P_5(2 \le m \le 5)$ – шукана імовірність. Хоча вона візуально нагадує ліву частину інтегральної формули Лапласа, останню використовувати не можна, бо $npq = 5 \cdot 0{,}3 \cdot 0{,}7 = 1{,}05 << 9$. Випадкова подія $(2 \le m \le 5)$ може відбутися тоді, коли або $(m = 2)$, або $(m = 3)$, або $(m = 4)$, або $(m = 5)$, тобто

$(2 \le m \le 5) = (m = 2) + (m = 3) + (m = 4) + (m = 5)$. Випадкові події-доданки справа є попарно несумісними, тому згідно з теоремою додавання імовірностей

$$P_5(2 \le m \le 5) = P(m = 2) + P(m = 3) + P(m = 4) + P(m = 5) =$$
$$= P_5(2) + P_5(3) + P_5(4) + P_5(5),$$

де в останній рівності було враховано, зокрема, те, що випадкова подія $(m = 2)$ – в п'яти випробуваннях подія A відбудеться рівно два рази, тобто $P(m = 2) = P_5(2)$. Використавши формулу Бернуллі, отримаємо:

$$P_5(3) = C_5^3(0{,}3)^3 \cdot (0{,}7)^2 = 0{,}1323;$$
$$P_5(4) = C_5^4(0{,}3)^4 \cdot 0{,}7 = 0{,}02835;$$
$$P_5(5) = C_5^5(0{,}3)^5 \cdot 0{,}7^0 = 0{,}00243.$$

Остаточно

$$P_5(2 \le m \le 5) = 0{,}3087 + 0{,}1323 + 0{,}028535 + 0{,}00243 = 0{,}47178.$$

Другий метод. Події $(m \ge 2)$ та $(m < 2)$ протилежні, тому

$$P_5(m \ge 2) = 1 - P_5(m < 2) = 1 - [P_5(0) + P_5(1)] =$$
$$= 1 - [C_5^0(0{,}3)^0(0{,}7)^5 + C_5^1(0{,}3)^1(0{,}7)^4] =$$
$$= 1 - (0{,}16807 + 0{,}36015) = 0{,}47178.$$

Висновок: другий метод значно швидше веде до мети, його ефективність ще більш відчутна при збільшенні числа доданків.

2) Подія A – вдала гра дилера. За умовою задачі $n = 5$, $p = 0{,}7$, $q = 0{,}3$. Найімовірніше число m_0 дилерів, які будуть грати вдало, знайдемо за подвійною нерівністю (3.4)

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

Підставивши значення в ліву та праву частини, знайдемо $3{,}2 \le m_0 \le 4{,}2$, звідки з врахуванням того, що m_0 – ціле число, остаточно отримаємо: $m_0 = 4$. Нарешті,

$$P_5(m_0) = P_5(4) = C_5^4(0{,}7)^4 \cdot 0{,}3 = 0{,}36015. \bullet$$

Для деяких задач імовірність появи події А в одному випробуванні знаходиться порівняно складніше.

Задача 3.2. Два станки з програмним управлінням виготовляють однотипні деталі, які надходять на спільний конвеєр. Їх продуктивності відносяться як 2 : 3, причому перший виготовляє 35% деталей вищої якості, якими комплектуються вироби на експорт, другий – 10%. Знайти імовірність того, що з 400 навмання відібраних з конвеєра деталей вищої якості виявилось: а) 75 деталей; б) хоча б 80; в) не більше 75.

○ Випробування – відбір деталі, за умовою $n = 400$. A – відібрана деталь має вищу якість. Знайдемо $P(A) = p$. Відібрана деталь (вищої якості чи ні) може бути виготовлена або першим станком (подія B_1), або другим (подія B_2). Ці гіпотези утворюють повну групу, а подія A може відбутися тільки після появи однієї із них. Тому $P(A)$ можна знайти за формулою повної імовірності

$$P(A) = P(B_1)P_{B_1}(A) + P(B_2)P_{B_2}(A).$$

Якщо за деякий проміжок часу перший станок виготовить $2a$ деталей, то другий – $3a$ деталей. За класичними означенням імовірності

$$P(B_1) = \frac{2a}{5a} = 0{,}4,\ \ P(B_2) = \frac{3a}{5a} = 0{,}6,$$

$$P_{B_1}(A) = \frac{35}{100} = 0{,}35,\ \ P_{B_2}(A) = \frac{10}{100} = 0{,}1.$$

Підставивши знайдені імовірності у формулу повної імовірності, отримаємо

$P(A) = 0{,}4 \cdot 0{,}35 + 0{,}6 \cdot 0{,}1 = 0{,}2$. Отже, $p = 0{,}2$, $q = 1 - p = 0{,}8$.

а) Потрібно знайти $P_{400}(75)$. Оскільки $n = 400$ – велике, p та q немалі і виконується нерівність $npq = 400 \cdot 0{,}2 \cdot 0{,}8 = 64 > 9$, то потрібно вибрати локальну формулу Лапласа, яка в даному випадку дасть високу точність наближення. Згідно із (3.5)

$$x = \frac{m - np}{\sqrt{npq}} = \frac{75 - 400 \cdot 0{,}2}{\sqrt{400 \cdot 0{,}2 \cdot 0{,}8}} = -0{,}63,$$

$\varphi(-0{,}63) = \varphi(0{,}63) = 0{,}3271$ (за табл. 1 додатків),

$$P_{400}(75) \approx \frac{1}{\sqrt{400 \cdot 0{,}2 \cdot 0{,}8}}\varphi(-0{,}63) = \frac{0{,}3271}{8} = 0{,}0410.$$

б) Для знаходження імовірності $P_{400}(m \geq 80) = P_{400}(80 \leq m \leq 400)$ (див. зауваження до алгоритму) використаємо інтегральну формулу Лапласа, оскільки $npq = 64 > 9$. У відповідності із (3.7)

$$x_1 = \frac{m_1 - np}{\sqrt{npq}} = \frac{80 - 400 \cdot 0{,}2}{\sqrt{400 \cdot 0{,}2 \cdot 0{,}8}} = 0;\ \ x_2 = \frac{m_2 - np}{\sqrt{npq}} = \frac{400 - 400 \cdot 0{,}2}{\sqrt{400 \cdot 0{,}2 \cdot 0{,}8}} = 40,$$

$$P_{400}(80 \leq m \leq 400) \approx \Phi(40) - \Phi(0) = 0{,}5 - 0 = 0{,}5.$$

При знаходженні $\Phi(40)$ враховувалося те, що $\Phi(x) = 0{,}5$ для $x > 5$, а $\Phi(0)$ знаходилося за табл. 3 додатків.

в) Імовірність $P_{400}(m \leq 75) = P_{400}(0 \leq m \leq 75)$ знову обчислюємо за інтегральною формулою Лапласа:

$$x_1 = \frac{0 - 400 \cdot 0{,}2}{8} = -10; \; x_2 = \frac{75 - 400 \cdot 0{,}2}{8} = -0{,}63,$$

$$P_{400}(0 \le m \le 75) \approx \Phi(-0{,}63) - \Phi(-10) = \Phi(10) - \Phi(0{,}63) =$$

$$= 0{,}5 - 0{,}2357 = 0{,}2643.$$

В останніх рівностях використана непарність $\Phi(x)$. ●

Зауваження. При знаходженні імовірностей в а) та в) можна було досягти вищої точності, оскільки точні значення x та x_2 рівні –0,625, а $\varphi(0{,}62) = 0{,}3292$, $\Phi(0{,}62) = 0{,}2324$. Для цього потрібно здійснити лінійну інтерполяцію:

$$\varphi(0{,}625) = (\varphi(0{,}62) + \varphi(0{,}63))/2 = (0{,}3292 + 0{,}0{,}3271)/2 = 0{,}32815,$$

$$\Phi(0{,}625) = (\Phi(0{,}62) + \Phi(0{,}63)/2 = (0{,}2324 + 0{,}2357)/2 = 0{,}23405.$$

Проте доцільність такого підходу визначається вимогами до точності відповіді, поставленими конкретною задачею.

Наскільки суттєво враховувати **всі** умови використання тієї чи іншої формули, ілюструє розв'язування наступної задачі.

Задача 3.3. У приватній страховій компанії було застраховано 15 000 громадян приблизно одного віку і однієї соціальної групи. Імовірність смерті застрахованого протягом року в середньому складає 0,004. Кожен застрахований вносить на 1 січня 64 грн. страхових, а на випадок смерті його родичі дістануть від страхової компанії 10 000 грн. Знайти імовірність того, що компанія за підсумками року:

1) понесе збитки; 2) дістане прибуток, який не менший від 60 000 грн.

○ Випробування – страхування громадянина. A – смерть застрахованого протягом року. За умовою задачі $n = 15000$, $P(A) = p = 0{,}004$.

1) Від страхування громадян страхова компанія отримала $15\,000 \cdot 64 = 960\,000$ грн. і понесе збитки, якщо не менше всієї цієї суми буде виплачено родинам потерпілих протягом року (навіть якщо **тільки** ця сума буде виплачена, то діяльність компанії буде збитковою по цьому виду страхування з урахуванням поточних витрат протягом року). Ця випадкова (складна) подія відбудеться, якщо $m \ge 96$, де m – число відбуття події A, $96 = 960\,000 : 10\,000$. Тобто, потрібно знайти $P_{15\,000}(m \ge 96) = P_{15\,000}(96 \le m \le 15\,000)$. Подія A – рідкісна (малоімовірна), бо $p = 0{,}004 \le 0{,}1$. Проте використовувати формулу Пуассона (після застосування теореми додавання імовірностей) не можна, оскільки

$\lambda = np = 15\,000 \cdot 0{,}004 = 60$ (значно більше 9). Разом з тим, $npq = 59{,}76 > 25$. Тому можна використати інтегральну формулу Лапласа, яка дасть дуже високу точність. Знаходимо:

$$x_1 = \frac{m_1 - np}{\sqrt{npq}} = \frac{96 - 15\,000 \cdot 0{,}004}{\sqrt{59{,}76}} = 4{,}66; \quad x_2 = \frac{15\,000 - 60}{\sqrt{59{,}76}} = 1932{,}73,$$

$$P_{15\,000}(96 \le m \le 15\,000) \approx \Phi(1932{,}73) - \Phi(4{,}66) =$$
$$= 0{,}5 - 0{,}499997 = 0{,}000003.$$

2) Припустимо спочатку, що всі основні витрати компанії покриваються за рахунок інших видів страхування. Тоді прибуток, який не менший від 60 000 грн., компанія отримає у випадку, коли на виплату страхових сум у випадку смерті клієнтів вона витратить не більше 960 000 – 60 000 = 900 000 грн. Ця випадкова подія відбувається, якщо $m \le 90$. Тоді за інтегральною формулою Лапласа

$$P_{15\,000}(m \le 90) = P_{15\,000}(0 \le m \le 90) \approx \Phi(x_2) - \Phi(x_1),$$

де
$$x_1 = \frac{m_1 - np}{\sqrt{npq}} = \frac{0 - 60}{\sqrt{59{,}76}} = -7{,}76; \quad x_2 = \frac{90 - 60}{\sqrt{59{,}76}} = 3{,}88.$$

Тобто,

$$P_{15\,000}(0 \le m \le 90) \approx \Phi(3{,}88) - \Phi(-7{,}76) =$$
$$= 0{,}49995 + 0{,}5 = 0{,}99995.$$

Нехай тепер a – сумарні поточні видатки компанії за рік при здійсненні вказаного виду страхування. Тоді прибуток, не менший від 60 000 грн., компанія отримає тоді, коли відбудеться подія $m \le k$, де k – ціла частина числа $(900\,000 - a)/10\,000$. Далі за інтегральною формулою Лапласа можна знайти шукану імовірність $P_{15\,000}(0 \le m \le k)$. Вона вже буде суттєво меншою від попередньої імовірності. ●

Задача 3.4. При скануванні текстового матеріалу в середньому на кожну тисячу символів два помилкові. Знайти імовірність того, що після сканування тексту обсягом в 2 500 символів виявиться помилкових: а) шість символів; б) хоча б шість.

○ Випробування – сканування символу тексту, подія A – отримання помилкового символу. За умовою $n = 2500$, $p = P(A) = 0{,}002$.

а) Число появи події $m = 6$. Для знаходження $P_{2500}(6)$ скористаємося формулою Пуассона, оскільки n – велике, $p = 0{,}002 << 0{,}1$, $\lambda = np = 5 < 9$.

$$P_n(m) \approx \frac{\lambda^m e^{-\lambda}}{m!}, P_{2500}(6) \approx \frac{5^6 e^{-5}}{6!} = 0{,}14622.$$

Останнє значення знайдене для функції $\frac{\lambda^m e^{-\lambda}}{m!}$ двох змінних λ та m в табл. 2 додатків для значень $\lambda = 5$, $m = 6$.

б) Випадкова подія $(m \geq 6)$, імовірність якої треба знайти, зображається через прості події таким чином:

$$(m \geq 6) = (m = 6) + (m = 7) + ... + (m = 2500).$$

Використовувати теорему додавання імовірностей в такому випадку практично неможливо в зв'язку з тим, що в правій частині є 2495 доданків (!). З другого боку, протилежною до події $(m \geq 6)$ є подія $(m < 6)$, для якої виконується рівність

$$(m < 6) = (m = 0) + (m = 1) + (m = 2) + (m = 3) + (m = 4) + (m = 5),$$

звідки після використання теореми додавання імовірностей отримуємо

$$P_{2500}(m < 6) = P_{2500}(0) + P_{2500}(1) + P_{2500}(2) + P_{2500}(3) + P_{2500}(4) + P_{2500}(5).$$

Кожний із доданків обчислюється за формулою Пуассона. В даному випадку ці імовірності знаходяться за табл. 2 додатків для $\lambda = 5$ та $m = 0, 1, ..., 5$. Тобто

$$P_{2500}(m < 6) = 0{,}00674 + 0{,}03369 + 0{,}08422 + 0{,}14037 + + 0{,}17547 + 0{,}17547 = 0{,}61596.$$

Із урахуванням протилежності подій

$$P_{2500}(m \geq 6) = 1 - P_{2500}(m < 6) = 1 - 0{,}61596 = 0{,}38404.$$ ●

Задача 3.5. Імовірність того, що виріб задовольняє вимогам вищого сорту, дорівнює 0,8.

1) За місяць ВТК заводу перевірено 400 виробів. Знайти імовірність того, що відносна частота виготовлення виробу вищого сорту відхилиться від його імовірності по модулю не більше від 0,09.

2) Скільки виробів треба перевірити, щоб з імовірністю 0,95 можна було очікувати відхилення відносної частоти виготовлення виробу вищого сорту від його імовірності по модулю не більше від 0,04?

3) За наступні два місяці ВТК перевірив 900 виробів. Знайти з імовірністю 0,95 межі, в яких буде знаходитися число m виробів вищого ґатунку серед перевірених.

○ Випробування – перевірка виробу. Подія A – виріб задовольняє якостям вищого сорту.

За умовою задачі $n = 400$, $p = P(A) = 0{,}8$, $q = 0{,}2$. Використовуючи формулу (3.8)

$$P\left(\left|m/n - p\right| \le \varepsilon\right) \approx 2\Phi\left(\varepsilon\sqrt{n/(pq)}\right),$$

де за умовою $\varepsilon = 0{,}09$, отримаємо

$$P\left(\left|m/400 - 0{,}8\right| \le 0{,}09\right) \approx 2\Phi\left(0{,}09\sqrt{400/(0{,}8\cdot 0{,}2)}\right) =$$

$$= 2\Phi(4{,}5) = 2\cdot 0{,}499997 = 0{,}999994.$$

2) Згідно з умовою задачі $p = 0{,}8$, $q = 0{,}2$, $\varepsilon = 0{,}04$ і $P\left(\left|m/n - 0{,}8\right| \le 0{,}04\right) = 0{,}95.$ Потрібно знайти n.

З формули (3.8) знайдемо рівність

$$2\Phi\left(0{,}04\sqrt{n/(0{,}8\cdot 0{,}2)}\right) = 0{,}95,$$

звідки

$$\Phi\left(0{,}1\sqrt{n}\right) = 0{,}475.$$

За таблицею значень функції Лапласа $\left(\Phi(1{,}96) = 0{,}475\right)$ відтворимо аргумент, для якого виконується остання рівність: $0{,}1\sqrt{n} = 1{,}96.$ Далі $\sqrt{n} = 19{,}5,$ $n = 384{,}16.$

Враховуючи те, що n – ціле, а також поведінку похибки в наближеній рівності (3.8), остаточно отримуємо: $n \ge 385$.

3) За умовою $n = 900$, $p = 0{,}8$, $q = 0{,}2$, $P\left(\left|m/n - 0{,}8\right| \le \varepsilon\right) = 0{,}95.$

Потрібно знайти межі для числа m. Знайдемо спочатку ε, використавши згідно із формулою (3.8) рівність

$$2\Phi\left(\varepsilon\sqrt{900/(0{,}8\cdot 0{,}2)}\right) = 0{,}95 \quad \text{або} \quad \Phi(75\varepsilon) = 0{,}475.$$

За табл. 3 додатків знайдемо $\Phi(1{,}96) = 0{,}475$, тому $75\varepsilon = 1{,}96$, звідки $\varepsilon \approx 0{,}03$.

Таким чином, з імовірністю 0,95 відхилення відносної частоти числа виробів вищої якості від імовірності 0,8 задовольняє нерівність

$$\left|m/800 - 0{,}8\right| \le 0{,}03 \quad \text{або} \quad 0{,}77 \le m/800 \le 0{,}83,$$

і остаточно $616 \le m \le 664$. ●

Повернемося до розгляду задачі 1.2, спрощений варіант розв'язування якої було наведено в § 1.

Задача 3.6. Монету кидають двічі на горизонтальну тверду поверхню. Знайти імовірність того, що хоча б раз випаде герб.

○ Випробування – кидання монети, подія A – поява герба. Потрібно знайти $P_2(m \geq 1)$. Подія $(m \geq 1)$ протилежна до події $(m = 0)$. Тому $P_2(m \geq 1) = 1 - P_2(0) = 1 - C_2^0 \cdot (1/2)^0 \cdot (1/2)^2 = 3/4$. Отриманий результат співпадає із відповіддю задачі 1.2 ●

Наступна задача, з одного боку, ілюструє хибність припущень, продиктованих інтуїцією, а з другого – показує необхідність аналітичного обгрунтування прийняття рішень.

Задача 3.7. Підприємство отримало спецзамовлення на виготовлення 320 виробів, які працюватимуть в екстремальних умовах. Проте за існуючим технологічним процесом в середньому два вироби з кожних десяти не відповідають вимогам спецзамовлення. Тому керівництво підприємства розглядає питання про збільшення планових показників: виготовити 400 виробів, "компенсуючи" тим самим появу виробів, які бракуватимуться із врахуванням вимог спецзамовлення. При цьому виготовлені понад план вироби є високоліквідними.

1) Дати оцінку рішення керівництва підприємства.
2) Скільки потрібно виготовити всіх виробів, щоб з імовірністю, не меншою від 0,9999, виконати спецзамовлення?

○ Випробування – виготовлення виробу. Подія A – виріб задовольняє вимогам спецзамовлення.

1) За умовою n=400, $p=P(A)$=0,8, q=0,2.

Шукану імовірність $P_{400}(m \geq 320) = P_{400}(320 \leq m \leq 400)$ обчислимо за інтегральною формулою Лапласа ($npq = 400 \cdot 0{,}8 \cdot 0{,}2 = 64 >> 9$). Тоді

$$x_1 = \frac{m_1 - np}{\sqrt{npq}} = \frac{320 - 400 \cdot 0{,}8}{\sqrt{400 \cdot 0{,}8 \cdot 0{,}2}} = 0;\ x_2 = \frac{400 - 320}{\sqrt{64}} = 10,$$

$$P_{400}(320 \leq m \leq 400) \approx \Phi(10) - \Phi(0) = 0{,}5 .$$

Висновок: Вказаний варіант вирішення проблеми є невдалим, оскільки імовірність виконання замовлення дорівнює 0,5 – неприпустимо мала.

2) Нехай n – невідоме. За умовою задачі

$$P_n(320 \leq m \leq n) \geq 0{,}9999 .$$

Ліву частину нерівності знайдемо за інтегральною формулою Лапласа, використавши непарність функції Лапласа:

$$x_1 = \frac{320 - 0{,}8n}{\sqrt{n \cdot 0{,}8 \cdot 0{,}2}} = \frac{800 - 2n}{\sqrt{n}};\ x_2 = \frac{n - 0{,}8n}{0{,}4\sqrt{n}} = 0{,}5\sqrt{n};$$

$$P_n(320 \le m \le n) \approx \Phi(0{,}5\sqrt{n}) - \Phi[2(400 - n)/\sqrt{n}] = \Phi(0{,}5\sqrt{n}) +$$
$$+ \Phi[2(n - 400)/\sqrt{n}].$$

Нехай n>400, тоді $\Phi(0{,}5\sqrt{n}) = 0{,}5$ і початкова нерівність набере такого виду:

$$\Phi[2(n - 400)/\sqrt{n}] \ge 0{,}4999\ ,$$

За табл.3 додатків 0,4999=Ф(3,72). Тому отримуємо систему нерівностей

$$\begin{cases} \Phi[2(n - 400)/\sqrt{n}] \ge \Phi(3{,}72), \\ n > 400, \end{cases}$$

яка з урахуванням монотонного зростання функції Лапласа рівносильна системі

$$\begin{cases} 2(n - 400)/\sqrt{n} \ge 3{,}72, \\ n > 400. \end{cases}$$

Позначивши $\sqrt{n} = y$, першу нерівність запишемо в такому вигляді

$$y^2 - 1{,}86y - 400 \ge 0\ ,$$

звідки $y \in (-\infty; -19{,}091661] \cup [20{,}95161; \infty)$.

Тому остання система нерівностей рівносильна такій

$$\begin{cases} \sqrt{n} \ge 20{,}95161, \\ n > 400, \end{cases}$$

звідки $n \ge 438{,}96996$. Враховувавши цілочисельність n, отримаємо відповідь: для того, щоб з ймовірністю, не меншою від 0,9999, виконати спецзамовлення, підприємству потрібно виготовити не менше, ніж 439 виробів. ●

Задача 3.8. Цех отримав замовлення на термінове виготовлення 105 виробів. Проте в середньому кожні три вироби із десяти виготовлених вимагають тривалої доводки і тому не можуть бути включені у партію виробів термінового замовлення.

1) Знайти, скільки потрібно виготовити цеху виробів, щоб число 105 було найімовірнішим. 2) Враховуючи знайдене

число виробів, оцінити можливість виконання замовлення. При цьому вважати, що продукція є високоліквідною.

○ Випробування – виготовлення виробу. Подія A – виріб стандартний (не вимагає доводки). За умовою $p=P(A)=0{,}7$, тоді $q=0{,}3$.

1) Знайдемо число випробувань n, використавши інформацію про те, що найімовірніше число появи події A дорівнює 105, тобто $m_0=105$. Згідно із співвідношенням (3.4)

$$np-q\le m_0\le np+p\,.$$

Підставляючи дані задачі, отримаємо систему нерівностей для знаходження n:

$$\begin{cases}0{,}7n-0{,}3\le 105,\\ 0{,}7n+0{,}7\ge 105,\end{cases}\Leftrightarrow\begin{cases}0{,}7n\le 105{,}3\\ 0{,}7n\ge 104{,}3\end{cases}\Leftrightarrow$$

$$\Leftrightarrow 149\le n\le 150{,}429.$$

Оскільки n – ціле, то остаточно отримаємо, що $n=149$ або $n=150$.

2) Знайдемо імовірність виконання замовлення при виготовленні 149 виробів (при цьому врахуємо можливість випуску більшої кількості виробів від замовленої, оскільки за умовою понадпланова продукція буде реалізована згодом):

$$P_{149}(m\ge 105)=P_{149}(105\le m\le 149)\,.$$

Використаємо інтегральну формулу Лапласа

$$(npq=149\cdot 0{,}7\cdot 0{,}3=31{,}29>>9):$$

$$x_1=\frac{105-149\cdot 0{,}7}{\sqrt{31{,}29}}=0{,}125\,,\;\; x_2=\frac{149-149\cdot 0{,}7}{\sqrt{31{,}29}}=7{,}99\,,$$

$$P_{149}(105\le m\le 149)\approx\Phi(7{,}99)-\Phi(0{,}125)=0{,}5-0{,}0474=0{,}45026\,.$$

При цьому використаємо лінійну інтерполяцію:

$$\Phi(0{,}125)=\Phi(0{,}12)+\frac{1}{2}[\Phi(0{,}13)-\Phi(0{,}12)]\,.$$

Якщо ж $n=150$, тоді

$$x_1=\frac{105-150\cdot 0{,}7}{\sqrt{150\cdot 0{,}7\cdot 0{,}3}}=0\,,\;\; x_2=\frac{150-150\cdot 0{,}7}{5{,}6125}=8{,}02\,,$$

$$P_{150}(105\le m\le 150)\approx\Phi(8{,}02)-\Phi(0)=0{,}5\,.$$ ●

Задача 3.9. Статистично встановлено, що із 100 виготовлених деталей 90 – стандартні. Скільки потрібно виготовити деталей, щоб з імовірністю 0,96856 можна було очікувати, що не менше 270 деталей будуть стандартними?

○ Випробування – виготовлення деталі. Подія A – деталь стандартна. За умовою задачі $p=P(A)=0{,}9$, тоді $q=0{,}1$, $P(270 \le m \le n) = 0{,}96856$. Потрібно знайти число випробувань n.

За інтегральною формулою Лапласа

$$x_1 = \frac{270 - 0{,}9n}{\sqrt{n \cdot 0{,}9 \cdot 0{,}1}} = (900 - 3n)\big/\sqrt{n}\,;\; x_2 = \frac{n - 0{,}9n}{0{,}3\sqrt{n}} = \sqrt{n}\big/3\,,$$

$$P(270 \le m \le n) \approx \Phi(\sqrt{n}\big/3) - \Phi\left[(900 - 3n)\big/\sqrt{n}\right].$$

Оскільки функція Лапласа зростаюча і $n \ge 270$, то

$$\Phi(\sqrt{n}\big/3) \ge \Phi(\sqrt{270}\big/3) = \Phi(5{,}477) \approx 0{,}5\,.$$

Враховуючи умову задачі, отримаємо рівняння

$$0{,}5 - \Phi[3(300 - n)\big/\sqrt{n})] = 0{,}96856$$

або (функція Лапласа непарна)

$$\Phi[3(n - 300)\big/\sqrt{n})] = 0{,}46856\,.$$

За табл.3 додатків $0{,}46856 = \Phi(1{,}86)$, тому

$$3(n - 300)\big/\sqrt{n} = 1{,}86\,.$$

Розв'язавши це рівняння, як квадратне відносно $\sqrt{n}$, отримаємо, що $\sqrt{n} = 17{,}633282$, звідки $n=310{,}93263$. Отже, шукане число виготовлених деталей $n=311$. ●

Задача 3.10. Підприємство відправило замовнику 20000 стандартних виробів. Середнє число виробів, які пошкоджуються при транспортуванні, складає 0,02%. Знайти імовірність того, що замовник отримає непошкоджених виробів: а) 19997; б) хоча б 19997.

○ Випробування – транспортування одного виробу. Подія A – виріб залишиться стандартним після транспортування. За умовою $n=20000$, $p=P(A)=0{,}9998$.

а) Потрібно знайти $P_{20000}(19997)$. Оскільки $npq = 20000 \cdot 0{,}9998 \cdot 0{,}0002 = 3{,}99992 << 9$, то використання локальної формули Лапласа призведе до великої похибки. З другого боку, для партії 20000 виробів випадкові події "19997 стандартних (непошкоджених) виробів" і "3 пошкоджені вироби" – рівносильні. Тому $P_{20000}(19997)= P_{20000}(3)$, де в останній імовірності вже в якості події A фігурує подія $\overline{A}$ і те, що $\lambda = np = 4 < 9$, а n велике, дозволяє використати формулу Пуассона (3.6):

$$P_{20000}(3) \approx \frac{4^3 \cdot e^{-4}}{3!}.$$

За табл.2 додатків знайдемо значення функції $P(m;\lambda)$ при m=3, $\lambda = 4$; P(3; 4)=0,19537. Отже, шукана імовірність дорівнює 0,19537.

б) Використаємо наведений вище перехід від події A до події $\overline{A}$ - пошкодження виробу внаслідок транспортування. Тоді шукана імовірність позначиться $P_{20000}(m \le 3)$, де p=0,0002.

$$P_{20000}(m \le 3) = P_{20000}(0) + P_{20000}(1) + P_{20000}(2) + P_{20000}(3).$$

Кожний із доданків справа обчислимо за формулою Пуассона, використавши табл.2 додатків (знаходимо значення функції $P(m;\lambda)$ для $\lambda = 4$ і m=0, 1, 2, 3):

$$P_{20000}(m \le 3) = 0{,}01832 + 0{,}07326 + 0{,}14653 + 0{,}19537 = 0{,}43348. \bullet$$

Зауваження. Використання локальної формули Лапласа в п. а) дає імовірність 0,17604, а інтегральної в п. б) – 0,28579.

ЗАДАЧІ ДЛЯ САМОСТІЙНОГО РОЗВ'ЯЗУВАННЯ

Завдання №4

1. Три станки з програмним управлінням виготовляють однотипні деталі, які подаються на спільний конвеєр. Їх продуктивності відносяться як 2:5:3, причому перший виготовляє 83% деталей вищої якості, якими комплектуються вироби на експорт, другий – 84%, а третій – 88%. Знайти імовірність того, що з 800 навмання відібраних з конвеєра деталей вищої якості виявиться: а) 675; б) хоча б 675.

2. Цех отримав замовлення на термінове виготовлення 37 виробів. Кожні два вироби із десяти виготовлених вимагають тривалої доводки і тому не можуть бути включені у партію виробів термінового замовлення. 1) Скільки потрібно виготовити цеху виробів, щоб 37 було найімовірнішим числом? 2) Яка імовірність виконання замовлення в такій постановці задачі?

3. Обленерго обслуговує 1200 найпотужніших споживачів електроенергії. Перебої у подачі енергії протягом доби виникають з імовірністю 0,0025. Знайти імовірність того, що протягом доби надійде не менше п'яти, але не більше восьми повідомлень про перебої. Порівняти знайдену імовірність із імовірністю, що відповідає найімовірнішому числу повідомлень про перебої за добу.

4. Для нормальної роботи гуртової бази на лінії має бути не менше 4 вантажних бусів, а їх є 7. Імовірність для кожного з них не вийти на лінію дорівнює 0,05. Знайти імовірність того, що найближчого дня гуртова база буде працювати нормально.

5. Для студентського гуртожитку закуплено 6 телевізорів. Імовірність того, що будь-який із них витримає гарантійний термін, дорівнює 0,9. Знайти імовірність того, що протягом гарантійного терміну з ладу вийдуть: а) два телевізори; б) принаймні два. Знайти найімовірніше число телевізорів, що витримають гарантійний термін.

6. Статистично встановлено, що 70% продукції складають вироби вищої категорії якості. Підприємство отримало замовлення на термінову поставку на експорт 560 виробів вищої категорії якості. Вся продукція користується високим попитом на внутрішньому ринку, тому на момент отримання замовлення продукції на складі не було. Скільки потрібно виготовити всіх виробів, щоб з імовірністю, не меншою від 0,999, виконати термінове замовлення?

7. Імовірність появи події в кожному із 300 незалежних випробувань дорівнює 0,9. Знайти таке додатне число ε, щоб з імовірністю 0,96 абсолютна величина відхилення відносної частоти появи події від її імовірності 0,9 не перевищила ε.

8. Проводиться порівняння зон покриття мобільного зв’язку двох основних операторів. Для цього було залучено 620 телефонів першого оператора і 650 – другого. Відомо, що зв’язок першого оператора підтримується у 90% зони, а другого – 85%. Знайти найімовірнішу кількість телефонів кожного оператора, які мали зв’язок, а також імовірність таких кількостей.

9. Наклад виданого посібника складає 15000 екземплярів. Імовірність того, що навмання взятий посібник виявиться неправильно зброшурованим, дорівнює 0,0002. Знайти імовірність того, що: 1) хоча б 14997 книг будуть зброшуровані правильно; 2) наклад містить чотири браковані книжки.

10. Імовірність того, що пасажир запізниться до відправлення поїзда, дорівнює 0,01. Для 760 пасажирів поїзда знайти імовірність найбільш імовірного числа тих пасажирів, які запізняться.

11. Технологічний процес підприємства дозволяє одержати 90% виробів вищого ґатунку. Знайти найімовірніше число виробів

вищого ґатунку серед 300 виготовлених підприємством виробів, а також імовірність появи цього числа виробів.

12. Імовірність того, що навмання взята електрична лампочка відпрацює передбачений стандартом термін, дорівнює 0,95. Знайти імовірність того, що з 400 придбаних лампочок хоча б 370 відпрацюють передбачений стандартом час, а також найімовірніше число таких лампочок.

13. Два рівносильні партнери грають в шахи. Знайти, що імовірніше для одного з них: виграти не менше двох партій з чотирьох чи не менше трьох партій із п'яти. Партії, які завершилися внічию, до уваги не беруться.

14. Імовірність того, що мале підприємство за проміжок часу T збанкрутує, дорівнює 0,3. Знайти імовірність того, що із десяти малих підприємств за час T: а) збанкрутує більше двох; б) продовжать свою діяльність щонайменше три.

15. Контролер перевіряє однотипні деталі на стандартність. Імовірність того, що деталь є стандартною, складає 0,8. Знайти межі, в яких з імовірністю 0,95 знаходиться число стандартних деталей серед 400 перевірених.

16. Відомо, що в партії однотипних деталей брак в середньому складає 5%. Скільки необхідно перевірити деталей, щоб з імовірністю 0,954 відхилення відносної частоти від імовірності браку не перебільшило 4%?

17. Три з 20 000 квитків лотереї є виграшними. Серед працівників підприємства розповсюджено 400 квитків лотереї. Знайти імовірність того, що серед них виявиться: а) 2 виграшних квитки; б) хоча б 2 виграшних.

18. У гуртожитку мешкає 60 відсотків студентів стаціонару. Яка імовірність того, що з 10 випадково вибраних студентів не більше 8 проживає у гуртожитку?

19. Імовірність того, що стодоларова купюра фальшива, дорівнює 0,002. 1). Знайти найімовірніше число фальшивих купюр серед 400, а також імовірність такої кількості купюр. 2). Знайти імовірність того, що з 400 купюр хоча б одна виявиться фальшивою.

20. Для розвинутих країн Заходу частка тіньового бізнесу складає 1%. Яка імовірність того, що серед 200 зареєстрованих за рік фірм

таким бізнесом займається хоча б дві?

21. Деяка компанія володіє мережею дилерів на біржі. Імовірність того, що дилер буде грати вдало, становить 0,8. Знайти імовірність того, що з шести дилерів у збитках виявиться: а) два; б) хоча б два.

22. Знайти імовірність появи події в кожному із 100 незалежних випробувань, якщо найімовірніше число появи події під час випробувань складає 25.

23. Імовірність того, що випадково відібраний із партії прилад вимагає додаткового регулювання, дорівнює 0,05. Якщо при вибірковій перевірці партії приладів виявиться, що не менше 6% відібраних приладів вимагають регулювання, тоді вся партія повертається для доробки. Знайти імовірність того, що партія буде повернена, якщо для контролю із партії відібрали 500 приладів.

24. Відомо, що три чверті населення міста користується послугами кабельного телебачення. Яка імовірність того, що серед 300 мешканців такими послугами користується хоча б 220?

25. Відділом технічного контролю заводу встановлено, що в середньому 4 вироби із 100 є бракованими. Знайти межі в яких з імовірністю 0,95 буде знаходитись число m бракованих виробів серед 485 виготовлених.

26. Встановлено, що 5% імпортних телевізорів виходять з ладу через перепади напруги електромережі. Яка імовірність того, що з п'яти придбаних телевізорів хоча б три не вийдуть з ладу?

27. На дорогах України лише 80% автомобільних шин витримують гарантійний термін. Знайти найімовірніше число шин, що не витримують гарантійного терміну, з партії 500 шин, а також імовірність такої кількості шин.

28. Завод відправив на гуртову базу 30000 стандартних виробів. Відсоток пошкоджених при транспортуванні виробів за цим маршрутом складає 0,02%. Знайти імовірність того, що замовник отримає виробів: 1) хоча б 2 пошкоджених; 2) хоча б 29997 неушкоджених.

29. Відомо, що в технологічному процесі виготовлення мікросхем забезпечується 98% продукції, яка відповідає технологічним вимогам. Яка імовірність того, що з 200 мікросхем бракованих виявиться не менше трьох.

30. Відомо, що серед готівкової маси 0,5% купюр є непридатними до наступного використання. Знайти імовірність того, що серед 2400 купюр виручки магазину непридатними для наступного використання є хоча б дві купюри.

31. За результатами перевірок податковою інспекцією встановлено, що в середньому одне мале підприємство із чотирьох має порушення фінансової дисципліни. Знайти імовірність того, що з 520 зареєстрованих в регіоні малих підприємств мають порушення фінансової дисципліни: а) найімовірніше число підприємств; б) не менше 135.

32. Страхова компанія виплачує страхову суму в середньому по 9% укладених угод. Знайти найімовірніше число настання страхового випадку із виплатою страхової суми, якщо укладено 450 угод. Яка імовірність цього числа виплат страхової суми?

33. У середньому 30% акцій видавничих фірм протягом року стають збитковими. Яка імовірність того, що серед 120 акцій цих фірм збитковими будуть менше 40?

34. Оцінити імовірність p появи події A в кожному із 59 незалежних випробувань, якщо найімовірніше число появи події A в цих випробуваннях дорівнює 35.

35. Кожний десятий пасажир громадського транспорту має документ про пільговий проїзд. Контролер перевіряє проїзні документи у п’яти навмання вибраних пасажирів. Яка імовірність того, що хоча б один із перевірених пасажирів має документ про пільговий проїзд?

36. Серед автомобілів, що ввозяться в Україну, 85% складають легківки. Протягом дня на митницю прибуло 10 автомобілів. Яка імовірність того, що не більше 9 з них легківки?

37. 2. За статистичними даними 20% усіх затримок рейсів авіакомпанії відбувається з вини служби перевезень. Протягом тижня з різних причин із затримкою було виконано 34 рейси. Знайти найбільш імовірне число рейсів, затриманих із вини служби перевезень, і обчислити відповідну ймовірність.

38. Абоненти мобільного зв'язку не отримують відправлені повідомлення з імовірністю 0,1. Знайти ймовірність того, що серед 300 відправлених повідомлень буде рівно 90 отриманих.

39. Проводиться дослідження для з'ясування питомої ваги захворюваності на грип серед інших захворювань. Скільки лікарняних слід включити в дослідження, щоб відхилення відносної частоти захворюваності на грип від ймовірності 0,4 при одному випробуванні не перевищило числа 0,06 (за абсолютною величиною) з ймовірністю 0,9973.

40. На контроль надійшла велика партія виробів. Відомо, що 5 % усіх виробів не відповідає стандарту. Знайти найбільш імовірну кількість нестандартних виробів серед 6 перевірених і відповідну їй імовірність.

41. Авіакомпанія виконує протягом місяця 500 рейсів. Імовірність повного комерційного завантаження кожного рейсу дорівнює 0,6. Знайти ймовірність того, що протягом місяця з повним комерційним завантаженням буде виконано не менше половини рейсів.

42. Ймовірність того, що відвідувачу в їдальні знадобиться перша страва, дорівнює 0,7. Знайти ймовірність того, що із 100 відвідувачів першу страву замовлять не менше 65 і не більше 80 відвідувачів.

43. У камері Вільсона в середньому реєструється 15 елементарних частинок за годину. Визначити ймовірність того, що протягом 20 хвилин буде зареєстровано хоча б 2 частинки.

44. На авіарейс продано 44 квитки. Ймовірність того, що пасажир відмовиться від рейсу, дорівнює 0,1. Знайти ймовірність того, що рівно 4 пасажирів відмовляться від рейсу.

45. Імовірність того, що покупець, який завітав до взуттєвого магазину, здійснить покупку, дорівнює в середньому 0,1. Яка ймовірність того, що із 400 покупців, що завітали до магазину, здійснять покупку від 100 до 180 покупців?

46. Енергетична компанія обслуговує 1000 споживачів електроенергії. Перебої у подачі енергії протягом доби виникають з імовірністю 0,005. Яка ймовірність того, що протягом доби надійде не більше 4, але не менше 9 повідомлень про перебої?

47. Візуально спостерігати в заданому пункті штучний супутник Землі можна з імовірністю 0,1 щоразу, коли він пролітає над цим пунктом. Скільки разів має пролетіти супутник над пунктом

спостереження, щоб з імовірністю, не меншою за 0,998, вдалося здійснити принаймні 5 спостережень?

48. Оптова база постачає 30 крамниць, від кожної з яких може надходити заявка на черговий день з імовірністю 0,6, незалежно від заявок від інших крамниць. Знайти найбільш імовірну кількість заявок на день і ймовірність одержання цієї кількості заявок.

49. Локальна мережа складається із 50 комп'ютерів. Імовірність виникнення збоїв у роботі протягом доби для кожного з них дорівнює 0,02. Яка ймовірність того, що протягом доби збої виникнуть не більше ніж у 2 комп'ютерах?

50. Фірма виконує поліграфічні роботи, причому 10 % замовлень припадає на виготовлення візитних карток. Знайти ймовірність того, що серед 50 клієнтів 15 замовлять візитні картки.

51. Піцерія отримує 60 % замовлень на фірмову піцу. Знайти ймовірність того, що серед 50 замовлень на фірмову піцу буде рівно половина.

52. До агентства з нерухомості звертаються з приводу оренди квартир з імовірністю 0,7. Яка ймовірність того, що серед 9 довільно вибраних заявок буде не менше 5 щодо оренди?

53. До технічного водопроводу приєднано 80 підприємств, кожне з яких з імовірністю 0,8 у даний момент здійснює забір води з магістралі. Визначити ймовірність того, що в цей момент забір води здійснює 60 підприємств.

§ 4. ДИСКРЕТНІ ВИПАДКОВІ ВЕЛИЧИНИ ТА ЇХ ЧИСЛОВІ ХАРАКТЕРИСТИКИ

Випадкові величини та їх види. Закон розподілу імовірностей дискретної випадкової величини. Основні розподіли дискретних (цілочисельних) випадкових величин (рівномірний, біноміальний, пуассонівський, геометричний, гіпергеометричний). Найпростіший потік подій. Дії над випадковими величинами. Приклади знаходження законів розподілу випадкових величин. Числові характеристики дискретних випадкових величин та їх властивості (математичне сподівання, дисперсія, середнє квадратичне відхилення, початкові та центральні моменти). Числові характеристики основних законів розподілу.

КОРОТКІ ТЕОРЕТИЧНІ ВІДОМОСТІ ТА ПРИКЛАДИ РОЗВ'ЯЗУВАННЯ ЗАДАЧ

Випадковою називається величина, яка при випробуванні набирає єдиного значення із всіх можливих з деякою імовірністю, тобто наперед невідомо, яке конкретне можливе значення вона набере, оскільки це залежить від випадкових причин.

Дискретною (**перервною**) називається випадкова величина, можливі значення якої є ізольованими числами. Число можливих значень дискретної випадкової величини може бути скінченним або зчисленним. В останньому випадку можна встановити взаємно-однозначну відповідність між можливими значеннями і натуральними числами 1, 2, 3, ..., n, ...

Неперервною називається випадкова величина, можливі значення якої заповнюють суцільно деякий скінченний або нескінченний проміжок. Очевидно, що число можливих значень кожної неперервної величини нескінченне.

Зауваження. Означення неперервної випадкової величини має попередній характер. Уточнення буде зроблено в наступному параграфі.

Інформації про множину можливих значень недостатньо для повного описання випадкової величини (різні величини можуть мати однакові можливі значення). Потрібно ще знати, з якими імовірностями набираються можливі значення випадковою величиною.

Законом розподілу імовірностей дискретної випадкової величини називається відповідність між можливими значеннями та імовірностями, з якими вони набираються випадковою величиною.

Таблична форма задання закону розподілу має такий вид

$$\begin{array}{c|cccc} X & x_1 & x_2 & \dots & x_n \\ \hline P & p_1 & p_2 & \dots & p_n \end{array}, \qquad (4.1)$$

де $p_i = P(X = x_i)$, $i = \overline{1,n}$. Оскільки в одному випробуванні випадкова величина набирає тільки одне із своїх можливих значень, то випадкові події $(X = x_1)$, $(X = x_2)$, ..., $(X = x_n)$ утворюють повну групу. Тому сума їх імовірностей дорівнює одиниці:

$$p_1 + p_2 + \dots p_n = 1. \qquad (4.2)$$

Ця рівність називається **умовою нормування**.

Якщо множина можливих значень дискретної випадкової величини зчисленна: $x_1, x_2, ..., x_n, ...$, то ряд $\sum\limits_{t=1}^{\infty} p_i$ збігається і його сума дорівнює одиниці.

Задача 4.1. Складальник навмання бере дві деталі із контейнера, в якому знаходиться 20 деталей, серед яких 4 нестандартні. Знайти закон розподілу числа нестандартних деталей серед відібраних.

○ Нехай X – число нестандартних деталей серед двох відібраних. Можливі значення X – 0, 1, 2. Знайдемо відповідні імовірності, використовуючи класичне означення:

$$p_1 = P(X = 0) = m/n = C_{16}^2 / C_{20}^2 = 12/19,$$

$$p_2 = P(X = 1) = C_{16}^1 C_4^1 / C_{20}^2 = 32/95,$$

$$p_3 = P(X = 2) = C_4^2 / C_{20}^2 = 3/95.$$

Перевірка: $p_1 + p_2 + p_3 = 12/19 + 32/95 + 3/95 = 1.$

Шуканий закон розподілу має такий вид:

$$\begin{array}{c|ccc} X & 0 & 1 & 2 \\ \hline P & 12/19 & 32/95 & 3/95 \end{array}. \quad \bullet$$

Основні розподіли дискретних (цілочисельних) випадкових величин

Закон розподілу дискретної випадкової величини X можна задати також **аналітично**, тобто з допомогою формули $p_i = P(X = x_i) = g(x_i)$, $i = \overline{1,n}$. Всі нижче наведені закони задаються аналітично.

Рівномірний розподіл

Задача 4.2. У зв'язці є п'ять ключів, з яких тільки один підходить до замка. Скласти закон розподілу числа ключів, що випробовуються при відкриванні замка, якщо ключ, що був у випробуванні, у наступних випробуваннях участі не бере. Знайти імовірність того, що число випробувань не перевищить двох.

○ Позначимо: X – число ключів, що випробовуються при відкриванні замка; A_i – випадкова подія, яка полягає в тому, що i-тий ключ відкриє замок $(i = \overline{1,\ 5})$. Можливі значення X: 1, 2, 3, 4, 5. Знайдемо відповідні імовірності закону розподілу, попередньо з'ясувавши структуру випадкових подій (X = 1), (X = 2), ..., (X = 5):

$$(X = 1) = A_1,\quad (X = 2) = \overline{A}_1 A_2,\quad (X = 3) = \overline{A}_1 \overline{A}_2 A_3,$$
$$(X = 4) = \overline{A}_1 \overline{A}_2 \overline{A}_3 A_4,\quad (X = 5) = \overline{A}_1 \overline{A}_2 \overline{A}_3 \overline{A}_4 A_5.$$

Використавши теорему добутку для залежних подій, отримаємо:

$$p_1 = 1/5,\ \ p_2 = P(\overline{A}_1) P_{\overline{A}_1}(A_2) = \frac{4}{5} \cdot \frac{1}{4} = \frac{1}{5},$$
$$p_3 = P(\overline{A}_1) P_{\overline{A}_1}(\overline{A}_2) P_{\overline{A}_1 \overline{A}_2}(A_3) = \frac{4}{5} \cdot \frac{3}{4} \cdot \frac{1}{3} = \frac{1}{5},$$
$$p_4 = \frac{1}{5},\ \ p_5 = \frac{1}{5}.$$

Шуканий розподіл має вид

X	1	2	3	4	5
P	$1/5$	$1/5$	$1/5$	$1/5$	$1/5$

Випадкова подія ($X \le 2$) полягає в тому, що число випробувань ключів не перевищить двох. Ясно, що ($X \le 2$) = (X = 1) + (X = 2). В силу теореми додавання імовірностей для несумісних подій $P(X \le 2) = P(X = 1) + P(X = 2) = 2/5$. ●

Характерною особливістю отриманого закону розподілу є рівність всіх імовірностей. Відповідна випадкова величина називається **рівномірно розподіленою**. В загальному випадку цілочисельна випадкова величина **розподілена за рівномірним законом** (**рівномірно розподілена**), якщо імовірності в законі розподілу мають такий вид: $p_k = P(X = k) = 1/n,\ k = \overline{1, n}.$

Біноміальний розподіл

Задача 4.3. Імовірність появи події A в кожному із n повторних випробувань дорівнює p. Скласти закон розподілу випадкової величини X – числа появи події в n випробуваннях.

○ Можливі значення величини X такі: $x_1 = 0$, $x_2 = 1$, $x_3 = 2$, ..., $x_{n+1} = n$. Нехай m будь-яке із цих чисел. Тоді $(X = m)$ – випадкова подія, яка полягає в тому, що в n повторних незалежних випробуваннях ($P(A) = p$ залишається незмінною для кожного випробування) подія A відбудеться рівно m разів. Тоді її імовірність можна позначити $P_n(m)$ (в термінах § 3) і знайти за формулою Бернуллі: $P(X = m) = P_n(m) = C_n^m p^m q^{n-m}$, де $m = 0, 1, 2, ..., n$. Отже, аналітичний вираз закону розподілу імовірностей даної випадкової величини X має такий вид:

$$p_{m+1} = P(X = m) = C_n^m p^m q^{n-m}, \qquad (4.3)$$

де $m = 0, 1, 2, ..., n$; $q = 1 - p$. ●

Закон розподілу цілочисельної випадкової величини, імовірності якого знаходяться за формулою Бернуллі (4.3), називається **біноміальним**.

Задача 4.4. Імовірність того, що підприємець при перетині кордону декларує не весь товар, дорівнює 0, 4. За зміну митний контроль пройшло 4 підприємці. Скласти закон розподілу числа підприємців, які декларують весь товар.

○ Позначимо: X – число підприємців, які задекларують весь товар при перетині кордону, A – підприємець декларує весь товар. За умовою задачі $p = P(A) = 0,6$, $q = 0,4$, $n = 4$. Можливі значення X: 0, 1, 2, 3, 4. Відповідні імовірності знайдемо, використовуючи співвідношення (4.3):

$$p_1 = P(X = 0) = C_4^0 (0,6)^0 (0,4)^4 = 0,0256;$$

$$p_2 = P(X = 1) = C_4^1 (0,6)^1 (0,4)^3 = 0,1536;$$

$$p_3 = P(X = 2) = C_4^2 (0,6)^2 (0,4)^2 = 0,3456;$$

$$p_4 = P(X = 3) = C_4^3 (0,6)^3 (0,4)^1 = 0,3456;$$

$$p_5 = P(X = 4) = C_4^4 (0,6)^4 (0,4)^0 = 0,1296.$$

Шуканий закон розподілу є біноміальним і має вид:

X	0	1	2	3	4
P	0,0256	0,1536	0,3456	0,3456	0,1296

. ●

Пуассонівський розподіл

Нехай X – випадкова величина, визначена в задачі 4.3, де n – велике, $p << 0{,}1$, $\lambda = np \le 9$. Тоді імовірності (4.3) в законі розподілу доцільно шукати за формулою Пуассона:

$$p_{m+1} = P(X = m) = \frac{\lambda^m e^{-\lambda}}{m!}, \quad m = 0,\ 1,\ 2,\ \ldots,\ n,$$

а відповідний розподіл називається пуассонівським.

Зауваження. В пуассонівському розподілі суттєво, що $n \to \infty$.

Геометричний розподіл

Задача 4.5. Проводяться повторні випробування, в кожному з яких $P(A) = p > 0$. Випробування закінчуються, як тільки відбудеться подія A. Скласти закон розподілу випадкової величини X – числа випробувань, які потрібно провести до першої появи події A.

○ Можливі значення X – 1, 2, 3, ..., n, ... Подія A може відбутися або в 1-му, або в 2-му, ..., або в n-му, ..., випробуваннях, тобто $(X = 1) = A$, $(X = 2) = \overline{A}A$, $(X = 3) = \overline{A}\overline{A}A$, ..., $(X = n) = \overline{A}\overline{A}\ldots\overline{A}A$, ... Враховуючи незалежність випробувань і теорему множення імовірностей, отримаємо:

$$p_1 = p,\ \ p_2 = qp,\ \ p_3 = q^2 p,\ \ \ldots,\ \ p_n = q^{n-1} p,\ \ldots\ . \quad \bullet \tag{4.4}$$

Закон розподілу імовірностей, який визначається послідовністю імовірностей (4.4), називається **геометричним**. Назва закону зумовлена тим, що послідовність (4.4) утворює нескінченну **геометричну прогресію** з першим членом p і знаменником $q \in (0{,}1)$.

Гіпергеометричний розподіл

Задача 4.6. З партії N виробів, серед яких M стандартних ($M < N$), навмання вибирається n виробів ($n \le M$) без повернення кожного із них. Знайти закон розподілу випадкової величини X – числа стандартних виробів серед відібраних.

○ Можливі значення X: 0, 1, 2, 3, ..., n. Нехай m одне із цих можливих значень. Тоді випадкова подія ($X = m$) полягає в тому, що серед n відібраних виробів m стандартні і $n - m$ нестандартних. Імовірність цієї події знайдемо за класичним означенням:

$$P(X = m) = C_M^m \cdot C_{N-M}^{n-m} \big/ C_N^n, \quad m = 0,\ 1,\ 2,\ \ldots,\ n. \quad \bullet \tag{4.5}$$

Закон розподілу імовірностей, який визначається співвідношенням (4.5), називається **гіпергеометричним**.

Самостійно розглянути випадок $M < n < N$.

Найпростіший потік подій

Потоком подій називається послідовність подій, які відбуваються у випадкові моменти часу.

Найпростішим називається **потік подій**, який має властивості **стаціонарності**, **відсутності післядії** та **ординарності**.

Інтенсивністю λ потоку називається середнє число подій, які відбуваються за одиницю часу.

Математичною моделлю найпростішого потоку подій є формула Пуассона

$$P_t(m) = (\lambda t)^m e^{-\lambda t} / m!, \qquad (4.6)$$

з допомогою якої можна знайти імовірність появи m подій найпростішого потоку за проміжок часу довжиною t.

Задача 4.7. Середнє число обривів нитки на прядильному верстаті за 1 хв. дорівнює двом. Знайти імовірність того, що за 3 хв. число обривів нитки становитиме: а) 4; б) менше чотирьох; в) не менше чотирьох. Припускається, що потік обривів нитки найпростіший.

○ а) За умовою $\lambda = 2$, $t = 3$, $m = 4$. За формулою (4.6) імовірність того, що за 3 хв. число обривів нитки становитиме 4, $P_3(4) = 6^4 e^{-6}/4! = 0{,}13385$. Числове значення знаходиться з допомогою таблиці значень для формули Пуассона (табл. 2 додатків) для $\boldsymbol{\lambda = 6}$, $m = 4$ (порівняйте формулу (4.6) із формулою Пуассона!).

б) Випадкова подія $(m < 4) = (m = 0) + (m = 1) + (m = 2) + (m = 3)$, звідки із використанням теореми додавання імовірностей попарно несумісних подій

$$P_3(m < 4) = P_3(0) + P_3(1) + P_3(2) + P_3(3) =$$
$$= 0{,}00248 + 0{,}01487 + 0{,}04462 + 0{,}08924 = 0{,}15121.$$

в) Події $(m \ge 4)$ та $(m < 4)$ протилежні, тому

$$P_3(m \ge 4) = 1 - P_3(m < 4) = 1 - 0{,}15121 = 0{,}84879.$$ ●

Дії над випадковими величинами

Визначимо **добуток сталої величини *C* на дискретну випадкову величину *X***, задану законом розподілу (4.1), як дискретну випадкову величину CX, закон розподілу якої має вид

$$\begin{array}{c|cccc} CX & Cx_1 & Cx_2 & \dots & Cx_n \\ \hline P & p_1 & p_2 & \dots & p_n \end{array}. \qquad (4.7)$$

Квадрат випадкової величини *X*, тобто X^2 – це нова дискретна випадкова величина, яка описується законом розподілу

$$\begin{array}{c|cccc} X^2 & x_1^2 & x_n^2 & \dots & x_n^2 \\ \hline P & p_1 & p_2 & \dots & p_n \end{array}. \qquad (4.8)$$

Нехай випадкова величина X задана розподілом (4.1), а Y – законом розподілу $\begin{array}{c|cccc} Y & y_1 & y_2 & \cdots & y_m \\ \hline P & g_1 & g_2 & \cdots & g_m \end{array}$. Ці величини називаються **незалежними**, якщо випадкові події $(X = x_i)$, $(Y = y_j)$ незалежні при довільних $i = \overline{1,n}$ та $j = \overline{1,m}$. Поняття незалежності випадкових величин поширюється на довільне скінченне число випадкових величин. В протилежному випадку випадкові величини називаються **залежними**.

Дискретні випадкові величини $X_1, X_2, ..., X_k$ називаються **незалежними у сукупності (взаємно незалежними)**, якщо закон розподілу **кожної** із них не змінюється, якщо довільна випадкова величина або будь-які групи цих величин наберуть яке завгодно із своїх можливих значень.

Визначимо **суму випадкових величин** X та Y як випадкову величину $X + Y$, можливі значення якої рівні сумам кожного можливого значення X з кожним можливим значення Y, а імовірності можливих значень $X + Y$ для незалежних величин X та Y дорівнюють добуткам імовірностей доданків; для залежних величин – добуткам імовірностей одного доданку на умовну імовірність другого. Якщо деякі суми $x_i + y_j$ виявляються рівними між собою, тоді імовірність можливого значення суми дорівнює сумі відповідних імовірностей.

Задача 4.8. Випадкові величини X та Y задані законами розподілу $\begin{array}{c|cc} X & 2 & 5 \\ \hline P & 0{,}8 & 0{,}2 \end{array}$, $\begin{array}{c|ccc} Y & 1 & 3 & 6 \\ \hline P & 0{,}3 & 0{,}5 & 0{,}2 \end{array}$. Знайти $X + Y$, якщо X та Y незалежні величини.

○ Якщо X та Y задані законами розподілу відповідно $\begin{array}{c|cc} X & x_1 & x_2 \\ \hline P & p_1 & p_2 \end{array}$, $\begin{array}{c|ccc} Y & y_1 & y_2 & y_3 \\ \hline P & g_1 & g_2 & g_3 \end{array}$, то за означенням випадкова величина $X + Y$ описується законом розподілу

$X+Y$	x_1+y_1	x_1+y_2	x_1+y_3	x_2+y_1	x_2+y_2	x_2+y_3
P	p_1g_1	p_1g_2	p_1g_3	p_2g_1	p_2g_2	p_2g_3

Підставивши дані, отримуємо:

$X+Y$	$2+1$	$2+3$	$2+6$	$5+1$	$5+3$	$5+6$
P	$0{,}8\cdot 0{,}3$	$0{,}8\cdot 0{,}5$	$0{,}8\cdot 0{,}2$	$0{,}2\cdot 0{,}3$	$0{,}2\cdot 0{,}5$	$0{,}2\cdot 0{,}2$

Можливі значення 2 + 6 та 5 + 3 рівні, тому відповідні імовірності 0,16 та 0,1 додаємо, залишивши в остаточному законі розподілу $X+Y$ одне можливе значення 8:

$X+Y$	3	5	6	8	11
P	0,24	0,4	0,06	0,26	0,04

●

Визначимо **добуток незалежних випадкових величин** X та Y як випадкову величину XY, можливі значення якої рівні добуткам кожного можливого значення X на кожне можливе значення Y; імовірності можливих значень добутку XY дорівнюють добуткам імовірностей можливих значень співмножників. У випадку рівності добутків x_iy_j імовірність можливого значення XY рівна сумі відповідних імовірностей.

Наприклад, якщо X та Y визначені розподілами задачі 4.8, то закон розподілу XY має такий вид:

XY	2	5	6	12	15	30
P	0,24	0,06	0,4	0,16	0,1	0,04

Приклади знаходження законів розподілу випадкових величин

Задача 4.9. Імовірність того, що виготовлений виріб вимагає додаткового регулювання, дорівнює p. Контролер перевіряє якість партії виробів, навмання вибираючи виріб. Якщо він вимагає додаткового регулювання, то наступні випробування припиняються, а вся партія відправляється на доробку. Якщо ж виріб стандартний, то контролер бере наступний виріб, тощо. Згідно із інструкцією контролер перевіряє не більше п'яти виробів. 1) Скласти закон розподілу числа перевірених контролером виробів. 2) Знайти імовірність доробки всієї партії виробів.

○ 1) Позначимо: X – число виробів, перевірених контролером, A_i – i-тий відібраний виріб вимагає додаткового регулювання ($i=\overline{1,5}$). За умовою $P(A_i)=p$, $i=\overline{1,5}$. Тоді $\overline{A_i}$ – i-тий виріб стандартний. $P(\overline{A_i})=1-p=q$, $i=\overline{1,5}$.

Можливі значення X: 1, 2, 3, 4, 5. Знайдемо імовірності, з якими X набирає ці значення. Випадкова подія (X=1) відбудеться тоді, коли перший відібраний прилад вимагатиме додатково регулювання, тобто (X=1)=A_1, звідки $p_1=P(X=1)=P(A_1)=p$. Відбуття події (X=2) означає, що перший виріб стандартний і другий вимагає регулювання: $(X=2)=\overline{A_1}\cdot A_2$. Використовуючи теорему множення імовірностей, отримаємо: $p_2=P(X=2)=P(\overline{A_1}\cdot A_2)=qp$. Аналогічно знаходимо, що $p_3=P(X=3)=q^2p$, $p_4=P(X=4)=q^3p$.

Нарешті, подія (X=5) відбудеться або тоді, коли чотири перші вироби стандартні, а п'ятий вимагає регулювання (партія відправляється на доробку), або коли всі п'ять виробів стандартні (партія пропускається), тобто

$$(X=5)=\overline{A_1}\overline{A_2}\overline{A_3}\overline{A_4}A_5+\overline{A_1}\overline{A_2}\overline{A_3}\overline{A_4}\overline{A_5},$$

звідки $p_5=P(X=5)=q^4p+q^5=q^4(p+q)=q^4$, бо $p+q=1$.

Остаточно шуканий закон розподілу має такий вид:

X	1	2	3	4	5
P	p	qp	q^2p	q^3p	q^4

Для перевірки з'ясуємо, чи виконується умова нормування:

$$\sum_1^5 p_i=p+qp+q^2p+q^3p+q^4=p+qp+q^2p+q^3(p+q)=$$

$$=p+qp+q^2(p+q)=p+q(p+q)=p+q=1.$$

2) Партія пропускається контролером в єдиному випадку, коли всі п'ять перевірених виробів є стандартними, імовірність цієї випадкової події дорівнює q^5. Тому імовірність протилежної події (партія виробів відправляється на доробку) дорівнює $1-q^5$ або $1-(1-p)^5$. ●

Задача 4.10. В урні знаходяться шість більярдних куль з номерами від 1 до 6. Скласти закон розподілу випадкової величини X – суми номерів двох навмання витягнутих куль.

○ Нехай X – сума номерів двох навмання взятих більярдних куль. З'ясуємо, які можливі значення величини X. Куля з номером 1 може утворити такі суми з номерами решти куль: 3, 4, 5, 6, 7. Куля з номером 2 – 5, 6, 7, 8 (при цьому 2 сумується тільки із більшими номерами, бо сума з 1 вже врахована); куля з номером 3 – 7, 8, 9; з 4 – 9, 10 і з 5 – 11. Отже, можливі значення X – це цілі числа від 3 до 11 включно.

Знайдемо відповідні імовірності, користуючись класичним означенням імовірності. Випробування – відбір двох куль. Порядок розташування куль у двійці відібраних несуттєвий, бо номери куль додаються. Тому загальне число наслідків випробування $n = C_6^2 = 15$. Кожній із випадкових подій (X=3), (X=4), (X=10), (X=11) сприяє тільки один наслідок випробування, тобто їх імовірності дорівнюють 1/15. Кожній із подій (X=5), (X=6), (X=8), (X=9) сприяють вже два наслідки випробування, тому їх імовірності – 2/15. Нарешті, події (X=7) сприяють три наслідки, і P(X=7)=3/15=1/5.

Таким, чином, шуканий закон розподілу має такий вид:

X	3	4	5	6	7	8	9	10	11
P	1/15	1/15	2/15	2/15	3/15	2/15	2/15	1/15	1/15

Очевидно, що умова нормованості виконується. ●

Задача 4.11. Дослідження роботи блоків чотирьох типів в умовах перепаду напруги показало, що імовірність безвідмовної роботи протягом часу T кожного із типів складає відповідно 0,8; 0,9; 0,7 та 0,75. Відбираються чотири блоки кожного типу. Скласти закон розподілу випадкової величини X – числа блоків, які безвідмовно працюватимуть у вказаних умовах.

○ Задачу можна розв’язати не одним способом.

Перший спосіб.

Нехай, події A_i – блок i-го типу працює безвідмовно, $i = \overline{1,4}$,

$\overline{A_i}$ – блок i-го типу вийшов з ладу.

За умовою $P(A_1)$=0,8, $P(A_2)$=0,9, $P(A_3)$=0,7, $P(A_4)$=0,75, $P(\overline{A_1}) = 0{,}2$, $P(\overline{A_2}) = 0{,}1$, $P(\overline{A_3}) = 0{,}3$, $P(\overline{A_4}) = 0{,}25$.

Можливі значення X – 0, 1, 2, 3, 4.

Знайдемо відповідні імовірності:

$P(X = 0) = P(\overline{A_1}\,\overline{A_2}\,\overline{A_3}\,\overline{A_4}) = 0{,}2 \cdot 0{,}1 \cdot 0{,}3 \cdot 0{,}25 = 0{,}0015$;

$P(X=1)=P(A_1\overline{A}_2\overline{A}_3\overline{A}_4+\overline{A}_1A_2\overline{A}_3\overline{A}_4+\overline{A}_1\overline{A}_2A_3\overline{A}_4+\overline{A}_1\overline{A}_2\overline{A}_3A_4)=$
$0{,}8\cdot 0{,}1\cdot 0{,}3\cdot 0{,}25+0{,}2\cdot 0{,}9\cdot 0{,}3\cdot 0{,}25+0{,}2\cdot 0{,}1\cdot 0{,}7\cdot 0{,}25+0{,}2\cdot 0{,}1\cdot 0{,}3\cdot 0{,}75=$
$=0{,}006+0{,}0135+0{,}0035+0{,}0045=0{,}0275;$
$P(X=2)=P(A_1A_2\overline{A}_3\overline{A}_4+A_1\overline{A}_2A_3\overline{A}_4+A_1\overline{A}_2\overline{A}_3A_4+\overline{A}_1A_2A_3\overline{A}_4+\overline{A}_1A_2\overline{A}_3A_4+$
$\overline{A}_1\overline{A}_2A_3A_4)=0{,}1685;$
$P(X=3)=P(A_1A_2A_3\overline{A}_4+A_1A_2\overline{A}_3A_4+A_1\overline{A}_2A_3A_4+\overline{A}_1A_2A_3A_4)=0{,}4245;$
$P(X=4)=P(A_1A_2A_3A_4)=0{,}378$

Остаточно отримаємо шуканий закон розподілу:

X	0	1	2	3	4
P	0,0015	0,0275	0,1685	0,4245	0,378

. (*)

Другий спосіб. Позначимо випадкові величини: X_k – число блоків k-го типу, які безвідмовно працюватимуть в умовах перепаду напруги ($k=\overline{1,4}$).

Оскільки тип представлений тільки одним блоком, то X_k може набирати два можливих значення: 0 (блок вийшов з ладу) і 1 (блок безвідмовно працюватиме). Із врахуванням умови задачі отримаємо закони розподілу.

X_1	0	1
P	0,2	0,8

,

X_2	0	1
P	0,1	0,9

,

X_3	0	1
P	0,3	0,7

,

X_4	0	1
P	0,25	0,75

.

Число (**загальне**) блоків, які безвідмовно працюватимуть у вказаних умовах, складається із числа блоків кожного типу, що витримають випробування:

$X=X_1+X_2+X_3+X_4$.

Позначимо $Y=X_1+X_2$; $Z=Y+X_3$; $X=Z+X_4$.
За аналогією із розв'язуванням задачі 4.8 послідовно отримаємо:

$Y=X_1+X_2$	0	1	2
P	0,02	0,26	0,72

,

$Z=Y+X_3$	0	1	2	3
P	0,006	0,092	0,398	0,504

і, нарешті, закон розподілу для $X=Z+X_4$, який співпадає із (*).

Третій спосіб. Закон розподілу для X можна отримати ще й чисто механічно, перемноживши двочлени (біноми):
$\Upsilon_4(z)=(0{,}8z+0{,}2)(0{,}9z+0{,}1)(0{,}7z+0{,}3)(0{,}75z+0{,}25)$.

В результаті отримаємо
$\Upsilon_4(z)=0{,}378z^4+0{,}4245z^3+0{,}1685z^2+0{,}0275z+0{,}0015$.

При цьому кожний із п'яти коефіцієнтів при z^k (k=0,1,2,3,4) дорівнює імовірності $P(X=k)$ (порівняйте із (*)). ●

Зауваження. Функція

$$\Upsilon_n(z)=\prod_{i=1}^{n}(p_i z+q_i)=(p_1 z+q_1)(p_2 z+q_2)\cdot\ldots\cdot(p_n z+q_n)\ ,$$

коефіцієнти розкладу якої по степеням z^k(k=0,1,2,…,n) дорівнюють імовірностям того, що при випробуванні випадкова величина X набере можливе значення k, називається **твірною функцією величини *X***. Твірна функція використовується також при знаходженні імовірностей $P_n(m)$ для n повторних випробувань таких, що в першому випробуванні випадкова подія A відбувається з імовірністю p_1, в другому – p_2, …, в n-му випробуванні – p_n, а імовірності непояви події A відповідно рівні $q_1, q_2, \ldots, q_n$. В такому випадку $Y_n(z)$ називається **твірною функцією імовірностей $P_n(m)$**.

Задача 4.12. У лотереї, яка проводиться з приводу презентації фірми, на 4000 білетів розігруються п'ять речей, вартість яких складає відповідно 100, 300, 400, 500 та 700 грн. Вартість одного білета – 5 грн. Скласти закон розподілу суми чистого виграшу для особи, яка має два білети.

○ Хай випадкова величина X – сума виграшу особи, яка має два лотерейні білети. Знайдемо можливі значення X. Якщо обидва білети невиграшні, тоді особа втрачає 2·5=10, тобто виграш складає –10. Нехай один білет невиграшний, а другий виграшний. Тоді із врахуванням вартості білетів можливі варіанти виграшу: 90, 290, 390, 490 та 690. Нарешті, обидва білети можуть бути виграшні, при цьому можливі варіанти: 100+300-10; 100+400-10; 100+500-10; 100+700-10; 300+400-10; 300+500-10; 300+700-10; 400+500-10; 400+700-10; 500+700-10.

Отже, можливі значення X в порядку розглянутих варіантів такі: а) –10; б)90, 290, 390, 490, 690; в) 390, 490, 590, 790, 690, 790, 990, 890, 1090, 1190.

Відмітимо, що серед них є співпадаючі, а також порушений порядок зростання. Поділ на групи зумовлений тим, що в межах кожної групи імовірності того, що X набере можливе значення, рівні.

Випробування полягає у відборі (придбанні) двох білетів. Всіх наслідків випробування є $C_{2000}^{2}=1999000$. Невиграшних білетів є 1995. Тому

$$P(X=-10)=C_{1995}^{2}\Big/C_{2000}^{2}=1989015\Big/C_{2000}^{2}.$$

Кожне із можливих значень групи б) набирається величиною X тоді, коли один білет буде виграшним із конкретною (однією) вартістю виграшу, а другий – невиграшний. Тому

$$P(X=90)=P(X=290)=P(X=390)=P(X=490)=P(X=690)=$$
$$=C_{1}^{1}\cdot C_{1995}^{1}\Big/C_{2000}^{2}=1995\Big/C_{2000}^{2}.$$

Кожне із можливих значень групи в) величина X набирає, коли обидва білети будуть виграшними, причому пара виграшів буде єдиною (за виключенням можливого значення 790, яке двічі повторюється). Тому

$$P(X=390)=P(X=490)=P(X=590)=P(X=690)=P(X=890)=$$
$$=P(X=990)=P(X=1090)=P(X=1190)=1\Big/C_{2000}^{2}.$$

$$P(X=790)=2\Big/C_{2000}^{2}.$$

Тепер врахуємо те, що можливі значення 390, 490, 690 зустрічаються в групах б) та в). Використавши теорему додавання імовірностей, отримаємо:

$$P(X=390)=1995\Big/C_{2000}^{2}+1\Big/C_{2000}^{2}=1996\Big/C_{2000}^{2},$$

$$P(X=490)=P(X=490)=1996\Big/C_{2000}^{2}.$$

Таким чином, шуканий закон розподілу імовірностей випадкової величини X має такий вид:

X	-10	90	290	390	490	590	690	790	890	990	1090	1190
P	$\frac{1989015}{C_{2000}^{2}}$	$\frac{1995}{C_{2000}^{2}}$	$\frac{1995}{C_{2000}^{2}}$	$\frac{1996}{C_{2000}^{2}}$	$\frac{1996}{C_{2000}^{2}}$	$\frac{1}{C_{2000}^{2}}$	$\frac{1996}{C_{2000}^{2}}$	$\frac{2}{C_{2000}^{2}}$	$\frac{1}{C_{2000}^{2}}$	$\frac{1}{C_{2000}^{2}}$	$\frac{1}{C_{2000}^{2}}$	$\frac{1}{C_{2000}^{2}}$

Додавши всі імовірності і врахувавши, що $C_{2000}^{2}=1999000$, можна переконатися у виконанні умови нормування. ●

Зауваження. Розглянемо можливість використання для попередньої задачі другого методу розв’язування задачі 4.11. Тобто, нехай X_1 – величина чистого виграшу на 1-й білет, а X_2 – на 2-й. Тоді природно очікувати, що загальний виграш на два білети буде складатися із виграшу на 1-й та 2-й білети, тобто $X=X_1+X_2$. Для задачі 4.11 величини X_1 та X_2 були незалежними. Закон розподілу імовірностей для величини X_1 має такий вид:

X_1	-5	95	295	395	495	695
P	$\frac{1995}{2000}$	$\frac{1}{2000}$	$\frac{1}{2000}$	$\frac{1}{2000}$	$\frac{1}{2000}$	$\frac{1}{2000}$

Спроба ж записати закон розподілу для X_2 наштовхується на труднощі: він залежатиме (буде змінюватися) від того, яке можливе значення набрала величина X_1. Проаналізуйте два випадки: ($X_1 = -5$) та (X_1=295).

Задача 4.13. В першій скрині міститься 4 білих і 6 чорних куль, в другій – 7 білих і 3 чорні кулі. Із другої скрині навмання беруть дві кулі і перекладають в першу, перемішують вмістиме, а потім навмання беруть одну кулю і перекладають в другу скриню. Скласти закон розподілу числа чорних куль в першій та у другій скринях.

○ Знайдемо закон розподілу випадкової величини X – числа чорних куль у другій скрині. Позначимо випадкові події: A_i – i-та взята куля з ІІ-ої скрині чорна (i=1,2), B – взята куля з І-ої урни куля чорна (і перекладена в ІІ-у). Найменше можливе значення випадкової величини X отримається тоді, коли з ІІ-ої урни взято дві чорні кулі і з І-ої перекладено білу, тобто відбувається випадкова подія $A_1A_2\overline{B}$, де $\overline{B}$ – біла куля з І-ої урни. В такому випадку X набере можливого (найменшого) значення 1. Отже, враховуючи залежність подій A_1, A_2 та $\overline{B}$, отримаємо:

$$P(X=1) = P(A_1A_2\overline{B}) = P(A_1)P_{A_1}(A_2)P_{A_1A_2}(\overline{B}) = \frac{3}{10}\cdot\frac{2}{9}\cdot\frac{4}{12} = \frac{1}{45}.$$

Випадкова величина X набирає значення 2 в кожному із трьох випадків:

1) з ІІ-ої скрині взято першу кулю чорну, другу – білу і покладено з І-ої білу ($A_1\overline{A}_2\overline{B}$);
2) з ІІ-ої скрині взято першу білу кулю, другу – чорну і покладено з І-ої білу ($\overline{A}_1A_2\overline{B}$);
3) з ІІ-ої скрині взято обидві кулі чорні і покладено з І-ої чорну (A_1A_2B).

Тому

$$P(X=2) = P(A_1\overline{A}_2\overline{B} + \overline{A}_1A_2\overline{B} + A_1A_2B) =$$

$$= P(A_1)P_{A_1}(\overline{A}_2)P_{A_1\overline{A}_2}(\overline{B}) + P(\overline{A}_1)P_{\overline{A}_1}(A_2)P_{\overline{A}_1A_2}(\overline{B}) + P(A_1)P_{A_1}(A_2)P_{A_1A_2}(B) =$$

$$= \frac{3}{10}\cdot\frac{7}{9}\cdot\frac{5}{12} + \frac{7}{10}\cdot\frac{3}{9}\cdot\frac{5}{12} + \frac{3}{10}\cdot\frac{2}{9}\cdot\frac{8}{12} = \frac{7}{12} + \frac{7}{12} + \frac{2}{45} = \frac{43}{180}.$$

Міркуючи аналогічно, отримаємо:

$$P(X=3)=P(A_1\overline{A}_2B+\overline{A}_1A_2B+\overline{A}_1\overline{A}_2\overline{B})=$$
$$=\frac{3}{10}\cdot\frac{7}{9}\cdot\frac{7}{12}+\frac{7}{10}\cdot\frac{3}{9}\cdot\frac{7}{12}+\frac{7}{10}\cdot\frac{6}{9}\cdot\frac{6}{12}=\frac{49}{360}+\frac{49}{360}+\frac{7}{30}=\frac{91}{180};$$
$$P(X=4)=P(\overline{A}_1\overline{A}_2B)=\frac{7}{10}\cdot\frac{6}{9}\cdot\frac{6}{12}=\frac{7}{30}.$$

Отже, шуканий закон розподілу для X має такий вигляд:

X	1	2	3	4
P	$\frac{1}{45}$	$\frac{43}{180}$	$\frac{91}{180}$	$\frac{7}{30}$

Оскільки

$$\sum_{i=1}^{4}p_i=\frac{1}{45}+\frac{43}{180}+\frac{91}{180}+\frac{7}{30}=1,$$

то умова нормованості виконується.

Нехай Y – число чорних куль в першій скрині. Попередні міркування можна було б використати для знаходження закону розподілу Y.

Проте всі виконані процедури залишають незмінним сумарне число чорних куль обох скринь, тобто виконується рівність $X+Y=9$. Тому $Y=9-X$.

Якщо X набирає значення 1, то Y набуватиме можливе значення 8, при цьому випадкові події ($X=1$) та ($Y=8$) є рівносильними, тобто $P(X=1)=P(Y=8)=1/45$.

Аналогічно отримуємо імовірності відбуття події ($Y=7$),($Y=6$) та ($Y=5$).

Отже, закон розподілу випадкової величини Y задається такою таблицею:

Y	5	6	7	8
P	$\frac{7}{30}$	$\frac{91}{180}$	$\frac{43}{180}$	$\frac{1}{45}$

●

Числові характеристики дискретних випадкових величин та їх властивості

Закон розподілу імовірностей дає повну інформацію про дискретну випадкову величину. Проте в багатьох практично важливих випадках економісту-досліднику буває достатньо знати одне або кілька чисел, пов’язаних із випадковою величиною, які дають менш повне, але більш наочне описання випадкової величини. Такі числа, які сумарно описують випадкову величину, називаються її **числовими характеристиками**.

Математичне сподівання

Математичним сподіванням дискретної випадкової величини X називається сума добутків всіх її можливих значень на відповідні імовірності:

$$M(X) = x_1p_1 + x_2p_2 + \ldots + x_np_n. \tag{4.9}$$

Якщо величина X може набирати зчисленну множину значень (наприклад, як у випадку геометричного або пуассонівського розподілу), то при умові, що цей ряд абсолютно збіжний,

$$M(X) = \sum_{i=1}^{\infty} x_i p_i. \tag{4.10}$$

Імовірнісний зміст математичного сподівання: ***M*(*X*) приблизно дорівнює** (тим точніше, чим більше число випробувань) **середньому арифметичному спостережених значень випадкової величини**.

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 двох випадкових величин дорівнює сумі математичних сподівань доданків**: $M(X + Y) = M(X) + M(Y)$.

Наслідок. Математичне сподівання суми кількох випадкових величин дорівнює сумі математичних сподівань доданків.

4. **Математичне сподівання добутку двох незалежних випадкових величин дорівнює добутку їх математичних сподівань**: $M(XY) = M(X) \cdot M(Y)$.

Наслідок. Математичне сподівання добутку кількох незалежних у сукупності випадкових величин дорівнює добутку їх математичних сподівань.

Задача 4.14. Існує три методи експрес-контролю партії виробів. Кожний із них безпомилково виявляє якісні вироби, а при ідентифікації некондиційних допускає помилки, тобто вони визнаються якісними. Число таких виробів для кожного із методів є відповідно випадковими величинами X, Y та Z, закони розподілу яких мають такий вид:

$$\begin{array}{c|cccc} X & 0 & 1 & 3 & 4 \\ \hline P & 0{,}5 & 0{,}4 & 0{,}06 & 0{,}04 \end{array}, \quad \begin{array}{c|cccc} Y & 0 & 1 & 2 & 3 \\ \hline P & 0{,}6 & 0{,}2 & 0{,}1 & 0{,}1 \end{array}, \quad \begin{array}{c|ccccc} Z & 0 & 1 & 2 & 3 & 4 \\ \hline P & 0{,}8 & 0{,}1 & 0{,}04 & 0{,}03 & 0{,}03 \end{array}.$$

Який із методів експрес-контролю кращий?

○ Нехай a – кількість стандартних виробів у партії. Тоді згідно з умовою задачі кожен із методів контролю відповідно дасть $a + X$, $a + Y$, $a + Z$ стандартних деталей після перевірки всієї партії, де кожний другий доданок у сумах – це помилка при ідентифікації бракованих виробів. Мінімальне значення її (у середньому) визначить кращий метод. Характеристикою середнього значення є математичне сподівання. Тому знайдемо

$$M(X) = 0 \cdot 0{,}5 + 1 \cdot 0{,}4 + 3 \cdot 0{,}06 + 4 \cdot 0{,}04 = 0{,}74;$$
$$M(Y) = 0 \cdot 0{,}6 + 1 \cdot 0{,}2 + 2 \cdot 0{,}1 + 3 \cdot 0{,}1 = 0{,}7;$$
$$M(Z) = 0 \cdot 0{,}8 + 1 \cdot 0{,}1 + 2 \cdot 0{,}04 + 3 \cdot 0{,}03 + 4 \cdot 0{,}03 = 0{,}39.$$

Отже, кращим методом експрес-контролю є третій. ●

Дисперсія

Для оцінки розкиду можливих значень випадкової величини навколо середнього значення (математичного сподівання) використовують, зокрема, числову характеристику, яку називають **дисперсією**.

Дисперсією (розкидом) випадкової величини X називається математичне сподівання квадрату відхилення X від $M(X)$:

$$D(X) = M[X - M(X)]^2. \qquad (4.11)$$

Згідно із цим означенням дисперсія характеризує середню величину розкиду можливих значень випадкової величини навколо її матсподівання (середньої) в квадратних одиницях.

Якщо X розподілена за законом (4.1), то з урахуванням закону (4.8) випадкова величина $[X - M(X)]^2$ має такий закон розподілу:

$[X-M(X)]^2$	$[x_1-M(X)]^2$	$[x_2-M(X)]^2$	...	$[x_n-M(X)]^2$
P	p_1	p_2	...	p_n

Використавши означення дисперсії та математичного сподівання, отримаємо формулу для обчислення $D(X)$:

$$D(X) = M[X-M(X)]^2 = [x_1 - M(X)]^2 p_1 + [x_2 - M(X)]^2 p_2 + \\ + \ldots + [x_n - M(X)]^2 p_n = \sum_{i=1}^{n} [x_i - M(X)]^2 p_i. \qquad (4.11^*)$$

Зменшення обсягу обчислень досягається за рахунок використання **розрахункової формули** для обчислення дисперсії:

$$D(X) = M(X^2) - [M(X)]^2. \qquad (4.12)$$

Застосувавши означення математичного сподівання до закону розподілу (4.8), отримаємо **числову реалізацію розрахункової формули**:

$$D(X) = \sum_{i=1}^{n} x_i^2 p_i - [M(X)]^2. \qquad (4.12^*)$$

Властивості дисперсії.

1. **Дисперсія сталої дорівнює нулю**: $D(C) = 0$.

2. **Сталий множник можна виносити за знак дисперсії, піднісши його до квадрату**: $D(CX) = C^2 D(X)$.

3. **Дисперсія суми двох незалежних випадкових величин дорівнює сумі дисперсій цих величин**: $D(X + Y) = D(X) + D(Y)$.

Наслідок. Дисперсія суми кількох незалежних у сукупності випадкових величин дорівнює сумі дисперсій цих величин.

4. **Дисперсія різниці двох незалежних випадкових величин дорівнює сумі їх дисперсій**: $D(X - Y) = D(X) + D(Y)$.

Зауваження. Навіть якщо X та Y незалежні, то в загальному випадку виконується нерівність $D(XY) \neq D(X)\,D(Y)$.

Середнє квадратичне відхилення

Незручність використання дисперсії в деяких випадках зумовлена тим, що вона має розмірність, яка дорівнює квадрату розмірності випадкової величини. Наприклад, якщо можливі значення X вимірюються в кг, то $D(X)$ в (кг)^2.

Середнім квадратичним відхиленням випадкової величини X називається квадратний корінь із дисперсії:

$$\sigma(X) = \sqrt{D(X)}. \qquad (4.13)$$

$\sigma(X)$ характеризує середню величину розкиду можливих значень X навколо $M(X)$ (середньої) в лінійних одиницях.

Розглянемо типові задачі.

Задача 4.15. Знайти середнє квадратичне відхилення випадкової величини $Z = 2X - 5Y - 30$, якщо X та Y – незалежні випадкові величини, $D(X) = 0{,}25$, $D(Y) = 0{,}2$.

○ Рівність випадкових величин зумовлює рівність їх дисперсій: $D(Z) = D(2X - 5Y - 30)$. Використання властивостей дисперсії та умов задачі дає такий ланцюжок рівностей:

$$D(Z) = D(2X + (-5Y) + (-30)) = D(2X) + D(-5Y) + D(-30) =$$

$= 2^2 D(X) + (-5)^2 D(Y) + 0 = 4 \cdot 0{,}25 + 25 \cdot 0{,}2 = 6.$

Тоді згідно із (4.13) $\sigma(Z) = \sqrt{D(Z)} = \sqrt{6} \approx 2{,}45.$ ●

Задача 4.16. Знайти закон розподілу дискретної випадкової величини на основі числових характеристик цієї величини та певної інформації про закон її розподілу:

а) $\begin{array}{c|ccc} X & 3 & 5 & x_3 \\ \hline P & 0{,}1 & 0{,}6 & p_3 \end{array}$, $M(X) = 5{,}7;$

б) $\begin{array}{c|ccc} Y & -2 & 1 & 2 \\ \hline P & p_1 & p_2 & p_3 \end{array}$, $M(Y) = 0{,}9$ $M(Y^2) = 3{,}1;$

в) $\begin{array}{c|cc} Z & z_1 & z_2 \\ \hline P & 0{,}8 & p_2 \end{array}$, $M(Z) = 1{,}2$ $\sigma(Z) = 0{,}4.$

○ а) Сума імовірностей в законі розподілу дорівнює одиниці, тому $0{,}1 + 0{,}6 + p_3 = 1$, звідки $p_3 = 0{,}3$. Згідно з означенням (4.9) математичного сподівання і умовою задачі $M(X) = 3 \cdot 0{,}1 + 5 \cdot 0{,}6 + x_3 \cdot 0{,}3 = 5{,}7$ або $0{,}3x_3 = 4$, звідки $x_3 = 40/3$. Знайдений закон розподілу випадкової величини має такий вид: $\begin{array}{c|ccc} X & 3 & 5 & 40/3 \\ \hline P & 0{,}1 & 0{,}6 & 0{,}3 \end{array}$.

б) За аналогією із випадком а) і використавши рівність $M(Y^2) = y_1^2 p_1 + y_2^2 p_2 + y_3^2 p_3$, отримаємо таку систему рівнянь відносно невідомих p_1, p_2, p_3:

$$\begin{cases} p_1 + p_2 + p_3 = 1, \\ -2p_1 + p_2 + 2p_3 = 0{,}9, \\ 4p_1 + p_2 + 4p_3 = 3{,}1. \end{cases}$$

Перше рівняння розв'яжемо відносно p_1, а друге помножимо на 2 і додамо до третього. В результаті система набуде такого виду:

$$\begin{cases} p_1 = 1 - p_2 - p_3, \\ -2p_1 + p_2 + 2p_3 = 0{,}9, \\ 3p_2 + 8p_3 = 4{,}9. \end{cases} \Leftrightarrow \begin{cases} p_1 = 1 - p_2 - p_3, \\ 3p_2 + 4p_3 = 2{,}9, \\ 3p_2 + 8p_3 = 4{,}9. \end{cases} \Leftrightarrow \begin{cases} p_1 = 1 - p_2 - p_3, \\ 3p_2 + 4p_3 = 2{,}9, \\ 4p_3 = 2. \end{cases}$$

Остаточно $p_1 = 0{,}2$, $p_2 = 0{,}3$, $p_3 = 0{,}5$.

в) Оскільки $p_2 = 1 - 0{,}8 = 0{,}2$, то для знаходження двох невідомих z_1 і z_2 складемо систему рівнянь:

$$\begin{cases} 0{,}8z_1 + 0{,}2z_2 = 1{,}2, \\ \sqrt{0{,}8z_1^2 + 0{,}2z_2^2 - (1{,}2)^2} = 0{,}4. \end{cases} \Leftrightarrow \begin{cases} z_2 = 6 - 4z_1, \\ 0{,}8z_1^2 + 0{,}2z_2^2 = 1{,}6. \end{cases} \Leftrightarrow \begin{cases} z_2 = 6 - 4z_1, \\ z_1^2 - 2{,}4z_1 + 1{,}4 = 0. \end{cases}$$

Звідки $z_1^{(1)} = 1,\ z_1^{(2)} = 1{,}4,\ z_2^{(1)} = 2,\ z_2^{(2)} = 0{,}4.$

Отже, умовам задачі задовольняють два закони розподілу імовірностей:

$$\begin{array}{c|cc} Z^{(1)} & 1 & 2 \\ \hline P & 0{,}8 & 0{,}2 \end{array};\quad \begin{array}{c|cc} Z^{(2)} & 1{,}4 & 0{,}4 \\ \hline P & 0{,}8 & 0{,}2 \end{array}. \quad \bullet$$

Задача 4.17. Маршрут руху інкасаторського автомобіля від банку до його філії перетинає п'ять перехресть, регульованих світлофорами, які з імовірностями відповідно 0,2, 0,5, 0,8, 0,4, 0,5 дозволяють рух без зупинки. Знайти середнє число зупинок спецавтомобіля на перехрестях по цьому маршруту, якщо світлофори працюють незалежно один від інших (відсутня «зелена лінія»).

○ Позначимо: X – число зупинок автомобіля на перехрестях по всьому маршруту, X_i – число зупинок при проходженні i-ого перехрестя, де $i = 1, 2, ..., 5$. Тоді загальне число зупинок буде складатися із числа зупинок на кожному світлофорі: $X = X_1 + X_2 + X_3 + X_4 + X_5$. За властивістю математичного сподівання $M(X) = \sum_{i=1}^{5} M(X_i)$, при цьому $M(X)$ характеризує середнє число зупинок по всьому маршруту. Можливі значення X_1: 0, 1. Випадкова подія $(X_1 = 1)$ полягає в тому, що на першому світлофорі автомобіль зупинився, а $(X_1 = 0)$ – не зупинився. За умовою задачі $P(X_1 = 0) = 0{,}2$, тоді $P(X_1 = 1) = 1 - 0{,}2 = 0{,}8$. Отже, закон розподілу випадкової величини X_1 має такий вид: $\begin{array}{c|cc} X_1 & 0 & 1 \\ \hline P & 0{,}2 & 0{,}8 \end{array}$, звідки $M(X_1) = 0 \cdot 0{,}2 + 1 \cdot 0{,}8 = 0{,}8$. Аналогічно знаходимо: $M(X_2) = 0{,}5$, $M(X_3) = 0{,}2$, $M(X_4) = 0{,}6$, $M(X_5) = 0{,}5$. Остаточно $M(X) = 0{,}8 + 0{,}5 + 0{,}2 + 0{,}6 + 0{,}5 = 2{,}6$. ●

Задача 4.18. Автоматична лінія після сервісної наладки випускає стандартні вироби з імовірністю 0,98. Чергова наладка здійснюється після появи першого нестандартного виробу. Знайти середнє число виробів, виготовлених між двома черговими сервісними наладками лінії.

○ Нехай випадкова величина X – число виробів, виготовлених лінією між її двома черговими сервісними наладками. Всі ці вироби стандартні за виключенням останнього нестандартного, поява якого і зумовлює здійснення чергової наладки при умові, що перший виріб не виявився бракованим. Позначимо: A – поява нестандартного виробу. За умовою задачі $p = P(A) = 1 - 0{,}98 = 0{,}02$ залишається незмінною для кожного із випробувань (виготовлення виробу автоматичною лінією). Тоді можна стверджувати (див. задачу 4.5), що X розподілена за геометричним законом і її закон розподілу має такий вид:

$$\begin{array}{c|cccccc} X & 1 & 2 & 3 & \ldots & n & \ldots \\ \hline P & p & qp & q^2 p & \ldots & q^{n-1} p & \ldots \end{array},$$

де $p = 0{,}02$, $q = 0{,}98$.

Шукане середнє число виробів, виготовлених між двома черговими наладками лінії, дорівнює $M(X)$. Використовуючи формулу (4.10), а також те, що $1 + q + q^2 + \ldots + q^n + \ldots = 1/(1-q)$, отримаємо

$$M(X) = p + 2qp + 3q^2 p + \ldots + nq^{n-1} p + \ldots = p\sum_{n=1}^{\infty} nq^{n-1} =$$

$$= p\left(\frac{d}{dx}\sum_{n=1}^{\infty} x^n\right)\Bigg|_{x=q} = p\left(\frac{d}{dx}\left(\frac{1}{1-x}\right)\right)\Bigg|_{x=q} = p \cdot \frac{1}{(1-q)^2} = \frac{1}{p} = \frac{1}{0{,}02} = 50.$$

Відмітимо, що правомірність почленного диференціювання степеневого ряду забезпечується його рівномірною збіжністю, оскільки він є сумою нескінченної геометричної прогресії із знаменником $q = 0{,}98 < 1$. ●

Початкові та центральні моменти

Узагальненням розглянутих вище основних числових характеристик випадкової величини, які описують середину розподілу (математичне сподівання) і розкид (дисперсія та середнє квадратичне відхилення), є поняття моментів випадкової величини. В теорії імовірностей розглядаються моменти двох видів: початкові та центральні.

Початковим моментом порядку *k* дискретної випадкової величини X називається математичне сподівання величини X^k:

$$\nu_k = M(X^k) = \sum x_i^k p_i. \tag{4.14}$$

Зокрема $\nu_1 = M(X)$, $\nu_2 = M(X^2)$, тому розрахункова формула для обчислення дисперсії набирає такого виду:

$$D(X) = \nu_2 - \nu_1^2.$$

Крім моментів самої величини X доцільно (особливо для неперервної величини) розглядати моменти відхилення $X - M(X)$.

Центральним моментом порядку *k* називається математичне сподівання величини $(X - M(X))^k$:

$$\mu_k = M[(X - M(X))^k] = \sum [x_i - M(X)]^k p_i. \qquad (4.15)$$

Зокрема, $\mu_1 = M[X - M(X)] = 0$, $\mu_2 = M[(X - M(X))^2] = D(X)$.

Зв'язки між початковими та центральними моментами отримуються із таких співвідношень:

$$\mu_2 = \nu_2 - \nu_1^2,\ \mu_3 = \nu_3 - 3\nu_2\nu_1 + 2\nu_1^3,\ \mu_4 = \nu_4 - 4\nu_3\nu_1 + 6\nu_2\nu_1^2 - 3\nu_1^4.$$

Моменти більш високих порядків використовуються рідко.

Зауваження. Назва моменту центральним зумовлена розглядом степенів відхилень від $M(X)$ – «центру» можливих значень, а початковим – відхилень від нуля ($M(X) = M(X - 0)$, або «початку».

Задача 4.19. Дискретна випадкова величина X задана законом розподілу

X	1	3
P	0,6	0,4

. Знайти початкові та центральні моменти першого, другого і третього порядків.

○ За формулами (4.14) для k = 1, 2, 3 отримуємо початкові моменти 1-го, 2-го та 3-го порядків:

$$\nu_1 = M(X) = 1 \cdot 0{,}6 + 3 \cdot 0{,}4 = 1{,}8,$$

$$\nu_2 = 1^2 \cdot 0{,}6 + 3^2 \cdot 0{,}4 = 4{,}3,$$

$$\nu_3 = 1^3 \cdot 0{,}6 + 3^3 \cdot 0{,}4 = 11{,}4.$$

Центральні моменти цих порядків знаходимо за допомогою формул (4.15):

$$\mu_1 = M[X - M(X)] = 0,$$

$$\mu_2 = (1 - 1{,}8)^2 \cdot 0{,}6 + (3 - 1{,}8)^2 \cdot 0{,}4 = 0{,}96,$$

$$\mu_3 = (1 - 1{,}8)^3 \cdot 0{,}6 + (3 - 1{,}8)^3 \cdot 0{,}4 = 0{,}384. \quad \bullet$$

Числові характеристики основних законів розподілу

Задача 4.20. Знайти числові характеристики випадкової величини, розподіленої за **біноміальним законом**.

○ Дана випадкова величина X – число появи події A в n повторних незалежних випробуваннях, в кожному з яких $P(A) = p$ (див. задачу 4.3). Нехай X_k – число появ події A в одному k-му випробуванні

$(k = 1,2...,n)$. Тоді $X = X_1 + X_2 + ... + X_n$, а закон розподілу X_k має такий вид:

$$\begin{array}{c|cc} X_k & 0 & 1 \\ \hline P & q & p \end{array}, \quad k = 1,2...,n.$$

Знайдемо числові характеристики величини X_k:

$$M(X) = \sum_{i=1}^{2} x_i p_i = 0 \cdot q + 1 \cdot p = p,$$

$$D(X) = \sum_{i=1}^{n} [x_i - M(X)]^2 p_i = (0-p)^2 q + (1-p)^2 p = p^2 q + q^2 p =$$

$$= pq(p+q) = pq,$$

оскільки $p+q=1$.

Використавши відповідні властивості математичного сподівання і дисперсії (при цьому враховується незалежність випадкових величин X_1, X_2, ..., X_n), отримаємо:

$$M(X) = M(X_1 + X_2 + ... + X_n) = \sum_{k=1}^{n} M(X_k) = \sum_{k=1}^{n} p = np,$$

$$D(X) = D(X_1 + X_2 + ... + X_n) = \sum_{k=1}^{n} D(X_k) = \sum_{k=1}^{n} pq = npq.$$

Отже,

$$M(X) = np, \quad D(X) = npq. \; \bullet \tag{4.16}$$

Використання отриманих формул ілюструє

Задача 4.21. Фірма розглядає питання про обсяг запасів деталей одного типу на пункті гарантійного обслуговування. Статистично встановлено, що імовірність виходу з ладу протягом гарантійного терміну деталі цього типу дорівнює 0,1. Скільки в середньому потрібно мати на пункті деталей такого типу і яка величина розкиду можливих кількостей, якщо в регіоні обслуговування пункту зареєстровано 180 виробів, комплектуючою яких є ця деталь?

○ Нехай A – вихід з ладу деталі протягом гарантійного терміну, випробування – робота деталі протягом цього часу (в складі виробу). Число випробувань $n = 180$. Вони є незалежними, бо $P(A) = 0,1$ для кожного із випробувань. Нехай X – сумарне число деталей, які вийдуть із ладу. Тоді середнє число деталей, які потрібно замінити, визначається математичним сподіванням величини X. Згідно з (4.16) $M(X) = 0,1 \cdot 180 = 18$. Величину розкиду доцільно характеризувати середнім квадратичним відхиленням, оскільки одиниці вимірювання його – лінійні. Тому за другою формулою (4.16) $\sigma(X)=\sqrt{D(X)}=\sqrt{npq}=\sqrt{16{,}2}\approx 4{,}02$. Самостійно перевірте, що найімовірніше число деталей, які вийдуть із ладу за гарантійний термін, також дорівнює 18. ●

Задача 4.22. Знайти математичне сподівання і дисперсію випадкової величини, розподіленої за законом Пуассона.

○ Згідно із означенням величина *X* розподілена за законом Пуассона, якщо вона набирає можливі значення 0, 1, 2, ..., *m*, ... з імовірностями

$$p_{m+1}=P(X=m)=\frac{\lambda^m e^{-\lambda}}{m!},\ m=0,1,2,\ldots\ .$$

Використавши рівності $m!=(m-1)!\cdot m$; $0!=1$ та розклад

$$e^{\lambda}=1+\lambda+\frac{\lambda^2}{2!}+\ldots+\frac{\lambda^m}{m!}+\ldots,$$

знайдемо математичне сподівання:

$$M(X)=\sum_{i=1}^{\infty}x_i p_i=\sum_{m=0}^{\infty}m\cdot\frac{\lambda^m e^{-\lambda}}{m!}=\sum_{m=1}^{\infty}\frac{\lambda^m}{(m-1)!}\cdot e^{-\lambda}=$$

$$=\lambda\cdot e^{-\lambda}\sum_{m=1}^{\infty}\frac{\lambda^{m-1}}{(m-1)!}=\lambda\cdot e^{-\lambda}(1+\lambda+\frac{\lambda^2}{2!}+\ldots)=\lambda\cdot e^{-\lambda}e^{\lambda}=\lambda.$$

Для обчислення дисперсії використаємо розрахункову формулу (4.12): $D(X)=M(X^2)-[M(X)]^2$. Тоді

$$M(X^2)=\sum_{i=1}^{\infty}x_i^2 p_i=\sum_{m=0}^{\infty}m^2\cdot\frac{\lambda^m e^{-\lambda}}{m!}=e^{-\lambda}\sum_{m=1}^{\infty}m\frac{\lambda^m}{(m-1)!}=$$

$$=e^{-\lambda}\sum_{m=1}^{\infty}\frac{[(m-1)+1]}{(m-1)!}\cdot\lambda^m=e^{-\lambda}\sum_{m=2}^{\infty}\frac{\lambda^2\lambda^{m-2}}{(m-2)!}+e^{-\lambda}\sum_{m=1}^{\infty}\frac{\lambda\cdot\lambda^{m-1}}{(m-1)!}=$$

$$=(e^{-\lambda}\lambda^2+e^{-\lambda}\lambda)(1+\lambda+\frac{\lambda^2}{2!}+\ldots)=(\lambda^2+\lambda)\cdot e^{-\lambda}e^{\lambda}=\lambda^2+\lambda;$$

$$D(X)=(\lambda^2+\lambda)-\lambda^2=\lambda\ .$$ ●

Задача 4.23. Знайти дисперсію рівномірно розподіленої випадкової величини.

○ Рівномірний закон розподілу має такий вид:

X	1	2	3	...	5
P	$1/n$	$1/n$	$1/n$	...	$1/n$

.

При знаходженні числових характеристик використаємо формули:

$$\sum_{k=1}^{n} k = \frac{n(n+1)}{2}, \ \sum_{k=1}^{n} k^2 = \frac{n(n+1)(2n+1)}{6}.$$

За формулами (4.9) і (4.12*)

$$M(X) = \sum_{k=1}^{n} x_k p_k = \sum_{k=1}^{n} k \cdot \frac{1}{n} = \frac{1}{n}\sum_{k=1}^{n} k = \frac{n+1}{2},$$

$$D(X) = \sum_{k=1}^{n} x_k^2 p_k - [M(X)]^2 = \sum_{k=1}^{n} k^2 \cdot \frac{1}{n} - (\frac{n+1}{2})^2 = \frac{1}{n} \cdot \frac{n(n+1)(2n+1)}{6} -$$

$$- \frac{(n+1)^2}{4} = (n+1)(\frac{2n+1}{6} - \frac{n+1}{4}) = \frac{n^2-1}{12}. \ \bullet$$

Задача 4.24. Обчислити дисперсію випадкової величини, розподіленої за геометричним законом .

○ Згідно із означенням закон розподілу випадкової величини, геометрично розподіленої, має вид

X	1	2	3	...	n	...
P	p	qp	$q^2 p$	...	$q^{n-1} p$	...

.

В задачі 4.18 отримана формула $M(X) = 1/p$.

Тоді за розрахунковою формулою (4.12)

$$D(X) = \sum_{k=1}^{\infty} x_k^2 p_k - [M(X)]^2 = \sum_{k=1}^{\infty} k^2 q^{k-1} p - \frac{1}{p^2} = p\sum_{k=1}^{\infty} k^2 q^{k-1} - \frac{1}{p^2}.$$

Обчислимо суму отриманого ряду:

$$\sum_{k=1}^{\infty} k^2 q^{k-1} = 1 + 2^2 q + 3^2 q^2 + ... + n^2 q^{n-1} + ... =$$

$$= (x + 2x^2 + 3x^3 + ... + nx^n + ...)'\Big|_{x=q} =$$

$$= \left[x(1 + 2x + 3x^2 + ... + nx^{n-1} + ...)\right]'\Big|_{x=q} =$$

$$= \left[x(x + x^2 + x^3 + ... + x^n + ...)'\right]'\Big|_{x=q} =$$

$$= \left[x\left(\frac{x}{1-x}\right)'\right]'\Big|_{x=q} = \left[\frac{x}{(1-x)^2}\right]'\Big|_{x=q} =$$

$$= \frac{(1-x)^2 + x \cdot 2(1-x)}{(1-x)^4}\Big|_{x=q} = \frac{1+x}{(1-x)^3}\Big|_{x=q} = \frac{1+q}{(1-q)^3} = \frac{1+q}{p^3}.$$

Тоді остаточно отримаємо

$$D(X) = p \cdot \frac{1+q}{p^3} - \frac{1}{p^2} = \frac{1}{p^2}(1+q-1) = \frac{q}{p^2} . \ \bullet$$

ЗАДАЧІ ДЛЯ САМОСТІЙНОГО РОЗВ'ЯЗУВАННЯ

Завдання №5

1. Імовірність випуску бракованої деталі дорівнює 0,01. Контролер перевіряє якість партії деталей, навмання вибираючи деталь. Якщо вона бракована, то наступні випробування припиняються, а вся партія бракується. Якщо ця деталь стандартна, то контролер бере наступну деталь, тощо. Згідно з інструкцією контролер перевіряє не більше чотирьох деталей. Скласти закон розподілу числа перевірених деталей та знайти середнє число цих деталей.

2. Випадкова величина X може набирати два можливих значення: x_1 з імовірністю 0,8 і x_2 – з імовірністю 0,2, при цьому $x_2 > x_1$. Знайти закон розподілу випадкової величини X, якщо відомо, що $M(X) = 2{,}4$, $\sigma(X) = 0{,}8$.

3. Знайти закон розподілу числа появ грані з п'ятьма очками в результаті п'яти кидань гральної кістки, а також числові характеристики цього числа.

4. Спортсмен виконав шість серій пострілів по 10 пострілів в кожній. Імовірність влучання при кожному пострілі дорівнює 0,8. Знайти математичне сподівання дискретної випадкової величини X – числа серій пострілів, в кожній з яких виявиться рівно дев'ять влучань.

5. Знайти математичне сподівання величини $Z = X \cdot Y$, якщо відомо, що X та Y – незалежні випадкові величини, закон розподілу яких:

$$\begin{array}{c|cc} X & 2 & 3 \\ \hline P & \ldots & 0{,}3 \end{array}, \quad \begin{array}{c|cc} Y & 1 & 1{,}5 \\ \hline P & 0{,}8 & \ldots \end{array}.$$

6. У скрині знаходиться п'ять однакових за розміром куль, з яких три – червоного кольору. Навмання відібрано три кулі. Записати закон розподілу числа куль червоного кольору серед відібраних. Знайти дисперсію цього числа куль.

7. Відомі можливі значення дискретної випадкової величини X: $x_1 = 1$, $x_2 = 2$, $x_3 = 3$, а також $M(X) = 2{,}3$; $M(X^2) = 5{,}9$. Знайти закон розподілу величини X і середнє квадратичне відхилення X.

8. Із гвинтівки виконуються постріли по цілі до першого влучання. Імовірність попадання для кожного пострілу дорівнює 0,6. Скласти закон розподілу випадкової величини X – числа витрачених набоїв. Знайти імовірність одного влучання в ціль, якщо в розпорядженні є тільки три набої.

9. Випадкові величини X та Y розподілені за такими законами:

$$\begin{array}{c|cc} X & 2 & x_2 \\ \hline P & 0{,}7 & \ldots \end{array}, \quad \begin{array}{c|cc} Y & -2 & 1 \\ \hline P & \ldots & 0{,}8 \end{array}.$$

Математичне сподівання випадкової величини $Z = X + Y$ дорівнює 3,6. Знайти закони розподілу імовірностей X та Y, а також імовірність випадкової події $3 \le Z \le 4$.

10. Імовірність того, що сумарні витрати електроенергії на підприємстві за один робочий день не перевищать $M = 20\,000$ кВт/год, дорівнює 0,9. Скласти закон розподілу числа робочих днів тижня (п’ятиденного), протягом кожного із яких витрати електроенергії не перевищуватимуть M. Знайти середнє число таких днів.

11. Імовірність того, що лучник влучить в мішень при одному пострілі, дорівнює 0,8. Скласти закон розподілу дискретної випадкової величини X – числа влучних пострілів із трьох проведених, а також $M(X)$ та $\sigma(X)$.

12. На дорогах України лише 70% автомобільних шин витримують гарантійний термін. Скласти закон розподілу числа шин, що витримають гарантійний термін, із п’яти придбаних. Оцінити середнє число таких шин та розкид можливих значень.

13. Три стрілки виконують по чотири постріли. Імовірність влучання для першого стрільця становить 0,8, для другого – 0,9, для третього – 0,7. Знайти математичне сподівання і середнє квадратичне відхилення загального числа пробоїн у мішені.

14. Випадкова величина X – число підприємців із кожних десяти, які декларують не весь товар при перетині кордону, розподілена за таким законом: $\begin{array}{c|cccc} X & 0 & 1 & 2 & 3 \\ \hline P & 0{,}4 & 0{,}3 & 0{,}2 & \ldots \end{array}$. Знайти числові характеристики випадкової величини X.

15. Випробовується пристрій, який складається із п'яти незалежно працюючих ланок. Імовірність виходу з ладу для кожної з ланок відповідно дорівнює 0,1; 0,7; 0,3; 0,6; 0,5. Знайти математичне сподівання і квадратичне відхилення числа ланок, що вийдуть із ладу.

16. При виробництві мікросхем 5% продукції складає брак. Знайти закон розподілу випадкової величини X – числа стандартних мікросхем серед п'яти навмання відібраних, а також $M(X)$ та $\sigma(X)$.

17. Імовірності зростання вартості кожного із чотирьох видів сировини за прогнозний період складають відповідно 0,2, 0,8, 0,1, 0,5. Скласти закон розподілу випадкової величини X – числа видів сировини, для яких не відбудеться зростання ціни за цей період, а також знайти $M(X)$ та $\sigma(X)$.

18. Статистично встановлено, що в середньому два вироби із десяти вимагають додаткового регулювання. Контролер перевіряє якість партії виготовлених виробів, навмання вибираючи виріб. Якщо він вимагає додаткового регулювання, то наступні випробування припиняються, а вся партія відправляється на доробку. Якщо ж виріб стандартний, то контролер бере наступний виріб і т. д. За інструкцією контролер повинен перебрати не більше шести виробів. 1) Скласти закон розподілу числа перевірених контролером виробів. 2) Знайти імовірність того, що перевірена партія виробів буде відправлена на доробку. 3) Скільки в середньому виробів доведеться перевірити контролерові?

19. Записати закон розподілу випадкової величини X – числа випадінь грані із чотирма очками при п'яти киданнях грального кубика. Знайти $M(X)$ та $\sigma(X)$.

20. Випадкові величини X та Y незалежні і часткова інформація про їх закони розподілу має вид:

$$\begin{array}{c|c|c} X & x_1 & 2 \\ \hline P & \ldots & 0{,}7 \end{array}, \quad \begin{array}{c|c|c} Y & -2 & 1 \\ \hline P & 0{,}2 & \ldots \end{array}.$$

Математичне сподівання випадкової величини $Z = XY$ дорівнює 1,04. Знайти закони розподілу імовірностей X та Y, а також імовірність випадкової події $Z > 1$.

21. Скласти закон розподілу випадкової величини X – числа вгаданих номерів в спортлото «5 із 36». Знайти $M(X)$ та $\sigma(X)$.

22. Імовірність того, що пасажир запізниться на поїзд, дорівнює 0,02. Знайти $\sigma(X)$, де X – кількість пасажирів, які запізнюються на поїзд, якщо відомо, що найімовірніше число пасажирів, які запізняться, дорівнює 35.

23. В студентській групі (30 студентів, серед яких 12 дівчат і 18 юнаків) розігруються п'ять путівок на лижну базу в Карпатах. Скласти закон розподілу числа юнаків, яким дістануться путівки, а також математичне сподівання цього числа.

24. У скрині знаходиться п'ять куль з номерами від 1 до 5. Навмання витягують дві кулі. Скласти закон розподілу випадкової величини X – суми номерів витягнутих куль. Знайти $\sigma(X)$.

25. За даними відділу маркетингу підприємства з імовірностями 0,8, 0,6, 0,2 прогнозується підвищення попиту на кожний із трьох видів продукції. Скласти закон розподілу числа видів продукції, для яких прогнозується підвищення попиту, а також середнє число таких видів продукції.

26. Підприємство використовує чотири види сировини. Імовірність зриву поставок кожної із них відповідно дорівнює 0,1; 0,05; 0,01; 0,08. Скласти закон розподілу випадкової величини X – числа видів сировини, поставка яких буде зірвана. Знайти середнє число таких видів сировини, а також оцінити розкид можливих значень X.

27. При виробництві бракованого виробу підприємство зазнає збитків. Закон розподілу X (тис. грн.) прибутків за деякий період має такий вид:

X	–20	–10	10	30
P	0,1	0,2	0,6	...

. Знайти середню величину прибутку за цей період і оцінку розподілу значень прибутку довкола середньої в лінійних одиницях.

28. Екзаменатор задає студенту додаткові питання до тих пір, поки той правильно відповідає. Як тільки число правильних відповідей досягне п'яти або студент дасть неправильну відповідь, тоді екзаменатор припиняє задавати питання. Імовірність правильної відповіді на кожне із додаткових питань дорівнює 0,75. Знайти середнє число додаткових питань, які задасть екзаменатор.

29. Статистично встановлено, що з кожних десяти проведених зустрічей з акціонерами шість завершуються укладанням угод. Планується проведення п'яти зустрічей з акціонерами. Скласти

закон розподілу числа укладених угод, а також знайти середнє квадратичне відхилення цього числа.

30. На двох конвеєрах виготовляються однакові вироби. Відомі закони розподілу числа стандартних виробів, виготовлених за одиницю часу на першому конвеєрі (X) та другому (Y):

X	0	1	2
P	0,05	0,15	0,8

,

Y	0	1	2
P	0,03	0,07	...

.

Потрібно:1) скласти закон розподілу загального числа вироблених за одиницю часу стандартних виробів; 2) середнє квадратичне відхилення цього числа виробів.

31. В магазин надійшла партія з 10 телевізорів, 6 з яких вимагають регулювання зображення. Майстер, шукаючи перший телевізор, що вимагає регулювання, перевіряє їх почергово, і, знайшовши такий апарат, припиняє наступний пошук. Знайти математичне сподівання і дисперсію випадкової величини – числа перевірених телевізорів.

32. Випадкова величина Y задана законом розподілу

Y	–2	1	4
P	0,2	0,5	…

.

Знайти $\sigma(Z)$, якщо $Z = 5X - 4Y + 8$, X та Y незалежні випадкові величини, $\sigma(X) = 2$.

33. Лісник, який має 5 набоїв із снодійною речовиною, стріляє по косулі. Імовірність влучання при першому пострілі дорівнює 0,9, а при кожному наступному зменшується на 0,15. Знайти математичне сподівання і дисперсію випадкової величини – числа набоїв, використаних лісником.

34. Знайти середнє квадратичне відхилення числа появ події в 200 повторних незалежних випробуваннях, якщо математичне сподівання числа появ цієї події у двох випробуваннях дорівнює 0,4.

35. Два бухгалтери виконують складні однотипні розрахунки. Імовірність помилки для першого у звітній відомості дорівнює 0,1, а для другого – 0,08. Скласти закон розподілу числа безпомилкових відомостей, якщо кожний із них заповнив по дві відомості. Знайти середнє число таких відомостей.

36. Скласти закон розподілу випадкової величини X – числа вгаданих номерів в національній лотереї «6 із 45». Знайти $M(X)$ та $D(X)$.

37. Знайти математичне сподівання випадкової величини $Z = X + Y$, якщо X та Y незалежні і задані законами розподілу

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

,

Y	2	5
P	0,4	0,6

.

38. Випадкові величини X та Y незалежні, а $Z = 2X - 5Y + 4$. Знайти $M(Z)$ і $\sigma(Z)$, якщо $M(X) = 3$, $M(Y) = 2$, $D(X) = 1$, $D(Y) = 0{,}4$.

39. Відомо, що серед готівкової маси 0,5% купюр є непридатними до наступного використання. Скласти закон розподілу величини X – числа нестандартних купюр серед шести навмання взятих, а також знайти $M(X)$ та $\sigma(X)$.

40. Випадкова величина X – дохід фірми в євро на прогнозований період, Y – курс євро відносно гривні на цей же період, при цьому X та Y незалежні випадкові величини і їх закони розподілу мають вид:

X	1500	2000
P	0,6	0,4

,

Y	30,1	30,4	30,8
P	0,3	0,5	0,2

Знайти: 1) імовірність того, що дохід в гривнях на прогнозний період виявиться не меншим від 58500; 2) середню величину очікуваного доходу в гривнях, а також дисперсію цієї величини.

§ 5. НЕПЕРЕРВНІ ВИПАДКОВІ ВЕЛИЧИНИ ТА ЇХ ЧИСЛОВІ ХАРАКТЕРИСТИКИ

Функція розподілу імовірностей і її властивості. Густина розподілу імовірностей та її властивості. Числові характеристики неперервних випадкових величин (математичне сподівання, дисперсія та середнє квадратичне відхилення, мода та медіана випадкової величини, початкові та центральні моменти, асиметрія і ексцес).

КОРОТКІ ТЕОРЕТИЧНІ ВІДОМОСТІ

Функцією розподілу імовірностей називається функція $F(x)$ детермінованого (невипадкового) аргументу x, яка чисельно дорівнює імовірності того, що в результаті випробування випадкова величина X набере значення, менше від x, тобто

$$F(x) = P(X < x). \tag{5.1}$$

Іноді функцію розподілу називають ще інтегральною.

Уточнимо означення неперевної випадкової величини: випадкову величину називають **неперервною**, якщо її функція розподілу імовірностей є неперервною на області її визначення, а похідна від функції розподілу неперервна у всіх точках, за виключенням, можливо, скінченного числа точок на довільному скінченному інтервалі.

Властивості функції розподілу імовірностей

Властивість 1. Область визначення функції розподілу – $R^1 = (-\infty; \infty)$, а область значень – відрізок [0; 1].

Властивість 2. $F(x)$ – неспадна функція, тобто для будь-якої пари чисел x_1, x_2 з нерівності $x_2 > x_1$ випливає нерівність $F(x_2) \geq F(x_1)$.

Наслідок 1. Імовірність того, що випадкова величина при випробуванні набере можливого значення з проміжку [a; b), дорівнює приросту функції розподілу на цьому проміжку:

$$P(a \leq X < b) = F(b) - F(a). \tag{5.2}$$

Наслідок 2. Імовірність того, що неперервна випадкова величина X набере при випробуванні одне конкретне можливе значення, дорівнює нулю.

Зауваження. Рівність $P(X = x_1) = 0$, де x_1 – конкретне можливе значення величини X, не означає, що подія $X = x_1$ є неможливою (на відміну від класичного означення імовірності). Нагадайте геометричне означення імовірності.

Наслідок 2 з використанням формули (5.2) та теореми додавання імовірностей дозволяє отримати такий ланцюжок рівностей:

$$P(a < X < b) = \big(P(a \le X \le b) = P(a \le X < b) = P(a < X \le b)\big) = \\ = F(b) - F(a). \tag{5.3}$$

Властивість 3. Якщо всі можливі значення випадкової величини належать відрізку [a; b], то $F(x) = 0$ для всіх $x \le a$ і $F(x) = 1$ для всіх $x > b$.

Наслідок. Якщо можливими значеннями неперервної випадкової величини є всі дійсні числа, тоді мають місце такі граничні співвідношення:

$$\lim_{x \to -\infty} F(x) = 0, \qquad \lim_{x \to \infty} F(x) = 1.$$

Густиною розподілу імовірностей неперервної випадкової величини X називається функція $f(x)$, яка дорівнює першій похідній від функції розподілу $F(x)$:

$$f(x) = F'(x). \tag{5.4}$$

Із означення (5.4) слідує, що $F(x)$ є первісною для густини розподілу. Нижче буде наведено формулу для знаходження $F(x)$, якщо відома функція $f(x)$. Таким чином, задання однієї із функцій $f(x)$ та $F(x)$ дозволяє знайти іншу. Тому в літературі ці функції називають **законами розподілу неперервної випадкової величини**.

Властивості густини розподілу

Властивість 1. Область визначення функції $f(x) - R^1$, а область значень проміжок [0, ∞).

Властивість 2. Невласний інтеграл від густини розподілу в межах від –∞ до ∞ дорівнює одиниці:

$$\int_{-\infty}^{\infty} f(x)dx = 1. \tag{5.5}$$

Рівність (5.5) називають **умовою нормування** неперервної випадкової величини.

Зауваження. Якщо $[\alpha, \beta]$- мінімальний проміжок, в якому містяться **всі** можливі значення неперервної випадкової величини, тоді умова нормування набирає такого виду:

$$\int_{\alpha}^{\beta} f(x)dx = 1. \qquad (5.5^*)$$

Мають місце такі формули:

$$P\left(a \overset{(\le)}{<} X \overset{(\le)}{<} b \right) = \int_{a}^{b} f(x)dx, \qquad (5.6)$$

$$F(x) = P(X < x) = P(-\infty < X < x) = \int_{-\infty}^{x} f(x)dx. \qquad (5.7)$$

Відмітимо, що рівність (5.5) є аналогом рівності одиниці суми імовірностей із закону розподілу дискретної випадкової величини.

З'ясуємо **імовірнісний зміст** густини розподілу імовірностей. Означення $f(x)$ (5.4) та похідної функції, а також рівність (5.3) дають такі рівності

$$f(x) = \lim_{\Delta x \to 0} \frac{F(x+\Delta x) - F(x)}{\Delta x} = \lim_{\Delta x \to 0} \frac{P(x < X < x+\Delta x)}{\Delta x}, \qquad (5.8)$$

які дозволяють записати наближену рівність

$$P(x < X < x+\Delta x) \approx f(x)\Delta x. \qquad (5.9)$$

з точністю до нескінченно малих вищого порядку відносно Δx.

Права частина (5.8) пояснює назву функції $f(x)$ за аналогією із змістом густини маси, розподіленої на відрізку, а наближена рівність (5.9) – «вклад» величини $f(x)$ у величину імовірності події $(x < X < x + \Delta x)$ для кожного відрізка довжиною Δx.

Математичним сподіванням неперервної випадкової величини X, всі можливі значення якої належать скінченному відрізку $[a; b]$, називається визначений інтеграл

$$M(X) = \int_{a}^{b} xf(x)dx. \qquad (5.10)$$

Якщо можливі значення X належать R^1, тоді

$$M(X) = \int_{-\infty}^{\infty} xf(x)dx, \qquad (5.10^*)$$

де за припущенням невласний інтеграл збігається абсолютно.

Як і для випадку дискретної величини, **дисперсією неперервної випадкової величини** називається математичне сподівання квадрату її відхилення від математичного сподівання. Із врахуванням означення $M(X)$:

$$D(X)=\int_a^b [x-M(X)]^2 f(x)dx, \qquad (5.11)$$

якщо всі можливі значення X належать скінченному відрізку $[a; b]$, і

$$D(X)=\int_{-\infty}^{\infty} [x-M(X)]^2 f(x)dx, \qquad (5.11^*)$$

якщо можливі значення X заповнюють R^1.

Розрахункові формули для обчислення дисперсії:

$$D(X)=\int_a^b x^2 f(x)dx-[M(X)]^2, \qquad (5.12)$$

$$D(X)=\int_{-\infty}^{\infty} x^2 f(x)dx-[M(X)]^2. \qquad (5.12^*)$$

Середнє квадратичне відхилення неперервної випадкової величини визначається, як і для випадку дискретної, рівністю

$$\sigma(X)=\sqrt{D(X)}. \qquad (5.13)$$

Модою $Mo(X)$ дискретної випадкової величини X називається те її можливе значення, якому відповідає найбільша імовірність її появи. Для неперервної випадкової величини X модою $Mo(X)$ називається можливе значення X, якому відповідає локальний максимум густини розподілу. Випадкова величина може мати кілька мод. У цьому випадку вона називається **многомодальною**. Зустрічаються також розподіли, що не мають максимуму. Такі розподіли називаються **антимодальними**.

Медіаною $Me(X)$ неперервної випадкової величини X називається те її можливе значення, для якого виконується рівність $P(X < Me(X)) = P(X > Me(X))$.

Початковим моментом *k*-ого порядку неперервної випадкової величини X, заданої густиною розподілу $f(x)$, називається математичне сподівання величини X^k, тобто

$$\nu_k=\int_{-\infty}^{\infty} x^k f(x)dx, \quad k=1,\ 2,\ 3,\ \ldots\ . \qquad (5.14)$$

Центральним моментом *k*-ого порядку неперервної випадкової величини X називається математичне сподівання величини $[X-M(X)]^k$:

$$\mu_k=\int_{-\infty}^{\infty} [x-M(X)]^k f(x)dx, \quad k=1,\ 2,\ 3,\ \ldots\ . \qquad (5.15)$$

Зокрема, якщо всі можливі значення X належать проміжку (a, b), то

$$\nu_k = \int_a^b x^k f(x)dx, \qquad \mu_k = \int_a^b [x - M(X)]^k f(x)dx.$$

Очевидно, що $\nu_1 = M(X)$, $\mu_1 = 0$, $\mu_2 = D(X)$.

Формули зв'язку центральних та початкових моментів, які наведені в § 4, залишаються в силі і для випадку неперервних випадкових величин.

Центральний момент третього порядку характеризує **асиметрію закону розподілу випадкової величини**, тобто порушення симетрії кривої розподілу. Якщо $\mu_3 = 0$, то говорять, що випадкова величина симетрично поділена відносно $M(X)$ (крива розподілу симетрична відносно прямої $x = M(X)$. Оскільки μ_3 має вимірність випадкової величини в кубі, то впроваджують безрозмірну величину – коефіцієнт асиметрії:

$$A_s = \mu_3 / \sigma^3.$$

Центральний момент четвертого порядку використовується для характеристики гостровершинності чи плосковершинності закону розподілу (кривої розподілу). Ці властивості описуються з допомогою **ексцесу**:

$$E_s = \mu_4 / \sigma^4 - 3,$$

де число 3 віднімається в зв'язку з тим, що для найпоширенішого нормального закону розподілу (див. § 6) $\mu_4 / \sigma^4 = 3$, а отже, $E_s = 0$. Таким чином, ексцес характеризує гостровершинність чи плосковершинність в порівнянні із **нормальною кривою** (кривою розподілу нормального закону).

РОЗВ'ЯЗУВАННЯ ТИПОВИХ ЗАДАЧ

Задача 5.1. Дискретна випадкова величина X задана законом розподілу

X	1	3	6	8
P	0,2	0,4	0,3	0,1

. Знайти функцію розподілу і побудувати її графік.

○ Якщо $x \le 1$, то у відповідності із властивістю 3 $F(x) = 0$.

Нехай $x \in (1; 3]$. Тоді випадкова подія $(X < x) = (X = 1)$, оскільки 1 – єдине можливе значення, яке менше від x. А тому згідно із (5.1) для $(1; 3]$ $F(x) = P(X < x) = P(X = 1) = 0,2$.

Якщо $x \in (3;6]$, тоді подія $(X < x)$ відбувається тоді, і тільки тоді, коли або $X = 1$, або $X = 3$, тобто має місце така рівність $(X < x) = (X = 1) + (X = 3)$, де доданки справа є несумісними випадковими подіями. Використання теореми додавання імовірностей і означення (5.1) дозволяє знайти $F(x)$ на цьому проміжку:

$$F(x) = P(X < x) = P(X = 1) + P(X = 3) = 0{,}2 + 0{,}4 = 0{,}6.$$

Якщо $x \in (6;8]$, то за аналогією із попереднім проміжком

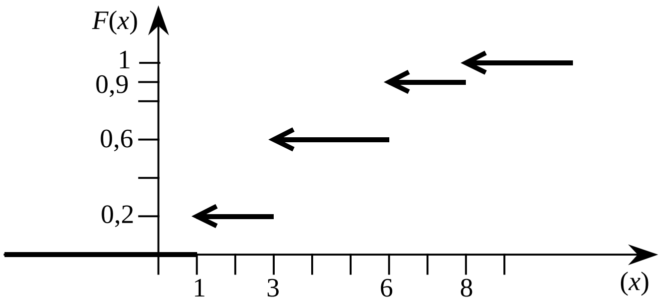
$$F(x) = P(X < x) = P(X = 1) + P(X = 3) + P(X = 6) = 0{,}9.$$

Нарешті якщо $x > 8$, то подія $(X < x)$ – достовірна, і $F(x) = 1$.

Отже, функція розподілу має такий вид:

$$F(x) = \begin{cases} 0, & \text{якщо } x \le 1, \\ 0{,}2, & \text{якщо } 1 < x \le 3, \\ 0{,}6, & \text{якщо } 3 < x \le 6, \\ 0{,}9, & \text{якщо } 6 < x \le 8, \\ 1, & \text{якщо } x > 8, \end{cases}$$

а її графік зображено на мал. 5.1.



Мал. 5.1.

де стрілками відмічені правосторонні розриви. ●

Задача 5.2. Випадкова величина X задана функцією розподілу

$$F(x) = \begin{cases} 0 & \text{при } x \le -2, \\ x^2/4 + x + 1 & \text{при } -2 < x \le 0, \\ 1 & \text{при } x > 0. \end{cases}$$

Знайти імовірність того, що: 1) в результаті випробування X набере значення з інтервалу $(-1; 0)$; 2) у чотирьох випробуваннях X тричі набере можливого значення з проміжку $[-4; -1)$.

○ 1) Для знаходження $P(-1 < X < 0)$ скористаємося однією із формул (5.3), поклавши $a = -1$, $b = 0$. Тоді

$$P(-1 < X < 0) = F(0) - F(-1) =$$
$$= 0^2/4 + 0 + 1 - ((-1)^2/4 - 1 + 1) = 1 - 1/4 = 0{,}75.$$

2) Попередньо знайдем імовірність відбуття події $(-4 \le X < -1)$ в одному випробуванні, врахувавши те, що -4 належить першому проміжку задання функції, в кожній точці якого $F(x) = 0$. Знову за формулою (5.3)

$$P(-4 \le X < -1) = F(-1) - F(-4) = 0{,}25 - 0 = 0{,}25.$$

Друге випробування, як і решта наступних, дасть ту ж саму імовірність. Отже, повторні випробування є незалежними відносно події $(-4 \le X < -1)$ і за формулою Бернуллі ($n = 4$ – мале), де $p = 0{,}25$, $q = 0{,}75$, знайдемо:

$$P_4(3) = C_4^3 (0{,}25)^3 0{,}75 = 0{,}1875. \bullet$$

Зауваження. Типова помилка студентів при використанні формули (5.3) – неправильний вибір аналітичного виразу, за яким обчислюється $F(x)$ при знаходженні $F(b)$ та (або) $F(a)$. Тобто попередньо не враховується, до якого інтервалу (в задачі 5.2 їх є три) належить a та b.

Задача 5.3. Функція розподілу імовірностей неперервної випадкової величини має такий вид:

$$F(x) = \begin{cases} 0 \quad , & \text{при } x \le 1, \\ ax + b \ , & \text{при } 1 < x \le 5, \\ 1 \quad , & \text{при } x > 5 \end{cases}$$

Знайти значення сталих a і b та побудувати графік функції *F(x)*.

○ Оскільки X - неперервна випадкова величина, то *F(x)* є непервною функцією для всіх $x \in R^1$. Тому

$$F(1) = \lim_{x\to 1+0} F(x) = \lim_{x\to 1+0} (ax+b) = a + b \quad \text{або } a+b=0.$$

Аналогічно

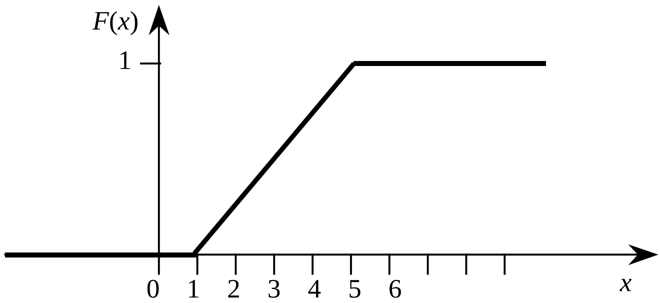
$$F(1) = a \cdot 5 + b = \lim_{x\to 5-0} F(x) = 1 \quad \text{або } 5a+b=1.$$

Отже, невідомі сталі a та b задовольняють систему рівнянь

$$\begin{cases} a + b = 0, \\ 5a + b = 1, \end{cases}$$

розв'язавши яку отримаємо a=0,25, b=–0,25.

Графік *F(x)* зображено на мал.5.2. ●



Мал.5.2.

Задача 5.4. Випадкова величина X задана густиною розподілу

$$f(x)=\begin{cases}0 & \text{при } x\le -\pi/2,\\ a\cos x & \text{при } -\pi/2<x\le \pi/2,\\ 0 & \text{при } x>\pi/2.\end{cases}$$

Знайти параметр a і функцію розподілу імовірностей, а також імовірність того, що при випробуванні величина X набере значення з інтервалу (–π/4, π/3).

○ Параметр a знайдемо з рівності (5.6):

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

Розіб'ємо невласний інтеграл на три інтеграли у відповідності із інтервалами задання густини розподілу:

$$\int_{-\infty}^{\infty} f(x)dx=\int_{-\infty}^{-\pi/2} f(x)dx+\int_{-\pi/2}^{\pi/2} f(x)dx+\int_{\pi/2}^{\infty} f(x)dx=$$

$$=\int_{-\infty}^{-\pi/2} 0dx+\int_{-\pi/2}^{\pi/2} a\cos xdx+\int_{\pi/2}^{\infty} 0dx=a\int_{-\pi/2}^{\pi/2}\cos xdx=$$

$$=a\sin x\Big|_{-\pi/2}^{\pi/2}=a(\sin(\pi/2)-\sin(-\pi/2))=2a.$$

Отже, параметр a задовольняє рівняння $2a=1$, звідки $a=0{,}5$. При знаходженні $F(x)$ врахуємо це значення параметра і використаємо рівність (5.7). Відмітимо, що вид густини розподілу в даній задачі вказує на те, що функція розподілу на трьох проміжках також буде задаватися різними аналітичними виразами. В зв'язку із цим нехай $x\le -\pi/2$. Тоді

$$F(x)=\int_{-\infty}^{x} f(x)dx=\int_{-\infty}^{x} 0dx=0.$$

Для $x\in(-\pi/2;\ \pi/2]$

$$F(x)=\int_{-\infty}^{x} f(x)dx=\int_{-\infty}^{-\pi/2} f(x)dx+\int_{-\pi/2}^{x} f(x)dx=\int_{-\infty}^{-\pi/2} 0dx+\int_{-\pi/2}^{x} 0{,}5\cos x dx=$$

$$=0{,}5\sin x\Big|_{-\pi/2}^{x}=0{,}5(\sin x-\sin(-\pi/2))=0{,}5(1+\sin x).$$

Нарешті, якщо $x>\pi/2$, то

$$F(x)=\int_{-\infty}^{x} f(x)dx=\int_{-\infty}^{-\pi/2} 0dx+\int_{-\pi/2}^{\pi/2} 0{,}5\cos x dx+\int_{\pi/2}^{x} 0dx=0{,}5\sin x\Big|_{-\pi/2}^{\pi/2}=1.$$

Таким чином, функція розподілу для даної величини X має такий вигляд:

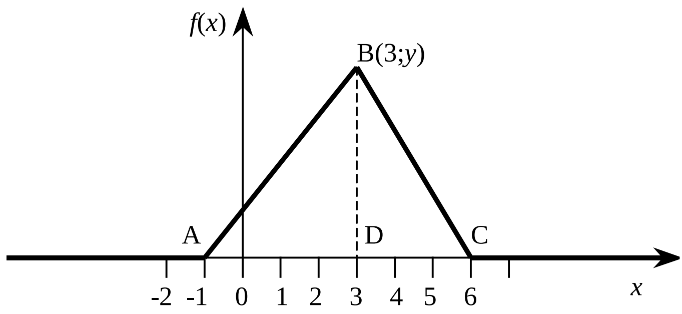
$$F(x)=\begin{cases}0 & \text{при } x\le-\pi/2,\\ 0{,}5(1+\sin x) & \text{при } -\pi/2<x\le\pi/2,\\ 1 & \text{при } x>\pi/2.\end{cases}$$

Використаємо формулу (5.6) для знаходження шуканої імовірності:

$$P(-\pi/4<X<\pi/3)=\int_{-\pi/4}^{\pi/3} f(x)dx=\int_{-\pi/4}^{\pi/3} 0{,}5\cos x dx=0{,}5\sin x\Big|_{-\pi/4}^{\pi/3}=$$

$$=0{,}5(\sin(\pi/3)-\sin(-\pi/4))=0{,}5\left(\sqrt{3}/2+\sqrt{2}/2\right)=\left(\sqrt{3}+\sqrt{2}\right)/4. \quad \bullet$$

Задача 5.5. Графік густини розподілу неперервної випадкової величини X зображено на мал.5.3. Знайти аналітичні вирази густини розподілу та функції розподілу імовірностей цієї величини. Побудувати графік $F(x)$ і обчислити $P(1<X<4)$, використавши інформацію про $F(x)$ та $f(x)$.



Мал.5.3.

○ Знайдемо ординату точки *B*, використовуючи умову нормування (5.5*):

$$\int_{-1}^{6} f(x)dx = S_{\Delta ABC} = \frac{1}{2} AC \cdot BD = \frac{1}{2} 7y = 1,$$

звідки $y = \frac{2}{7}$.

На проміжку [-1,3] $f(x)$ дорівнює ординаті відрізка прямої, що проходить через точки $A(-1;0)$ та $B(3;2/7)$. Знайдемо рівняння прямої *AB*:

$$\frac{y - y_A}{y_B - y_A} = \frac{x - x_A}{x_B - x_A} \Rightarrow \frac{y - 0}{\frac{2}{7} - 0} = \frac{x - (-1)}{3 - (-1)} \Rightarrow y = \frac{x+1}{14}.$$

Отже, $f(x) = \frac{x+1}{14}$, якщо $x \in [-1;3]$.

На проміжку [3;6] $f(x)$ дорівнює ординаті відрізка прямої, що проходить через точки $B(3;\frac{2}{7})$ і $C(6;0)$.

Рівняння прямої *BC*:

$$\frac{y - y_B}{y_C - y_B} = \frac{x - x_B}{x_C - x_B} \Rightarrow \frac{y - \frac{2}{7}}{0 - \frac{2}{7}} = \frac{x-3}{6-3} \Rightarrow y = \frac{2(6-x)}{21}.$$

Отже, $f(x) = \frac{2}{21}(6-x)$, якщо $x \in [3;6]$.

В підсумку отримаємо такий аналітичний вираз для густини розподілу:

$$f(x) = \begin{cases} 0\ , & \text{при } x \le -1, \\ \frac{x+1}{14}, & \text{при } -1 < x \le 3, \\ \frac{2(6-x)}{21}, & \text{при } 3 < x \le 6, \\ 0\ , & \text{при } x > 6. \end{cases} \qquad (5.16)$$

Для знаходження функції розподілу $F(x)$ використаємо формулу (5.7). Нехай $x \le -1$, тоді $f(x)=0$ і

$$F(x)=\int_{-\infty}^{x} f(x)dx=\int_{-\infty}^{x} 0dx=0.$$

Якщо $x\in(-1;\,3]$, то

$$F(x)=\int_{-\infty}^{x} f(x)dx=\int_{-\infty}^{-1} f(x)dx+\int_{-1}^{x} f(x)dx=F(-1)+\int_{-1}^{x}\frac{x+1}{14}dx=$$

$$=0+\frac{1}{14}\int_{-1}^{x}(x+1)d(x+1)=\frac{1}{28}(x+1)^2\Big|_{-1}^{x}=\frac{1}{28}(x+1)^2.$$

Для $x\in(3;\,6]$

$$F(x)=\int_{-\infty}^{x} f(x)dx=\int_{-\infty}^{3} f(x)dx+\int_{3}^{x} f(x)dx=F(3)+\int_{3}^{x}\frac{2(6-x)}{21}dx=$$

$$=\frac{1}{28}(3+1)^2-\frac{2}{21}\int_{3}^{x}(6-x)d(6-x)=\frac{4}{7}-\frac{1}{21}(6-x)^2\Big|_{3}^{x}=$$

$$=\frac{4}{7}-\frac{1}{21}(6-x)^2+\frac{3}{7}=1-\frac{(6-x)^2}{21}.$$

Нарешті, якщо $x>6$, то

$$F(x)=\int_{-\infty}^{x} f(x)dx=\int_{-\infty}^{6} 0dx+\int_{6}^{x} f(x)dx=F(6)+\int_{6}^{x} 0dx=1-\frac{(6-6)^2}{21}+0=1.$$

Таким чином, функція розподілу для даної випадкової величини X має такий вигляд:

$$F(x)=\begin{cases}0\ , & \text{якщо } x\le -1,\\ \dfrac{1}{28}(x+1)^2, & \text{якщо } -1<x\le 3,\\ 1-\dfrac{(6-x)^2}{21}, & \text{якщо } 3<x\le 6,\\ 1\ , & \text{якщо } x>6.\end{cases}$$

Графік функції $F(x)$ зображено на мал.5.4.

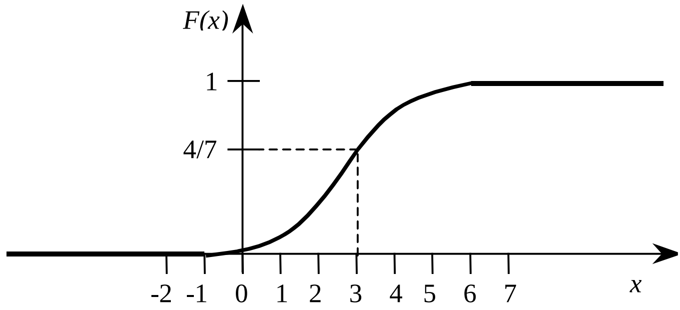
Згідно з формулою (5.3)

$P(1<X<4)=F(4)-F(1)$.

Значення аргументів 4 і 1 належать відповідно третьому та другому інтервалам, тому

$$F(4)=1-\frac{(6-4)^2}{21}=1-\frac{4}{21}=\frac{17}{21},$$

$$F(1)=\frac{1}{28}(1+1)^2=\frac{1}{7},$$



Мал.5.4.

$$P(1 < X < 4) = \frac{17}{21} - \frac{1}{7} = \frac{14}{21} = \frac{2}{3}.$$

Цю же саму імовірність знайдемо з допомогою формули (5.6):

$$P(1 < X < 4) = \int_1^4 f(x)dx = \int_1^3 f(x)dx + \int_3^4 f(x)dx =$$

$$= \int_1^3 \frac{(x+1)}{14}dx + \int_3^4 \frac{2(6-x)}{21}dx = \frac{1}{14}\int_1^3 (x+1)d(x+1) - \frac{2}{21}\int_3^4 (6-x)d(6-x) =$$

$$= \frac{1}{14}\frac{(x+1)^2}{2}\bigg|_1^3 - \frac{2}{21}\frac{(6-x)^2}{2}\bigg|_3^4 = \frac{1}{28}[(3+1)^2 - (1+1)^2] - \frac{1}{21}[(6-4)^2 - (6-3)^2] =$$

$$= \frac{3}{7} + \frac{5}{21} = \frac{2}{3}.$$ ●

Задача 5.6. Випадкова величина *X* задана густиною розподілу

$$f(x) = \begin{cases} 2\cos 2x, & \text{якщо} \quad x \in (0; \pi/4) \\ 0, & \text{якщо} \quad x \notin (0; \pi/4) \end{cases}$$

Знайти: а) інтервал, в якому знаходяться всі можливі значення *X*; б) $\sigma(X)$; в) моду *Mo*(*X*); г) медіану *Me*(*X*).

○ а) Згідно із наближеною рівністю (5.9) неперервна випадкова величина *X* набирає можливих значень при випробуванні в кожному інтервалі, на якому густина розподілу відмінна від нуля. Для нашого випадку $f(x) \neq 0 \ \forall \ x \in (0; \pi/4)$. Отже, в інтервалі $(0; \pi/4)$ знаходяться всі можливі значення *X*.

б) Для обчислення $\sigma(X)$ потрібно знайти спочатку *M*(*X*) та *D*(*X*). При знаходженні цих характеристик скористаємося інтегруванням частинами:

$$M(X)=\int\limits_0^{\pi/4} x2\cos 2x dx = 2\int\limits_0^{\pi/4} x\cos 2x dx =$$

$$u=x \qquad du=dx$$
$$dv=\cos 2x dx \qquad v=\frac{1}{2}\sin 2x$$

$$=2\left(x\cdot\frac{1}{2}\sin 2x\Big|_0^{\pi/4} - \frac{1}{2}\int\limits_0^{\pi/4}\sin 2x dx \right)=$$

$$=\frac{\pi}{4}\sin\frac{\pi}{2}+\frac{1}{2}\cos 2x\Big|_0^{\pi/4} = (\pi-2)/4.$$

Дисперсію знайдемо за розрахунковою формулою (5.12):

$$M\left(X^2\right)=\int\limits_0^{\pi/4} x^2 2\cos 2x dx = 2\int\limits_0^{\pi/4} x^2\cos x dx =$$

$$u=x^2 \qquad du=2x dx$$
$$dv=\cos 2x dx \qquad v=\frac{1}{2}\sin 2x$$

$$=2\left[\left(\frac{x^2}{2}\sin 2x\right)\Big|_0^{\pi/4} - \int\limits_0^{\pi/4} x\sin 2x dx\right]=$$

$$u=x \qquad du=dx$$
$$dv=\sin 2x dx \qquad v=-\frac{1}{2}\cos 2x$$

$$=\frac{\pi^2}{16}+x\cos 2x\Big|_0^{\pi/4} - \int\limits_0^{\pi/4}\cos 2x dx = \frac{\pi^2}{16}-\frac{1}{2}\sin 2x\Big|_0^{\pi/4} = \frac{\pi^2}{16}-\frac{1}{2}.$$

$$D(X)=M(X^2)-\left[M(X)\right]^2 = \left(\pi^2/16-1/2\right)-(\pi-2)^2/16=$$
$$=(\pi-3)/4.$$

$$\sigma(X)=\sqrt{D(X)}=\sqrt{\pi-3}/2 \ .$$

в) Оскільки функція $f(x)$ = 2cos 2x у відкритому інтервалі $(0,\pi/4)$ не має максимуму, то X моди не має.

г) Знайдемо медіану Me, виходячи із її означення: $P(X < Me) = P(X > Me)$. Події $(X < Me)$ та $(X \geq Me)$ протилежні, тому

$P(X < Me) + P(X \geq Me) = 1$. Але у відповідності із наслідком 2 вл. 2 функції розподілу $P(X = Me) = 0$, тому $P(X < Me) = 1/2$. Проте

$$P(X < Me) = P(-\infty < X < Me) =$$

$$= \int_{-\infty}^{Me} f(x)dx = \int_{-\infty}^{0} 0dx + \int_{0}^{Me} 2\cos 2x dx = \sin 2x \Big|_{0}^{Me} = \sin(2Me).$$

В результаті отримуємо рівняння відносно Me:

$\sin 2\,Me = 1/2$, звідки $2\,Me = \arcsin(1/2) = \pi/6$.

Остаточно $Me = (\arcsin 1/2)/2 = \pi/12$. ●

Задача 5.7. Для випадкової величини X із задачі 5.5 знайти: а) середнє квадратичне відхилення; б) моду $Mo(X)$; в) медіану $Me(X)$.

○ Згідно з умовою густина розподілу випадкової величини X визначається виразом (5.16), суттєвою особливістю якого є те, що $f(x)$ визначається різними аналітичними виразами, відмінними від нуля, на різних інтервалах.

а) Для обчислення $\sigma(X)$ знайдемо спочатку $M(X)$ та $D(X)$ за формулами (5.10) та (5.12):

$$M(X) = \int_{-1}^{6} xf(x)dx = \int_{-1}^{3} xf(x)dx + \int_{3}^{6} xf(x)dx =$$

$$= \int_{-1}^{3} x\frac{x+1}{14}dx + \int_{3}^{6} x\frac{2(6-x)}{21}dx = \frac{1}{14}\int_{-1}^{3}(x^2+x)dx + \frac{2}{21}\int_{3}^{6}(6x-x^2)dx =$$

$$= \frac{1}{14}\left(\frac{x^3}{3}+\frac{x^2}{2}\right)\Big|_{-1}^{3} + \frac{2}{21}\left(6\cdot\frac{x^2}{2}-\frac{x^3}{3}\right)\Big|_{3}^{6} = \frac{8}{3};$$

$$M(X^2) = \int_{-1}^{6} x^2 f(x)dx = \int_{-1}^{3} x^2\frac{x+1}{14}dx + \int_{3}^{6} x^2\frac{2(6-x)}{21}dx =$$

$$= \frac{1}{14}\int_{-1}^{3}(x^3+x^2)dx + \frac{2}{21}\int_{3}^{6}(6x^2-x^3)dx =$$

$$D(X) = M(X^2) - [M(X)]^2 = \frac{55}{6} - \left(\frac{8}{3}\right)^2 = \frac{37}{18};$$

$$\sigma(X) = \sqrt{D(X)} = \sqrt{\frac{37}{18}} \approx 1{,}434.$$

б) Із мал.5.3, на якому зображений графік густини розподілу, видно, що максимальне значення $f(x)$ досягає в т. x=3. Отже,

$Mo(X)$=3.

в) Всі можливі значення даної випадкової величини знаходяться або на проміжку (-1; 3], або на (3; 6], але на кожному із них густина розподілу задається різними аналітичними виразами. З другого боку, медіана задовольняє (див. задачу 5.6 п. *в*) рівнянню $P(X<Me)$=0,5.

Тому якщо $P(X \le 3) \ge 0{,}5$, то $Me(X) \in (1;3]$, якщо ж $P(X \le 3) < 0{,}5$, то $Me(X) \in (3;6]$.

Використавши формулу (5.3), отримаємо

$$P(X \le 3) = P(-\infty < X \le 3) = F(3) - F(-\infty) = \frac{1}{28}(x+1)^2\Big|_{x=3} - 0 =$$

$$= \frac{16}{28} = \frac{4}{7} > 0{,}5.$$

Отже, $Me(X) \in (1;3]$. Тоді для знаходження *Me* отримаємо рівняння:

$$P(X < Me) = F(Me) - F(-\infty) = \frac{1}{28}(Me+1)^2 - 0 = 0{,}5,$$

або $(Me+1)^2$=14,

звідки $Me = \sqrt{14} - 1 \approx 2{,}742$. ●

ЗАДАЧІ ДЛЯ САМОСТІЙНОГО РОЗВ'ЯЗУВАННЯ

Завдання №6

В задачах 1-24 випадкова величина *X* задана густиною розподілу *f*(*x*). Потрібно:

1) знайти функцію розподілу імовірностей;

2) обчислити середнє квадратичне відхилення величини *X*;

3) побудувати графіки функції розподілу імовірностей та густини розподілу;

4) знайти імовірність того, що в трьох випробуваннях випадкова величина *X* двічі набере значення з інтервалу (α, β).

1. $f(x) = \begin{cases} 0{,}5\cos 0{,}5x, & x \in (0, \pi), \\ 0, & x \notin (0, \pi). \end{cases}$

$\alpha = 0, \beta = 4.$

2. $f(x) = \begin{cases} -2x + 4, & x \in [1,2), \\ 0, & x \notin [1,2). \end{cases}$

$\alpha = -4, \beta = 1{,}5.$

3. $f(x)=\begin{cases} x-2{,}5, & x\in(3{,}4], \\ 0, & x\notin(3{,}4]. \end{cases}$

$\alpha=3{,}5, \beta=12.$

4. $f(x)=\begin{cases} 2x-4, & x\in(2{,}3], \\ 0, & x\notin(2{,}3]. \end{cases}$

$\alpha=2{,}5, \beta=4.$

5. $f(x)=\begin{cases} 2x+2, & x\in(-1{,}0], \\ 0, & x\notin(-1{,}0]. \end{cases}$

$\alpha=-0{,}6, \beta=0{,}5.$

6. $f(x)=\begin{cases} 8x-8, & x\in(1{,}3/2], \\ 0, & x\notin(1{,}3/2]. \end{cases}$

$\alpha=1{,}5, \beta=3{,}1.$

7. $f(x)=\begin{cases} 2x-2, & x\in(1{,}2], \\ 0, & x\notin(1{,}2]. \end{cases}$

$\alpha=1{,}5, \beta=6.$

8. $f(x)=\begin{cases} x/2+1, & x\in(-2{,}0], \\ 0, & x\notin(-2{,}0]. \end{cases}$

$\alpha=-3, \beta=-1.$

9. $f(x)=\begin{cases} 1/4, & x\in(0{,}4], \\ 0, & x\notin(0{,}4]. \end{cases}$

$\alpha=2, \beta=5.$

10. $f(x)=\begin{cases} 2x/9, & x\in[0{,}3), \\ 0, & x\notin[0{,}3). \end{cases}$

$\alpha=2, \beta=7.$

11. $f(x)=\begin{cases} 4x+3, & x\in(-0{,}5;0], \\ 0, & x\notin(-0{,}5;0]. \end{cases}$

$\alpha=-6, \beta=-0{,}3.$

12. $f(x)=\begin{cases} 2x/3+4/3, & x\in[-1{,}0), \\ 0, & x\notin[-1{,}0). \end{cases}$

$\alpha=-13, \beta=-1.$

13. $f(x)=\begin{cases} -0{,}5\sin x, & x\in(\pi,2\pi], \\ 0, & x\notin(\pi,2\pi]. \end{cases}$

$\alpha=3\pi/2, \beta=3\pi.$

14. $f(x)=\begin{cases} 2(x-2), & x\in(2{,}3], \\ 0, & x\notin(2{,}3]. \end{cases}$

$\alpha=-4, \beta=2{,}2.$

15. $f(x)=\begin{cases} x/50, & x\in(0{,}10], \\ 0, & x\notin(0{,}10]. \end{cases}$

$\alpha=7, \beta=16.$

16. $f(x)=\begin{cases} 2x-6, & x\in(3{,}4], \\ 0, & x\notin(3{,}4]. \end{cases}$

$\alpha=3{,}3, \beta=8.$

17. $f(x)=\begin{cases} 2x-8, & x\in(4{,}5], \\ 0, & x\notin(4{,}5]. \end{cases}$

$\alpha=5, \beta=18.$

18. $f(x)=\begin{cases} 2x-10, & x\in(5{,}6], \\ 0, & x\notin(5{,}6]. \end{cases}$

$\alpha=5{,}4, \beta=10.$

19. $f(x)=\begin{cases} 2x-12, & x\in(6{,}7], \\ 0, & x\notin(6{,}7]. \end{cases}$

$\alpha=2, \beta=6{,}1.$

20. $f(x)=\begin{cases} 2x-16, & x\in(8{,}9], \\ 0, & x\notin(8{,}9]. \end{cases}$

$\alpha=8{,}5, \beta=10.$

21. $f(x)=\begin{cases}2x+4, & x\in(-2,-1],\\ 0, & x\notin(-2,-1].\end{cases}$

$\alpha=-7, \beta=-1{,}5.$

22. $f(x)=\begin{cases}2x, & x\in(0,1],\\ 0, & x\notin(0,1].\end{cases}$

$\alpha=0{,}4, \beta=3.$

23. $f(x)=\begin{cases}2x+8, & x\in(-4,-3],\\ 0, & x\notin(-4,-3].\end{cases}$

$\alpha=-3{,}2, \beta=8.$

24. $f(x)=\begin{cases}2x+10, & x\in(-5,-4],\\ 0, & x\notin(-5,-4].\end{cases}$

$\alpha=-4{,}5, \beta=0.$

В задачах 25-40 задана функція $f(x)$. Потрібно:

1) знайти значення параметра a, при якому вона стає густиною розподілу деякої випадкової величини X;

2) знайти функцію розподілу імовірностей для X;

3) обчислити $\sigma(X)$;

4) побудувати графіки функції розподілу та густини розподілу.

25. $f(x)=\begin{cases}a(x-3), & x\in(3,4],\\ 0, & x\notin(3,4].\end{cases}$

26. $f(x)=\begin{cases}a(x-4), & x\in(4,5],\\ 0, & x\notin(4,5].\end{cases}$

27. $f(x)=\begin{cases}a\cos 0{,}5x, & x\in(0,\pi],\\ 0, & x\notin(0,\pi].\end{cases}$

28. $f(x)=\begin{cases}a(x-2), & x\in(1,2],\\ 0, & x\notin(1,2].\end{cases}$

29. $f(x)=\begin{cases}a(x-4), & x\in(4,5],\\ 0, & x\notin(4,5].\end{cases}$

30. $f(x)=\begin{cases}a(x-5), & x\in(5,6],\\ 0, & x\notin(5,6].\end{cases}$

31. $f(x)=\begin{cases}a(x-6), & x\in(6,7],\\ 0, & x\notin(6,7].\end{cases}$

32. $f(x)=\begin{cases}a(x+1), & x\in(-1,0],\\ 0, & x\notin(-1,0].\end{cases}$

33. $f(x)=\begin{cases}a(x-1), & x\in(1,3/2],\\ 0, & x\notin(1,3/2].\end{cases}$

34. $f(x)=\begin{cases}a(x-1), & x\in(1,2],\\ 0, & x\notin(1,2].\end{cases}$

35. $f(x)=\begin{cases}a(x-1/2), & x\in(-2,0],\\ 0, & x\notin(-2,0].\end{cases}$

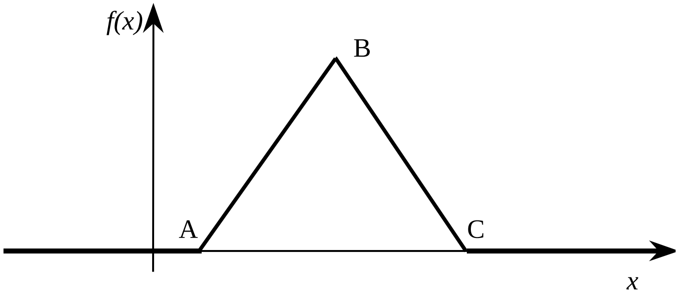
36. $f(x)=\begin{cases}a\,x/9, & x\in[0,3),\\ 0, & x\notin[0,3).\end{cases}$

37. $f(x)=\begin{cases}a(x+2), & x\in[-1,0),\\ 0, & x\notin[-1,0).\end{cases}$

38. $f(x)=\begin{cases}a\sin x, & x\in(\pi,2\pi],\\ 0, & x\notin(\pi,2\pi].\end{cases}$

39. $f(x)=\begin{cases} a(x-2), & x\in(2,3], \\ 0, & x\notin(2,3]. \end{cases}$ 40. $f(x)=\begin{cases} ax^{-1/2}, & x\in(0,1], \\ 0, & x\notin(0,1]. \end{cases}$

В задачах 41^{*}-55^{*} графік густини розподілу неперервної випадкової величини X зображений на мал.5.5



Мал.5.5.

На підставі інформації про координати точок A, B та C: 1) знайти аналітичні вирази густини розподілу та функції розподілу імовірностей цієї величини; 2) побудувати графік функції $F(x)$; 3) обчислити середнє квадратичне відхилення, моду та медіану.

Номер задачі	Координати точок		
	A	*B*	*C*
41	(1;0)	(3;*y*)	(5;0)
42	(-1;0)	(2;*y*)	(4;0)
43	(1;0)	(4;*y*)	(6;0)
44	(1;0)	(3;*y*)	(4;0)
45	(2;0)	(3;*y*)	(5;0)
46	(2;0)	(4;*y*)	(5;0)
47	(-2;0)	(0;*y*)	(3;0)
48	(-1;0)	(3;*y*)	(5;0)
49	(1;0)	(3;*y*)	(6;0)
50	(2;0)	(4;*y*)	(6;0)
51	(-2;0)	(1;*y*)	(4;0)
52	(-2;0)	(2;*y*)	(5;0)
53	(-3;0)	(0;*y*)	(2;0)
54	(-3;0)	(0;*y*)	(4;0)
55	(-3;0)	(1;*y*)	(2;0)

§ 6. ОСНОВНІ ЗАКОНИ НЕПЕРЕРВНИХ ВИПАДКОВИХ ВЕЛИЧИН

Нормальний закон (імовірнісний зміст параметрів розподілу; нормальна крива; імовірність попадання у заданий інтервал; знаходження імовірності заданого відхилення). Закон рівномірного розподілу. Показниковий закон. Гамма-розподіл та розподіл Ерланга. Розподіл χ^2 (*хі*-квадрат).

КОРОТКІ ТЕОРЕТИЧНІ ВІДОМОСТІ

Випадкова величина називається **нормально розподіленою** (**розподіленою за нормальним законом**), якщо її густина розподілу має такий вид

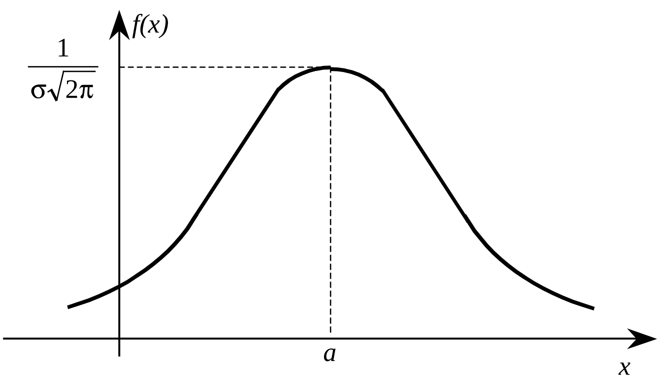
$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{\frac{-(x-a)^2}{2\sigma^2}}. \qquad (6.1)$$

Густина розподілу (6.1) повністю визначається двома параметрами: a та σ. Імовірнісний зміст цих параметрів визначається такими рівностями:

$$a = M(X), \qquad \sigma = \sqrt{D(X)}. \qquad (6.2)$$

Нормальний закон розподілу випадкової величини з параметрами a=0, σ=1 називається **стандартним** або **нормованим**.

Графік густини нормального розподілу називається нормальною кривою (крива Гаусса) (мал.6.1).



Мал. 6.1.

Імовірність того, що нормально розподілена випадкова величина при випробуванні набере значення з інтервалу (α, β), обчислюється за формулою

$$P(\alpha < X < \beta) = \Phi\left(\frac{\beta - a}{\sigma}\right) - \Phi\left(\frac{\alpha - a}{\sigma}\right). \tag{6.3}$$

В практично важливих задачах виникає необхідність в знаходженні імовірності того, що відхилення нормально розподіленої випадкової величини від a по абсолютній величині менше від заданого додатнього числа ε, тобто $P\left(|X - a| < \varepsilon\right)$.

Ця імовірність обчислюється за формулою

$$P\left(|X - a| < \varepsilon\right) = 2\Phi\left(\frac{\varepsilon}{\sigma}\right), \tag{6.4}$$

яка, зокрема, показує, що імовірність відхилення для одного і того ж ε буде тим більша, чим меншим буде σ.

Випадкова величина називається **розподіленою за рівномірним законом (рівномірно розподіленою)**, якщо її густина розподілу має такий вид

$$f(x) = \begin{cases} C = \text{const}, & \text{якщо } x \in [a, b], \\ 0, & \text{якщо } x \notin [a, b]. \end{cases}$$

Знайдемо значення сталої C, використавши другу властивість густини розподілу. Геометричний зміст цієї властивості означає рівність одиниці площі, обмеженої кривою розподілу, тобто $C(b - a) = 1$, звідки $C = 1/(b - a)$. Отже, густина рівномірно розподіленої випадкової величини має такий вид:

$$f(x) = \begin{cases} 1/(b - a), & \text{при } x \in [a, b], \\ 0, & \text{при } x \notin [a, b]. \end{cases} \tag{6.5}$$

Відрізок $[a, b]$ називається **відрізком концентрації рівномірного розподілу**.

Випадкова величина X розподіляється за **показниковим (експоненціальним) законом (показниково розподілена)**, якщо її густина розподілу імовірностей має такий вид

$$f(x) = \begin{cases} \lambda e^{-\lambda x}, & \text{якщо } x \geq 0, \\ 0, & \text{якщо } x < 0, \end{cases} \tag{6.6}$$

де стала $\lambda > 0$ називається параметром експоненціального розподілу.

Моделювання складних явищ та процесів (зокрема, економічних) передбачає використання адекватних математичних моделей. Показниковий розподіл є частинним випадком так званого гамма-розподілу, який успішно використовується в техніко-економічному моделюванні і який, зокрема, достатньо добре описує час безвідмовної роботи різних технічних пристроїв та обладнання.

Невід’ємна неперервна випадкова величина X має **гамма-розподіл імовірностей**, якщо густина розподілу має такий вид:

$$f(x)=\begin{cases}\dfrac{\lambda^{\alpha}}{\Gamma(\alpha)}x^{\alpha-1}e^{-\lambda x}, & \text{якщо } x\ge 0,\\ 0, & \text{якщо } x<0,\end{cases} \tag{6.7}$$

де $\alpha>0$, $\lambda>0$ – параметри розподілу, $\Gamma(\alpha)=\int\limits_0^{\infty} z^{\alpha-1}e^{-z}dz$ – гамма-функція (інтеграл Ейлера).

Наведемо властивості гамма-функції, корисні при користуванні гамма-розподілом:

1) $\Gamma(1)=1$; 2) $\Gamma(\alpha+1)=\alpha\Gamma(\alpha)$; 3) якщо $\alpha=n$ – ціле невід’ємне число, то $\Gamma(n+1)=n(n-1)(n-2)\ldots 2\cdot 1=n!$; 4) $\Gamma(1/2)=\sqrt{\pi}$.

Із врахуванням властивості 1) видно, що при $\alpha=1$ з (6.7) отримується (6.6), тобто при $\alpha=1$ отримується показниковий розподіл.

Використання властивостей гамма-функції дозволяє знайти числові характеристики гамма-розподіленої випадкової величини X:

$$M(X)=\alpha/\lambda,\qquad D(X)=\alpha/\lambda^2,\qquad \sigma(X)=\sqrt{\alpha}/\lambda. \tag{6.8}$$

Якщо параметр α набуває цілих значень ($\alpha=k=1, 2, 3, \ldots$), то гамма-розподіл перетворюється на розподіл Ерланга k-го порядку:

$$f(x)=\begin{cases}\dfrac{\lambda(\lambda x)^{k-1}}{(k-1)!}e^{-\lambda x}, & \text{якщо } x\ge 0,\\ 0, & \text{якщо } x<0.\end{cases}$$

Розподіл Ерланга знаходить важливе застосування в теорії масового обслуговування.

Числові характеристики цього розподілу отримуються із співвідношень (6.8), де $\alpha=k$.

Частинним випадком гамма-розподілу при $\alpha=k/2$, де $k=1, 2, 3, \ldots$, а $\lambda=1/2$, є так званий розподіл χ^2 (*хі*-квадрат), значення якого в математичній статистиці важко переоцінити; параметр k називається в цьому випадку числом ступенів вільності розподілу χ^2. Сформулюємо інший підхід до визначення розподілу χ^2.

Нехай задано k незалежних випадкових величин $X_1, X_2, \ldots, X_k$, кожна з яких має **нормований нормальний закон розподілу імовірностей**, тобто $M(X_i) = 0$, $\sigma(X_i) = 1$, $i = \overline{1, k}$. Тоді випадкова величина $X = \sum_{i=1}^{k} X_i^2$ матиме розподіл χ^2 із k ступенями вільності, тобто густина розподілу імовірностей X матиме такий вид:

$$f(x) = \begin{cases} 0, & \text{якщо } x < 0, \\ \dfrac{1}{2^{k/2}\,\Gamma(k/2)} x^{(k/2)-1} e^{-x/2}, & \text{якщо } x \geq 0. \end{cases} \qquad (6.9)$$

Числові характеристики випадкової величини, розподіленої за законом χ^2, визначаються рівностями:

$$M(X) = k, \quad D(X) = 2k, \qquad \sigma(X) = \sqrt{2k}.$$

Із збільшення числа ступенів вільності розподіл χ^2 «повільно» наближається до нормального.

РОЗВ'ЯЗУВАННЯ ТИПОВИХ ЗАДАЧ

Задача 6.1. Коробки із шоколадом упаковуються автоматично, при цьому середня маса однієї коробки становить 1,04 кг. Відомо, що тільки 2,5% коробок мають масу, меншу від 1 кг. Припускаючи, що маса коробок розподілена нормально, знайти середнє квадратичне відхилення.

○ Позначимо: X – маса навмання взятої коробки із шоколадом. За умовою X – нормально розподілена випадкова величина, $M(X) = 1{,}04$ і $P(X < 1) = 0{,}025$. Протилежною до події $(X < 1)$ є подія $(1 \leq X < \infty)$, тому $P(1 \leq X < \infty) = 1 - P(X < 1) = 1 - 0{,}025 = 0{,}975$. Для знаходження σ використаємо формулу (6.3), де $a = 1{,}04$, $\alpha = 1$, $\beta = \infty$, тобто

$$P(1 \leq X < \infty) = 0{,}975 = \Phi\left(\frac{\infty - 1{,}04}{\sigma}\right) - \Phi\left(\frac{1 - 1{,}04}{\sigma}\right),$$

або

$$0{,}975 = 0{,}5 + \Phi\left(\frac{0{,}04}{\sigma}\right),$$

оскільки $\Phi(x)$ – непарна функція і $\Phi(x) = 0{,}5$ для $x > 5$.

Остаточно рівняння набуває такого вигляду

$$\Phi\left(\frac{0{,}04}{\sigma}\right) = 0{,}475.$$

За табл. 3 додатків Φ(1,96) = 0,475, тому

$$0{,}04/\sigma = 1{,}96, \text{ звідки } \sigma = 0{,}04/1{,}96 \approx 0{,}0204. \quad \bullet$$

Задача 6.2. На автоматичному токарному станку виготовляють болти, номінальна довжина яких 40 мм. В процесі роботи станка спостерігаються випадкові відхилення від вказаного розміру, які розподілені за нормальним законом з математичним сподіванням 0 і середнім квадратичним відхиленням 1 мм. При контролі бракуються всі болти, розміри яких відрізняються від номінального більше, ніж на допуск в 2 мм. Знайти імовірність того, що навмання взятий болт виявиться бракованим.

○ Нехай випадкова величина X – відхилення розміру навмання взятого болта від номінального. За умовою задачі X розподілена за нормальним законом і $M(X) = 0$, $\sigma(X) = 1$. Потрібно знайти $P\left(|X| > 2\right)$. Але випадкова подія $\left(|X| > 2\right)$ протилежна до події $\left(|X| \le 2\right)$, тому $P\left(|X| > 2\right) = 1 - P\left(|X| \le 2\right)$.

Для знаходження імовірності в правій частині рівності можна використати формулу (6.4) з такими значеннями параметрів: $a = 0$, $\varepsilon = 2$, $\sigma = 1$.

Тоді $\quad P\left(|X| \le 2\right) = 2\Phi(2) = 2 \cdot 0{,}4772 = 0{,}9544,$

$$P\left(|X| > 2\right) = 1 - 0{,}9544 = 0{,}0456. \quad \bullet$$

Задача 6.3. Випадкова величина X розподілена за нормальним законом із математичним сподіванням $a = 0$ і середнім квадратичним відхиленням σ. При якому значенні σ імовірність того, що при випробуванні величина X набуде можливого значення з інтервалу (1, 2), буде максимальною?

○ За формулою (6.3)

$$P(1 < X < 2) = \Phi\left(\frac{2}{\sigma}\right) - \Phi\left(\frac{1}{\sigma}\right) = \frac{1}{\sqrt{2\pi}}\left(\int_0^{2/\sigma} e^{-x^2/2}dx - \int_0^{1/\sigma} e^{-x^2/2}dx\right).$$

Ця імовірність є функцією однієї змінної σ і для знаходження її максимального значення знайдемо критичні точки, продиференціювавши обидві частини і прирівнявши до нуля похідну:

$$\frac{d}{d\sigma}P(1< X < 2) = \frac{1}{\sqrt{2\pi}}\left(e^{-(2/\sigma)^2/2}\,(2/\sigma)' - e^{-(1/\sigma)^2/2}\,(1/\sigma)' \right) =$$

$$= \frac{1}{\sqrt{2\pi}}\left(e^{-2/\sigma^2}(-2/\sigma^2) - e^{-1/(2\sigma^2)}(-1/\sigma^2)\right) = 0.$$

Звідси $2e^{-2/\sigma^2} = e^{-1/2\sigma^2}$. Прологарифмувавши ліву і праву частину за основою e, після простих перетворень отримаємо:

$$\sigma = \sqrt{3/(2\ln 2)} \approx \sqrt{3/(2\cdot 0{,}693)} \approx 1{,}471.$$

Можна перевірити, що $\frac{d}{d\sigma}P(1< X < 2)$ при проходженні через знайдену точку змінює знак з «+» на «–», а тому $\sigma \approx 1{,}471$ надає максимального значення імовірності випадкової події $(1 < X < 2)$. ●

Задача 6.4. Випадкова величина *X* розподілена за нормальним законом із математичним сподіванням 7,6. Імовірність того, що при випробуванні *X* набере значення з інтервалу (7,65; 7,68), дорівнює 0,25. Знайти Р(7,52<*X*<7,55).

○ Нормальна крива є симетричною відносно прямої x=7,6, а інтервали однакової довжини (7,52; 7,55) і (7,65; 7,68) – симетричні відносно точки x=7,6. Тому рівними є площі криволінійних трапецій, основами яких є вказані відрізки і які обмежені зверху нормальною кривою. За формулою (5.6) $P(7{,}52 < X < 7{,}55) = \int_{7{,}52}^{7{,}55} f(x)dx$, де визначений інтеграл – площа криволінійної трапеції із основою (7,52; 7,55), яка обмежена зверху частиною нормальної кривої. Із рівності площ розглянутих криволінійних трапецій отримуємо шукану імовірність: P(7,52<*X*<7,55)=0,25. ●

Задача 6.5. Випадкова величина *Y* розподілена за нормальним законом з числовими характеристиками $M(Y)$=15, $D(Y)$=9. Знайти, при яких значеннях t виконується рівність

$$P\left[|Y - M(Y)| < t\sqrt{D(Y)}\right] = P\left[|Y - M(Y)| \ge t\sqrt{D(Y)}\right]. \quad (6.10)$$

○ Випадкові події $\left[|Y - M(Y)| < t\sqrt{D(Y)}\right]$ та $\left[|Y - M(Y)| \ge t\sqrt{D(Y)}\right]$ протилежні, тому

$$P\left[|Y - M(Y)| < t\sqrt{D(Y)}\right] + P\left[|Y - M(Y)| \ge t\sqrt{D(Y)}\right] = 1,$$

звідки $P\left[\left|Y-M(Y)\right|\geq t\sqrt{D(Y)}\right]=1-P\left[\left|Y-M(Y)\right|<t\sqrt{D(Y)}\right].$ (6.11)

Підставивши (6.11) в праву частину (6.10), отримаємо

$$P\left[\left|Y-M(Y)\right|<t\sqrt{D(Y)}\right]=1-P\left[\left|Y-M(Y)\right|<t\sqrt{D(Y)}\right]$$

або

$$P\left[\left|Y-M(Y)\right|<t\sqrt{D(Y)}\right]=0{,}5\,.$$

З урахуванням умов задачі остання рівність набере такого виду

$$P(\left|Y-15\right|<3t)=0{,}5\,.$$

Використаємо формулу (6.4)

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

де a=M(X), $\sigma=\sqrt{D(Y)}$.

В даному випадку $\varepsilon=3t,\sigma=3.$ Тому для знаходження невідомого значення t отримаємо рівняння

$$2\Phi\left(\frac{3t}{3}\right)=0{,}5 \text{ або } \Phi(t)=0{,}25\,\textbf{.}$$

За табл.3 додатків знаходимо $t\approx 0{,}675$, використавши лінійну інтерполяцію. ●

Задача 6.6. Випадкова величина X розподілена за нормальним законом із середнім значенням 10. Знайти P(8<X<20), якщо P(6<X<14)=0,5.

○ За умовою задачі параметр a=10, а параметр σ - невідомий. Для знаходження його використаємо рівність P(6<X<14)=0,5, формулу (6.3) і непарність функції Лапласа:

$$P(6<X<14)=\Phi\left(\frac{14-10}{\sigma}\right)-\Phi\left(\frac{6-10}{\sigma}\right)=\Phi\left(\frac{4}{\sigma}\right)-\Phi\left(\frac{-4}{\sigma}\right)=2\Phi\left(\frac{4}{\sigma}\right);$$

$$2\Phi\left(\frac{4}{\sigma}\right)=0{,}5\Rightarrow\Phi\left(\frac{4}{\sigma}\right)=0{,}25\Rightarrow\frac{4}{\sigma}=0{,}675\Rightarrow\sigma=\frac{4}{0{,}675}\approx 5{,}93.$$

Тоді за формулою (6.3)

$$P(8<X<20)=\Phi\left(\frac{20-10}{5{,}93}\right)-\Phi\left(\frac{8-10}{5{,}93}\right)=\Phi(1{,}69)-\Phi(-0{,}34)=$$

$$=\Phi(1{,}69)+\Phi(0{,}34)=0{,}45449+0{,}13307=0{,}58756.$$ ●

Задача 6.7. Швейна фабрика виготовляє костюми, орієнтуючись на покупців конкретного регіону. Покладаючи, що ріст чоловіків певної вікової групи цього регіону є нормально розподіленою випадковою величиною X із параметрами a=174см і σ=5см, знайти: 1) густину розподілу імовірностей величини X; 2) частки костюмів третього росту (171-176см) і четвертого росту (176-181см), які потрібно передбачити в загальному обсязі виробництва для даної вікової групи.

○ 1) Згідно з умовою задачі густина розподілу імовірностей даної випадкової величини має вид (6.1):

$$f(x)=\frac{1}{5\sqrt{2\pi}}e^{-(x-174)^2/50}.$$

2) Нехай одиниця – загальний обсяг виробництва костюмів (всіх ростів). Тоді частка костюмів третього росту визначиться як імовірність $P(171<X<176)$, яку знайдемо за формулою (6.3):

$$P(171<X<176)=\Phi\left(\frac{176-174}{5}\right)-\Phi\left(\frac{171-174}{5}\right)=\Phi(0{,}4)+\Phi(0{,}6)=$$

$$=0{,}15542+0{,}22575=0{,}38117.$$

Аналогічно знаходимо частку костюмів четвертого росту:

$$P(176<X<181)=\Phi\left(\frac{181-174}{5}\right)-\Phi\left(\frac{176-174}{5}\right)=\Phi(1{,}4)-\Phi(0{,}4)=$$

$$=0{,}41924-0{,}15542=0{,}26382.$$

Отримані результати можна інтерпретувати таким чином: костюми третього росту повинні складати 38,1% всієї продукції, а четвертого – 26,4%. ●

Задача 6.8. Густина розподілу випадкової величини X має такий вид:

$$f(x)=\frac{1}{5\sqrt{2\pi}}e^{-(x-30)^2/50}.$$

1) Для можливих значень x_1=26 і x_2=38 знайти кількість середніх квадратичних відхилень від математичного сподівання, а також $P(X<x_1)$ та $P(X>x_2)$. 2) Знайти інтервал, в який з імовірністю 0,9861 потрапить при випробуванні можливе значення X.

○ Враховуючи формулу (6.1) і дані задачі, робимо висновок, що величина X розподілена за нормальним законом, при цьому $M(X)=a=30$, σ=5. Тоді

$$x_1 = 26 = 30 - 4 = 30 - \frac{4}{5} \cdot 5 = 30 - 0{,}8\sigma,$$

$$x_2 = 38 = 30 + 8 = 30 + \frac{8}{5} \cdot 5 = 30 + 1{,}6\sigma.$$

Отже, можливе значення x_1 відхиляється від математичного сподівання вліво на величину 0,8σ, а можливе значення x_2 – вправо на величину 1,6σ. Використавши формулу (6.3), непарність функції Лапласа, рівність $\Phi(x)=0{,}5$, для $x>5$, а також табл.3 додатків, отримаємо:

$$P(X < x_1) = P(-\infty < X < 26) = \Phi\left(\frac{26-30}{5}\right) - \Phi\left(\frac{-\infty-30}{5}\right) =$$

$$= \Phi(-0{,}8) + 0{,}5 = -0{,}28814 + 0{,}5 = 0{,}21186;$$

$$P(X > x_2) = P(38 < X < \infty) = \Phi\left(\frac{\infty-30}{5}\right) - \Phi\left(\frac{38-30}{5}\right) =$$

$$= 0{,}5 - \Phi(1{,}6) = 0{,}5 - 0{,}44520 = 0{,}0548.$$

2) Згідно з правилом трьох сигм практично всі можливі значення величини X зосереджені в інтервалі $(a–3\sigma;\ a+3\sigma)$, симетричному відносно a. Тому будемо вважати, що шуканий інтервал має вид $(30–\varepsilon,\ 30+\varepsilon)$, де ε – невідоме число.

Згідно з умовою $P(a–\varepsilon<X<a+\varepsilon)=0{,}9861$

або $P(|X-30| < \varepsilon) = 0{,}9861$,

за формулою (6.4)

$$P(|X-30| < \varepsilon) = 2\Phi\left(\frac{\varepsilon}{\sigma}\right) = 2\Phi\left(\frac{\varepsilon}{5}\right).$$

Тому

$$2\Phi\left(\frac{\varepsilon}{5}\right) = 0{,}9861 \Rightarrow \Phi\left(\frac{\varepsilon}{5}\right) = 0{,}49305\,.$$

За табл.3 додатків $\varepsilon/5=2{,}46$ звідки $\varepsilon = 5 \cdot 2{,}46 = 12{,}3$.

Отже, шуканий інтервал має вид

(30–12,3; 30+12,3)

або остаточно (17,7; 42,3). ●

Зауваження. Знайдений інтервал, в який з імовірністю 0,9861 потрапить при випробуванні можливе значення випадкової величини X, – неєдиний. Справді, нехай число α таке, що $P(X>\alpha)=0{,}9861$. Використавши формулу (6.3), отримаємо

$$P(X>\alpha)=P(\alpha<X<\infty)=\Phi\left(\frac{\infty-30}{5}\right)-\Phi\left(\frac{\alpha-30}{5}\right), \text{ звідки}$$

$$\Phi\left(\frac{30-\alpha}{5}\right)=0,4861,\Rightarrow \frac{30-\alpha}{5}=2,2\Rightarrow \alpha=19.$$

При цьому використовується нерівність $\alpha<30$ (проаналізуйте чому!). Отже, інтервал (19; ∞) – теж шуканий. Аналогічно можна було б показати, що в інтервал (–∞; 41) з тією ж імовірністю попадає можливе значення випадкової величини.

Задача 6.9. Для рівномірно розподіленої випадкової величини X знайти: 1) функцію розподілу імовірностей та побудувати її графік; 2) математичне сподівання; 3) дисперсію.

○ Густину (6.5) рівномірного розподілу напишемо в такому вигляді

$$f(x)=\begin{cases}0, & x<a,\\ 1/(b-a), & a\le x\le b,\\ 0, & x>b.\end{cases}$$

1) Використаємо формулу (5.7)

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

Якщо $x<a$, тоді $F(x)=0$. Нехай $x\in[a, b]$, тоді

$$F(x)=\int_{-\infty}^{x} f(x)dx=\int_{-\infty}^{a} f(x)dx+\int_{a}^{x} f(x)dx=\int_{-\infty}^{a} 0dx+\int_{a}^{x}\frac{1}{b-a}dx=$$

$$=\frac{1}{b-a}x\Big|_{a}^{x}=\frac{x-a}{b-a}.$$

Якщо ж $x>b$, то

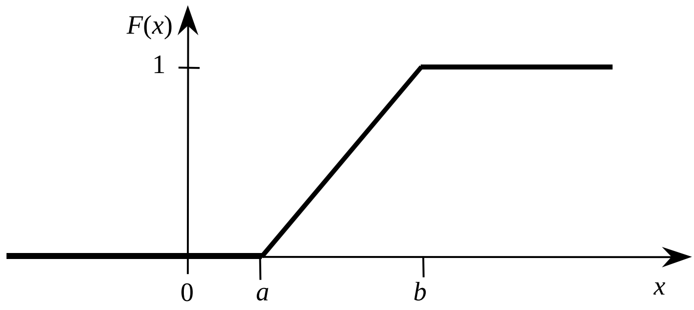
$$F(x)=\int_{-\infty}^{x} f(x)dx=\int_{-\infty}^{a} f(x)dx+\int_{a}^{b} f(x)dx+\int_{b}^{x} f(x)dx=$$

$$=\int_{-\infty}^{a} 0dx+\int_{a}^{b}\frac{dx}{b-a}+\int_{b}^{x}0dx=\frac{x}{b-a}\Big|_{a}^{b}=1.$$

Отже, шукана функція розподілу має такий вид

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

Її графік зображено на мал.6.2



Мал.6.2.

2) Для знаходження математичного сподівання використаємо формули (5.10*) та (6.5)

$$M(X)=\int_{-\infty}^{\infty}xf(x)dx=\int_{-\infty}^{a}xf(x)dx+\int_{a}^{b}xf(x)dx+\int_{b}^{\infty}xf(x)dx=$$

$$=\int_{-\infty}^{a}x0dx+\int_{a}^{b}x\frac{1}{(b-a)}dx+\int_{b}^{\infty}x0dx=0+\frac{1}{(b-a)}\int_{a}^{b}xdx+0=$$

$$=\frac{1}{(b-a)}\cdot\frac{x^2}{2}\bigg|_a^b=\frac{b^2}{2(b-a)}-\frac{a^2}{2(b-a)}=\frac{(b-a)(b+a)}{2(b-a)}=\frac{b+a}{2}.$$

3) Дисперсію знайдемо, використавши формули (5.12*) та (6.5)

$$D(X)=\int_{-\infty}^{\infty}x^2f(x)dx-[M(X)]^2=\int_{-\infty}^{a}x^2f(x)dx+\int_{a}^{b}x^2f(x)dx+$$

$$+\int_{b}^{\infty}x^2f(x)dx-\left(\frac{b+a}{2}\right)^2=\int_{-\infty}^{a}x^20dx+\int_{a}^{b}x^2\frac{1}{(b-a)}dx+\int_{b}^{\infty}x^20dx-$$

$$-\frac{(b+a)^2}{4}=0+\frac{1}{(b-a)}\int_{a}^{b}x^2dx+0-\frac{(b+a)^2}{4}=\frac{x^3}{3(b-a)}\bigg|_a^b-\frac{(b+a)^2}{4}=$$

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

$$=\frac{(b-a)^2}{12}.$$ ●

Задача 6.10. Перехрестя обладнане автоматичним світлофором, на якому зелене та червоне світло горить протягом 1 хв. та 0,75 хв. відповідно. Автомобіль під’їжджає до перехрестя у випадковий момент часу, не пов'язаний із роботою світлофора. Знайти імовірність того, що він проїде перехрестя, не зупиняючись.

○ Позначимо випадкову величину X – момент приїзду автомобіля до перехрестя. Її можливі значення знаходяться в проміжку (0; 1,75), довжина якого визначається періодом зміни кольорів на світлофорі. Зрозуміло, що можливість приїзду автомобілем в будь-який момент часу даного інтервалу з (0; 1,75) рівна можливості приїзду в момент часу іншого інтервалу такої ж довжини. Тому X – рівномірно розподілена з густиною розподілу (6.5), де $a = 0$, $b = 1{,}75$. Автомобіль проїде, не зупиняючись, перехрестя, якщо відбудеться випадкова подія $(0 < X < 1)$, тобто під час «дії» зеленого світла світлофора. Тоді за формулою (5.6)

$$P(0 < X < 1) = \int_0^1 f(x)dx = \frac{1}{1{,}75}\int_0^1 dx = 4/7. \quad \bullet$$

Задача 6.11. Знайти числові характеристики показниково розподіленої випадкової величини X.

○ За формулами (5.10*) та (6.6):

$$M(X) = \int_{-\infty}^{\infty} xf(x)dx = \int_{-\infty}^{0} x \cdot 0dx + \int_0^{\infty} x\lambda e^{-\lambda x}dx = \lambda\int_0^{\infty} xe^{-\lambda x}dx =$$

$$= \begin{vmatrix} u = x, & dv = e^{-\lambda x}dx \\ du = dx, & v = -\dfrac{1}{\lambda}e^{-\lambda x} \end{vmatrix} = \lambda\left\{-\frac{x}{\lambda}e^{-\lambda x}\Bigg|_0^{\infty} + \frac{1}{\lambda}\int_0^{\infty} e^{-\lambda x}dx\right\} = -\frac{1}{\lambda}e^{-\lambda x}\Bigg|_0^{\infty} = \frac{1}{\lambda}.$$

Для знаходження $D(X)$ скористаємось формулами (5.12*) та (6.6):

$$D(X) = \int_{-\infty}^{\infty} x^2 f(x)dx - [M(X)]^2 = \lambda\int_0^{\infty} x^2 e^{-\lambda x}dx - \frac{1}{\lambda^2};$$

$$\int_0^{\infty} x^2 e^{-\lambda x}dx = \begin{vmatrix} u = x^2, & du = 2xdx \\ dv = e^{-\lambda x}dx, & v = -\dfrac{1}{\lambda}e^{-\lambda x} \end{vmatrix} =$$

$$= -\frac{x^2}{\lambda}e^{-\lambda x}\Bigg|_0^{\infty} + 2\lambda^{-1}\int_0^{\infty} xe^{-\lambda x}dx = \begin{vmatrix} u = x, & du = dx \\ dv = e^{-\lambda x}dx, & v = -\dfrac{1}{\lambda}e^{-\lambda x} \end{vmatrix} =$$

$$=\frac{2}{\lambda}\left\{-\frac{x}{\lambda}e^{-\lambda x}\Big|_0^{\infty}+\frac{1}{\lambda}\int_0^{\infty}e^{-\lambda x}dx\right\}=\frac{-2}{\lambda^3}e^{-\lambda x}\Big|_0^{\infty}=\frac{2}{\lambda^3};$$

$$D(X)=\lambda\frac{2}{\lambda^3}-\frac{1}{\lambda^2}=\frac{1}{\lambda^2};\qquad \sigma(X)=\frac{1}{\lambda}.\quad \bullet$$

Задача 6.12. Статистично встановлено, що середній час очікування наступного покупця, який підійде до каси, дорівнює 0,2 хвилини. Час чекання касиром чергового покупця можна вважати випадковою величиною, розподіленою за показниковим законом. Для заміни стрічки касового апарату потрібно 2 хвилини. Знайти імовірність того, що за час заміни стрічки не утвориться черга покупців.

○ Нехай X – час чекання касиром чергового покупця. За умовою X розподілена за показниковим законом і $M(X)$=0,2. Але $M(X)=\frac{1}{\lambda}$ (див. задачу 6.11). Тому $1/\lambda=0{,}2$ звідки λ=5.

Нехай випадкова подія A – за час заміни стрічки касового апарату не утвориться черга. Тоді протилежна до неї подія

$\overline{A}=[M(X)\le X\le 2]$.

Але оскільки $M(X)$=0,2, то за формулою (5.6)

$$P(\overline{A})=P[M(X)\le X\le 2]=P(0{,}2\le X\le 2)=\int_{0,2}^{2}f(x)dx=$$

$$=\int_{0,2}^{2}5e^{-5x}dx=-\int_{0,2}^{2}e^{-5x}d(-5x)=-e^{-5x}\Big|_{0,2}^{2}=-e^{-10}+e^{-1}=e^{-1}-e^{-10}.$$

Отже, $P(A)=1-e^{-1}+e^{-10}$. ●

ЗАДАЧІ ДЛЯ САМОСТІЙНОГО РОЗВ'ЯЗУВАННЯ

Завдання №7

1. Аналіз статистичних даних показав, що місячний дохід працюючих мешканців одного з регіонів має нормальний розподіл із середнім значенням 9,2 тис. грн. і середнім квадратичним відхиленням 1,1 тис. грн. Знайти імовірність того, що місячний дохід навмання взятого працюючого мешканця виявиться:
   а) в межах 8,8 тис. грн. та 9,5 тис. грн.;

б) не меншим, ніж 9,5 тис. грн.;
в) не більшим 10 тис. грн.

2. Випадкова величина X розподілена за нормованим нормальним законом, а величина Y – за рівномірним законом на проміжку [0; 1]. Порівняти $P(0<X<1)$ та $P(0<Y<1)$.

3. Хвилинна стрілка електронного годинника переміщується стрибком через кожну хвилину. Дійсний час у фіксований момент часу буде випадковою величиною, що має рівномірний закон розподілу. Звіряється час за електронним годинником, заокруглюючи положення хвилинної стрілки до найближчої хвилинної поділки циферблата. Яка імовірність того, що похибка заокруглення часу не перевищить 15 с?

4. Випадкова величина X задана густиною розподілу

$$f(x) = \frac{1}{4\sqrt{2\pi}} e^{-(x-10)^2/32}.$$

Знайдіть інтервал, в який з імовірністю 0,9876 потрапить при випробуванні можливе значення X.

5. Неперервна випадкова величина X рівномірно розподілена на проміжку [1; 3]. Знайти: 1) функцію розподілу імовірностей і побудувати її графік; 2) $M(X)$ і $\sigma(X)$; 3) імовірність того, що X при випробуванні набере значення з інтервалу (1; 2).

6. Випадкова величина X розподілена за нормальним законом з числовими характеристиками $M(X)$=15; $D(X)$=25. Знайти, при яких значеннях $\varepsilon>0$ виконується рівність.

$$2P\left[\left|X - M(X)\right| < \varepsilon\sqrt{D(x)}\right] = P\left[\left|X - M(X)\right| \geq \varepsilon\sqrt{D(X)}\right].$$

7. Жирність молока корів фермерських господарств району (у відсотках) є нормально розподіленою випадковою величиною з математичним сподіванням 3,9 і дисперсією 0,0225. Записати густину розподілу цієї величини. Знайти імовірність того, що найближча партія молока матиме жирність, не меншу від 4,1.

8. Швейне підприємство виготовляє на експорт костюми. Вважаючи, що ріст чоловіків певної вікової групи країни-імпортера є нормально розподіленою випадковою величиною X із параметрами a=172 см і σ=5 см, знайти:
1) густину розподілу імовірностей величини X;

2) частки костюмів третього росту (169-174 см) і четвертого росту (174-179 см), які потрібно передбачити в загальному обсязі виробництва для даної вікової групи.

9. Випадкова величина розподілена за нормальним законом із середнім значенням 20. Знайти імовірність того, що при випробуванні ця величина набере значення з інтервалу (7;28), якщо імовірність попадання її можливого значення (при випробуванні) в інтервал (17; 23) дорівнює 0,5.

10. Випадкова величина X має показниковий закон розподілу з параметром λ. Знайти $P(X > a)$ та $P(a < X < 2a)$, де $a = const > 0$.

11. Випадкова величина X задана функцією розподілу

$$F(x) = \begin{cases} 0 & \text{при } x \le 0, \\ 0{,}2x & \text{при } 0 < x \le 5, \\ 1 & \text{при } x > 5. \end{cases}$$

Потрібно знайти: а) густину розподілу $f(x)$ і побудувати графіки функцій $F(x)$ і $f(x)$; б) імовірність того, що при випробуванні випадкова величина X набере значення з інтервалу: 1) (0; 3); 2) (0; 5); 3) (1,4; 5); 4) (6; 10). Як називається розподіл цієї величини?

12. Середнє значення доходів робітників за рік складає 80 тис. грн., а дисперсія – 4 тис. грн.^{2}. Для навмання вибраного робітника знайти імовірність того, що він заробляє: а) не менше 78 тис. грн.; б) в межах від 84 до 95 тис. грн.

13. Довжина Х виготовленої верстатом-автоматом деталі є випадкова величина, що описується густиною розподілу

$$f(x) = \frac{1}{2\sqrt{2\pi}} e^{-(x-100)^2/8}.$$

Знайти: 1) імовірність браку відібраної деталі, якщо розмір допускають рівним $100 \pm 0{,}4$ (мм); 2) яку точність довжини деталі, виготовленої верстатом-автоматом, можна гарантувати із імовірністю 0,95.

14. Соняшникова олія “Олейна” розливається в пляшки автоматично, при цьому вага нетто в середньому дорівнює 920 г. Відомо, що тільки 1,5% вмістиме пляшок має вагу, меншу від 915 г. Припускаючи, що вага олії в пляшці розподілена за нормальним законом, знайти середнє квадратичне відхилення.

15. Припускається, що гранична міцність виготовленої партії стального дроту діаметром 1,4 мм є нормально розподіленою випадковою величиною X з математичним сподіванням 150 кг/мм^2 і середнім квадратичним відхиленням 6 кг/мм^2.

Потрібно: 1) знайти густину розподілу імовірностей випадкової величини X; 2) знайти імовірність того, що при випробуванні X набере значення з інтервалу (145; 160); 3) визначити, яке граничне відхилення в одну або іншу сторону границі міцності досліджуваного зразка дроту від математичного сподівання можна трактувати із імовірністю 0,9901.

16. Відсоток вмісту золи у вугіллі є нормально розподіленою величиною з математичним сподіванням, рівним 15%, і середнім квадратичним, рівним 3%. Знайти імовірність того, що в навмання взятій пробі вугілля виявиться від 12 до 16% золи.

17. Довжина болтів, що виготовляються на станку-автоматі, є нормально-розподіленою випадковою величиною із математичним сподіванням $M(X) = 5{,}6$ см. Імовірність того, що навмання взятий болт має розмір з інтервалу (5,65; 5,67), дорівнює 0,2. Знайти $P(5{,}53 < X < 5{,}55)$.

18. Випадкова величина Y розподілена за показниковим законом із параметром $\lambda=2$. Знайти $P\left[\left|Y - M(Y)\right| < \sigma(Y)\right]$.

19. Густина розподілу випадкової величини X має такий вид:

$$f(x) = \frac{1}{6\sqrt{2\pi}} e^{-(x-50)^2/72}.$$

Для можливих значень $x_1 = 45$, $x_2 = 62$ знайти кількість середніх квадратичних відхилень від математичного сподівання, а також $P(X < x_1)$ і $P(X > x_2)$.

20. Ріст дорослого чоловіка задовільно описується нормальним законом розподілу. За статистичними даними середній ріст чоловіків складає 174 см, а середнє квадратичне відхилення дорівнює 6 см. Знайти імовірність того, що ріст навмання взятого чоловіка буде відрізнятись від середнього зросту не менше, ніж на 7 см.

21. Випадкова величина рівномірно розподілена на інтервалі (-1;11). Знайти точку, в якій її функція розподілу імовірностей дорівнює 1/3.

22. Короткотермінова прогнозна ціна акції певного виду може бути змодельована з допомогою нормального закону розподілу з математичним сподіванням 30 гр. од. і середнім квадратичним відхиленням 0,8 гр. од. 1) Знайти імовірність того, що ціна акції виявиться: а) не вищою від 31,4 гр. од.; б) не нижчою від 31,8 гр. од.; в) в межах від 29,8 гр. од. до 30,8 гр. од.. 2) Використовуючи правило трьох сигм, знайти межі, в яких буде знаходитися ціна акції.

23. Час X безвідмовної роботи станка має показниковий розподіл. Імовірність того, що станок не відмовить за 5 годин роботи, дорівнює 0,60653. Знайти $M(X)$, $D(X)$, $M(X^2)$.

24. Тролейбуси надходять на зупинку з інтервалом 4 хв. Припускаючи, що час чекання тролейбуса на зупинці має рівномірний закон розподілу, знайти: 1) функцію розподілу імовірностей; 2) імовірність того, що час чекання тролейбуса не перевищить 1 хв.

25. Густина розподілу випадкової величини X задана формулою

$$f(x)=\begin{cases} Cx, & \text{при } x\in[0,1], \\ 0, & \text{при } x\notin[0,1]. \end{cases}$$

Знайти: а) сталу C; б) функцію розподілу $F(x)$; в) побудувати графіки функцій $f(x)$ та $F(x)$; г) знайти $P(X > 1/2)$ та $P(1/4 < X < 3/4)$.

26. Випадкова величина X розподілена за рівномірним законом на проміжку [-1; 1]. Знайти

$$P\left[|X-M(X)|<\sqrt{D(X)}\right].$$

27. Густина розподілу випадкової величини X має такий вигляд:

$$f(x)=\frac{1}{4\sqrt{2\pi}}e^{-(x-10)^2/32}.$$

Знайти хоча б один інтервал, в який з імовірністю 0,975 потрапить при випробуванні можливе значення X.

28-40. Верстат-автомат виготовляє деталі, які вважаються придатними, якщо відхилення X від проектного розміру за абсолютною величиною не перевищує ε мм. Вважаючи X нормально розподіленою випадковою величиною із середнім квадратичним відхиленням σ мм, знайти: 1) імовірність того, що навмання взята деталь виявиться придатною; 2) найімовірніше число придатних деталей із n перевірених; 3) інтервал, в який з імовірністю 0,9973 потрапляють можливі значення X.

Номер задачі	ε	σ	n	Номер задачі	ε	σ	n
28	0,5	0,1	200	35	1,6	0,5	300
29	0,6	0,2	100	36	0,5	0,2	400
30	0,7	0,3	300	37	0,6	0,15	350
31	0,8	0,4	400	38	0,8	0,3	450
32	0,4	0,1	200	39	1,1	0,4	500
33	1,2	0,6	150	40	1,3	0,5	150
34	0,9	0,5	250	41	1,4	0,4	200

§ 7. СИСТЕМИ ВИПАДКОВИХ ВЕЛИЧИН (БАГАТОВИМІРНІ ВИПАДКОВІ ВЕЛИЧИНИ)

Закон розподілу імовірностей двовимірної дискретної випадкової величини. Функція розподілу двовимірної випадкової величини та її властивості. Густина розподілу імовірностей двовимірної випадкової величини та її властивості. Умовні закони розподілу. Залежні і незалежні випадкові величини. Умовне математичне сподівання. Рівняння регресії. Числові характеристики системи двох випадкових величин. Кореляційний момент. Коефіцієнт кореляції. Система довільного скінченного числа випадкових величин. Кореляційна матриця. Нормальний закон розподілу двовимірної випадкової величини.

КОРОТКІ ТЕОРЕТИЧНІ ВІДОМОСТІ

Системою випадкових величин називається сукупність випадкових величин, яка розглядається як єдине ціле.

Нехай X та Y є випадковими величинами, які або в одному і тому ж випробуванні набирають своїх можливих значень, або розглядаються одночасно. Будемо позначати (X, Y) систему цих величин (двовимірну випадкову величину), де кожну із величин X та Y називатимемо складовими або компонентами. Геометрично (X, Y) можна інтерпретувати як випадкову точку $A(X, Y)$ на площині $x0y$ або випадковий вектор $\overrightarrow{0A}$.

Розглянемо дискретну двовимірну випадкову величину (X, Y), тобто величину, складові якої є дискретними величинами. Множина можливих значень Ω_{XY} такої випадкової величини містить скінченне або зчисленне число точок (x_i, y_j), де $i, j = \overline{1, \infty}$, тобто

$$\Omega_{XY} = \{(x_1, y_1); (x_1, y_2); \ldots; (x_2, y_1); (x_2, y_2); \ldots; (x_3, y_1); (x_3, y_2); \ldots\}.$$

Позначимо:

$$p_{ij} = P\big((X = x_i) \cdot (Y = y_j)\big) = P(X = x_i; Y = y_j), \qquad (7.1)$$

тобто p_{ij} – це імовірність того, що при випробуванні складова X набирає можливого значення x_i і одночасно з цим складова Y набирає значення y_j, де $(x_i, y_j) \in \Omega_{XY}$, крапка – множення подій.

Законом розподілу імовірностей або **розподілом дискретної двовимірної випадкової величини** (X, Y) називається відповідність між парами чисел $(x_i, y_j) \in \Omega_{XY}$ та імовірностями p_{ij}.

Закон розподілу дискретної двовимірної величини (X, Y) задається таблицею, вид якої для випадку скінченного числа точок множини Ω_{XY} визначається табл. 7.1.

Таблиця 7.1

$X \setminus Y$	y_1	y_2	...	y_j	...	y_n
x_1	p_{11}	p_{12}	...	p_{1j}	...	p_{1n}
x_2	p_{21}	p_{22}	...	p_{2j}	...	p_{2n}
...	...	...	...	...	...	...
x_i	p_{i_1}	p_{i_2}	...	p_{ij}	...	p_{in}
...	...	...	...	...	...	...
x_m	p_{m_1}	p_{m_2}	...	p_{mj}	...	p_{mn}

Перший стовпець таблиці містить всі можливі значення складової X, а перший рядок – всі можливі значення складової Y. В клітинці, яка розміщена на перетині рядка «x_i» та стовпця «y_j» вказана імовірність p_{ij} того, що вектор (X, Y) набере значення (x_i, y_j), де $i = \overline{1, m}$, $j = \overline{1, n}$.

Всі можливі (випадкові) події $(X = x_i; Y = y_j)$ при $i = 1, 2, \ldots m$, $j = 1, 2, \ldots n$ утворюють повну групу, а тому $\sum_{i=1}^{m}\sum_{j=1}^{n} p_{ij} = 1$. Цю рівність називають **умовою нормування** для системи (X, Y).

Знання закону розподілу системи двох дискретних випадкових величин дозволяє знайти закони розподілу її складових. Справді, з урахуванням повноти групи подій $(Y = y_1)$, $(Y = y_2)$, ..., $(Y = y_n)$ для $i = \overline{1, m}$ має місце рівність

$$(X = x_i) = (X = x_i)(Y = y_1) + (X = x_i)(Y = y_2) + \ldots + (X = x_i)(Y = y_n),$$

доданки якої є попарно несумісними подіями. Використання теореми додавання імовірностей і позначень (7.1) дає рівність

$$P(X = x_i) = p_{i1} + p_{i2} + \ldots + p_{in}, \qquad (i = \overline{1, m}). \qquad (7.2)$$

Отже, для знаходження $P(X = x_i)$ потрібно додати імовірності рядка «x_i» табл. 7.1. Аналогічно просумувавши імовірності стовпця «y_j», отримаємо:

$$P(Y = y_j) = \sum_{i=1}^{m} p_{ij} \qquad (j = \overline{1, n}). \qquad (7.3)$$

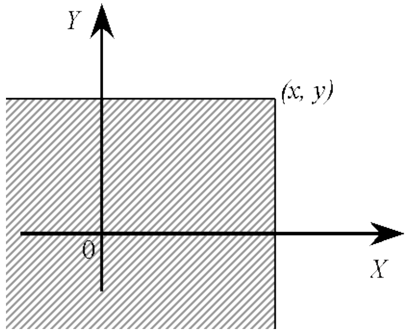
Розглянемо двовимірну випадкову величину (X, Y), складові якої можуть бути дискретними або неперервними величинами і множина можливих значень якої є Ω_{XY}.

Функцією розподілу (імовірностей) двовимірної випадкової величини (X, Y) називається функція $F(x, y)$ детермінованих аргументів

x та y, яка дорівнює імовірності того, що при випробуванні X набере значення, менше від x, і при цьому Y набере значення, менше від y:

$$F(x, y) = P\{(X < x) \cdot (Y < y)\} = P(X < x, Y < y). \quad (7.4)$$

З геометричної точки зору функція розподілу є імовірність попадання випадкової точки (X, Y) в нескінченний квадрант з вершиною в точці (x, y), який знаходиться лівіше і нижче цієї точки (мал. 7.1).



Мал. 7.1.

Властивості функції розподілу імовірностей двовимірної випадкової величини.

1°. $0 \le F(x, y) \le 1$.

2°. $F(x, y)$ є неспадною функцією по кожному аргументу, тобто

$$F(x_2, y) \ge F(x_1, y), \text{ якщо } x_2 > x_1;$$
$$F(x, y_2) \ge F(x, y_1), \text{ якщо } y_2 > y_1.$$

3°. Для $F(x, y)$ виконуються граничні співвідношення

$$\lim_{x \to -\infty} F(x, y) = \lim_{y \to -\infty} F(x, y) = \lim_{\substack{x \to -\infty \\ y \to -\infty}} F(x, y) = 0, \lim_{\substack{x \to +\infty \\ y \to +\infty}} F(x, y) = 1.$$

4°. Якщо один із аргументів $F(x, y)$ прямує до $+\infty$, то вона перетворюється у функцію розподілу випадкової величини, яка відповідає іншому аргументу:

$$F(x, +\infty) = F_1(x), F(+\infty, y) = F_2(y),$$

де $F_1(x)$ і $F_2(y)$ – відповідно функції розподілу випадкових величин X та Y.

5°. $F(x, y)$ неперервна зліва по кожному аргументу.

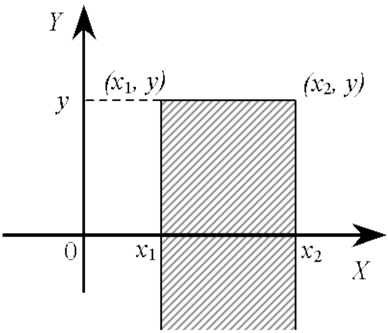
Властивості 1°-5° є **характеристичними**. Це означає, що кожна функція $F(x, y)$, яка володіє цими властивостями, є функцією розподілу двовимірної випадкової величини (X, Y).

Використання розглянутих властивостей і теореми додавання імовірностей дозволяє отримати такі твердження.

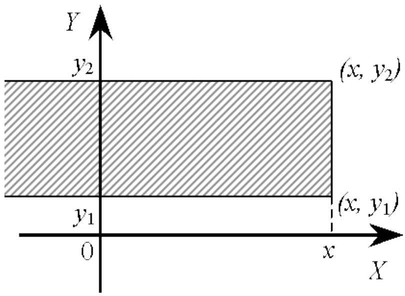
Наслідок 1. Імовірність попадання при випробуванні випадкової точки (X, Y) в нескінченну напівсмугу дорівнює прирosту функції розподілу по одному із аргументів (мал. 7.2, 7.3):

$$P\{(x_1 < X < x_2)\cdot(Y < y)\} = F(x_2, y) - F(x_1, y),$$

$$P\{(X < x)\cdot(y_1 < Y < y_2)\} = F(x, y_2) - F(x, y_1).$$



Мал. 7.2.

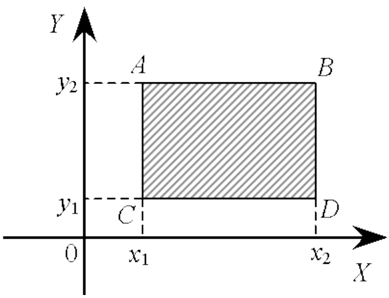


Мал. 7.3.

Наслідок 2. Імовірність попадання при випробуванні випадкової точки (X, Y) в прямокутник, сторони якого паралельні координатним осям, обчислюється за формулою

$$P\{(x_1 < X < x_2)\cdot(y_1 < Y < y_2)\} = \\ = F(x_2, y_2) - F(x_1, y_2) - \left(F(x_2, y_1) - F(x_1, y_1)\right). \quad (7.5)$$

Вказана імовірність знаходиться наступним чином: від імовірності попадання випадкової точки в півсмугу AB віднімається імовірність попадання випадкової точки в півсмугу CD (мал. 7.4).



Мал.7.4.

Густиною сумісного розподілу імовірностей $f(x, y)$ двовимірної неперервної випадкової величини (X, Y) називається друга мішана частинна похідна від функції розподілу:

$$f(x,y)=\frac{\partial^2 F(x,y)}{\partial x\partial y}. \tag{7.6}$$

Геометрично функцію $f(x, y)$ можна зобразити певною поверхнею, яку називають **поверхнею розподілу**.

Імовірність попадання неперервної двовимірної випадкової величини (X, Y) в довільну область D визначається за формулою

$$P\big((X,Y)\in D\big)=\iint_D f(x,y)dxdy. \tag{7.7}$$

Функцію розподілу двовимірної випадкової величини за відомою густиною розподілу можна знайти за формулою

$$F(x,y)=\int_{-\infty}^{y}\int_{-\infty}^{x} f(\xi,\zeta)d\xi d\zeta. \tag{7.8}$$

Властивості густини розподілу імовірностей двовимірної неперервної випадкової величини.

1°. $f(x,y)\ge 0.$

2°. $\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x,y)dxdy=1.$

Функції розподілу імовірностей складових X та Y двовимірної неперервної випадкової величини (X, Y), якщо відомий її закон розподілу, описуються такими функціями розподілу:

$$F_1(x)=\int_{-\infty}^{x}d\xi\int_{-\infty}^{\infty} f(\xi,\zeta)d\zeta, \tag{7.9}$$

$$F_2(y)=\int_{-\infty}^{y}d\zeta\int_{-\infty}^{\infty} f(\xi,\zeta)d\xi. \tag{7.10}$$

Густини розподілів імовірностей складових X та Y двовимірної неперервної випадкової величини (X, Y), якщо відома її густина розподілу, описуються функціями:

$$f_1(x)=\int_{-\infty}^{\infty} f(x,y)dy, \tag{7.11}$$

$$f_2(y)=\int_{-\infty}^{\infty} f(x,y)dx. \tag{7.12}$$

Розглянемо систему двох дискретних величин (X, Y), закон розподілу якої наведений у табл. 7.1.

Умовним розподілом складової Y при $X = x_i$ $(i = \overline{1, m})$ називається сукупність умовних імовірностей

$$p(y_1 \mid x_i), p(y_2 \mid x_i), ..., p(y_n \mid x_i),$$

обчислених у припущенні, що випадкова подія ($X = x_i$) відбулася.

Умовним розподілом складової X при $Y = y_j$ $(j = \overline{1, n})$ називається сукупність умовних імовірностей

$$p(x_1 \mid y_j), p(x_2 \mid y_j), ..., p(x_m \mid y_j),$$

обчислених у припущенні, що випадкова подія ($Y = y_j$) відбулася.

В загальному випадку умовні закони розподілу складової Y визначаються рівностями

$$p(y_j \mid x_i) = p_{ij} \Big/ \sum_{j=1}^{n} p_{ij}, \quad j = \overline{1, n}; i = \overline{1, m}. \tag{7.13}$$

а складової X:

$$p(x_i \mid y_j) = p_{ij} \Big/ \sum_{i=1}^{m} p_{ij}, \quad i = \overline{1, m}; j = \overline{1, n}. \tag{7.14}$$

Зауваження. Сума імовірностей умовного розподілу як для складової X, так і для складової Y, дорівнює одиниці, тобто

$$\sum_{i=1}^{m} p(x_i \mid y_j) = 1, \quad j = \overline{1, n}; \qquad \sum_{j=1}^{n} p(y_j \mid x_i) = 1, \quad i = \overline{1, m}.$$

Ці властивості умовних розподілів використовуються для контролю обчислень.

Нехай (X, Y) – неперервна двовимірна випадкова величина.

Вираз для **умовної функції розподілу складової** X має вигляд

$$F(x \mid y) = P\{(X < x) \mid (Y = y)\} = \int_{-\infty}^{x} f(\xi, y) d\xi / f_2(y),$$

а складової Y:

$$F(y \mid x) = P\{(Y < y) \mid (X = x)\} = \int_{-\infty}^{y} f(x, \xi) d\xi / f_1(x),$$

де $f_1(x)$ та $f_2(y)$ визначається рівностями (7.11) та (7.12) відповідно.

Умовні густини розподілу імовірностей (умовні закони розподілу) складових X та Y неперервної двовимірної випадкової величини (X, Y) обчислюються за формулами:

$$\varphi(x \mid y) = \frac{f(x, y)}{f_2(y)}; \tag{7.15}$$

$$\psi(y\,|\,x)=\frac{f(x,y)}{f_1(x)}. \qquad (7.16)$$

Теорема множення густин імовірностей. Густина розподілу імовірностей неперервної двовимірної випадкової величини (*X, Y*) дорівнює густині розподілу однієї із складових, помноженій на умовну густину розподілу іншої складової, обчислену при умові, що перша складова набрала задане значення:

$$f(x,y)=f_1(x)\psi(y\,|\,x) \quad \text{або} \quad f(x,y)=f_2(y)\varphi(x\,|\,y).$$

Критерії незалежності випадкових величин

Теорема 1. Для того, щоб дискретні випадкові величини *X* та *Y*, які є складовими двовимірної випадкової величини (*X, Y*), були незалежними, необхідно і достатньо, щоб функція розподілу величини (*X, Y*) дорівнювала добутку функцій розподілу складових:

$$F(x, y) = F_1(x)\, F_2(y).$$

Теорема 2. Для того, щоб неперервні випадкові величини *X* та *Y*, що є складовими двовимірної випадкової величини (*X, Y*), були незалежними, необхідно і достатньо, щоб густина розподілу системи (*X, Y*) дорівнювала добутку густин розподілу складових *X* та *Y*:

$$f(x, y) = f_1(x)\, f_2(y).$$

Теорема 3. Якщо двовимірна випадкова величина (*X, Y*) дискретна, то для незалежності її складових необхідно і достатньо виконання рівностей

$$p_{ij} = P(X = x_i)\, P(Y = y_j),$$

для довільних x_i та y_j.

Умовним математичним сподіванням дискретної випадкової величини Y при $X = x$ (де x – певне можливе значення випадкової величини X) називається сума добутків можливих значень Y на відповідні умовні імовірності:

$$M(Y\,|\,X=x)=\sum_{j=1}^{n} y_j\, p(y_j\,|\,x). \qquad (7.17)$$

Аналогічно визначається математичне сподівання дискретної випадкової величини X при $Y = y$:

$$M(X\,|\,Y=y)=\sum_{i=1}^{m} x_i\, p(x_i\,|\,y). \qquad (7.18)$$

Для неперервних випадкових величин *X* та *Y*

$$M(Y \mid X = x) = \int_{-\infty}^{\infty} y\psi(y \mid x)dy, \qquad (7.17^*)$$

$$M(X \mid Y = y) = \int_{-\infty}^{\infty} x\varphi(x \mid y)dx, \qquad (7.18^*)$$

де $\psi(y|x)$ та $\varphi(x|y)$ відповідні умовні густини розподілів системи (X, Y).

Із наведених означень випливає, що при зміні значення x, взагалі кажучи, змінюється і $M(Y|X = x)$. Це означає, що $M(Y|X = x)$ можна розглядати як функцію аргумента x, тобто

$$M(Y \mid X = x) = g(x). \qquad (7.19)$$

Функція $g(x)$ називається **функцією регресії *Y* на *X***, рівняння (7.19) називається **рівнянням регресії *Y* на *X***, а графік функції $g(x)$ – **лінією регресії *Y* на *X***.

Аналогічно умовне математичне сподівання $M(X|Y = y)$ можна розглядати як функцію аргументу y:

$$M(X \mid Y = y) = q(y). \qquad (7.20)$$

Функція $q(y)$ називається **функцією регресії *X* на *Y***, рівняння (7.20) називається **рівнянням регресії *X* на *Y***, а графік функції $q(y)$ – **лінією регресії *X* на *Y***.

Числові характеристики системи двох випадкових величин

Нехай (X, Y) – двовимірна дискретна випадкова величина, закон розподілу якої заданий табл. 7.1. Тоді **математичне сподівання для кожної складової** визначається рівностями:

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

Математичне сподівання для складових системи двох неперервних випадкових величин визначається за формулами:

$$M(X) = \int_{-\infty}^{\infty} xf_1(x)dx = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} xf(x, y)dxdy,$$

$$M(Y) = \int_{-\infty}^{\infty} yf_2(y)dy = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} yf(x, y)dxdy,$$

де $f(x, y)$ – густина розподілу імовірностей системи (X, Y), $f_1(x)$ та $f_2(y)$ – густини розподілів складових X та Y відповідно.

Дисперсії складових системи (X, Y) визначаються формулами:

$$D(X)=\begin{cases}\sum\limits_{i=1}^{m}\sum\limits_{j=1}^{n}\left[(x_i-M(X)\right]^2 p_{ij} & \text{- для дискретної в.в. } (X,Y);\\ \int\limits_{-\infty}^{\infty}\int\limits_{-\infty}^{\infty}\left[(x_i-M(X)\right]^2 f(x,y)dxdy & \text{- для неперервної в.в. } (X,Y);\end{cases}$$

$$D(X)=\begin{cases}\sum\limits_{i=1}^{m}x_i^2 P(X=x_i)-\left[M(X)\right]^2 & \text{- для дискретної в.в. } (X,Y), \text{ якщо відомі імовірності } P(X=x_i);\\ \int\limits_{-\infty}^{\infty}x^2 f_1(x)dx-\left[M(X)\right]^2 & \text{- для неперервної в.в. } (X,Y), \text{ якщо відома густина } f_1(\text{x});\end{cases}$$

$$D(Y)=\begin{cases}\sum\limits_{i=1}^{m}\sum\limits_{j=1}^{n}\left[(y_i-M(Y)\right]^2 p_{ij} & \text{- для дискретної в.в. } (X,Y);\\ \int\limits_{-\infty}^{\infty}\int\limits_{-\infty}^{\infty}\left[(y-M(Y)\right]^2 f(x,y)dxdy & \text{- для неперервної в.в. } (X,Y);\end{cases}$$

$$D(Y)=\begin{cases}\sum\limits_{i=1}^{m}y_j^2 P(Y=y_j)-\left[M(X)\right]^2 & \text{- для дискретної в.в. } (X,Y), \text{ якщо відомі імовірності } P(Y=y_j);\\ \int\limits_{-\infty}^{\infty}y^2 f_2(y)dy-\left[M(Y)\right]^2 & \text{- для неперервної в.в. } (X,Y), \text{ якщо відома густина } f_2(\text{y}).\end{cases}$$

Сукупність математичних сподівань $M(X)$ і $M(Y)$ являє собою характеристику положення системи (X, Y). Геометрично – це координати «середньої» точки на площині, навколо якої здійснюється розсіювання значень системи (X, Y) як випадкової точки. Числові характеристики $D(X)$ та $D(Y)$ характеризують розсіювання випадкової точки в напрямку осей $0x$ та $0y$ відповідно.

Кореляційним моментом (**моментом зв'язку**) системи двох випадкових величин (X, Y) називається математичне сподівання добутку відхилень цих величин від своїх математичних сподівань:

$$K_{XY}=M\left[(X-M(X))(Y-M(Y))\right].$$

Використання властивостей математичного сподівання дозволяє записати кореляційний момент у такому вигляді:

$$K_{XY}=M(XY)-M(X)M(Y).$$

Для обчислення кореляційного момента дискретних випадкових величин користуються формулами:

$$\begin{aligned}K_{XY}&=\sum_{i=1}^{m}\sum_{j=1}^{n}\left(x_i-M(X)\right)\left(y_j-M(Y)\right)p_{ij},\\ K_{XY}&=\sum_{i=1}^{m}\sum_{j=1}^{n}x_i y_j p_{ij}-M(X)M(Y),\end{aligned}\tag{7.21}$$

а для неперервних випадкових величин – формулами:

$$K_{XY} = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\big(x - M(X)\big)\big(y - M(Y)\big) f(x,y)dxdy,$$
$$K_{XY} = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} xyf(x,y)dxdy - M(X)M(Y). \qquad (7.22)$$

Розмірність кореляційного момента дорівнює добутку розмірностей величин X та Y. Тому для одних і тих же величин величина K_{XY} має різні значення в залежності від того, в яких одиницях вимірюються ці величини. Для усунення цього недоліку вводиться нова числова характеристика – коефіцієнт кореляції.

Коефіцієнтом кореляції r_{XY} випадкових величин X та Y називається відношення кореляційного момента до добутку середніх квадратичних відхилень цих величин:

$$r_{xy} = K_{XY} \big/ (\sigma_x \sigma_y). \qquad (7.23)$$

Властивості коефіцієнта кореляції

1°. r_{xy} є безрозмірною величиною.

2°. $\left|r_{xy}\right| \le 1.$

3°. Якщо $r_{xy} = \pm 1,$ то між складовими X та Y випадкової величини (X, Y) існує лінійна функціональна залежність: $Y = \alpha_1 X + \alpha_0$, де α_1, α_0 – дійсні числа.

4°. Якщо $r_{xy} \ne 0,$ то складові X та Y двовимірної випадкової величини (X, Y) залежні.

Коефіцієнт кореляції r_{XY} характеризує силу лінійного зв'язку між випадковими величинами X та Y: чим ближче $\left|r_{xy}\right|$ до одиниці, тим тісніший лінійний зв'язок; чим ближче $\left|r_{xy}\right|$ до нуля, тим слабший лінійний зв'язок.

Випадкові величини X та Y називаються **корельованими**, якщо $r_{xy} \ne 0,$ і **некорельованими**, якщо $r_{xy} = 0.$

Згідно із властивістю 4° коефіцієнта кореляції **із корельованості двох випадкових величин випливає їх залежність**. Проте **із залежності випадкових величини ще не слідує їх корельованість**, тобто вони можуть бути або корельованими, або некорельованими.

З другого боку, **із незалежності двох випадкових величин випливає їх некорельованість**, але **із некорельованості випадкових величини ще не можна зробити висновок про їх незалежність**.

Система довільного скінченного числа випадкових величин

Функцією розподілу системи *n* випадкових величин $(X_1, X_2, ..., X_n)$ називається функція детермінованих аргументів $x_1, x_2, ..., x_n$, яка дорівнює імовірності сумісного виконання нерівностей $X_i < x_i \; (i = \overline{1, n})$:

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

де крапками позначено добуток (перетин) подій.

Ця функція є неспадною по кожному із аргументів x_i, при решті фіксованих. Вона прямує до нуля, якщо хоча б один аргумент прямує до $-\infty$.

Якщо із системи випадкових величин $X_1, X_2, ..., X_n$ виділити деяку підсистему, наприклад $X_1, X_2, ..., X_k$, де $k < n$, то функція розподілу для цієї підсистеми матиме вигляд $F(x_1, x_2, ..., x_k, \infty, ..., \infty)$. Зокрема, функцію розподілу для кожної із величин, що входять до складу системи, одержимо, якщо вся решта аргументів прямуватиме до $+\infty$. Наприклад, $F(x_1) = F(x_1, \infty, ..., \infty)$. Якщо всі аргументи функції розподілу прямують до ∞, то вона прямує до 1.

Густина розподілу імовірностей *n*-вимірної неперервної випадкової величини $(X_1, X_2, ..., X_n)$ визначається рівністю

$$f(x_1, x_2, ..., x_n) = \frac{\partial^n F(x_1, x_2, ..., x_n)}{\partial x_1 \partial x_2 ... \partial x_n}.$$

Ця функція невід'ємна і для неї виконується умова нормування

$$\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} ... \int_{-\infty}^{\infty} f(x_1, x_2, ..., x_n) dx_1 dx_2 ... dx_n = 1.$$

Густина розподілу для підсистеми $(X_1, X_2, ..., X_k)$, де $k < n$, визначається рівністю

$$f(x_1, x_2, ..., x_k) = \underbrace{\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} ... \int_{-\infty}^{\infty}}_{n-k} f(x_1, x_2, ..., x_n) dx_{k+1} dx_{k+2} ... dx_n.$$

Умовним законом розподілу підсистеми випадкових величин $(X_1, X_2, ..., X_k)$ називається такий закон, який отримується за умови, що решта з них $(X_{k+1}, X_{k+2}, ..., X_n)$ набули своїх можливих значень $x_{k+1}, x_{k+2}, ..., x_n$.

Умовна густина розподілу імовірностей визначається за формулою

$$f(x_1, x_2, ..., x_k \mid x_{k+1}, ..., x_n) = \frac{f(x_1, x_2, ..., x_n)}{f(x_{k+1}, x_{k+2}, ..., x_n)}.$$

Необхідною і достатньою умовою незалежності складових n-вимірної випадкової величини $(X_1, X_2, ..., X_n)$ є виконання рівності

$$f(x_1, x_2, ..., x_n) = f(x_1) f(x_2) ... f(x_n).$$

Числові характеристики *n*-вимірних випадкових величин.

Математичне сподівання неперервної величини X_i, що входить до системи, обчислюється за формулою

$$M(X_i) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} ... \int_{-\infty}^{\infty} x_i f(x_1, x_2, ..., x_n) dx_1 dx_2 ... dx_n, \quad (i = \overline{1, n}),$$

а дисперсія цієї ж величини (розрахункова формула) –

$$D(X_i) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} ... \int_{-\infty}^{\infty} x_i^2 f(x_1, x_2, ..., x_n) dx_1 dx_2 ... dx_n - \left[M(X_i)\right]^2.$$

Кореляційні моменти для кожної пари (X_i, Y_j) складових n-вимірної випадкової величини $(X_1, X_2, ..., X_n)$ обчислюються за формулою

$$K_{X_i Y_j} = K_{ij} = M\left\{(X_i - M(X_i))(X_j - M(X_j))\right\}.$$

Зокрема, якщо $i = j$, то $K_{ij} = D(X_i) = \sigma_i^2$.

Всі кореляційні моменти випадкової величини $(X_1, X_2, ..., X_n)$ розміщують у вигляді квадратної таблиці, яка називається **кореляційною матрицею системи *n* випадкових величин**:

$$\begin{pmatrix} K_{11} & K_{12} & K_{13} & ... & K_{1n} \\ K_{21} & K_{22} & K_{23} & ... & K_{2n} \\ ... & ... & ... & ... & ... \\ K_{n1} & K_{n2} & K_{n3} & ... & K_{nn} \end{pmatrix}.$$

З урахуванням рівностей $K_{ij} = K_{ji}$, $K_{ii} = \sigma_i^2$, кореляційна матриця набирає такого виду:

$$\begin{pmatrix} \sigma_1^2 & K_{12} & K_{13} & ... & K_{1n} \\ K_{12} & \sigma_2^2 & K_{23} & ... & K_{2n} \\ ... & ... & ... & ... & ... \\ K_{1n} & K_{2n} & K_{3n} & ... & \sigma_n^2 \end{pmatrix}.$$

У випадку некорельованості кожної пари системи $(X_1, X_2, ..., X_n)$ кореляційна матриця стає діагональною:

$$\begin{pmatrix} \sigma_1^2 & 0 & 0 & \dots & 0 \\ 0 & \sigma_2^2 & 0 & \dots & 0 \\ \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & 0 & \dots & \sigma_n^2 \end{pmatrix}.$$

Знаючи кореляційні моменти, можна визначити **коефіцієнти кореляції для кожної пари випадкових величин *n*-вимірної випадкової величини**. Такі коефіцієнти кореляції називаються парними і визначаються за формулою

$$r_{ij} = \frac{K_{ij}}{\sigma_i \sigma_j}.$$

Із парних коефіцієнтів кореляції утворюється квадратна матриця, яка називається **нормованою кореляційною матрицею**:

$$\begin{pmatrix} 1 & r_{12} & r_{13} & \dots & r_{1n} \\ r_{12} & 1 & r_{23} & \dots & r_{2n} \\ \dots & \dots & \dots & \dots & \dots \\ r_{1n} & r_{2n} & r_{3n} & \dots & 1 \end{pmatrix}.$$

Розподіл імовірностей системи (*X, Y*) називається нормальним (двовимірна випадкова величина розподілена за нормальним законом), якщо густина розподілу ймовірностей

$$f(x,y) = \frac{1}{2\pi\sigma_x\sigma_y\sqrt{1-r_{xy}^2}} e^{-\frac{1}{2(1-r_{xy}^2)}\left[\frac{(x-a)^2}{\sigma_x^2}+\frac{(y-b)^2}{\sigma_y^2} - 2r_{xy}\frac{(x-a)(y-b)}{\sigma_x\sigma_y}\right]},$$

де п’ять параметрів розподілу мають такий імовірнісний зміст: $a = M(X)$, $b = M(Y)$, $\sigma_x = \sqrt{D(X)}$, $\sigma_y = \sqrt{D(Y)}$, r_{xy} – коефіцієнт кореляції величин *X* та *Y*.

Відомо якщо система (*X, Y*) розподілена за нормальним законом із параметрами $a, b, \sigma_x, \sigma_y, r_{xy}$, то її складові також розподілені за нормальним законом з параметрами, рівними відповідно a, σ_x та b, σ_y.

Для складових нормально розподіленої двовимірної випадкової величини поняття незалежності та некорельованості є еквівалентними.

РОЗВ’ЯЗУВАННЯ ТИПОВИХ ЗАДАЧ

Задача 7.1. Якість продукції характеризується двома випадковими параметрами: *X* та *Y*. Закон розподілу випадкової величини

$Z = (X, Y)$ наведений в табл. 7.2. Знайти: а) закони розподілу складових випадкової величини $Z = (X, Y)$; б) умовний закон розподілу складової X при умові, що складова Y набрала значення $y_3 = 0{,}2$; в) умовний закон розподілу складової Y при умові, що $X = x_1 = 2$; г) умовне математичне сподівання випадкової величини Y при умові, що $X = x_2 = 3$.

Таблиця 7.2

$X \backslash Y$	0	0,1	0,2	0,3
2	0,2	0,05	0,1	0,05
3	0	0,1	0,15	0,15
4	0	0,15	0	0,05

○ а) Використаємо рівності (7.2):

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

Закон розподілу складової X має вид

X	2	3	4
P	0,4	0,4	0,2

.

За формулами (7.3)

$$P(Y = 0) = 0{,}2 + 0 + 0 = 0{,}2;$$
$$P(Y = 0{,}1) = 0{,}05 + 0{,}1 + 0{,}15 = 0{,}3;$$
$$P(Y = 0{,}2) = 0{,}1 + 0{,}15 + 0 = 0{,}25;$$
$$P(Y = 0{,}3) = 0{,}05 + 0{,}15 + 0{,}05 = 0{,}25,$$

звідки отримується закон розподілу складової Y:

Y	0	0,1	0,2	0,3
P	0,2	0,3	0,25	0,25

.

б) Умовні імовірності того, що при випробуванні величина X набере кожне із своїх можливих значень при умові, що складова Y набрала значення $y_3 = 0{,}2$, обчислимо за формулами (7.14), враховуючи те, що сума в знаменнику дорівнює сумі елементів третього стовпця табл. 7.2.

$$P(x_1 \mid y_3) = p_{13} / P(Y = y_3) = 0{,}1/0{,}25 = 0{,}4,$$
$$P(x_2 \mid y_3) = p_{23} / P(Y = y_3) = 0{,}15/0{,}25 = 0{,}6,$$
$$P(x_3 \mid y_3) = p_{33} / P(Y = y_3) = 0/0{,}25 = 0.$$

Запишемо шуканий умовний розподіл складової X:

X	2	3
$P(X \mid y_3)$	0,4	0,6

Відмітимо, що в отриманому умовному розподілі X при $Y = y_3$ фігурує тільки два можливих значення, оскільки подія ($X = x_3$) при $Y = y_3$ є неможливою. Якщо ж скласти умовний розподіл X при $Y = y_4$, то таких можливих значень X вже буде три (проаналізуйте!).

в) Для знаходження умовних імовірностей складової Y при умові, що складова X набрала значення $x_1 = 2$, знайдемо за формулами (7.13). При цьому суму в знаменнику знайдемо, додаючи елементи першого рядка табл. 7.2.

$$P(y_1 \mid x_1) = p_{11}/P(X = x_1) = 0{,}2/0{,}4 = 0{,}5,$$
$$P(y_2 \mid x_1) = p_{12}/P(X = x_1) = 0{,}05/0{,}4 = 0{,}125,$$
$$P(y_3 \mid x_1) = p_{13}/P(X = x_1) = 0{,}1/0{,}4 = 0{,}25,$$
$$P(y_4 \mid x_1) = p_{14}/P(X = x_1) = 0{,}05/0{,}4 = 0{,}125.$$

Отже, шуканий умовний розподіл складової Y має вигляд:

Y	0	0,1	0,2	0,3
$P(Y \mid x_1)$	0,5	0,125	0,25	0,125

г) Використаємо формулу (7.20), де $x = x_2$.

$$M(Y \mid X = x_2) = \sum_{j=1}^{4} y_j\, p(y_j \mid x_2).$$

Обчислимо умовні імовірності $p(y_j|x_2)$, $j = \overline{1,4}$, застосувавши формули (7.13) (див. п. в):

$$p(y_1 \mid x_2) = p_{21}/P(X = x_2) = 0/0{,}4 = 0,$$
$$p(y_2 \mid x_2) = p_{22}/P(X = x_2) = 0{,}1/0{,}4 = 0{,}25,$$
$$p(y_3 \mid x_2) = p_{23}/P(X = x_2) = 0{,}15/0{,}4 = 0{,}375,$$
$$p(y_4 \mid x_2) = p_{24}/P(X = x_2) = 0{,}15/0{,}4 = 0{,}375.$$

Отже,

$$M(Y \mid X = x_2) = 0 \cdot 0 + 0{,}1 \cdot 0{,}25 + 0{,}2 \cdot 0{,}375 + 0{,}3 \cdot 0{,}375 = 0{,}2125. \quad \bullet$$

Задача 7.2. Двовимірна випадкова величина задана функцією розподілу

$$F(x, y) = \begin{cases} 0, & \text{якщо } x \le 0 \text{ або } y \le 0, \\ (1 - e^{-x})(1 - e^{-2y}), & \text{якщо } x > 0 \text{ і } y > 0. \end{cases}$$

Знайти імовірність того, що при випробуванні випадкова величина (X, Y) набере значення з квадрату, вершини якого мають такі координати: (0, 0), (0, 1), (1, 0) і (1, 1).

○ Множина точок вказаного в умові квадрата визначається співвідношеннями $\{(0 \le X < 1)\cdot(0 \le Y < 1)\}$, а тому за формулою (7.5)

$$P\{(0 \le X < 1)\cdot(0 \le Y < 1)\} = F(1{,}1) - F(0{,}1) - \big(F(1{,}0) - F(0{,}0)\big) =$$

$$= (1-e^{-1})(1-e^{-2}) - (1-e^{0})(1-2^{-2}) - (1-e^{-1})(1-e^{0}) + 0 =$$

$$= (1-e^{-1})(1-e^{-2}) \approx 0{,}547. \ \bullet$$

Задача 7.3. Двовимірна випадкова величина (*X, Y*) розподілена із сталою густиною розподілу всередині квадрата *D*, вершини якого мають координати (0; 0), (0; 2), (2; 0), (2; 2).
Знайти густину імовірностей *f*(*x, y*) і функцію розподілу.

○ Згідно з умовою задачі густина розподілу імовірностей має такий вид

$$f(x,y) = \begin{cases} C, & \text{якщо } (x,y) \in D, \\ 0, & \text{якщо } (x,y) \notin D, \end{cases}$$

де *C* – стала. Знайдемо *C*, використавши властивість 2° густини розподілу

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x,y)dxdy = \iint_D Cdxdy = C\int_0^2 dx\int_0^2 dy = 2C\int_0^2 dx = 4C = 1,$$

звідки C = 1/4. Отже,

$$f(x,y) = \begin{cases} 0{,}25, & \text{якщо } (x,y) \in D, \\ 0, & \text{якщо } (x,y) \notin D. \end{cases}$$

Функцію розподілу знайдемо за формулою (7.8).

Нехай $x \le 0$ або $y \le 0$, тоді $f(x, y) = 0$ і

$$F(x,y) = P\{(X < x)\cdot(Y < y)\} = \int_{-\infty}^{x} d\xi \int_{-\infty}^{y} f(\xi,\zeta)d\zeta = 0.$$

Якщо $0 < x \le 2$ і $0 < y \le 2$, тоді

$$F(x,y) = P\{(X < x)\cdot(Y < y)\} = \int_{-\infty}^{x} d\xi \int_{-\infty}^{y} f(\xi,\zeta)d\zeta =$$

$$= \int_0^x d\zeta \int_0^y 0{,}25 d\xi = 0{,}25\int_0^x y d\xi = 0{,}25xy.$$

Нехай $x > 2$ і $y \in (0, 2]$, тоді

$$F(x,y)=P\{(X<x)\cdot(Y<y)\}=\int\limits_{-\infty}^{x}d\xi\int\limits_{-\infty}^{y}f(\xi,\zeta)d\zeta=$$

$$=\int\limits_{0}^{y}d\zeta\int\limits_{0}^{2}0{,}25d\xi=0{,}5\int\limits_{0}^{y}d\zeta=0{,}5y.$$

В цьому випадку враховувалося те, що $f(x, y) = 0$ за межами квадрата D.

Нехай $x\in(0, 2]$ і $y > 2$, тоді

$$F(x,y)=P\{(X<x)\cdot(Y<y)\}=\int\limits_{-\infty}^{x}d\xi\int\limits_{-\infty}^{y}f(\xi,\zeta)d\zeta=0{,}25\int\limits_{0}^{x}d\xi\int\limits_{0}^{2}d\zeta=0{,}5x.$$

Нарешті, якщо $x> 2$ і $y > 2$, то

$$F(x,y)=\int\limits_{-\infty}^{x}d\xi\int\limits_{-\infty}^{y}f(\xi,\zeta)d\zeta=0{,}25\int\limits_{0}^{2}d\xi\int\limits_{0}^{2}d\zeta=1.$$

Таким чином, шукана функція розподілу даної двовимірної випадкової величини має такий вид:

$$F(x,y)=\begin{cases}0, & \text{якщо } x\le 0 \text{ або } y\le 0,\\ 0{,}25xy, & \text{якщо } x\in(0{,}2] \text{ і } y\in(0{,}2],\\ 0{,}5y, & \text{якщо } x>2 \text{ і } y\in(0{,}2],\\ 0{,}5x, & \text{якщо } x\in(0{,}2] \text{ і } y>2,\\ 1, & \text{якщо } x>2 \text{ і } y>2.\end{cases}$$ ●

Задача 7.4. Густина розподілу імовірностей двовимірної випадкової величини (X, Y)

$$f(x,y)=\frac{\sqrt{2}}{\pi}e^{-x^2-2xy-3y^2}.$$

Знайти густину розподілу імовірностей складової Y.

○ Шукану густину знаходимо за формулою (7.12):

$$f_2(y)=\frac{\sqrt{2}}{\pi}\int\limits_{-\infty}^{\infty}e^{-\xi^2-2\xi y-3y^2}d\xi=\frac{\sqrt{2}}{\pi}\int\limits_{-\infty}^{\infty}e^{-(\xi+y)^2-2y^2}d\xi=$$

$$=\frac{\sqrt{2}}{\pi}e^{-2y^2}\int\limits_{-\infty}^{\infty}e^{-(\xi+y)^2}d(\xi+y).$$

Враховуючи інтеграл Пуассона $\int\limits_{-\infty}^{\infty}e^{-t^2}dt=\sqrt{\pi}$, остаточно отримаємо

$$f_2(y) = \frac{\sqrt{2}}{\pi} e^{-2y^2} \sqrt{\pi} = \sqrt{\frac{2}{\pi}} e^{-2y^2}. \quad \bullet$$

Задача 7.5. Густина розподілу імовірностей двовимірної випадкової величини (X, Y)

$$f(x, y) = \begin{cases} 4 - x - y, & \text{якщо } (x, y) \in D, \\ 0, & \text{якщо } (x, y) \notin D. \end{cases}$$

де D – квадрат, обмежений прямими $x = 0$, $x = 2$, $y = 0$, $y = 2$.

Знайти умовні закони розподілу складових, що входять у систему і з'ясувати, чи є залежними ці складові.

○ За формулами (7.11), (7.12) знайдемо густину розподілу імовірностей кожної складової. Для $x \in [0; 2]$

$$f_1(x) = \int_0^2 (4 - x - y) dy = \left((4 - x) y - \frac{y^2}{2} \right)\Bigg|_0^2 = 2(3 - x),$$

для $y \in [0; 2]$

$$f_2(y) = \int_0^2 (4 - x - y) dx = 2(3 - y).$$

Отже,

$$f_1(x) = \begin{cases} 2(3 - x), & \text{при } x \in [0,2], \\ 0, & \text{при } x \notin [0,2]. \end{cases}$$

$$f_2(y) = \begin{cases} 2(3 - y), & \text{при } y \in [0,2], \\ 0, & \text{при } y \notin [0,2]. \end{cases}$$

Використавши формули (7.15) та (7.16), отримаємо умовні густини розподілу імовірностей складових X та Y відповідно:

$$\varphi(x \mid y) = \frac{4 - x - y}{2(3 - y)}, \quad \text{якщо } (x, y) \in D,$$

$$\psi(y \mid x) = \frac{4 - x - y}{2(3 - x)}, \quad \text{якщо } (x, y) \in D.$$

Оскільки для $(x, y) \in D$

$$f_1(x) f_2(y) = 2(3 - x) \cdot 2(3 - y) \neq 4 - x - y = f(x, y),$$

то згідно з теоремою 2 випадкові величини X та Y є залежними. ●

Задача 7.6. Частка продукції заводу, що містить брак через дефект А, становить 3 %, а через дефект В – 4,5 %. Придатна продукція становить 95 %. Знайти кореляційний момент, коефіцієнти кореляції дефектів А і В, а також функцію розподілу імовірностей.

○ Розглянемо систему дискретних випадкових величин (*X, Y*) Їх значення дорівнюють відповідно 1, якщо продукція має дефект *A* або *B*, і нулю, якщо дефект відсутній. Можливі 4 комбінації значень змінних. Визначимо їхні ймовірності. За умовою придатна продукція становить 95 %, тому $P(X=0;Y=0)=0{,}95.$ Випадкова величина X набуває значення 1 з імовірністю 0,03, тоді $P(X=0)=1-0{,}03=0{,}97.$
Отже, $P(X=0;Y=1)=P(X=0)-P(X=0;\ Y=0)=0{,}97-0{,}95=0{,}02.$

Далі визначаємо такі ймовірності:
$P(X=1;Y=1)=P(Y=1)-P(X=0;\ Y=1)=0{,}045-0{,}02=0{,}025;$
$P(Y=0)=1-P(Y=1)=1-0{,}045=0{,}955.$

Запишемо результати обчислень у таблицю:

X \ Y	0	1
0	0,005	0,025
1	0,95	0,02

Знайдемо кореляційний момент за другою із формул (7.21). Для цього обчислимо значення величин, які входять до цієї формули:

$$M(XY)=\sum_{i=1}^{2}\sum_{j=1}^{2}x_i y_j p_{ij}=0{,}025;\ M(X)=0{,}03;\ M(Y)=0{,}045.$$

Тоді $K_{XY}=0{,}025-0{,}03\cdot 0{,}045=0{,}02365.$

Обчислимо дисперсії X і Y:

$M(X^2)=0{,}03;\quad D(X)=0{,}0291;\quad M(Y^2)=0{,}045;\quad D(Y)=0{,}042975.$

Тоді згідно із (7.26) $r_{XY}=\dfrac{0{,}02365}{\sqrt{0{,}0291\cdot 0{,}042975}}\approx 0{,}669.$

Знайдемо функцію розподілу $F(X,Y)$.

Якщо $x\le 0$ або $y\le 0$, то у відповідності із властивістю 3 $F(X,Y)=0$.

Нехай $x\in(0;1]$ і $y\in(0;1]$. Тоді випадкова подія $(X<x,Y<y)=(X=0,Y=0)$, оскільки 0 – єдине можливе значення, яке менше від *x* та *y*. А тому згідно із (7.4) для $x\in(0;1]$ і $y\in(0;1]$ $F(X,Y)=P(X<x,Y<y)=P(X=0,Y=0)=0{,}005$.

Якщо $x > 1$ і $y \in (0;1]$, тоді подія $(X < x, Y < y)$ відбувається тоді, і тільки тоді, коли одночасно $X = 0$ і $Y = 0$, або одночасно $X = 1$ і $Y = 0$, тобто має місце така рівність $(X < x, Y < y) = (X = 0, Y = 0) + {}$ ${}+(X = 1, Y = 0)$, де доданки справа є несумісними подіями. Використання теореми додавання імовірностей і означення (7.4) дозволяє знайти $F(X,Y)$ на цих проміжках:

$$F(X,Y) = P(X < x, Y < y) = P(X = 0, Y = 0) + P(X = 1, Y = 0) =$$
$$= 0{,}005 + 0{,}95 = 0{,}955 .$$

Якщо $x \in (0;1]$ і $y > 1$, то за аналогією із попередніми інтервалами

$$F(X,Y) = P(X < x, Y < y) = P(X = 0, Y = 0) + P(X = 0, Y = 1) = 0{,}03.$$

Нарешті якщо $x > 1$ і $y > 1$, то подія $(X < x, Y < y)$ – достовірна, і $F(X,Y) = 1$.

Отже, функцію розподілу можна записати у вигляді таблиці:

$X \backslash Y$	$y \le 0$	$0 \le y \le 1$	$y > 1$
$x \le 0$	0	0	0
$0 \le x \le 1$	0	0	0,03
$x > 1$	0	0,955	1

●

Задача 7.7. Двовимірна випадкова величина (*X, Y*) розподілена рівномірно в крузі $x^2 + y^2 \le r^2$. Довести, що X та Y – залежні випадкові величини, але некорельовані.

○ За умовою задачі густина розподілу величини (*X, Y*) має такий вид:

$$f(x,y) = \begin{cases} C = \text{const} & \text{при } x^2 + y^2 \le r^2, \\ 0 & \text{при } x^2 + y^2 > r^2. \end{cases}$$

Використавши властивість 2° густини розподілу імовірності, одержимо $C = 1/(\pi r^2)$. За формулами (7.11), (7.12) знайдемо густину розподілу кожної складової:

$$f_1(x) = \int_{-\infty}^{\infty} f(x,y)dy = \int_{-r}^{r} \left(1/(\pi r^2)\right)dy = \left(1/(\pi r^2)\right) y\Big|_{-r}^{r} = 2/(\pi r);$$

$$f_2(y) = \int_{-\infty}^{\infty} f(x,y)dx = 2/(\pi r).$$

Оскільки рівність $f(x,y) = f_1(x)\, f_2(y)$ не виконується, то робимо висновок, що *X* та *Y* є залежними випадковими величинами.

Кореляційний момент системи (X, Y) обчислимо за другою із формул (7.22):

$$K_{XY} = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} xyf(xy)dxdy - M(X)M(Y) =$$

$$= \frac{1}{\pi r^2}\int_{-r}^{r} xdx \int_{-\sqrt{r^2-x^2}}^{\sqrt{r^2-x^2}} ydy - \frac{1}{\left(\pi r^2\right)^2}\int_{-r}^{r} xdx \int_{-r}^{r} ydy = 0.$$

Отже, $r_{XY} = 0$ і величини X та Y є некорельованими. ●

ЗАДАЧІ ДЛЯ САМОСТІЙНОГО РОЗВ'ЯЗУВАННЯ

Завдання №8

1. Закон розподілу дискретної двовимірної випадкової величини (X, Y) заданий таблицею:

$X \backslash Y$	1	2	3
– 1	0,2	0,4	0,1
2	0,05	0,1	0,15

   Знайти: а) закони розподілу одновимірних випадкових величин X та Y; б) умовний закон розподілу складової X при умові $Y = 1$; в) умовний закон розподілу складової Y при умові $X = 2$; г) обчислити $P(Y<X)$.

2. Дискретну двовимірну випадкову величину (X, Y) задано законом розподілу

$X \backslash Y$	1	2	3
– 1	0,06	a	0,1
0	0,03	0,2	0,01
2	$2a$	0,05	0,1

   Знайти: а) параметр a; б) закони розподілу випадкових величин X і Y; в) $M(X)$ і $M(Y)$; г) $D(X)$ і $D(Y)$.

3. Система випадкових величин (X, Y) має такий розподіл імовірностей:

$X \backslash Y$	0	1
0	0,06	0,24
1	0,03	0,2
2	0,42	0,05

Знайти: а) функцію розподілу $F(x,y)$; б) функції розподілу $F_1(x)$ і $F_2(y)$; в) $P(1{,}3 < X < 5; -0{,}1 < Y < 1)$.

4. Систему випадкових величин (X,Y) задано законом розподілу:

X \ Y	0	1
– 1	0,3	0,2
0	0,2	0,2
1	0,05	0,05

Знайти: а) $M(X \mid Y=0)$; б) $M(Y \mid X=-1)$; в) $\sigma(X \mid Y=0)$; г) $\sigma(Y \mid X=-1)$.

5. П'ять телевізорів випробовують на надійність. Імовірність того, що окремо взятий телевізор витримає режим випробування, дорівнює 0,95. Нехай X – число телевізорів, які витримають випробування, а Y – число телевізорів, які не витримають їх. Побудувати закон спільного розподілу X і Y і обчислити K_{xy}, r_{xy}.
6. Відсоток бракованих деталей, виготовлених на заводі із дефектом А становить 7%, причому серед забракованих за ознакою А деталей у 3% випадків зустрічається дефект В, і серед деталей без дефекту А, дефект В зустрічається в 1% випадків. Знайти коефіцієнт кореляції між ознаками А та В.
7. У першій посудині міститься 5 стандартних і 2 бракованих виробів, у другій – 3 стандартних та 2 бракованих. Із першої посудини навмання беруть 3 вироби, а з другої – один виріб. Нехай X – поява числа стандартних виробів серед трьох навмання взятих із першої посудини, а Y – поява числа стандартних виробів під час виймання з другої посудини. Побудувати закон спільного розподілу X і Y. Встановити наявність кореляційної залежності між X і Y.
8. У кожної із двох конкуруючих фірм є по три стратегії розвитку фірми. Нехай X – номер стратегії, яку може вибрати перша фірма, а Y – номер стратегії вибраної другою фірмою. Експерти оцінюють ймовірність вибору пар стратегій згідно наведеної таблиці:

X \ Y	1	2	3
1	1/9	2/9	0
2	1/18	1/9	1/18
3	2/9	1/9	1/9

Визначити найбільш ймовірні стратегії для кожної фірми і коефіцієнт кореляції між номерами стратегій розвитку фірм.

В задачах 8-10 двовимірна випадкова величина (*X*, *Y*) задана функцією розподілу $F(x,y)$. Потрібно знайти:

1) функцію густини імовірностей системи (*X*, *Y*);

2) імовірність попадання випадкової точки (*X*; *Y*) в область *D*.

9. $F(x,y)=\begin{cases}1-3^{-x^2}-3^{-3y^2}+3^{-x^2-3y^2}, & x\geq 0 \text{ і } y\geq 0,\\ 0, & \text{в інших випадках,}\end{cases}$

$D=\{1\leq x\leq 2; 1\leq y\leq 2\}$.

10. $F(x,y)=\begin{cases}\sin x\sin y, & 0\leq x\leq \pi/2 \text{ і } 0\leq y\leq \pi/2,\\ 0, & \text{в інших випадках,}\end{cases}$

$D=\left\{\frac{\pi}{6}\leq x\leq\frac{\pi}{2}; \frac{\pi}{3}\leq y\leq\frac{\pi}{3}\right\}$.

11. $F(x,y)=\begin{cases}(1-e^{-2x})(1-e^{-3y}), & x\geq 0 \text{ і } y\geq 0,\\ 0, & \text{в інших випадках,}\end{cases}$

$D=\{0\leq x\leq 2; 0\leq y\leq 2\}$.

12. Двовимірна випадкова величина має густину розподілу $f(x,y)=c(xy+x^2)$ в області $D=\{0\leq x\leq 2; 0\leq y\leq 2\}$ і $f(x,y)=0$ по за областю. Знайти: 1) параметр *c*; 2) умовні щільності розподілу складових системи. Встановити залежність чи незалежність випадкових величин X та Y.

В задачах 13-20 двовимірна випадкова величина (*X*, *Y*) задана густиною розподілу $f(x,y)$. Потрібно:

1) знайти значення параметра с;

2) знайти функцію розподілу імовірностей системи (X, Y);

3) обчислити r_{xy}.

13. $f(x,y)=\begin{cases}c\sin(x+y), & x\in D,\\ 0, & x\notin D,\end{cases}$ $D=\left\{0\leq x\leq\frac{\pi}{2};\ 0\leq y\leq\frac{\pi}{2}\right\}$.

14. $f(x,y)=\begin{cases}c(x+y), & x\in D,\\ 0, & x\notin D,\end{cases}$ $D=\{0\leq x\leq 1;\ 0\leq y\leq 2\}$.

15. $f(x,y)=\begin{cases} c(x+y), & x\in D, \\ 0, & x\notin D, \end{cases}$ $D=\{0\le x\le 1;\ x\le y\le 1\}.$

16. $f(x,y)=\dfrac{c}{(16+x^2)(25+y^2)}$, $x,y\in R.$

17. $f(x,y)=\dfrac{c}{1+x^2+y^2+x^2y^2}$, $x,y\in R.$

18. $f(x,y)=\begin{cases} cxy, & x\in D, \\ 0, & x\notin D, \end{cases}$ $D=\{0\le x\le 1;\ 0\le y\le 1\}.$

19. $f(x,y)=\begin{cases} cxy, & x\in D, \\ 0, & x\notin D, \end{cases}$ $D=\{y=1+x^2; y=2; x\ge 0\}.$

20. $f(x,y)=\begin{cases} c(xy+y^2), & x\in D, \\ 0, & x\notin D, \end{cases}$ $D=\{0\le x\le 1;\ 0{,}5x\le y\le x\}.$

21. Нехай задано випадкову величину (*X, Y*), де *X* – час появи першого покупця у понеділок, а *Y* – час появи першого покупця у вівторок. Відомо, що $f(x,y)=e^{-x-2y}$, $x,y\ge 0$. Знайти: 1) $F(x,y)$; 2) $F_1(x)$, $F_2(y)$; 3) $f_1(x)$, $f_2(y)$; 4) $F(x\,|\,y)$, $F(y\,|\,x)$; 5) $\varphi(x\,|\,y)$, $\psi(y\,|\,x)$. Перевірити залежність випадкових величин *X* та *Y*.

22. Обсяги виготовленої продукції заводом за два квартали описуються двовимірною випадковою величиною (*X, Y*) з густиною розподілу імовірностей

$$f(x,y)=\begin{cases} c, & 50\le x\le 100, 10\le y\le 60, \\ 0, & \text{в інших випадках}. \end{cases}$$

Знайти: 1) константу с; 2) функцію розподілу $F(X,Y)$; 3) дослідити випадкові величини X і Y на незалежність.

23. Система випадкових величин (*X, Y*) розподілена зі сталою густиною всередині квадрату зі стороною 3. Записати вираз для густини розподілу та знайти функцію розподілу системи випадкових величин (*X, Y*).

24. Система випадкових величин (*X, Y*) розподілена зі сталою густиною всередині області $D=\{3\le x\le 6; 5\le y\le 8\}$ і $f(x,y)=0$ по за

областю. Визначити залежність чи незалежність випадкових величин X і Y, а також їх корельованість.

25. Система випадкових величин (X, Y) має рівномірний розподіл ймовірностей в області $D = \{-1 \le x \le 1; -1 \le y \le 1\}$. Поза цією областю густина розподілу ймовірностей системи випадкових величин дорівнює нулю. Записати аналітичний вираз для функції розподілу та функції густини. Знайти $M(X)$, $M(Y)$.

26. Значення X попиту на продукцію коливається під впливом випадкових факторів A і B. Середнє квадратичне відхилення в результаті дії A дорівнює 1,3, а B – 1,1. Коефіцієнт кореляції між відхиленнями дорівнює 0,14. Знайти σ_X як результат дії двох факторів.

В задачах 27-29 двовимірна випадкова величина (X, Y) задана густиною розподілу $f(x, y)$. Потрібно:

1) знайти густини розподілу імовірностей складових X і Y;

2) знайти умовні густини розподілу імовірностей складових X і Y;

3) визначити залежність, чи незалежність випадкових величин X та Y, а також їх корельованість.

27. $f(x, y) = \begin{cases} 1/(3\pi), \ x^2/4 + y^2/16 \le 1, \\ 0, \ x^2/4 + y^2/16 > 1. \end{cases}$

28. $f(x, y) = \begin{cases} 1/(2\pi)\left(3 - \sqrt{x^2 + y^2}\right), \ x^2 + y^2 \le 9, \\ 0, \ x^2 + y^2 > 9. \end{cases}$

29. $f(x, y) = \begin{cases} c = \text{const}, & x^2 + y^2 \le 4, \\ 0, & x^2 + y^2 > 4. \end{cases}$

30. Дві незалежні випадкові величини X та Y, які є складовими двовимірної випадкової величини (X, Y) задано законами розподілу:

$$f_1(x) = \begin{cases} 0, & x < 0, \\ 2e^{-2x}, & x \ge 0, \end{cases} \quad f_2(y) = \begin{cases} 0, & y < 0, \\ 3e^{-3y}, & y \ge 0. \end{cases}$$

Знайти густину розподілу та функцію розподілу системи випадкових величин (X, Y).

31. Задано систему двох неперервних величин X та Y. Відомо, що

$$f_2(y)=\begin{cases}0, & y<0,\\ 3e^{-3y}, & y\geq 0,\end{cases} \qquad \varphi(x\,|\,y)=\begin{cases}0, & y<0,\\ xe^{-xy}, & x\geq 0.\end{cases}$$

Знайти $f(x,y)$, $f_1(x)$, $\psi(y\,|\,x)$.

32. Систему дискретних випадкових величин (X, Y) задано законом розподілу:

X \ Y	–1	0	1
–1	a	0,3	$2a$
0	a	0,1	0,05
1	0,05	0,04	0,06

Знайти імовірності p_{11}, p_{33}, p_{21}. Побудувати кореляційну матрицю.

33. Задана густина ймовірностей:

$$f(x,y)=\begin{cases}2(x+y), & 0\leq x\leq 2, 0\leq y\leq 2,\\ 0, & \text{в інших випадках}.\end{cases}$$

Побудувати кореляційну матрицю.

34. Задано закон системи двох неперервних випадкових величин (X, Y):

$$f(x,y)=\begin{cases}0{,}25(x+y), & -3\leq x\leq 2, -1\leq y\leq 4,\\ 0, & \text{в інших випадках}.\end{cases}$$

Знайти: 1) функцію розподілу $F(x,y)$; 2) $P(-2<X<1; 0<Y<2)$.

35. Відомо, що $p(A)$ = 0,03, $p(B)$ = 0,05, $p(B|A)$ = 0,04. Обчислити коефіцієнт кореляції між A і B.

36. Випадкові події A і B мають однакову імовірність появи, яка дорівнює 0,6. Яка повинна бути умовна ймовірність $p(B|A)$, якщо відомо, що коефіцієнт кореляції між випадковими подіями A і B дорівнює 0,9?

37. За даною кореляційною матрицею трьохвимірної випадкової величини

$$K=\begin{pmatrix}49 & -14 & 12\\ -14 & 25 & -20\\ 12 & -20 & 36\end{pmatrix}$$

побудувати нормовану кореляційну матрицю.

38. Нехай ξ – число попадань у мішень при трьох пострілах, а η – число промахів. Обчислити коефіцієнт кореляції $r_{\xi\eta}$.

39. Задано функцію густини імовірностей

$$f(x, y) = \frac{\sqrt{3}}{2\pi} e^{-x^2 + xy - y^2}, \ \ x, y \in R.$$

Знайти функції регресії *X* на *Y* та *Y* на *X*.

40. Незалежні випадкові величини X та Y розподілені за нормальними законами:

$$f(x) = \frac{1}{2\sqrt{2\pi}} e^{-\frac{(x+2)^2}{8}}, \ \ f(y) = \frac{1}{3\sqrt{2\pi}} e^{-\frac{(y-3)^2}{18}}.$$

Обчислити: 1) $P(|X| < 3)$; 2) $P(|Y| > 1)$; 3) $P(|X| < 3; |Y| < 1)$.

41. Випадкова двовимірна величина (*X, Y*) розподілена за нормальним законом з параметрами $M(X) = 1$, $M(Y) = -2$, $D(X) = 2$, $D(Y) = 1$, $r_{xy} = 0$. Знайти ймовірність того, що випадкова точка (*X, Y*) потрапить в область $D = \{(x, y) : x^2 + y^2 \le 1\}$.

§ 8. ФУНКЦІЯ ВИПАДКОВИХ ВЕЛИЧИН

Функція одного випадкового аргументу та її математичне сподівання. Логарифмічний нормальний закон та χ-розподіл. Функції двох випадкових величин. Розподіл Ст'юдента, розподіл Фішера-Снедокора.

КОРОТКІ ТЕОРЕТИЧНІ ВІДОМОСТІ

Функція одного випадкового аргументу та її математичне сподівання

Якщо кожному можливому значенню випадкової величини (одновимірної) X відповідає одне можливе значення випадкової величини Y, то Y називається функцією випадкового аргумента X:

$$Y = \varphi(X). \tag{8.1}$$

Якщо **аргумент** X – **дискретна випадкова величина**, закон розподілу імовірностей якої має такий вид:

$$\begin{array}{c|cccc} X & x_1 & x_2 & \dots & x_n \\ \hline P & p_1 & p_2 & \dots & p_n \end{array},$$

то Y також дискретна випадкова величина.

Оскільки випадкові події $(X = x_i)$ та $(Y = \varphi(x_i))$, де $i = \overline{1, n}$, рівносильні, то їх імовірності рівні. Тому випадкова величина Y розподілена за таким законом:

$$\begin{array}{c|cccc} Y & \varphi(x_1) & \varphi(x_2) & \dots & \varphi(x_n) \\ \hline P & p_1 & p_2 & \dots & p_n \end{array}. \tag{8.2}$$

Якщо серед можливих значень Y є однакові, то в остаточному законі розподілу слід залишити одне із них, поставивши йому у відповідність суму імовірностей повторюваних значень Y.

Використавши означення математичного сподівання до закону розподілу (8.2), отримаємо:

$$M(Y) = \sum_{i=1}^{n} \varphi(x_i) p_i. \tag{8.3}$$

Нехай **аргумент** X **функції** (8.1) **є неперервною випадковою величиною**. Намітимо підхід до знаходження законів розподілу (функції розподілу та густини розподілу імовірностей) випадкової величини Y у загальному випадку функції φ, що дозволить оцінити труднощі, які виникають на цьому шляху, а також виділити простіші випадки, що часто зустрічаються на практиці.

Функція розподілу випадкової величини *Y*, яка пов'язана із величиною X функціональною залежністю (8.1) має вигляд

$$G(y) = P(Y < y) = \sum_i P(X \in \alpha_i(y)) = \sum_i \int_{c_i(y)}^{d_i(y)} f(x)dx, \qquad (8.4)$$

де $\alpha_i(y) = (c_i(y); d_i(y)) = \{x : \varphi(x) < y\}$, $f(x)$ – функція густини розподілу імовірностей випадкової величини X.

Відмітимо, що межі інтервалів $\alpha_i(y)$ можуть бути виражені як явні функції y для кожної конкретної функції $y = \varphi(x)$.

Густину розподілу імовірностей $g(y)$ величини Y можна знайти, продиференціювавши ліву і праву частини рівностей (8.4).

Розглянутий загальний випадок функції $y = \varphi(x)$ дозволяє очікувати, що випадок монотонної функції φ суттєво спростить знаходження законів розподілу випадкової величини Y.

Якщо $y = \varphi(x)$ – диференційована строго зростаюча або строго спадаюча функція, обернена до якої $x = \psi(y)$, то густина розподілу імовірностей випадкової величини Y має такий вид:

$$g(y) = f[\psi(y)]\,|\psi'(y)|. \qquad (8.5)$$

Відмітимо, що для знаходження $M(Y)$ не обов'язково знати густину розподілу величини Y. За аналогією із рівністю (8.3) можна отримати, що

$$M(Y) = \int_{-\infty}^{\infty} \varphi(x) f(x)dx.$$

Якщо ж $g(y)$ відома, тоді має місце така рівність:

$$M(Y) = \int_{-\infty}^{\infty} y g(y)dy.$$

Дисперсія функції розподілу випадкової величини *Y*, яка пов'язана із величиною X функціональною залежністю (8.1) визначається формулою:

$$D(Y) = \begin{cases} \sum_{i=1}^{n} [\varphi(x_i) - M(Y)]^2 p_i & \text{– для дискретної в.в.}(X); \\ \int_{-\infty}^{\infty} [\varphi(x) - M(Y)]^2 f(x)dx & \text{– для неперервної в.в.}(X). \end{cases}$$

Якщо ж відома $g(y)$, тоді має місце рівність:

$$D(Y) = \int_{-\infty}^{\infty} [y - M(Y)]^2 g(y)dy.$$

Випадкова величина Y називається **логарифмічно нормально** (**логнормально**) **розподіленою**, якщо її густина розподілу імовірностей має вид:

$$g(y) = \begin{cases} 0, & якщо\ y \le 0, \\ \dfrac{1}{y\sigma_x\sqrt{2\pi}} e^{-\frac{(\ln y - a_x)^2}{2\sigma_x^2}}, & якщо\ y > 0. \end{cases} \tag{8.6}$$

Випадкова величина Y має **розподіл** χ, якщо її густина розподілу імовірностей має вид:

$$g(y) = \begin{cases} 0, & якщо\ y \le 0, \\ \dfrac{1}{2^{k/2-1}\Gamma(k/2)} y^{k-1} e^{-y^2/2}, & якщо\ y > 0. \end{cases} \tag{8.7}$$

Функції двох випадкових величин

Розглянемо систему двох неперервних випадкових величин (X, Y) із густиною розподілу імовірностей $f(x, y)$. Нехай випадкова величина Z пов'язана із випадковими величинами X та Y функціональною залежністю $Z = \varphi(X, Y)$.

Функція розподілу випадкової величини Z визначається за формулою

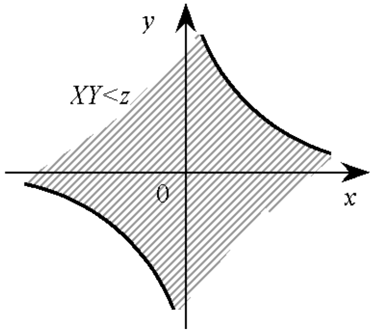
$$G(z) = P\big(\varphi(X,Y) < z\big) = P\big((X,Y) \in D\big) = \iint\limits_D f(x,y)dxdy, \tag{8.8}$$

де $D = \{(x,y) : \varphi(x,y) < z\}$.

Величина z входить в праву частину рівності (8.8) неявно через межі інтегрування. Явний вид функції $z = \varphi(x, y)$ дозволяє в принципі знайти ці межі як функцію z.

Густина розподілу імовірностей випадкової величини Z є похідною від функції розподілу: $g(z) = G'(z)$.

Якщо, наприклад, $Z = XY$, то для визначення функції розподілу і густини розподілу потрібно побудувати криву $xy = z$ на площині $x0y$ і вибрати ту її частину, для якої $XY < z$. В розглянутому випадку – це заштрихована частина на мал. 8.1.



Мал. 8.1.

Тоді

$$G(z) = \int_{-\infty}^{0} dx \int_{z/x}^{\infty} f(x,y)dy + \int_{0}^{\infty} dx \int_{\infty}^{z/x} f(x,y)dy;$$

$$g(z) = G'(z) = -\int_{-\infty}^{0} \frac{1}{x} f\left(x, \frac{z}{x}\right)dx + \int_{0}^{\infty} \frac{1}{x} f\left(x, \frac{z}{x}\right)dx.$$

Математичне сподівання та дисперсія функції двох випадкових величин $Z = \varphi(X, Y)$ визначаються формулами:

$$M(Z) = \begin{cases} \sum_{i=1}^{n}\sum_{j=1}^{m} \varphi(x_i, y_j)p_{ij} & \text{– для дискретних в.в. } X \text{ та } Y; \\ \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \varphi(x,y)f(x,y)dxdy & \text{– для неперервних в.в. } X \text{ та } Y; \end{cases}$$

$$D(Z) = \begin{cases} \sum_{i=1}^{n}\sum_{j=1}^{m} [\varphi(x_i, y_j) - M(Z)]^2 p_{ij} & \text{– для дискретних в.в. } X \text{ та } Y; \\ \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} [\varphi(x,y) - M(Z)]^2 f(x,y)dxdy & \text{– для неперервних в.в. } X \text{ та } Y, \end{cases}$$

де p_{ij} – імовірність того, що вектор (X, Y) набере значення (x_i, y_j) при $i = \overline{1,n}, \quad j = \overline{1,m}$.

Розподіл Ст'юдента, розподіл Фішера-Снедокора

Нехай випадкова величина X має розподіл $\chi/\sqrt{k}$, випадкова величина Y розподілена за нормованим нормальним законом із параметрами $M(Y) = 0$, $\sigma(Y) = 1$. Якщо X та Y незалежні випадкові величини, тоді **випадкова величина $Z = Y/X$ називається розподіленою за законом Ст'юдента (t-розподіл)**.

Густина розподілу імовірностей Ст'юдента має вид:

$$g(z)=\frac{1}{\sqrt{\pi k}}\cdot\frac{\Gamma\left(\frac{k+1}{2}\right)}{\Gamma(k/2)}\left(1+z^2/k\right)^{-(k+1)/2},\quad -\infty<z<\infty.$$

Розподіл Ст'юдента має k ступенів вільності і при зростанні k він швидко наближається до нормального.

Якщо Y та X – незалежні випадкові величини, розподілені за законом χ^2 із ступенями вільності k_1 та k_2 відповідно, тоді величина

$$F=\frac{Y/k_1}{X/k_2}$$

має розподіл, який називається **розподілом *F* Фішера-Снедокора із ступенями вільності k_1 та k_2.**

Густина цього розподілу

$$g(z)=\begin{cases}0, & \text{при } z\le 0,\\ C_0\dfrac{z^{(k_1-2)/2}}{(k_2+k_1z)^{(k_2+k_1)/2}} & \text{при } z>0,\end{cases}$$

де

$$C_0=\frac{\Gamma\left(\frac{k_1+k_2}{2}\right)k_1^{k_1/2}k_2^{k_2/2}}{\Gamma(k_1/2)\Gamma(k_2/2)}.$$

РОЗВ'ЯЗУВАННЯ ТИПОВИХ ЗАДАЧ

Задача 8.1. Випадкова величина X задана законом розподілу

X	-1	1	2
P	0,1	0,5	0,4

. Знайти $M(Y)$, де $Y=\varphi(X)=2X^2+1$.

○ Знайдемо можливі значення Y:

$\varphi(-1)=2(-1)^2+1=3,\ \ \varphi(1)=2\cdot 1^2+1=3,\ \ \varphi(2)=2\cdot 2^2+1=9.$

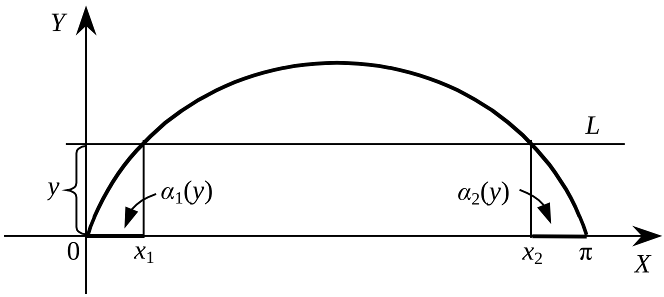
Згідно із (8.3) $M(Y)=3\cdot 0{,}1+3\cdot 0{,}5+9\cdot 0{,}4=5{,}4.$ ●

Задача 8.2. Випадкова величина X задана густиною розподілу імовірностей

$$f(x)=\begin{cases}1/\pi, & \text{якщо } x\in(0,\pi),\\ 0, & \text{якщо } x\notin(0,\pi).\end{cases}$$

Знайти закон розподілу випадкової величини $Y=\sin X$.

○ Побудуємо графік функції $y=\sin x$ на відрізку $[0,\pi]$ (мал. 8.2.).



Мал. 8.2.

Використовуючи формулу (8.4), отримаємо

$$G(y) = \int_0^{x_1} f(x)dx + \int_{x_2}^{\pi} f(x)dx.$$

Оскільки $x_1 = \arcsin y$, $x_2 = \pi - \arcsin y$, бо $x_2 \in \left(\frac{\pi}{2}, \pi\right)$, то функція розподілу

$$G(y) = \int_0^{\arcsin y} \frac{1}{\pi} dx + \int_{\pi - \arcsin y}^{\pi} \frac{1}{\pi} dx = \frac{2}{\pi} \arcsin y, \text{ де } y \in (0,1).$$

Густина розподілу імовірностей

$$g(y) = G'(y) = \left(\frac{2}{\pi} \arcsin y\right)' = \frac{2}{\pi\sqrt{1-y^2}} \text{ для } y \in (0,1).$$

Отже, функція розподілу випадкової величини Y

$$G(y) = \begin{cases} 0, & \text{якщо } y \le 0, \\ 2\pi^{-1} \arcsin y, & \text{якщо } 0 < y < 1, \\ 1, & \text{якщо } y \ge 1. \end{cases}$$

а густина розподілу

$$g(y) = \begin{cases} \dfrac{2}{\pi\sqrt{1-y^2}}, & \text{якщо } y \in (0,1), \\ 0, & \text{якщо } y \notin (0,1). \end{cases}$$ •

Задача 8.3. Випадкова величина X розподілена за нормальним законом із параметрами a та σ.

Знайти закон розподілу випадкової величини $Y = bX + c$, де b та c детерміновані величини (сталі).

○ Оскільки функція $y = bx + c$ є монотонною (при $b > 0$ монотонно зростає, при $b < 0$ монотонно спадає), то згідно із формулою (8.5) $\psi(y) = (y-c)/b$, $\psi'(y) = 1/b$, густина розподілу випадкової величини *Y* має такий вид:

$$g(y) = f\left((y-c)/b\right)\left|1/b\right| = \frac{1}{|b|\sigma\sqrt{2\pi}} e^{-\frac{[y-(ab+c)]^2}{2b^2\sigma^2}}.$$

Отже, дана випадкова величина *Y* є нормально розподіленою із параметрами $M(Y) = ab + c$ і $\sigma(Y) = |b|\sigma$. ●

ЗАДАЧІ ДЛЯ САМОСТІЙНОГО РОЗВ'ЯЗУВАННЯ

Завдання №9

В задачах 1-8 випадкова величина *X* задана законом розподілу

X	-2	-1	0	1	2	3
P	0,05	0,2	0,1	0,3	0,1	0,25

Потрібно для випадкової величини *Y*:

1) побудувати ряд розподілу імовірностей;
2) обчислити числові характеристики;
3) побудувати функцію розподілу імовірностей.

1. $Y = X + 2.$ 2. $Y = |X - 1|.$ 3. $Y = \sqrt{|X|}.$ 4. $Y = 3X^2.$

5. $Y = \sin^2 \frac{\pi}{3} X.$ 6. $Y = 3^{|X|}.$ 7. $Y = e^X.$ 8. $Y = \lg|X|.$

В задачах 9-14 випадкова величина *X* задана густиною розподілу *f*(*x*). Потрібно для випадкової величини *Y* :

1) знайти густину розподілу імовірностей;
2) знайти функцію розподілу імовірностей;
3) обчислити $M(Y)$; $D(Y)$; $\sigma(Y)$.

9. $f(x) = \begin{cases} 1/\pi, & x \in (-\pi/2, \pi/2), \\ 0, & x \notin (-\pi/2, \pi/2). \end{cases}$

$Y = \cos X.$

10. $f(x) = \begin{cases} \cos x, & x \in [0, \pi/2], \\ 0, & x \notin [0, \pi/2]. \end{cases}$

$Y = X^4.$

11. $f(x)=\begin{cases}\sin x, & x\in[0,\pi/2],\\ 0, & x\notin[0,\pi/2].\end{cases}$

$Y=X^2.$

12. $f(x)=\begin{cases}\cos x, & x\in[-\pi/2,\pi/2],\\ 0, & x\notin[-\pi/2,\pi/2].\end{cases}$

$Y=X^3$

13. $f(x)=\begin{cases}xe^{-x^2/2}, & x>0,\\ 0, & x\le 0.\end{cases}$

$Y=e^{-X^2}.$

14. $f(x)=\begin{cases}5\sqrt[5]{x^3}, & x\in(0,\,1],\\ 0, & x\notin(0,\,1].\end{cases}$

$Y=X^2.$

В задачах 15-18 випадкова величина *X* рівномірно розподілена на відрізку $[\alpha,\beta]$. Знайти густину і функцію розподілу імовірностей випадкової величини *Y*.

15. $Y=X^3.$ $\alpha=0,\beta=3.$

16. $Y=\sqrt{X+1}.$ $\alpha=0,\beta=8.$

17. $Y=2\sin X.$ $\alpha=0,\beta=\pi.$

18. $Y=\sin X.$ $\alpha=-\pi/2,\beta=\pi/2.$

§ 9. ЗАКОН ВЕЛИКИХ ЧИСЕЛ

Лема та нерівність Чебишева. Теорема Чебишева (стійкість середніх). Теорема Бернуллі (стійкість відносних частот). Центральна гранична теорема Ляпунова.

КОРОТКІ ТЕОРЕТИЧНІ ВІДОМОСТІ

Лема Чебишева. Якщо всі можливі значення випадкової величини Y невід'ємні, тоді імовірність того, що вона при випробуванні набере значення, більше від додатного числа b, не більша від дробу, чисельник якого – математичне сподівання від Y, а знаменник – число b:

$$P(Y > b) \le M(Y)/b. \tag{9.1}$$

Нерівність Чебишева. Імовірність того, що відхилення випадкової величини X від її математичного сподівання за абсолютною величиною не більше від додатного числа ε, не менша, ніж $1 - D(X)/\varepsilon^2$:

$$P\left(\left|X - M(X)\right| \le \varepsilon\right) \ge 1 - D(X)/\varepsilon^2. \tag{9.2}$$

Зауваження. Якщо випадкова величина X розподілена за біноміальним законом (X – число появи події в n повторних незалежних випробуваннях), тоді нерівність Чебишева набере такого виду:

$$P\left(\left|m - np\right| \le \varepsilon\right) \ge 1 - \frac{npq}{\varepsilon^2} \tag{9.3}$$

або

$$P\left(\left|\frac{m}{n} - p\right| \le \varepsilon\right) \ge 1 - \frac{pq}{n\varepsilon^2}, \tag{9.4}$$

де m – число появи події A в n повторних незалежних випробуваннях, $p = P(A)$, m/n – відносна частота появи події A.

Теорема Чебишева. Якщо $X_1, X_2, ..., X_n$ попарно незалежні випадкові величини, дисперсії яких рівномірно обмежені $\left(D(X_i) \le C, i = \overline{1, n}\right)$, тоді для довільного ε > 0 і достатньо великого n імовірність події

$$\left|\frac{X_1 + X_2 + ... + X_n}{n} - \frac{M(X_1) + M(X_2) + ... + M(X_n)}{n}\right| \le \varepsilon \tag{9.5}$$

буде як завгодно близькою до одиниці.

При доведені теореми встановлюється правильність нерівності

$$P\left(\left|\frac{X_1+X_2+...+X_n}{n}-\frac{M(X_1)+M(X_2)+...+M(X_n)}{n}\right|\le\varepsilon\right)\ge$$
$$\ge 1-C\big/(n\varepsilon^2). \qquad (9.6)$$

Теорема Бернуллі. Якщо в кожному із n **повторних випробувань імовірність** p **появи події** A **стала, тоді як завгодно близька до одиниці імовірність того, що відхилення відносної частоти події** A **від імовірності** p **за абсолютною величиною буде як завгодно малим, якщо число випробувань достатньо велике.**

З нерівності (9.6) можна отримати нерівність

$$P\left(\left|{m}\big/{n}-p\right|\le\varepsilon\right)\ge 1-\frac{0{,}25}{n\varepsilon^2}. \qquad (9.7)$$

В теоремі Чебишева не використовувалася інформація про закони розподілу випадкових величин. Разом з тим для задач теорії і практики важливим є таке питання: за яким законом розподіляється сума достатньо великого числа випадкових величин?

Випадкову величину

$$Z_n=\left[\sum_{i=1}^{n}X_i-\sum_{i=1}^{n}M(X_i)\right]\Bigg/\sqrt{\sum_{i=1}^{n}D(X_i)}. \qquad (9.8)$$

будемо називати **нормованою сумою** або **центрованою випадковою величиною**.

Теорема Ляпунова. Якщо $X_1, X_2, ..., X_n$ **– незалежні випадкові величини, які мають математичні сподівання** m_i, **дисперсії** σ_i^2 **і скінченні абсолютні центральні моменти третього порядку** $\left|\mu_3(X_i)\right|$ **що задовольняють умову**

$$\lim_{n\to\infty}\left(\sum_{i=1}^{n}\left|\mu_3(X_i)\right|\right)\Big/\left(\sum_{i=1}^{n}\sigma_i^2\right)^{3/2}=0, \qquad (9.9)$$

тоді при необмеженому збільшенні n **закон розподілу нормованої суми (9.8) збігається за імовірністю до нормального закону з параметрами** $a = 0$, $\sigma = 1$.

Зміст умови (9.9) полягає в тому, що дисперсія кожної випадкової величини $X_i, i=\overline{1,n},$ складає лише малу частину в загальній дисперсії суми $\sum_{i=1}^{n} X_i$.

Наслідок теореми Ляпунова. Якщо виконуються умови теореми Ляпунова, тоді випадкова величина $\sum_{i=1}^{n} X_i$ для великих n з достатнім ступенем точності розподілена за нормальним законом з параметрами

$$a=\sum_{i=1}^{n} m_i,\quad \sigma^2=\sum_{i=1}^{n}\sigma_i^2.$$

В практичних задачах центральну граничну теорему Ляпунова часто використовують для обчислення імовірності того, що сума кількох випадкових величин набере значення, яке належить вказаному інтервалу.

Підставою такого використання є так званий

Частковий наслідок теореми Ляпунова. Якщо X_1, X_2, ..., X_n – незалежні випадкові величини, у яких існують рівні математичні сподівання $M(X_i)=m$, дисперсіїї $D(X_i)=\sigma^2$ і абсолютні центральні моменти третього порядку $M\left[\left|X_i-m\right|^3\right]=\left|\mu_3\right|\quad i=\overline{1,n}$, то закон розподілу суми $Y_n=X_1+X_2+...+X_n$ при $n\to\infty$ необмежено наближається до нормального закону з параметрами $a=nm$, $\boldsymbol{\sigma}^2=n\sigma^2$.

РОЗВ'ЯЗУВАННЯ ТИПОВИХ ЗАДАЧ

Задача 9.1. Випадкова величина задана законом розподілу

X	0,3	0,6
P	0,2	0,8

Використовуючи нерівність Чебишева, оцінити імовірність того, що X відхилиться від свого математичного сподівання на величину, яка за абсолютною величиною не перевищить 0,2. Перевірити точність отриманої оцінки.

○ Знайдемо $M(X)$ та $D(X)$

$$M(X)=0{,}3\cdot 0{,}2+0{,}6\cdot 0{,}8=0{,}54,$$

$$D(X) = M(X^2) - [M(X)]^2 = 0{,}3^2 \cdot 0{,}2 + 0{,}6^2 \cdot 0{,}8 - (0{,}54)^2 = 0{,}0144.$$

Використавши нерівність (9.2) з ε = 0,3, отримаємо

$$P\left(|X - 0{,}54| \le 0{,}2\right) \ge 1 - 0{,}0144/0{,}04 = 0{,}64.$$

Випадкова подія $|X - 0{,}54| \le 0{,}2$ рівносильна події $0{,}34 \le X \le 0{,}74$, яка відбувається з урахуванням закону розподілу, тоді і тільки тоді, коли X=0,6. Тому

$$P\left(|X - 0{,}54| \le 0{,}2\right) = P(0{,}34 \le X \le 0{,}74) = P(X = 0{,}6) = 0{,}8.$$

Отже, нижня межа оцінки імовірності дорівнює 0,64, а справжнє значення імовірності дорівнює 0,8. ●

Задача 9.2. Імовірність появи події в кожному випробуванні дорівнює 0,25. Використовуючи нерівність Чебишева, оцінити імовірність того, що число X появи події міститься в межах від 150 до 250 включно, якщо буде проведено 800 випробувань. Перевірити точність отриманої оцінки.

○ Використаємо нерівність (9.3), де $X = m$, $np = 800 \cdot 0{,}25 = 200$ і попередньо треба знайти ε. Випадкову подію $(150 \le m \le 250)$ можна записати в такій рівносильній формі: $\left(|m - 200| \le 50\right)$.

А тому в (9.3) ε = 50 і

$$P(150 \le m \le 250) = P\left(|m - 200| \le 50\right) \ge 1 - \frac{800 \cdot 0{,}25 \cdot 0{,}75}{50^2} = 0{,}94.$$

Тобто, остаточно $P(150 \le m \le 250) \ge 0{,}94$.

Перевіримо точність отриманого результату, використавши інтегральну формулу Лапласа (3.7)

($npq = 800 \cdot 0{,}25 \cdot 0{,}75 = 150 >> 9$):

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

$$x_2 = \frac{m_2 - np}{\sqrt{npq}} = \frac{250 - 800 \cdot 0{,}25}{\sqrt{150}} = \frac{50}{12{,}247} = 4{,}08,$$

$$x_1 = \frac{m_1 - np}{\sqrt{npq}} = \frac{150 - 200}{\sqrt{150}} = \frac{-50}{12{,}247} = -4{,}08,$$

$$P_n(150 \le m \le 250) \approx \Phi(4{,}08) - \Phi(-4{,}08) = 2\Phi(4{,}08) = 2 \cdot 0{,}4999992 =$$
$$=0{,}9999984.$$

Отже, нижня межа оцінки імовірності дорівнює 0,94, а справжнє значення імовірності дорівнює 0,9999984. ●

Задача 9.3. Дисперсія кожної із 2500 незалежних випадкових величин не перевищує 5. Оцінити імовірність того, що абсолютна величина відхилення середнього арифметичного цих випадкових величин від середнього арифметичного їх математичних сподівань не перевищить 0,4.

○ Для задачі виконуються обидві умови теореми Чебишева, а тому можна використати нерівність (9.6), де n = 2500, ε = 0,4, C = 5. Тоді

$$P\left(\left|\left(\sum_{i=1}^{2500} X_i - \sum_{i=1}^{2500} M(X_i)\right)\Big/2500\right| \le 0{,}4\right) \ge 1 - \frac{5}{2500(0{,}4)^2} = \frac{79}{80}.$$

Отже, шукана імовірність оцінюється знизу числом 79/80. ●

Задача 9.4. Відомо, що 80% виробів механічного цеху є першосортними. Оцінити імовірність того, що відносна частоту виробів першого сорту серед 20 000 виготовлених відрізнятиметься від імовірності виготовлення виробу першого сорту не більше, ніж на 0,02 в той чи інший бік. Перевірити точність отриманої оцінки.

○ Подія A – виготовлений виріб першосортний. Тоді $P(A)$ = 0,8, n = 20 000, ε = 0,02. Використаємо нерівність (9.7):

$$P\left(\left|m\big/n - 0{,}8\right| \le 0{,}02\right) \ge 1 - \frac{0{,}25}{20000 \cdot 0{,}0004} = 31\big/32.$$

Таким чином, шукана імовірність оцінюється знизу числом 31/32.

Для перевірки точності отриманої оцінки використаємо формулу (3.8):

$$P\left(|m/n - p| \le \varepsilon\right) \approx 2\Phi\left(\varepsilon\sqrt{n/(pq)}\right),$$

ε=0,02, n=20000, p=0,8, q=0,2,

$$P\left(|m/n - 0{,}8| \le 0{,}02\right) \approx 2\Phi\left(0{,}02\sqrt{\frac{20000}{0{,}8 \cdot 0{,}2}}\right) = 2\Phi(7{,}07) = 2 \cdot 0{,}5 = 1.$$

Оскільки npq=3200 і 7,07>5, то похибки у наближеній рівності (3.8) і $\Phi(7{,}07) \approx 0{,}5$ є надзвичайно малими, тобто подія $\left(|m/n - 0{,}8| \le 0{,}02\right)$ є майже достовірною. ●

Задача 9.5. Випадкові величини X_1, X_2, …, X_n, … є незалежними, причому для k=1, 2, …, n, … закон розподілу величини X_k має такий вид:

X_k	$-\sqrt{k}$	0	$\sqrt{k}$
P	$1/(4k)$	$1-2/k$	$1/(4k)$

Чи можна використати теорему Чебишева до цієї послідовності випадкових величин?

○ Можливість використання теореми Чебишева означає виконання обох умов теореми відносно послідовності випадкових величин. Згідно з умовою задачі перша з них виконується. Перевіримо виконання другої умови – рівномірну обмеженість дисперсій випадкових величин. Для цього знайдемо $D(X_k)$, де k=1, 2, …, n, … .

Оскільки $M(X_k)=\frac{-\sqrt{k}}{4k}+0\cdot(1-\frac{2}{k})+\frac{\sqrt{k}}{4k}=0$, то за розрахунковою формулою

$$D(X_k)=M(X_k^2)-[M(X_k)]^2=\left(-\sqrt{k}\right)^2\cdot\frac{1}{4k}+0\cdot\left(1-\frac{2}{k}\right)+\left(-\sqrt{k}\right)^2\cdot\frac{1}{4k}=$$

$=0{,}5.$

Отже, дисперсії випадкових величин X_1, X_2, …, X_n, … обмежені в сукупності сталою C=0,5, а тому до даної послідовності випадкових величин можна використати теорему Чебишева. ●

Задача 9.6. Розмір виплати кожному клієнту банку випадковий. Статистичні дані цього банку показали, що середня виплата одному клієнту складає 1000 грн., а середнє квадратичне відхилення цієї ж величини – 360 грн. Вважаючи, що виплати окремим клієнтам є незалежними, знайти, скільки повинно бути готівки в банку, щоб з імовірністю, не меншою 0,95, грошей вистачило б на обслуговування 80 клієнтів. Знайти цю саму величину у додатковому припущенні, що абсолютний центральний момент третього порядку для виплати кожного клієнту є сталим.

○ Нехай випадкова величина X_i – виплата i-му клієнту, де $i=\overline{1,80}$; x – невідома кількість готівки в банку для обслуговування 80 клієнтів. За умовою $M(X_i)$=1000 і $\sigma(X_i)$=360 для $i=\overline{1,80}$,

$$P(X_1+X_2+...+X_{80}\le x)\ge 0{,}95 \tag{9.10}$$

Оскільки інформація про закон розподілу випадкової величини $X_1+X_2+…+X_{80}$ відсутня, то для знаходження невідомого числа x

використаємо нерівність Чебишева для величини $\overline{X}=(X_1+X_2+...+X_{80})/80$:

$$P\left(\left|\overline{X}-M(\overline{X})\right|\le\varepsilon\right)\ge 1-D(\overline{X})\big/\varepsilon^2. \qquad (9.11)$$

Враховуючи властивості математичного сподівання і дисперсії, незалежність випадкових величин $X_1, X_2, \ldots, X_{80}$, а також умову задачі, отримаємо:

$$M(\overline{X})=M\left(\sum_{i=1}^{80}X_i/80\right)=\left[\sum_{i=1}^{80}M(X_i)\right]\Big/80=\left(\sum_{i=1}^{80}1000\right)\Big/80=1000,$$

$$D(\overline{X})=D\left(\sum_{i=1}^{80}X_i/80\right)=\frac{1}{80^2}\sum_{i=1}^{80}D(X_i)=\frac{1}{80^2}\sum_{i=1}^{80}\sigma^2(X_i)=$$

$$=\frac{1}{80^2}\sum_{i=1}^{80}360^2=360^2\big/80=1620.$$

Випадкова подія $\left(\sum_{i=1}^{80}X_i\le x\right)$ рівносильна події $\left(\frac{X_1+X_2+...+X_{80}}{80}-1000\le\frac{x}{80}-1000\right)$

або $\left[\left|\overline{X}-M(\overline{X})\right|\le\frac{x}{80}-1000\right]$,

звідки

$$P(X_1+X_2+...+X_{80}\le x)=P\left[\left|\overline{X}-M(\overline{X})\right|\le\frac{x}{80}-1000\right].$$

Тому якщо ліва частина нерівності (9.11), де $\varepsilon=\frac{x}{80}-1000$, буде не меншою від 0,95, тоді виконається нерівність (9.10). Отже, для знаходження невідомого числа x отримаємо нерівність

$$1-\frac{1620}{\left(\frac{x}{80}-1000\right)^2}\ge 0{,}95\ ,$$

яка рівносильна нерівностям

$$\frac{1620}{\left(\frac{x}{80}-1000\right)^2}\le 0{,}05\Leftrightarrow\left(\frac{x}{80}-1000\right)^2\ge 32400\Leftrightarrow$$

$$\frac{x}{80}-1000\geq 180 \Leftrightarrow x\geq 94400.$$

Врахуємо тепер додаткову умову про сталість для кожного клієнта абсолютного центрального момента третього порядку величини виплати. Тоді згідно із спрощеним наслідком теореми Ляпунова випадкова величина $Y_{80}=\sum_{i=1}^{80} X_i$ наближено розподілена за нормальним законом із параметрами $a=\sum_{i=1}^{80} M(X_i)=80000$,

$$\sigma=\sqrt{\sum_{i=1}^{80}\sigma^2(X_i)}=\sqrt{80\cdot 360^2}=3219{,}94.$$

З одного боку

$$P(Y_{80}\leq x)=P(-\infty<Y_{80}\leq x)=\Phi\left(\frac{x-a}{\sigma}\right)-\Phi\left(\frac{-\infty-a}{\sigma}\right)=$$

$$=\Phi\left(\frac{x-80000}{3219{,}94}\right)+0{,}5,$$

а за умовою $P(Y_{80}\leq x)\geq 0{,}95.$ Отже, x можна знайти із нерівності

$$\Phi\left(\frac{x-80000}{3219{,}94}\right)\geq 0{,}45.$$

Використавши табл. 3 додатків і монотонне зростання функції Лапласа, отримаємо

$$\frac{x-80000}{3219{,}94}\geq 1{,}645,$$

звідки $x\geq 85296{,}801.$ ●

Задача 9.7. Кожна із 200 незалежних випадкових величин $X_1,X_2,\ldots,X_{200}$ рівномірно розподілена на проміжку [0; 0,8]. Знайти:
1) наближений закон розподілу для випадкової величини $Y=\sum_{i=1}^{200} X_i$; 2) P(Y<0,7).

○ Згідно із задачею 6.9 для рівномірно розподіленої на проміжку $[a;b]$ величини X

$$M(X)=\frac{b+a}{2},\ D(X)=\frac{(b-a)^2}{12}.$$

Тому $M(X_i)=0{,}4$, $D(X_i)=4/75$, $i=\overline{1{,}200}$.

Очевидно, що і абсолютні центральні моменти третього порядку для кожної із величин $X_1, X_2, \ldots, X_{200}$ також рівні. За спрощеним наслідком теореми Ляпунова випадкову величину Y можна вважати розподіленою за нормальним законом із параметрами

$$a = \sum_{i=1}^{200} M(X_i) = 200 \cdot 0{,}4 = 80,\ \sigma = \sqrt{\sum_{i=1}^{200} D(X_i)} = \sqrt{32/3} = 4\sqrt{2/3}.$$

У відповідності із рівністю (6.1) густина розподілу величини Y має такий вид

$$f(x) = \frac{\sqrt{3}}{8\sqrt{\pi}} e^{-\frac{3(x-80)^2}{64}}.$$

2) За формулою (6.3)

$$P(Y > 0{,}7) = P(0{,}7 < Y < \infty) = \Phi\left(\frac{\infty - a}{\sigma}\right) - \Phi\left(\frac{0{,}7 - a}{\sigma}\right) =$$

$$= 0{,}5 - \Phi\left(\frac{0{,}7 - 80}{3{,}27}\right) = 0{,}5 + \Phi(24{,}25) \approx 1. \ \bullet$$

Задача 9.8. Випробування розробленого пристрою показали, що причиною його виходу з ладу є відмова певного елемента. Дослідження ідентичних елементів дозволило встановити, що час безвідмовної роботи кожного з них розподілений за показниковим законом, а середній час роботи складає 0,5год. Для збільшення надійності роботи пристрою його модернізували, обладнавши його ста ідентичними елементами таким чином, що при виході з ладу одного елемента миттєво включається наступний тощо. Знайти наближену імовірність того, що пристрій після модернізації безвідмовно пропрацює не менше 30 год.

○ Випадкова величина X_i – час безвідмовної роботи i-ого елемента $i = \overline{1,100}$ – за умовою задачі розподілена за показниковим законом, при цьому $M(X_\text{i})$=0,5год. Згідно із задачею 9.11 $M(X_i) = \frac{1}{\lambda}$,

$D(X_i) = \frac{1}{\lambda^2}$, $i = \overline{1,100}$, тому $\frac{1}{\lambda} = \frac{1}{2}$, звідки λ=2.

Стосовно випадкових величин $X_1, X_2, \ldots, X_{100}$ виконуються умови спрощеного наслідка теореми Ляпунова, згідно з яким величину

$Y=\sum_{i=1}^{100} X_i$ можна вважати розподіленою за нормальним законом з параметрами

$$a=M(Y)=\sum_{i=1}^{100} M(X_i)=0{,}5\cdot 100=50,$$

$$\sigma=\sigma(Y)=\sqrt{\sum_{i=1}^{100} D(X_i)}=\sqrt{\frac{1}{4}\cdot 100}=5.$$

Тоді за формулою (6.3)

$$P(Y\ge 30)\approx P(30\le Y<\infty)=\Phi\left(\frac{\infty-50}{5}\right)-\Phi\left(\frac{30-50}{5}\right)=$$

$$=0{,}5+\Phi(4)=0{,}5+0{,}4999683=0{,}9999683. \bullet$$

ЗАДАЧІ ДЛЯ САМОСТІЙНОГО РОЗВ'ЯЗУВАННЯ

Завдання №10

1. З якою імовірністю середнє арифметичне вимірів певної величини відповідає справжньому значенню цієї величини, якщо було здійснено 200 вимірювань із точністю 0,2 і при цьому середні квадратичні відхилення випадкових величин (результатів вимірювання) не перевищують 0,4?

2. Для визначення міцності волокон бавовни в отриманій партії відібрали 1000 зразків і випробували їх на розрив. При цьому виявилось, що середня міцність зразків дорівнює 5,8г/мм^2. Оцінити імовірність того, що похибка при визначенні міцності волокон не перевищить 0,2г/мм^2, якщо відомо, що середнє квадратичне відхилення міцності зразків не перевищує 1,44г/мм^2.

3. За статистичними даними в середньому 88% новонароджених доживають до 50 років. З допомогою нерівності Чебишева оцінити імовірність того, що з 1000 новонароджених частка тих, хто доживе до 50 років, буде відрізнятися від імовірності цієї ж події не більше ніж 0,04 за абсолютною величиною. Уточнити отриману оцінку.

4. Оцінити імовірність того, що відхилення довільної випадкової величини від її математичного сподівання буде не більшим трьох середніх квадратичних відхилень за абсолютною величиною. Знайти таку ж імовірність для рівномірного та показникового законів розподілу.

5. Випадкова величина $\overline{X}$ – середнє арифметичне 1000 незалежних випадкових величин, які мають один і той самий закон розподілу, при цьому середнє квадратичне відхилення кожної з них дорівнює 4. Яке максимальне відхилення величини $\overline{X}$ від її математичного сподівання можна очікувати з імовірністю 0,9732?

6. Середня зміни курсу акції певної компанії протягом одних біржових торгів складає 0,4%. Оцінити імовірність того, що на чергових торгах курс акції зміниться більше, ніж на 2%.

7. Протягом певного періоду на біржі зберігався відносно стабільний курс долара США до євро. На підставі біржової статистики за цей період складено закон розподілу випадкової величини. X – зміна курсу долара до євро (у відсотках):

X	–0,5	–0,25	0	0,25	0,5
P	0,1	0,2	0,5	0,15	0,05

З допомогою нерівності Чебишева оцінити імовірність того, що відбудеться зміна курсу долара до євро не більше ніж на 0,03% за абсолютною величиною. Знайти уточнене значення шуканої імовірності.

8. Дисперсія кожної із 2000 незалежних випадкових величин, які мають один і той самий закон розподілу, дорівнює 12. Оцінити імовірність того, що відхилення середнього арифметичного цих величин від середнього арифметичного їх математичних сподівань, взяте за абсолютною величиною, не перевищить 0,3.

9. При штампуванні 80% деталей виявляються першосортними. Користуючись нерівністю Чебишева, знайти, скільки треба взяти деталей, щоб з імовірністю, рівною 0,96, відносна частота деталей першого сорту серед них відрізнялася за модулем від імовірності отримання першосортної деталі не більше ніж на 0,04. Порівняти отриману кількість з уточненою.

10. Кількість води, яка використовується підприємством протягом доби для технічних потреб, є випадковою величиною, математичне сподівання якої дорівнює 120 м^3. Оцінити імовірність того, що найближчої доби витрати води підприємством виявляться: а) більшими від 140 м^3; б) не більшими 150 м^3.

11. Використовуючи нерівність Чебишева, оцінити імовірність того, що випадкова величина X відхилиться за модулем від свого математичного сподівання: а) не більше ніж чотири середніх

квадратичних відхилення; б) більше, ніж на три середніх квадратичних відхилення.

12. Імовірність виготовлення автоматом стандартної деталі дорівнює 0,96. Оцінити з допомогою нерівності Чебишева імовірність того, що число бракованих серед 4000 виготовлених деталей знаходиться в межах від 140 до 180 включно. Уточнити оцінювану імовірність з допомогою інтегральної формули Лапласа.

13. Для визначення середньої тривалості горіння електричних лампочок в партії із 300 однакових ящиків було випадковим чином відібрано по одній лампочці з кожного ящика. Оцінити імовірність того, що середня тривалість горіння відібраних 300 лампочок відрізняється від середньої тривалості горіння ламп у всій партії не більше ніж на 6 год. за абсолютною величиною, якщо відомо, що середнє квадратичне відхилення тривалості горіння ламп в кожному ящику не перевищує 8 год.

14. Статистичні дані рекламної компанії показали, що адресна реклама дає дві заявки в 50 випадків. Компанія розіслала 2000 рекламних проспектів. З'ясувати, чи можна з допомогою нерівності Чебишева оцінити імовірність того, що число заявок виявиться не меншим 60 і не більшим 90. Якщо використання нерівності Чебишева виявиться неможливим, то змінити верхню межу таким чином, щоб використання нерівності Чебишева стало можливим, а також оцінити відповідну імовірність. Уточнити отриманий результат з допомогою інтегральної формули Лапласа.

15. Імовірність вчасної реалізації одиниці продукції дорівнює 0,4. Оцінити імовірність того, що серед 400 підготовлених для реалізації одиниць продукції відхилення відносної частоти реалізації від імовірності 2/5 за абсолютною величиною буде не більшим від 0,2 та порівняти оцінювану імовірність із точним значенням.

16. Кожна із 120 незалежних випадкових величин X_1, X_2..., X_{120} має рівномірний розподіл на проміжку [0; 0,6]. Знайти: 1) наближений закон розподілу для випадкової величини $Y = \sum_{i=1}^{120} X_i$; 2) $P(Y>0,5)$.

17. Для деякого регіону 8% працездатного населення складають безробітні. З допомогою нерівності Чебишева оцінити імовірність

того, що рівень безробіття серед 40000 працездатних мешканців регіону буде міститися в межах від 6% до 10% включно.

18. Кожна із 100 незалежних випадкових величин X_1, X_2, ..., X_{100} розподілена за показниковим законом розподілу з параметром λ=10 для всіх X_i $i = \overline{1{,}100}$. Побудувати наближений закон розподілу випадкової величини $Y = \sum_{i=1}^{100} X_i$ та знайти $P(Y<5)$.

19. Імовірність того, що телевізор витримає гарантійний термін роботи, дорівнює 0,9 для всіх 100 телевізорів, які обслуговує гарантійна майстерня. Оцінити імовірність того, що число телевізорів, які витримають гарантійний термін роботи, буде в межах [85; 95], а також знайти точніше значення цієї ж імовірності.

20. Відомо, що 90% виробів є вищого ґатунку. Оцінити імовірність того, що відносна частота виробів вищого ґатунку серед 10000 виготовлених відрізнятиметься від імовірності виготовлення виробу такої ж якості за абсолютною величиною не більше ніж на 0,01. Порівняти отриману оцінку із більш точною.

21. Відомо, що $P(|X - M(X)| < \varepsilon) \geq 0{,}9$, $\sigma(X) = 0{,}2$. Використовуючи нерівність Чебишева знайти ε.

22. Середнє значення довжини заготовки деталі складає 60см, а дисперсія 0,25см^2. Використовуючи нерівність Чебишева, оцінити імовірність того, що навмання взята заготовка виявиться по довжині в межах від 59,6 до 60,4 включно. Уточнити оцінювану імовірність, якщо відомо, що довжина навмання взятої заготовки має нормальний закон розподілу.

23. Середній дохід на душу населення розподілений за нормальним законом розподілу з параметрами a та σ. Оцінити за нерівністю Чебишева $P(|X - a| > 2\sigma)$ та порівняти її з точним значенням.

24-33. Відомо, що частка продукції вищого сорту серед всієї продукції дорівнює b. Оцінити імовірність того, що m виробів вищого сорту серед відібраних відрізнятиметься від математичного сподівання цього числа за абсолютною величиною не більше, ніж на ε штук, а також знайти точне значення цієї імовірності.

Номер задачі	b	m	ε
24	7/8	70000	700
25	4/5	100000	1000
26	5/8	90000	900
27	0,9	80000	800
28	0,7	250000	2500
29	0,9	300000	3000
30	3/4	150000	1500
31	0,8	200000	2000
32	4/5	100000	1000
33	0,6	40000	400

34-40. Послідовність незалежних випадкових величин X_1, X_2..., X_n задана законом розподілу.

а) $\begin{array}{c|ccc} X_n & -k\sqrt{n} & 0 & k\sqrt{n} \\ \hline P & \dfrac{k}{n} & 1-\dfrac{2k}{n} & \dfrac{k}{n} \end{array}$, б) $\begin{array}{c|ccc} X_n & -k & 0 & k \\ \hline P & \dfrac{k}{n} & 1-\dfrac{2k}{n} & \dfrac{k}{n} \end{array}$,

де k – номер задачі (34, 35, ..., 40), n=1, 2, ..., n. Чи можна використати для цих величин теорему Чебишева?

ЧАСТИНА ДРУГА
МАТЕМАТИЧНА СТАТИСТИКА

§ 1.ВСТУП В МАТЕМАТИЧНУ СТАТИСТИКУ. ВИБІРКОВИЙ МЕТОД

Задачі математичної статистики. Генеральна та вибіркова сукупності. Способи утворення вибіркової сукупності. Статистичний розподіл вибірки. Емпірична функція розподілу та її властивості. Графічне зображення статистичних розподілів (полігон та гістограма). Числові характеристики вибірки. Метод добутків обчислення зведених характеристик вибірки. Числові характеристики сукупностей, що складаються із груп.

КОРОТКІ ТЕОРЕТИЧНІ ВІДОМОСТІ

Перша задача математичної статистики — вказати метод відбору і групування статистичних даних, а також знаходження числа необхідних випробувань (статистичних даних).

Друга задача математичної статистики полягає в розробці методів аналізу статистичних даних в залежності від цілей дослідження. Одна з них — оцінка невідомих:

— імовірності випадкової події;

— функції розподілу імовірностей (густини розподілу);

— параметрів розподілу, вид якого відомий;

— залежності випадкової величини від однієї або кількох випадкових величин.

Друга мета — це перевірка статистичних гіпотез про вид невідомого розподілу або про величину параметрів розподілу, вид якого відомий.

Генеральною називається вся сукупність однотипних об'єктів, яка підлягає вивченню. Множину цих об'єктів позначатимемо через Ω.

Об'єкти множини Ω можуть характеризуватися однією або кількома ознаками. Ці ознаки можуть бути кількісними або якісними.

Іноді проводять **суцільне** обстеження, тобто обстежують **кожний** елемент множини Ω відносно ознаки, яка досліджується. Проте на практиці суцільне обстеження застосовують порівняно рідко. Це зумовлено тим, що при перевірці об'єкт частково або повністю знищується, або дослідження об'єктів вимагає великих матеріальних витрат. Деколи провести суцільне обстеження фізично неможливо. В таких випадках природно відібрати із генеральної сукупності обмежене число об'єктів, дослідити їх на кількісну чи якісну ознаку, а потім робити висновки про всю генеральну сукупність.

Вибірковою сукупністю або просто **вибіркою** називається сукупність **випадково** відібраних об'єктів із генеральної сукупності. Вибірка утворює підмножину V множини Ω ($V \subset \Omega$).

Обсягом сукупності (генеральної або вибіркової) називається число об'єктів цієї сукупності. Надалі обсяг генеральної сукупності позначатимемо літерою N, а вибіркової — n.

Для правомірності висновків про досліджувану ознаку об'єктів генеральної сукупності на підставі опрацювання вибірки необхідно, щоб об'єкти вибірки правильно представляли генеральну сукупність, тобто вибірка повинна володіти властивістю репрезентативності (представницькості). Випадковість відбору об'єктів у вибіркову сукупність і використання закону великих чисел дозволяють вирішити питання про репрезентативність вибірки.

Точність результатів вибіркового спостереження, в кінцевому підсумку, буде залежати від способу відбору об'єктів, ступеня коливання досліджуваної ознаки в генеральній сукупності та від обсягу вибірки.

На практиці використовуються різні способи утворення вибірки, які принципово розподіляються на два види:

1) відбір, що не вимагає розчленування генеральної сукупності на частини (**простий (власне випадковий) відбір**);

2) відбір, при якому генеральна сукупність розбивається на частини (**типовий** відбір, **механічний** відбір, **серійний** відбір, **комбінований** відбір).

Простим випадковим (**власне випадковим**) називається такий відбір, при якому об'єкти відбираються по одному випадковим чином із усієї генеральної сукупності. Проста випадкова вибірка може бути **повторною** або **безповторною**. **Повторною** називається вибірка, при утворенні якої відібраний об'єкт (перед відбором наступного) повертається в генеральну сукупність. **Безповторною** називається вибірка, в процесі утворення якої відібраний об'єкт в генеральну сукупність не повертається.

Нехай досліджується кількісна ознака X об'єктів генеральної сукупності Ω. Для скорочення припустимо, що вона є одновимірною випадковою величиною. Після опрацювання n об'єктів вибіркової сукупності отримуються n чисел $x_1, x_2, \ldots, x_n$, які називаються

варіантами і утворюють **ряд варіант** або **простий статистичний ряд**.

Первинна обробка ряду варіант полягає у групуванні рівних варіант цього ряду. Ряд варіант розташуємо в порядку зростання і у відповідності з цим перенумеруємо їх. В результаті одержиться послідовність чисел $x_1 \le x_2 \le x_3 \le \ldots \le x_n$, яка називається варіаційним рядом. Якщо серед цієї послідовності є однакові варіанти, то ми їх знову перенумеруємо, залишаючи один і той самий номер однаковим варіантам. Нехай у варіаційному ряді варіанта x_1 повторюється n_1 разів, x_2 — n_2 разів, ..., x_k — n_k разів. Числа n_i називаються **частотами** (**абсолютними частотами**), а їх відношення до обсягу вибірки $n_i/n = w_i$ — **відносними частотами**. Із цих означень випливають рівності:

$$\sum_{i=1}^{k} n_i = n, \quad \sum_{i=1}^{k} w_i = 1. \tag{1.1}$$

Статистичним розподілом вибірки називається відповідність між варіантами та частотами або відносними частотами:

$$\begin{array}{c|cccc} x_i & x_1 & x_2 & \ldots & x_k \\ \hline n_i & n_1 & n_2 & \ldots & n_k \end{array}, \tag{1.2}$$

$$\begin{array}{c|cccc} x_i & x_1 & x_2 & \ldots & x_k \\ \hline w_i = n_i/n & w_1 & w_2 & \ldots & w_k \end{array}. \tag{1.3}$$

В більшості випадків статистичний розподіл вибірки у вигляді (1.2) або (1.3) використовується тоді, коли ряд варіант є реалізацією **дискретної** випадкової величини *X*. Якщо ж *X* є неперервною випадковою величиною, тоді статистичний розподіл вибірки задається у вигляді відповідності між інтервалами і частотами або відносними частотами тих варіант, які потрапляють у ці інтервали, тобто у вигляді таблиць:

$$\begin{array}{c|cccc} [x_i, x_{i+1}) & [x_1, x_2) & [x_2, x_3) & \ldots & [x_k, x_{k+1}] \\ \hline n_i & n_1 & n_2 & \ldots & n_k \end{array}, \tag{1.4}$$

$$\begin{array}{c|cccc} [x_i, x_{i+1}) & [x_1, x_2) & [x_2, x_3) & \ldots & [x_k, x_{k+1}] \\ \hline w_i = n_i/n & w_1 & w_2 & \ldots & w_k \end{array}. \tag{1.5}$$

Ці таблиці називаються **інтервальним статистичним розподілом вибірки**. При побудові інтервального статистичного розподілу на основі ряду варіант розглядається *k* інтервалів однакової довжини. Кількість інтервалів можна визначати наближено за формулою Стерджеса: $k = 1+3{,}322 \lg n$. При цьому $k \in [5;6]$ при $20 \le n \le 30$; $k \in [7;8]$ при $60 \le n \le 70$; $k \in [8;9]$ при $70 \le n \le 200$; $k \in [9;15]$ при $n > 200$.

Статистичні розподіли вибірки (1.3) та (1.5) називаються **емпіричними (дослідними) розподілами випадкової величини *X*** (кількісної ознаки об'єктів генеральної сукупності).

Одна із задач математичної статистики — оцінка (наближене знаходження) невідомої функції розподілу $F(x)$ імовірностей кількісної ознаки X об'єктів генеральної сукупності. За означенням $F(x) = P(X < x)$. В розпорядженні дослідника є статистичні дані, згруповані в статистичному розподілі частот або відносно частот. Тому, врахувавши властивість стійкості відносної частоти, доцільно імовірність події $(X < x)$ наближати відносною частотою цієї ж події.

Емпіричною функцією розподілу (функцією розподілу вибірки) називається функція $F^*(x)$ детермінованого аргумента x, яка дорівнює відносній частоті появи події $(X < x)$ для даної вибірки значень випадкової величини X, тобто

$$F^*(x) = W(X < x) = n_x/n, \tag{1.6}$$

де n_x — сума частот тих варіант, які менші від x,
n — обсяг вибірки.

На противагу $F^*(x)$ функцію $F(x)$ називають теоретичною функцією розподілу.

Із означення емпіричної функції випливають такі її властивості, аналогічні властивостям теоретичної функції розподілу:

1) $D(F^*) = R$, $E(F^*) = [0, 1]$;
2) $F^*(x)$ — неспадна функція;
3) якщо x_1 — найменша варіанта, тоді $F^*(x) = 0$ для $x \le x_1$; якщо x_k — найбільша варіанта, тоді $F^*(x) = 1$ для $x > x_k$.

В процесі аналізу статистичних даних суттєву роль відіграє геометрична ілюстрація цих даних. Для наочності будують різні графіки статистичних розподілів, зокрема полігон і гістограму.

Нехай статистичний розподіл вибірки визначається таблицями (1.2) або (1.3).

Полігоном частот (частотним многокутником) називається ламана, прямолінійні відрізки якої з'єднують сусідні точки (x_1, n_1), (x_2, n_2), …, (x_k, n_k). Для побудови полігона на осі абсцис відкладають варіанти, а на осі ординат — відповідні їм частоти.

Полігоном відносних частот називається ламана, прямолінійні відрізки якої з'єднують сусідні точки (x_1, w_1), (x_2, w_2), …, (x_k, w_k). При побудові полігона відносних частот на осі абсцис відкладають варіанти x_i, а на осі ординат — відповідні їм відносні частоти w_i.

У випадку неперервної ознаки доцільно будувати гістограму. Нехай розподіли (1.4) та (1.5) такі, що довжина кожного із частинних інтервалів дорівнює одному і тому ж числу h.

Гістограмою частот називається сходинкова фігура, що складається із прямокутників, основами яких є частинні інтервали довжиною h, а висоти дорівнюють відношенню n_i/h (**густина частоти**). Для побудови гістограми частот на осі абсцис відкладаються частинні інтервали, а над ними проводяться прямолінійні відрізки, паралельні осі абсцис на віддалі n_i/h. Площа i-го частинного прямокутника дорівнює $hn_i/h = n_i$, тобто сумі частот тих варіант, що потрапляють в i-ий інтервал. Тому площа гістограми частот дорівнює сумі всіх частот вибірки n.

Порівняння двох гістограм дозволяє зробити висновок про те, що виразність гістограми суттєво залежить від обрання довжини h частинних інтервалів.

Гістограмою відносних частот називається сходинкова фігура, що складається із прямокутників, основами яких є частинні інтервали довжиною h, а висоти дорівнюють відношенню w_i/h (**густини відносної частоти**). Для побудови гістограми відносних частот на осі абсцис відкладаються частинні інтервали, а над ними проводяться відрізки, паралельні осі абсцис на віддалі w_i/h. Площа i-ого частинного прямокутника дорівнює $hw_i/h = w_i$, тобто відносній частоті тих варіант, що потрапили в i-ий інтервал. Отже, площа гістограми відносних частот дорівнює одиниці.

Побудова статистичних розподілів вибірки (1.2) або (1.4) та їх графічне зображення — це тільки перший крок на шляху розв’язування задач математичної статистики. Наступний крок передбачає знаходження числових характеристик, які у компактній формі виражають найбільш суттєві особливості статистичного розподілу вибірки і слугують оцінками (наближеними значеннями) невідомих параметрів розподілу кількісної ознаки генеральної сукупності.

Середня вибіркова (**середня арифметична варіант**) статистичного розподілу (1.2) визначається формулою

$$\overline{x}_в = \left(\sum_{i=1}^{k} x_i n_i \right) \Big/ n. \qquad (1.7)$$

Якщо всі n варіант різні, тоді (1.7) набирає такого виду:

$$\overline{x}_в = \left(\sum_{i=1}^{n} x_i\right)\Big/ n. \qquad (1.7^*)$$

Якщо вибірка задається інтервальним статистичним розподілом частот (1.4), тоді при знаходженні $\overline{x}_в$ потрібно перейти до дискретного розподілу (1.2), "нові" варіанти якого є серединами інтервалів, а потім використати формулу (1.7).

Крім вказаної середньої, у статистиці застосовують ще й **структурні середні**, які не залежать від значень варіант, що розташовані на краях розподілу, а пов'язані із рядом частот. До структурних середніх належать медіана та мода.

Медіаною Me^* дискретного статистичного розподілу вибірки (1.2) називається таке число, яке ділить варіаційний ряд, що "породжує" цей розподіл, на дві рівні за кількістю варіант частини. Якщо число варіант непарне, тобто $n = 2m + 1$, тоді $Me^* = x_{m+1}$. Якщо ж обсяг вибірки є парним числом, тобто $n = 2m$, тоді медіана дорівнює середньому арифметичному "середньої" (медіанної) пари варіант:

$$Me^* = (x_m + x_{m+1})/2.$$

Медіаною для інтервального статистичного розподілу називається таке число Me^*, для якого виконується рівність:

$$F^*(Me^*) = 0,5, \qquad (1.8)$$

де $F^*(x)$ — емпірична функція цього розподілу.

Формула для обчислення медіани має такий вид:

$$Me^* = x_m + \frac{0,5 - F^*(x_m)}{F^*(x_{m+1}) - F^*(x_m)}(x_{m+1} - x_m), \qquad (1.9)$$

де $[x_{\mathrm{m}}, x_{\mathrm{m+1}})$ – так званий медіанний частинний інтервал ($1 \le m \le k$) для якого виконуються нерівності $F^*(x_m) < 0,5$, $F^*(x_{m+1}) > 0,5$.

Модою Mo^* дискретного статистичного розподілу (1.2) називається варіанта, якій відповідає найбільша частота.

Мода для інтервального статистичного розподілу обчислюється таким чином. Спочатку визначається модальний інтервал $[x_m, x_{m+1})$, тобто такий інтервал, для якого $n_m\big/h_m = \max\limits_{1\le i\le k}\{n_i/h_i\}$, де h_i — довжина частинного інтервалу $[x_m, x_{m+1})$, n_i — число варіант з цього інтервалу. Значення Mo^* міститься всередині модального інтервалу і обчислюється за інтерполяційною формулою

$$Mo^* = x_m + \frac{n_m - n_{m-1}}{2n_m - n_{m-1} - n_{m+1}} h_m. \qquad (1.10)$$

Розглянемо деякі числові характеристики розсіювання варіант навколо середньої вибіркової.

Дисперсією вибірковою статистичного розподілу (1.2) називається середнє арифметичне квадратів відхилень варіант від середньої вибіркової:

$$D_в = \frac{1}{n}\sum_{i=1}^{k}(x_i - \overline{x}_в)^2 n_i. \tag{1.11}$$

$D_в$ характеризує середню величину розкиду варіант навколо $\overline{x}_в$ в квадратних одиницях.

На практиці зручніше користуватися так званою **розрахунковою формулою для обчислення дисперсії**:

$$D_в = \overline{x^2} - (\overline{x}_в)^2 = \frac{1}{n}\sum_{i=1}^{k} x_i^2 n_i - (\overline{x}_в)^2. \tag{1.12}$$

Недоліком $D_в$ є її розмірність. Для виправлення цього недоліку використовується інша числова характеристика:

середнє квадратичне відхилення вибіркове

$$\sigma_в = \sqrt{D_в}. \tag{1.13}$$

Коливність окремих значень варіант характеризують показники варіації. Найпростішим із них є показник **розмаху варіації** R, який дорівнює різниці між найбільшою та найменшою варіантами розподілу: $R = x_{\max} - x_{\min}$. Розмах варіації використовується при статистичному вивченні якості продукції.

Якщо $\overline{x}_в$ відмінна від нуля, тоді для порівняння двох статистичних розподілів з точки зору їх розмірності відносно середньої вибіркової вводиться показник **коефіцієнт варіації**, який дорівнює відношенню середнього квадратичного відхилення до середньої вибіркової і виражений у відсотках:

$$V = \frac{\sigma_в}{\overline{x}_в} \cdot 100\%. \tag{1.14}$$

Узагальнюючими характеристиками статистичних розподілів є статистичні моменти розподілу.

Початковим емпіричним моментом *m*-го порядку v_m^* розподілу (1.2) називається середнє значення варіант у степені m:

$$v_m^* = \frac{1}{n}\sum_{i=1}^{k}(x_i)^m n_i. \tag{1.15}$$

Тоді:

при $m = 0$ $\quad \nu_0^* = \frac{1}{n}\sum_{i=1}^{k}(x_i)^0 n_i = 1;$

при $m = 1$ $\quad \nu_1^* = \frac{1}{n}\sum_{i=1}^{k} x_i n_i = \overline{x}$ — середнє вибіркове;

при $m = 2$ $\quad \nu_2^* = \frac{1}{n}\sum_{i=1}^{k} x_i^2 n_i = \overline{x^2}$ — середнє квадратичне;

при $m = 3$ $\quad \nu_3^* = \frac{1}{n}\sum_{i=1}^{k} x_i^3 n_i = \overline{x^3};$

при $m = 4$ $\quad \nu_4^* = \frac{1}{n}\sum_{i=1}^{k} x_i^4 n_i = \overline{x^4}.$

На практиці використовуються моменти перших чотирьох порядків.

З урахуванням рівності (1.12) отримується такий вираз дисперсії вибіркової через початкові емпіричні моменти першого та другого порядків: $D_в = \nu_2^* - (\nu_1^*)^2.$

Центральним емпіричним моментом *m*-го порядку розподілу (1.2) називається середня величина відхилення варіант (від середньої вибіркової) у степені m:

$$\mu_m^* = \frac{1}{n}\sum_{i=1}^{k}(x_i - \overline{x}_в)^m n_i. \qquad (1.16)$$

За означенням $\mu_0^* = 1,\ \mu_1^* = 0,$ за означенням $D_в$ (1.11) $\mu_2^* = D_в.$

На практиці використовуються центральні емпіричні моменти третього та четвертого порядку, які з допомогою бінома Ньютона можна виразити через відповідні початкові моменти:

$$\mu_3^* = \nu_3^* - 3\nu_2^*\nu_1^* + 2(\nu_1^*)^3; \qquad (1.17)$$

$$\mu_4^* = \nu_4^* - 4\nu_3^*\nu_1^* + 6\nu_2^*(\nu_1^*)^2 - 3(\nu_1^*)^4. \qquad (1.18)$$

Центральний емпіричний момент третього порядку використовується для обчислення **коефіцієнта асиметрії**

$$A_s^* = \mu_3^* / \sigma_в^3. \qquad (1.19)$$

Для строго симетричного розподілу варіант статистичного розподілу $A_s^* = 0.$ Вважається, якщо $\left|A_s^*\right| < 0{,}25,$ то асиметрія низька, якщо $\left|A_s^*\right| \le 0{,}5$ — середня, якщо $\left|A_s^*\right| > 0{,}5$ — висока.

Якщо $A_s^* < 0,$ то у статистичному розподілі вибірки переважають варіанти, значення яких менші від $\overline{x}_в.$ Така **асиметрія** називається

від'ємною або **лівосторонньою**. Коли $A_s^* > 0,$ тоді у варіаційному ряді переважають варіанти, значення яких більші від $\overline{x}_в$ (**додатна** або **правостороння асиметрія**).

Центральний емпіричний момент четвертого порядку μ_4^* використовується при обчисленні **ексцесу**:

$$E_s^* = \frac{\mu_4^*}{\sigma_в^4} - 3, \tag{1.20}$$

який дає оцінку крутості досліджуваного статистичного розподілу в порівнянні із нормальним.

Для нормально розподіленої ознаки $\mu_4^* / \sigma_в^4 = 3,$ а тому $E_s^* = 0.$ Якщо $E_s^* > 0,$ тоді розподіл вважається гостровершинним, якщо ж $E_s^* < 0,$ тоді — плосковершинним.

Досі розглядалися числові характеристики кількісної ознаки об'єктів. Наведемо означення числової характеристики, пов'язаної із якісною ознакою об'єктів вибірки.

Вибірковою часткою називається відношення числа m об'єктів вибірки з ознакою α до обсягу вибірки: $w = m/n$. Ознакою α може бути стандартність виробу, сортність продукції, стать людини тощо. За змістом w є відносною частотою випадкової події, яка полягає в тому, що навмання відібраний об'єкт із генеральної сукупності має ознаку α.

На практиці часто досліджують вибірки значних обсягів, при цьому варіанти можуть мати велике число знаків або в цілій частині, або в дробовій. В таких випадках знаходження числових характеристик вибірки проводиться з використанням комп'ютерної техніки. Разом з тим, в більшості випадків обсяг обчислень можна суттєво зменшити, використавши спеціальні методи. Одним із них є **метод добутків**. Цей метод дає можливість спрощено обчислювати емпіричні моменти різних порядків для статистичного розподілу (1.2) із **рівновіддаленими варіантами**, тобто варіантами, що утворюють арифметичну прогресію.

Під **зведеними характеристиками** вибірки будемо розуміти середню, дисперсію та початкові емпіричні моменти.

Отже, нехай статистичний розподіл

$$\begin{array}{c|cccc} x_i & x_1 & x_2 & \ldots & x_k \\ \hline n_i & n_1 & n_2 & \ldots & n_k \end{array} \tag{1.21}$$

має рівновіддалені варіанти, тобто для будь-якого $i = \overline{1, k-1}$ виконується рівність

$$x_{i+1} - x_i = h \neq 0.$$

Метод добутків передбачає здійснення переходу від початкових варіант розподілу (1.21) до **умовних** варіант, які визначаються рівностями

$$u_i = (x_i - C)/h, \quad i = \overline{1, k}, \tag{1.22}$$

де C — хибний нуль (новий початок відліку).

В якості C береться одна із варіант розподілу (1.21), яка є модою або медіаною.

Однією із переваг умовних варіант є їх цілочисельність.

Для умовних варіант розглянемо емпіричні моменти. **Умовним емпіричним моментом порядку *m*** називається початковий момент порядку m, обчислений для умовних варіант:

$$M_m^* = \frac{1}{n}\sum_{i=1}^{k} u_i^m n_i. \tag{1.23}$$

З урахуванням рівності (1.22) для $m_1 = 1$ отримаємо:

$$M_1^* = \frac{1}{n}\sum_{i=1}^{k}\frac{x_i - C}{h} n_i = \frac{1}{h}\left[\frac{1}{n}\sum_{i=1}^{k} x_i n_i - C\frac{1}{n}\sum_{i=1}^{k} n_i\right] = (\overline{x}_в - C)/h, \tag{1.24}$$

звідки

$$\overline{x}_в = M_1^* h + C. \tag{1.25}$$

Аналогічно отримуються центральні емпіричні моменти μ_m^* через умовні емпіричні моменти:

$$D_в = \left[\, M_2^* - (M_1^*)^2 \right] h^2, \tag{1.26}$$

$$\left.\begin{aligned} \mu_3^* &= \left[\, M_3^* - 3M_2^* M_1^* + 2(M_1^*)^3 \right] h^3, \\ \mu_4^* &= \left[\, M_4^* - 4M_3^* M_1^* + 6M_2^*(M_1^*)^2 - 3(M_1^*)^4 \right] h^4. \end{aligned}\right. \tag{1.27}$$

Таким чином, **знаходження** $\overline{x}_в$, $D_в$, μ_3^* **та** μ_4^* **зводиться до обчислення умовних емпіричних моментів першого-четвертого порядків**. Останні можна знайти за допомогою методу добутків.

Припустимо, що деяка сукупність розбита на групи, що **не перетинаються**, тобто кожний об’єкт сукупності **належить тільки одній групі**. Для кожної із груп можна знайти середню (арифметичну), яка називається **груповою середньою**. Виникає питання: яким чином пов’язані середня всієї сукупності (загальна середня) із середніми груповими. Відповідь на це питання дає наступне твердження.

Теорема 1.1. Загальна середня дорівнює середній арифметичній групових середніх, зваженій за обсягами цих груп:

$$\overline{x} = \left(\overline{x^{(1)}} N_1 + \overline{x^{(2)}} N_2 + \ldots + \overline{x^{(k)}} N_k \right) \Big/ n = \sum_{i=1}^{k} \overline{x^{(i)}} N_i \Big/ n, \quad (1.28)$$

де $\overline{x}$ — загальна середня; $\overline{x^{(i)}}$ — середня групова i-ої групи $(i = \overline{1, k})$; N_i — обсяг i-ої групи; $n = \sum_{i=1}^{k} N_i$.

Аналогічне питання вивчимо і стосовно дисперсії всієї сукупності S, попередньо розглянувши додаткові означення.

Груповою дисперсією j-ої групи S_j називається дисперсія D_j розподілу значень ознаки, що складає цю сукупність, відносно групової середньої $\overline{x^{(j)}}$:

$$D_j = \left(\sum_{i=1}^{m} (x_i - \overline{x^{(j)}})^2 n_i \right) \Big/ N_j,$$

де n_i — частоти варіант $x_i \in S_j$, $N_j = \sum_{i=1}^{m} n_i$ — обсяг групи S_j.

Внутрігруповою дисперсією називається середнє арифметичне всіх групових дисперсій, зважених по обсягах груп:

$$D_{внгр} = \left(\sum_{j=1}^{k} N_j D_j \right) \Big/ n, \quad (1.29)$$

де N_j — обсяг j-ої групи S_j; $n = \sum_{j=1}^{k} N_j$ — обсяг всієї сукупності S.

Міжгруповою дисперсією називається дисперсія групових середніх відносно загальної середньої:

$$D_{міжгр} = \left(\sum_{j=1}^{k} N_j (\overline{x^{(j)}} - \overline{x})^2 \right) \Big/ n, \quad (1.30)$$

де $\overline{x^{(j)}}$ — групова середня групи S_j; N_j — обсяг групи S_j; $\overline{x}$ — загальна середня (сукупності S), $n = \sum_{j=1}^{k} S_j$ — обсяг всієї сукупності S.

Загальною дисперсією називається дисперсія значень ознаки всієї сукупності S відносно загальної середньої:

$$D_{заг} = \left(\sum_{i=1}^{m} n_i (x_i - \overline{x})^2 \right) \Big/ n, \quad (1.31)$$

де n_i — частота значення x_i, $\overline{x}$ — загальна середня сукупності S, n — обсяг сукупності S.

Зв'язок між $D_{заг}$, $D_{внгр}$ та $D_{міжгр}$ встановлює:

Теорема 1.2. Якщо сукупність складається із кількох груп, що не перетинаються, тоді загальна дисперсія дорівнює сумі внутрігрупової і міжгрупової дисперсій:

$$D_{заг} = D_{внгр} + D_{міжгр}.$$

РОЗВ'ЯЗУВАННЯ ТИПОВИХ ЗАДАЧ

Задача 1.1. При вивченні питання про норми виробітку ткаль спостерігалася така частота обривів пряжі на однотипних ткацьких станках на різних проміжках часу однакової довжини t:

7, 1, 2, 1, 2, 5, 4, 4, 3, 2, 2, 6, 0, 1, 6,
5, 3, 2, 0, 1, 4, 3, 2, 1, 5, 3, 0, 4, 2, 3.

1) Скласти статистичний розподіл частот та відносних частот числа обривів пряжі на станках;
2) побудувати полігон частот та відносних частот;
3) знайти емпіричну функцію розподілу та побудувати її графік;
4) з'ясувати питання, чи можна використати метод добутків для знаходження зведених числових характеристик; у випадку позитивної відповіді виконати розрахунки зведених характеристик в 5) цим методом;
5) обчислити вибіркові: середню, дисперсію, середнє квадратичне відхилення, моду, медіану, розмах варіацій, коефіцієнт варіацій, коефіцієнт асиметрії та ексцес;
6) здійснити розбиття вибіркової сукупності на групи, що не перетинаються, з тим, щоб використати теореми 1.1 та 1.2 для обчислення $\overline{x}_в$ та $D_в$, співставивши отримані відповіді із результатами з 5).

○ 1) Обсяг вибірки $n = 30$. Даний ряд варіант запишемо у вигляді варіаційного ряду:

0, 0, 0, 1, 1, 1, 1, 1, 2, 2, 2, 2, 2, 2, 2, 3, 3, 3, 3, 3, 4, 4, 4, 4, 5, 5, 5, 6, 6, 7.

Варіанти: $x_1 = 0$, $x_2 = 1$, $x_3 = 2$, $x_4 = 3$, $x_5 = 4$, $x_6 = 5$, $x_7 = 6$, $x_8 = 7$.

Частоти: $n_1 = 3$, $n_2 = 5$, $n_3 = 7$, $n_4 = 5$, $n_5 = 4$, $n_6 = 3$, $n_7 = 2$, $n_8 = 1$.

В підсумку отримується статистичний розподіл частот:

x_i	0	1	2	3	4	5	6	7
n_i	3	5	7	5	4	3	2	1

.

Контроль: $\sum n_i = 3 + 5 + 7 + 5 + 4 + 3 + 2 + 1 = 30 = n$.

За формулою $w_i = {n_i}/{n}$ послідовно обчислюємо відносні частоти: $w_1 = 3/30$, $w_2 = 5/30$, $w_3 = 7/30$, $w_4 = 5/30$, $w_5 = 4/30$, $w_6 = 3/30$, $w_7 = 2/30$, $w_8 = 1/30$. Контроль: $\sum w_i =$ 3/30 + 5/30 + 7/30 + + 5/30 + 4/30 + 3/30 + 2/30 + 1/30= 1. Отже, статистичний розподіл відносних частот числа обривів пряжі на станках має такий вид:

x_i	0	1	2	3	4	5	6	7
w_i	3/30	5/30	7/30	5/30	4/30	3/30	2/30	1/30

.

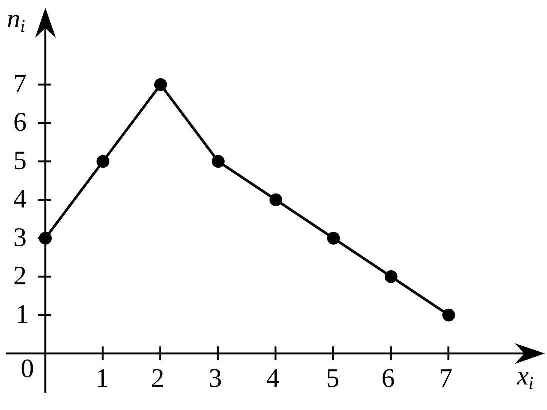
2) Полігон частот зобразимо на мал.1.1, а полігон відносних частот – на мал.1.2.

3) Обсяг вибірки $n = 30$. Якщо $x \le 0$, тоді немає жодної варіанти, меншої від x, тобто $n_x = 0$, а тому і $F^*(x) = n_x/30 = 0$.

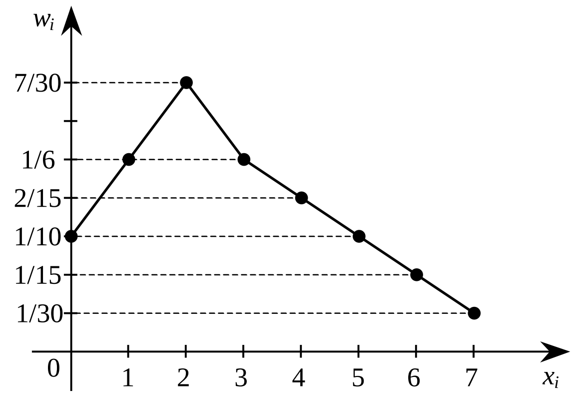
Нехай $x \in (0; 1]$. Тоді варіанта $x = 0$ є меншою від x, тому $n_x = 3$ і $F^*(x) = 3/30 = 0{,}1$.

Якщо x задовільняє подвійній нерівності $1 < x \le 2$, тоді меншими від x є варіанти 0 та 1, сума частот яких $n_x = 3 + 5 = 8$. Тому $F^*(x) = 8/30 = 4/15$ для $x \in (1; 2]$.

Якщо x таке, що виконується подвійна нерівність $2 < x \le 3$, тоді меншими від x є варіанти 0, 1, 2, суми частот яких $n_x = 3 + 5 + 7 = 15$. Тому для $x \in (2; 3]$ $F^*(x) = 15/30 = 0{,}5$.



Мал.1.1.



Мал.1.2.

Аналогічно знаходимо значення $F^*(x)$ для інтервалів (3; 4], (4; 5], (5; 6], (6; 7], (7; ∞]. В підсумку отримаємо шукану емпіричну функцію розподілу:

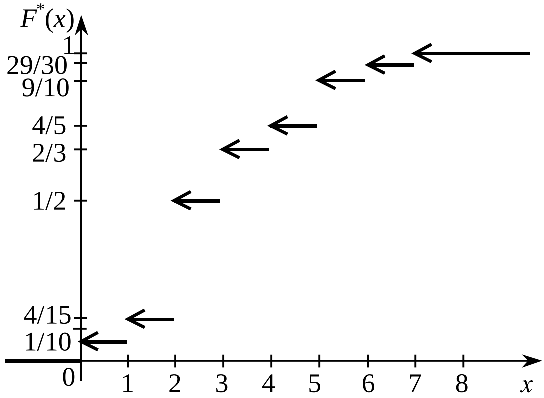
$$F^*(x)=\begin{cases} 0, & \text{якщо } x \le 0, \\ 1/10, & \text{якщо } 0 < x \le 1, \\ 4/15, & \text{якщо } 1 < x \le 2, \\ 1/2, & \text{якщо } 2 < x \le 3, \\ 2/3, & \text{якщо } 3 < x \le 4, \\ 4/5, & \text{якщо } 4 < x \le 5, \\ 9/10, & \text{якщо } 5 < x \le 6, \\ 29/30, & \text{якщо } 6 < x \le 7, \\ 1, & \text{якщо } x > 7. \end{cases}$$

Графік цієї функції наведено на мал. 1.3.

4) Метод добутків для знаходження зведених числових характеристик можна використати для статистичного розподілу із рівновіддаленими варіантами. В нашому випадку отриманий статистичний розподіл володіє цією властивістю.

5) Для знаходження зведених характеристик (середньої, дисперсії та початкових емпіричних моментів), використаємо розрахункову таблицю методу добутків, заповнюючи її в такому порядку:

1) Варіанти даного розподілу запишемо в перший стовпець.



Мал. 1.3.

2) Відповідні частоти поміщаємо у другий стовпець; суму частот (30) запишемо в нижню клітинку стовпця.

3) В якості хибного нуля C виберемо варіанту 3, яка знаходиться приблизно в середині варіаційного ряду. В третьому стовпці в рядку, що містить цю варіанту, записуємо 0; над ним записуємо послідовно –1, –2, –3, а під ним — числа 1, 2, 3, 4.

4) Добутки частот на умовні варіанти записуємо в четвертий стовпець; суму цих чисел $\left(\sum n_i u_i\right)$ поміщаємо в нижню клітинку стовпця.

5) В п'ятий стовпець запишемо добутки вмістимих клітинок третього та четвертого стовпців; їх суму $\left(\sum n_i u_i^2\right)$ поміщаємо в нижню клітинку стовпця.

6) В шостий контрольний стовпець записуємо добуток частот на квадрат умовних варіант, збільшений на одиницю. Сума їх $\left(\sum n_i (u_i+1)^2\right)$ записується в нижню клітинку. Суми другого, четвертого, п'ятого та шостого стовпців при правильних розрахунках повинні задовольняти рівність

$$\sum n_i (u_i+1)^2 = \sum n_i u_i^2 + 2\sum n_i u_i + n.$$

7) В сьомий стовпець записуються добутки вмістимих третього та п'ятого стовпців; їх сума $\left(\sum n_i u_i^3\right)$ поміщається в нижній клітинці стовпця.

8) У восьмий стовпець поміщаються добутки чисел сьомого та третього стовпців. Сума $\left(\sum n_i u_i^4\right)$ записується в нижній клітинці стовпця.

У підсумку отримаємо розрахункову табл. 1.1.

Таблиця 1.1

1	2	3	4	5	6	7	8
x_i	n_i	u_i	$n_i u_i$	$n_i u_i^2$	$n_i(u_i+1)^2$	$n_i u_i^3$	$n_i u_i^4$
0	3	–3	–9	27	12	–81	243
1	5	–2	–10	20	5	–40	80
2	7	–1	–7	7	0	–7	7
3	5	0	0	0	5	0	0
4	4	1	4	4	16	4	4
5	3	2	6	12	27	24	48
6	2	3	6	18	32	54	162
7	1	4	4	16	25	64	256
	$n =$ $= 30$		$\sum n_i u_i =$ $= -6$	$\sum n_i u_i^2 =$ $= 104$	$\sum n_i(u_i+1)^2 =$ $= 122$	$\sum n_i u_i^3 =$ $= 18$	$\sum n_i u_i^4 =$ $= 80$

Контроль:

$$\sum n_i u_i^2 + 2\sum n_i u_i + n = 104 + 2\cdot(-6) + 30 = 122,$$

$$\sum n_i(u_i+1)^2 = 122.$$

Обчислимо умовні моменти першого-четвертого порядків:

$$M_1^* = \left(\sum n_i u_i\right)/n = -6/30 = -0{,}2;$$

$$M_2^* = \left(\sum n_i u_i^2\right)/n = 104/30 \approx 3{,}4667;$$

$$M_3^* = \left(\sum n_i u_i^3\right)/n = 18/30 = 0{,}6;$$

$$M_4^* = \left(\sum n_i u_i^4\right)/n = 800/30 = 26{,}6667.$$

За умовою h = 1, за вибором хибного нуля C = 3.

Обчислимо вибіркові середню, дисперсію і середнє квадратичне відхилення, використавши рівності (1.25) та (1.26):

$$\overline{x}_в = M_1^* h + C = -0{,}2\cdot 1 + 3 = 2{,}8;$$

$$D_в = [M_2^* - (M_1^*)^2]h^2 = [3{,}4667 - (0{,}2)^2]\cdot 1^2 = 3{,}4267;$$

$$\sigma_в = \sqrt{D_в} = \sqrt{3{,}4267} = 1{,}8511.$$

Знайдемо центральні емпіричні моменти третього та четвертого порядків за формулами (1.27):

$$\mu_3^* = \left[\ M_3^* - 3M_2^* M_1^* + 2(M_1^*)^3 \right] h^3 =$$
$$= [0{,}6 - 3 \cdot 3{,}4667(-0{,}2) + 2(-0{,}2)^3] \cdot 1^3 = 2{,}66402;$$
$$\mu_4^* = \left[\ M_4^* - 4M_3^* M_1^* + 6M_2^*(M_1^*)^2 - 3(M_1^*)^4 \right] h^4 =$$
$$= [26{,}6667 - 4 \cdot 0{,}6 \cdot (-0{,}2) + 6 \cdot 3{,}4667 \cdot (-0{,}2)^2 - 3(-0{,}2)^4] \cdot 1^4 = 27{,}973908.$$

Мода $Mo^* = 2$, оскільки варіанті 2 відповідає найбільша частота 7.

Медіану Me^* можна знайти або за варіаційним рядом, отриманим в 1), або безпосередньо із статистичного розподілу. Сума частот перших трьох варіант цього розподілу дорівнює 15 (половині обсягу вибірки), а наступних п’яти — також 15. Тому медіана знаходиться між варіантами 2 та 3: $Me^* = (2+3)/2 = 2{,}5.$ Неспівпадання $\overline{x}_в$, Mo^* та Me^* свідчить про відсутність строгої симетричності розподілу.

Коефіцієнт варіації (згідно із формулою (1.14))

$$V = \frac{\sigma_в}{\overline{x}_в} \cdot 100\% = \frac{1{,}8511}{2{,}8} \cdot 100\% = 66{,}11\%.$$

Коефіцієнт асиметрії та ексцес обчислимо за формулами (1.19) та (1.20) відповідно:

$$A_s^* = \mu_3^* / \sigma_в^3 = 2{,}66402/(1{,}8511)^3 = 0{,}420;$$
$$E_s^* = \frac{\mu_4^*}{\sigma_в^4} - 3 = 27{,}973908/(1{,}8511)^4 - 3 = -0{,}6175.$$

Отже, даний статистичний розподіл числа розривів пряжі на станках характеризується середньою правосторонньою асиметрією і є плосковершинним.

Відмітимо, що для обчислення $\overline{x}_в$ та $D_в$ безпосередньо можна було використати формули (1.7) і (1.12):

$$\overline{x}_в = (\sum_{i=1}^{k} x_i n_i)/n = (0 \cdot 3 + 1 \cdot 5 + 2 \cdot 7 + 3 \cdot 5 + 4 \cdot 4 + 5 \cdot 3 + 6 \cdot 2 + 7 \cdot 1)/30 =$$
$$= 84/30 = 2{,}8;$$
$$D_в = \overline{x^2} - (\overline{x}_в)^2 = \frac{1}{n}\sum_{i=1}^{k} x_i^2 n_i - (\overline{x}_в)^2 =$$
$$= (0^2 \cdot 3 + 1^2 \cdot 5 + 2^2 \cdot 7 + 3^2 \cdot 5 + 4^2 \cdot 4 + 5^2 \cdot 3 + 6^2 \cdot 2 + 7^2 \cdot 1)/30 -$$
$$-(2{,}8)^2 = 338/30 - 7{,}84 = 3{,}4267.$$

Висновок: середнє число обривів пряжі на станках за проміжок часу складає 2,8, а середня величина розкиду чисел розривів пряжі навколо середньої 2,8 дорівнює 1,8511.

6) Розіб'ємо статистичний розподіл на дві групи, що не перетинаються :

перша група

x_i	0	1	2	3
n_i	3	5	7	5

;

друга група

x_i	4	5	6	7
n_i	4	3	2	1

.

Тоді обсяги груп: N_1 = 3+5 +7+5=20, N_2 = 4+3+2+1=10 і $n = N_1 + N_2 = 30$.

Обчислимо групові середні та загальну середню (в даному випадку середню вибіркову):

$$\overline{x^{(1)}} = \left(\sum x_i n_i \right) / N_1 = (0 \cdot 3 + 1 \cdot 5 + 2 \cdot 7 + 3 \cdot 5)/20 = 1{,}7;$$

$$\overline{x^{(2)}} = \left(\sum x_i n_i \right) / N_2 = (4 \cdot 4 + 5 \cdot 3 + 6 \cdot 2 + 7 \cdot 1)/10 = 5,$$

$$\overline{x} = \left(\sum N_i \overline{x^{(i)}} \right) \Big/ n = (20 \cdot 1{,}7 + 10 \cdot 5)/30 = 2{,}8.$$

Знайдемо групові дисперсії:

$$D_1 = \overline{x^2} - \left(\overline{x^{(1)}} \right)^2 = (0^2 \cdot 3 + 1^2 \cdot 5 + 2^2 \cdot 7 + 3^2 \cdot 5)/20 - (1{,}7)^2 = 1{,}01;$$

$$D_2 = \overline{x^2} - \left(\overline{x^{(2)}} \right)^2 = (4^2 \cdot 4 + 5^2 \cdot 3 + 6^2 \cdot 2 + 7^2 \cdot 1)/10 - 5^2 = 1.$$

Внутрігрупова дисперсія:

$$D_{внгр} = \frac{N_1 D_1 + N_2 D_2}{N_1 + N_2} = \frac{20 \cdot 1{,}01 + 10 \cdot 1}{30} = 1{,}0067;$$

міжгрупова дисперсія:

$$D_{міжгр} = \frac{N_1 \left(\overline{x^{(1)}} - \overline{x} \right)^2 + N_2 \left(\overline{x^{(2)}} - \overline{x} \right)^2}{n} = \frac{20(1{,}7 - 2{,}8)^2 + 10(5 - 2{,}8)^2}{30} = 2{,}42.$$

Згідно із теоремою 1.2 загальна дисперсія:

$$D_{заг} = D_{внгр} + D_{міжгр} = 1{,}0067 + 2{,}42 = 3{,}4267. \; \bullet$$

Розв'язування наступної задачі ілюструє:

1) перехід від дискретного розподілу до інтервального;

2) ефективність використання методу добутків при обчисленні зведених характеристик вибірки;

3) мізерність похибок при обчисленні числових характеристик вибірки, які виникають внаслідок переходу до інтервального розподілу.

Задача 1.2. За період між черговими переналадками обладнання здійснено контрольні виміри товщини (в міліметрах)

200 вкладишів шатунних підшипників. Отримані дані наведені в табл. 1.2.

Таблиця 1.2

1,754	1,739	1,743	1,764	1,733	1,736	1,743	1,742
1,728	1,732	1,731	1,752	1,737	1,747	1,758	1,737
1,724	1,737	1,733	1,713	1,740	1,740	1,729	1,740
1,751	1,739	1,747	1,748	1,730	1,750	1,740	1,732
1,740	1,730	1,748	1,757	1,741	1,733	1,743	1,745
1,748	1,723	1,737	1,748	1,741	1,751	1,714	1,750
1,744	1,748	1,758	1,756	1,727	1,731	1,738	1,753
1,735	1,738	1,743	1,729	1,743	1,737	1,731	1,734
1,741	1,742	1,744	1,756	1,744	1,752	1,739	1,740
1,729	1,745	1,742	1,753	1,743	1,734	1,731	1,734
1,732	1,732	1,746	1,748	1,755	1,738	1,742	1,729
1,731	1,725	1,729	1,745	1,739	1,754	1,752	1,720
1,750	1,734	1,749	1,738	1,747	1,757	1,751	1,746
1,723	1,736	1,746	1,744	1,759	1,728	1,751	1,750
1,746	1,759	1,748	1,740	1,735	1,745	1,740	1,746
1,737	1,726	1,743	1,755	1,740	1,726	1,745	1,744
1,735	1,746	1,739	1,732	1,758	1,744	1,754	1,724
1,742	1,750	1,761	1,758	1,753	1,757	1,720	1,733
1,738	1,728	1,758	1,732	1,763	1,733	1,745	1,766
1,745	1,743	1,734	1,733	1,755	1,756	1,769	1,750
1,740	1,762	1,738	1,742	1,740	1,740	1,760	1,752
1,746	1,728	1,743	1,718	1,738	1,762	1,728	1,734
1,753	1,751	1,748	1,735	1,739	1,729	1,754	1,736
1,762	1,748	1,738	1,726	1,757	1,738	1,726	1,720
1,751	1,734	1,724	1,741	1,752	1,732	1,738	1,739

1) Скласти інтервальний статистичний розподіл частот та відносних частот вибірки.

На основі отриманого інтервального статистичного розподілу:

2) побудувати гістограми частот та відносних частот; 3) знайти емпіричну функцію розподілу та побудувати її графік; 4) методом добутків обчислити зведені характеристики вибірки, моду, медіану, а також зробити висновки про симетричність чи асиметричність вибіркового розподілу та його гостровершинність чи плосковершинність на підставі коефіцієнта асиметрії та ексцесу; 5) порівняти числові характеристики $\overline{x}_в$, $D_в$ і

$\sigma_в$, знайдені за статистичними даними із табл.1.2 та обчисленими на підставі відповідного інтервального розподілу.

○ 1) Всі варіанти знаходяться у проміжку [1,713; 1,769] довжиною 0,056 мм. Розіб'ємо його на 8 рівних за довжиною інтервалів:
[1,713; 1,720), [1,720; 1,727), [1,727; 1,734), [1,734; 1,741),
[1,741; 1,748), [1,748; 1,755), [1,755; 1,762), [1,762; 1,769].
Кожну із варіант табл. 1.2 віднесемо до одного із цих частинних інтервалів і просумуємо число варіант, що потрапляють в перший, другий, …, восьмий інтервали. В результаті отримаємо:
$n_1 = 3$, $n_2 = 13$, $n_3 = 32$, $n_4 = 49$, $n_5 = 42$, $n_6 = 35$, $n_7 = 19$, $n_8 = 7$,
$n = \sum_{i=1}^{8} n_i = 200$.

Підсумком є інтервальний статистичний розподіл частот, наведений в табл. 1.3.

Таблиця 1.3

$[x_i; x_{i+1})$	n_i	$[x_i; x_{i+1})$	n_i
[1,713; 1,720)	3	[1,741; 1,748)	42
[1,720; 1,727)	13	[1,748; 1,755)	35
[1,727; 1,734)	32	[1,755; 1,762)	19
[1,734; 1,741)	49	[1,762; 1,769]	7

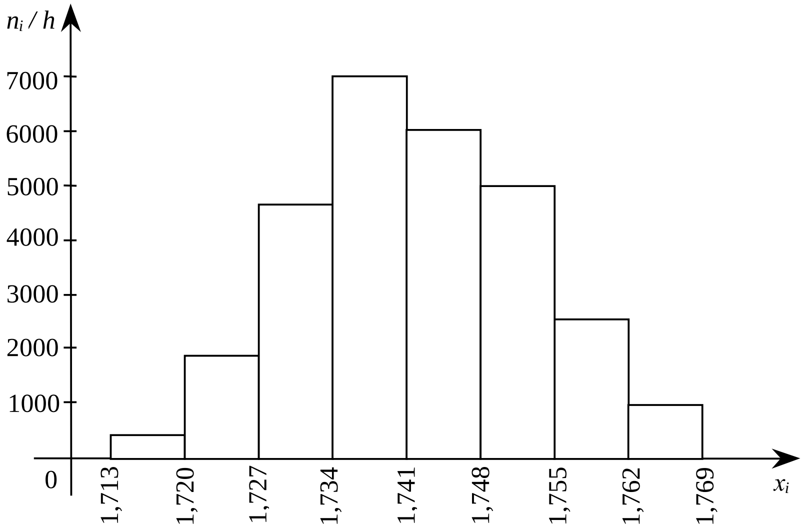
З урахуванням того, що обсяг вибірки $n = 200$, запишемо інтервальний статистичний розподіл відносних частот вибірки (табл. 1.4):

Таблиця 1.4

$[x_i; x_{i+1})$	w_i	$[x_i; x_{i+1})$	w_i
[1,713; 1,720)	3/200	[1,741; 1,748)	42/200
[1,720; 1,727)	13/200	[1,748; 1,755)	35/200
[1,727; 1,734)	32/200	[1,755; 1,762)	19/200
[1,734; 1,741)	49/200	[1,762; 1,769]	7/200

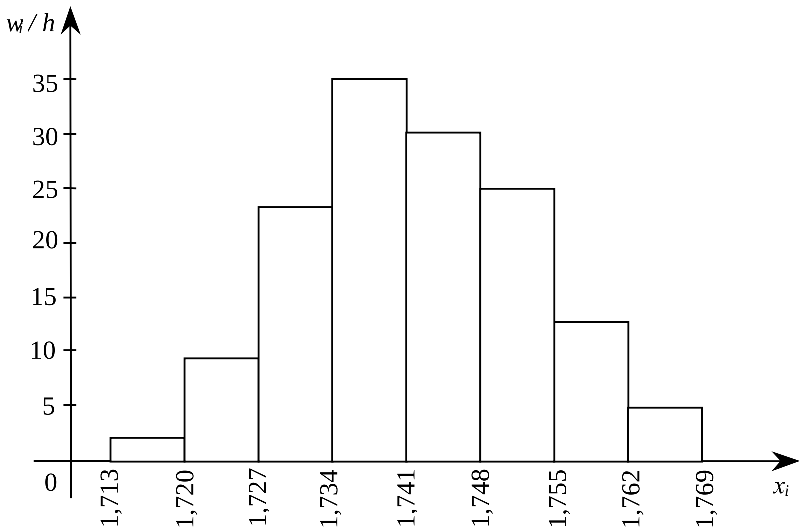
2) Гістограма частот розподілу із h=0,007 зображена на мал.1.4, а гістограма відносних частот на мал.1.5.

3) Досліджувана кількісна ознака X є непервною випадковою величиною. Тому і функція розподілу імовірностей $F(x)$, і емпірична функція $F^*(x)$ є неперервними функціями детермінованого аргумента x.



Мал.1.4.

Нехай $x \le 1{,}713$. Тоді $n_x = 0$, оскільки спостережених значень кількісної ознаки, менших від x, немає. Отже, і $F^*(x) = 0$ для всіх $x \le 1{,}713$.



Мал.1.5.

Знайдемо значення емпіричної функції розподілу для $x = 1{,}720$ — лівого кінця другого інтервалу. При цьому вважатимемо, що ми не можемо зробити цього для кожної внутрішньої точки першого інтервалу. При $x = 1{,}720$ $n_x = 3$ і $F^*(1{,}720) = 3/200 = 0{,}015$.

Для $x = 1{,}727$ $n_x = 3 + 13 = 16$, $F^*(1{,}727) = 16/200 = 0{,}08$.

Для $x = 1{,}734$ $n_x = 16 + 32 = 48$, $F^*(1{,}734) = 48/200 = 0{,}24$.

Аналогічно знаходимо:

$F^*(1{,}741) = (48 + 49)/200 = 97/200 = 0{,}485$;

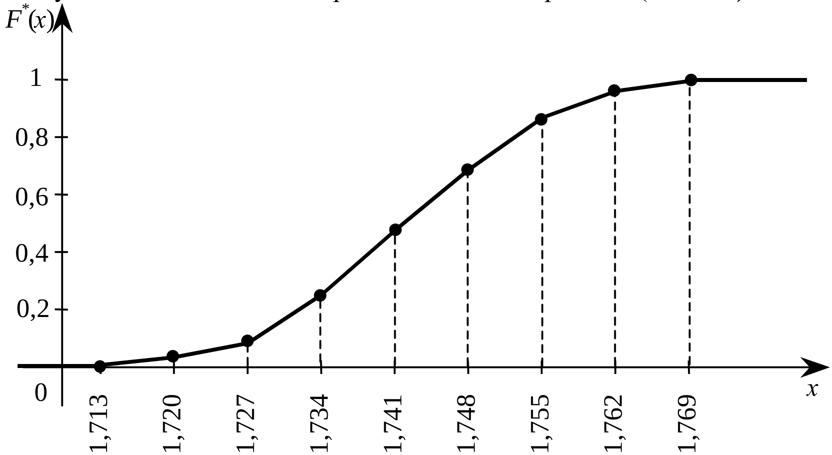
$F^*(1{,}748) = (97 + 42)/200 = 139/200 = 0{,}695$;

$F^*(1{,}755) = (139 + 35)/200 = 174/200 = 0{,}87$;

$F^*(1{,}762) = (174 + 19)/200 = 193/200 = 0{,}965$;

Нарешті, для $x > 1{,}769$ всі 200 спостережених значень кількісної ознаки менші від x, тобто $n_x = 200$. Тому $F^*(x) = 1$ для $x > 1{,}769$.

Побудуємо графік отриманої емпіричної функції розподілу: спочатку на інтервалах $(-\infty; 1{,}713]$ і $(1{,}769; \infty)$, а потім у вказаних точках. Для того, щоб показати неперервність зміни $F^*(x)$, отримані сусідні точки з'єднаємо прямолінійними відрізками (мал. 1.6).



Мал 1.6.

4) Для знаходження числових характеристик вибірки (в тому числі зведених), заданої інтервальним статистичним розподілом із п.1, перейдемо до дискретного розподілу :

x_i	1,7165	1,7235	1,7305	1,7375	1,7445	1,7515	1,7585	1,7655
n_i	3	13	32	49	42	35	19	7

При цьому "нові" варіанти є серединами частинних інтервалів.

Складемо розрахункову таблицю методу добутків (табл.1.5)

Таблиця 1.5

1	2	3	4	5	6	7	8
x_i	n_i	u_i	$n_i u_i$	$n_i u_i^2$	$n_i(u_i+1)^2$	$n_i u_i^3$	$n_i u_i^4$
1,7165	3	–3	–9	27	12	–81	243
1,7235	13	–2	–36	72	13	–144	288
1,7305	32	–1	–32	32	0	–32	32
1,7375	49	0	0	0	49	0	0
1,7445	42	1	42	42	168	42	42
1,7515	35	2	70	140	315	280	560
1,7585	19	3	57	171	304	513	1539
1,7655	7	4	28	112	175	448	1792
	$n =$ $= 200$		$\sum n_i u_i =$ $= 120$	$\sum n_i u_i^2 =$ $= 596$	$\sum n_i(u_i+1)^2 =$ $= 1036$	$\sum n_i u_i^3 =$ $= 1026$	$\sum n_i u_i^4 =$ $= 4496$

Контроль:

$$\sum n_i u_i^2 + 2\sum n_i u_i + n = 596 + 2\cdot 120 + 200 = 1036.$$

$$\sum n_i(u_i+1)^2 = 1036.$$

Обчислимо умовні моменти першого-четвертого порядків:

$$M_1^* = \left(\sum n_i u_i\right)/n = 120/200 = 0{,}6;$$

$$M_2^* = \left(\sum n_i u_i^2\right)/n = 596/200 = 2{,}98;$$

$$M_3^* = \left(\sum n_i u_i^3\right)/n = 16/200 = 5{,}13;$$

$$M_4^* = \left(\sum n_i u_i^4\right)/n = 4496/200 = 22{,}48.$$

За умовою h = 0,007, за вибором хибного нуля C = 1,7375.

Обчислимо вибіркові середню, дисперсію і середнє квадратичне відхилення, використавши рівності (1.30) та (1.31):

$$\bar{x}_в = M_1^* h + C = 0{,}6\cdot 0{,}007 + 1{,}7375 = 1{,}7417;$$

$$D_в = [M_2^* - (M_1^*)^2]h^2 = [2{,}98 - (0{,}6)^2](0{,}007)^2 = 0{,}0001283;$$

$$\sigma_в = \sqrt{D_в} = 0{,}0113269.$$

Знайдемо центральні емпіричні моменти третього та четвертого порядків за формулами (1.27):

$$\mu_3^* = \left[\, M_3^* - 3M_2^* M_1^* + 2(M_1^*)^3 \right] h^3 =$$

$$= [5{,}13 - 3\cdot 2{,}98\cdot 0{,}6 + 2(0{,}6)^3](0{,}007)^3 =$$

$$= (5{,}13 - 5{,}364 + 0{,}216)(0{,}007)^3 = -0{,}018 \cdot (0{,}007)^3.$$

$$\mu_4^* = \left[\, M_4^* - 4M_3^*M_1^* + 6M_2^*(M_1^*)^2 - 3(M_1^*)^4 \,\right] h^4 =$$

$$= [22{,}48 - 4 \cdot 5{,}13 \cdot 0{,}6 + 62{,}98 \cdot (0{,}6)^2 - 3(0{,}6)^4](0{,}007)^4 =$$

$$= (22{,}48 - 12{,}312 + 6{,}4368 - 0{,}3888)(0{,}007)^4 =$$

$$= 16{,}216 \cdot (0{,}007)^4.$$

Зауваження. Форма відповіді для μ_3^* та μ_4^* зумовлена, з одного боку, тим, що вони є "комп'ютерними нулями". По-друге, при обчисленні коефіцієнта асиметрії та ексцесу вони діляться на дуже малі числа: $\sigma_в^3$ та $\sigma_в^4$ $(\sigma_в \approx 0{,}011)$.

Оскільки всі частинні інтервали даного розподілу мають однакову довжину, то модальним є інтервал [1,734; 1,741), якому відповідає найбільша частота 49.

Значення Mo^* міститься всередині цього інтервалу і обчислюється за формулою (1.10):

$$Mo^* = x_m + \frac{n_m - n_{m-1}}{2n_m - n_{m-1} - n_{m+1}} h = 1{,}734 + \frac{49 - 32}{2 \cdot 49 - 32 - 42} \cdot 0{,}007 =$$

$$= 1{,}734 + \frac{17}{24} \cdot 0{,}007 \approx 1{,}73896.$$

Для обчислення медіани знайдемо спочатку медіанний частинний інтервал $[x_m; x_{m+1})$, для якого виконуються нерівності:

$$F^*(x_m) < 0{,}5, \qquad F^*(x_{m+1}) > 0{,}5.$$

Згідно із 3) F^*(1,741)=0,485<0,5, F^*(1,748)=0,695>0,5.

Отже, x_m = 1,741, x_{m+1} = 1,748. За формулою (1.9)

$$Me^* = x_m + \frac{0{,}5 - F^*(x_m)}{F^*(x_{m+1}) - F^*(x_m)}(x_{m+1} - x_m) =$$

$$= 1{,}741 + \frac{0{,}5 - 0{,}485}{0{,}695 - 0{,}485}(1{,}748 - 1{,}741) =$$

$$= 1{,}741 + \frac{0{,}015}{0{,}21} \cdot 0{,}007 = 1{,}741 + 0{,}0005 = 1{,}7415.$$

Коефіцієнт варіації обчислюємо за формулою (1.14):

$$V = \frac{\sigma_{в}}{\overline{x}_{в}} \cdot 100\% = \frac{0,0113269}{1,7417} \cdot 100\% = 0,65\%.$$

Коефіцієнт асиметрії та ексцес знайдемо за формулами (1.19) та (1.20) відповідно:

$$A_s^* = \mu_3^* / \sigma_в^3 = -0,018 \cdot (0,007)^3 / (0,0113269)^3 =$$
$$= -0,018(0,007/0,0113269)^3 \approx -0,0043.$$
$$E_s^* = \mu_4^* / \sigma_в^4 - 3 = 16,216 \cdot (0,007)^4 / (0,0113269)^4 - 3 \approx$$
$$\approx 2,365 - 3 = -0,6347.$$

На підставі отриманих даних можна зробити висновок, що даний інтервальний розподіл є практично симетричним із мізерною лівосторонньою асиметрією і плосковершинним.

5) Використання формул (1.7), (1.12) та (1.13) для початкового ряду варіант з табл. 1.2 дає такі результати:

$$\overline{x}_в = 1,7417, \qquad \sigma_в = 0,0113269.$$

Порівняння цих результатів із числовими характеристиками з 4) дозволяє зробити висновок про достатній рівень точності результатів, отриманих внаслідок переходу до інтервального розподілу.

ЗАДАЧІ ДЛЯ САМОСТІЙНОГО РОЗВ'ЯЗУВАННЯ

Завдання № 1

Задачі №№ 1-50. На телефонній станції досліджувалася величина X — кількість неправильних з'єднань за хвилину. Спостереження протягом 80 хв дали такі результати:

5, 1, 4, 0, 2, 4, 3, 6, 7, 3, 5, 5, 2, 5, 6, 7, 4, 3, 1, 0,
3, 1, 7, 2, 5, 0, 1, 2, 4, 6, 0, 3, 1, 7, 6, 5, 2, 4, 0, 1,
2, 0, 2, 1, 3, 4, 7, 5, 6, 3, 1, 7, 5, 4, 0, 1, 1, 2, 6, 5,
1, 2, 6, 5, 0, 2, 7, 4, 3, 0, 2, 1, 7, 2, 3, 4, 5, 7, 6, 5,
1, 6, 1, 4, 5, 0, 6, 7, 6, 3, 3, 6, 3, 1, 6, 7, 2, 0, 7, 5.

В якості вибіркової сукупності для варіанту № k, де $k = \overline{1,50}$, відібрати підряд 40 варіант, починаючи з k-ої від початку. Для отриманої вибірки:

1) скласти статистичний розподіл частот та відносних частот;
2) побудувати полігон частот та відносних частот;
3) знайти емпіричну функцію розподілу та побудувати її графік;
4) з'ясувати питання, чи можна використати метод добутків для знаходження зведених числових характеристик; у випадку

позитивної відповіді виконати розрахунки зведених характеристик в 5) цим методом;

5) обчислити вибіркові: середню, дисперсію, середнє квадратичне відхилення, моду, медіану, розмах варіацій, коефіцієнт варіацій, коефіцієнт асиметрії та ексцес;
6) здійснити розбиття вибіркової сукупності на групи, що не перетинаються, з тим, щоб використати теореми 1.1 та 1.2 з п. 1.9 для обчислення $\overline{x}_в$ та $D_в$, співставивши отримані відповіді із результатами з 5).

Завдання № 2

Для інтервальних статистичних розподілів, наведених в умовах задач №№ 1-50, а) побудувати гістограми частот та відносних частот; б) знайти емпіричну функцію розподілу та побудувати її графік; в) методом добутків обчислити зведені характеристики вибірки, моду, медіану, а також зробити висновки про симетричність чи асиметричність вибіркового розподілу та його гостровершинність чи плосковершинність на підставі коефіцієнта асиметрії та ексцесу.

Задачі №№ 1-10.

Дані про вибіркову перевірку ниток на міцність наведені у таблицях, в першому рядку яких розташовані частинні інтервали міцності ниток (кг), а в другому — кількість ниток з відповідного інтервалу.

№ 1

$[x_i;x_{i+1})$	[1,4; 1,6)	[1,6; 1,8)	[1,8; 2)	[2; 2,2)	[2,2; 2,4)	[2,4; 2,6)	[2,6; 2,8]
n_i	6	10	18	25	20	13	8

№ 2

$[x_i;x_{i+1})$	[1,6; 1,8)	[1,8; 2)	[2; 2,2)	[2,2; 2,4)	[2,4; 2,6)	[2,6; 2,8)	[2,8; 3]
n_i	2	7	13	38	22	11	7

№ 3

$[x_i;x_{i+1})$	[2; 2,2)	[2,2; 2,4)	[2,4; 2,6)	[2,6; 2,8)	[2,8; 3)	[3; 3,2)	[3,2; 3,4]
n_i	7	12	21	35	13	8	4

№ 4

$[x_i;x_{i+1})$	[1,7; 2)	[2; 2,3)	[2,3; 2,6)	[2,6; 2,9)	[2,9; 3,2)	[3,2; 3,5)	[3,5; 3,8]
n_i	5	8	17	29	21	13	7

№ 5

$[x_i; x_{i+1})$	[1,5; 1,6)	[1,6; 1,7)	[1,7; 1,8)	[1,8; 1,9)	[1,9; 2)	[2; 2,1)	[2,1; 2,2]
n_i	6	9	18	29	20	12	6

№ 6

$[x_i; x_{i+1})$	[1,8; 2)	[2; 2,2)	[2,2; 2,4)	[2,4; 2,6)	[2,6; 2,8)	[2,8; 3)	[3; 3,2]
n_i	6	11	25	27	20	8	3

№ 7

$[x_i; x_{i+1})$	[2,1; 2,2)	[2,2; 2,3)	[2,3; 2,4)	[2,4; 2,5)	[2,5; 2,6)	[2,6; 2,7)	[2,7; 2,8]
n_i	6	13	17	27	19	10	8

№ 8

$[x_i; x_{i+1})$	[1,9; 2,1)	[2,1; 2,3)	[2,3; 2,5)	[2,5; 2,7)	[2,7; 2,9)	[2,9; 3,1)	[3,1; 3,3]
n_i	5	10	18	28	20	13	6

№ 9

$[x_i; x_{i+1})$	[1,3; 1,5)	[1,5; 1,7)	[1,7; 1,9)	[1,9; 2,1)	[2,1; 2,3)	[2,3; 2,5)	[2,5; 2,7]
n_i	5	9	18	30	20	12	6

№ 10

$[x_i; x_{i+1})$	[1,7; 1,9)	[1,9; 2,1)	[2,1; 2,3)	[2,3; 2,5)	[2,5; 2,7)	[2,7; 2,9)	[2,9; 3,1]
n_i	6	10	18	25	20	13	8

Задачі №№ 11-20.

Результати вибіркового спостереження за часом обробки однієї деталі робітниками наведені в таблицях, в першому рядку яких розташовані частинні інтервали часу (у хв), а в другому — число робітників, час роботи яких потрапив у відповідний інтервал.

№ 11

$[x_i; x_{i+1})$	[4; 4,4)	[4,4; 4,8)	[4,8; 5,2)	[5,2; 5,6)	[5,6; 6)	[6; 6,4)	[6,4; 6,8]
n_i	3	8	21	31	19	14	4

№ 12

$[x_i; x_{i+1})$	[5; 5,4)	[5,4; 5,8)	[5,8; 6,2)	[6,2; 6,6)	[6,6; 7)	[7; 7,4)	[7,4; 7,8]
n_i	2	6	10	35	20	10	7

№ 13

$[x_i; x_{i+1})$	[3,8; 4)	[4; 4,2)	[4,2; 4,4)	[4,4; 4,6)	[4,6; 4,8)	[4,8; 5)	[5; 5,2]
n_i	2	8	25	34	20	8	3

№ 14

$[x_i; x_{i+1})$	[6; 6,4)	[6,4; 6,8)	[6,8; 7,2)	[7,2; 7,6)	[7,6; 8)	[8; 8,4)	[8,4; 8,8]
n_i	2	7	20	35	19	12	5

№ 15

$[x_i; x_{i+1})$	[7; 7,2)	[7,2; 7,4)	[7,4; 7,6)	[7,6; 7,8)	[7,8; 8)	[8; 8,2)	[8,2; 8,4]
n_i	4	10	18	30	20	12	6

№ 16

$[x_i; x_{i+1})$	[4; 4,4)	[4,4; 4,8)	[4,8; 5,2)	[5,2; 5,6)	[5,6; 6)	[6; 6,4)	[6,4; 6,8]
n_i	6	12	17	33	20	10	2

№ 17

$[x_i; x_{i+1})$	[5,5; 5,9)	[5,9; 6,3)	[6,3; 6,7)	[6,7; 7,1)	[7,1; 7,5)	[7,5; 7,9)	[7,9; 8,3]
n_i	6	10	17	28	20	11	8

№ 18

$[x_i; x_{i+1})$	[6,5; 6,9)	[6,9; 7,3)	[7,3; 7,7)	[7,7; 8,1)	[8,1; 8,5)	[8,5; 8,9)	[8,9; 9,3]
n_i	3	10	20	32	16	12	7

№ 19

$[x_i; x_{i+1})$	[7,5; 7,9)	[7,9; 8,3)	[8,3; 8,7)	[8,7; 9,1)	[9,1; 9,5)	[9,5; 9,9)	[9,9; 10,3]
n_i	8	12	17	29	18	10	6

№ 20

$[x_i; x_{i+1})$	[5; 5,4)	[5,4; 5,8)	[5,8; 6,2)	[6,2; 6,6)	[6,6; 7)	[7; 7,4)	[7,4; 7,8]
n_i	4	10	18	33	17	12	6

Задачі №№ 21-30.

Дослідження тривалості роботи (в тис. год.) електричних лампочок наведені в таблицях.

№ 21

$[x_i; x_{i+1})$	[2,1; 2,2)	[2,2; 2,3)	[2,3; 2,4)	[2,4; 2,5)	[2,5; 2,6)	[2,6; 2,7)	[2,7; 2,8]
n_i	2	8	22	40	12	10	6

№ 22

$[x_i; x_{i+1})$	[1,8; 2)	[2; 2,2)	[2,2; 2,4)	[2,4; 2,6)	[2,6; 2,8)	[2,8; 3)	[3; 3,2]
n_i	4	9	21	35	18	8	5

№ 23

$[x_i; x_{i+1})$	[1,6; 1,8)	[1,8; 2)	[2; 2,2)	[2,2; 2,4)	[2,4; 2,6)	[2,6; 2,8)	[2,8; 3]
n_i	2	8	15	35	20	12	8

№ 24

$[x_i; x_{i+1})$	[1,5; 1,7)	[1,7; 1,9)	[1,9; 2,1)	[2,1; 2,3)	[2,3; 2,5)	[2,5; 2,7)	[2,7; 2,9]
n_i	4	11	18	30	21	10	6

№ 25

$[x_i; x_{i+1})$	[2; 2,1)	[2,1; 2,2)	[2,2; 2,3)	[2,3; 2,4)	[2,4; 2,5)	[2,5; 2,6)	[2,6; 2,7]
n_i	5	10	17	32	20	12	4

№ 26

$[x_i; x_{i+1})$	[2,4; 2,6)	[2,6; 2,8)	[2,8; 3)	[3; 3,2)	[3,2; 3,4)	[3,4; 3,6)	[3,6; 3,8]
n_i	5	12	18	28	19	13	5

№ 27

$[x_i; x_{i+1})$	[2,3; 2,5)	[2,5; 2,7)	[2,7; 2,9)	[2,9; 3,1)	[3,1; 3,3)	[3,3; 3,5)	[3,5; 3,7]
n_i	3	9	15	33	21	13	9

№ 28

$[x_i; x_{i+1})$	[2; 2,2)	[2,2; 2,4)	[2,4; 2,6)	[2,6; 2,8)	[2,8; 3)	[3; 3,2)	[3,2; 3,4]
n_i	2	9	16	34	22	12	5

№ 29

$[x_i; x_{i+1})$	[1,4; 1,6)	[1,6; 1,8)	[1,8; 2)	[2; 2,2)	[2,2; 2,4)	[2,4; 2,6)	[2,6; 2,8]
n_i	4	12	18	33	19	11	3

№ 30

$[x_i; x_{i+1})$	[1,4; 1,6)	[1,6; 1,8)	[1,8; 2)	[2; 2,2)	[2,2; 2,4)	[2,4; 2,6)	[2,6; 2,8]
n_i	4	12	18	33	19	11	3

Задачі №№ 31-40.

Результати вибіркового вимірювання діаметрів валиків наведені у таблицях, перший рядок яких містить частинні інтервали діаметрів у мм, другий — число валиків, діаметри яких потрапили у відповідний інтервал.

№ 31

$[x_i; x_{i+1})$	[4,02; 4,04)	[4,04; 4,06)	[4,06; 4,08)	[4,08; 4,1)	[4,1; 4,12]
n_i	6	9	20	11	4

№ 32

$[x_i; x_{i+1})$	[2,06; 2,08)	[2,08; 2,1)	[2,1; 2,12)	[2,12; 2,14)	[2,14; 2,16]
n_i	5	8	22	11	4

№ 33

$[x_i; x_{i+1})$	[3,08; 3,1)	[3,1; 3,12)	[3,12; 3,14)	[3,14; 3,16)	[3,16; 3,18]
n_i	6	8	23	10	3

№ 34

$[x_i; x_{i+1})$	[3,12; 3,16)	[3,16; 3,2)	[3,2; 3,24)	[3,24; 3,28)	[3,28; 3,32]
n_i	4	7	25	9	5

№ 35

$[x_i; x_{i+1})$	[2,12; 2,16)	[2,16; 2,2)	[2,2; 2,24)	[2,24; 2,28)	[2,28; 2,32]
n_i	5	7	27	8	3

№ 36

$[x_i; x_{i+1})$	[3,28; 3,3)	[3,3; 3,32)	[3,32; 3,34)	[3,34; 3,36)	[3,36; 3,38]
n_i	2	9	28	10	1

№ 37

$[x_i; x_{i+1})$	[5,12; 5,16)	[5,16; 5,2)	[5,2; 5,24)	[5,24; 5,28)	[5,28; 5,32]
n_i	5	9	23	11	2

№ 38

$[x_i; x_{i+1})$	[4,92; 4,94)	[4,94; 4,96)	[4,96; 4,98)	[4,98; 5)	[5; 5,02]
n_i	5	9	21	11	4

№ 39

$[x_i; x_{i+1})$	[3,42; 3,46)	[3,46; 3,5)	[3,5; 3,54)	[3,54; 3,58)	[3,58; 3,62]
n_i	2	7	30	8	3

№ 40

$[x_i; x_{i+1})$	[4,56; 4,6)	[4,6; 4,64)	[4,64; 4,68)	[4,68; 4,72)	[4,72; 4,76]
n_i	5	11	24	9	1

Задачі №№ 41-50.

Результати випробувань на міцність стальних дротів однакової довжини наведені в таблицях, перший рядок яких містить частинні інтервали розривних зусиль (кг/мм^2), другий — число дротів, розривні зусилля яких потрапили у відповідний інтервал.

№ 41

$[x_i; x_{i+1})$	[30; 32)	[32; 34)	[34; 36)	[36; 38)	[38; 40)	[40; 42)	[42; 44]
n_i	6	11	18	29	19	12	5

№ 42

$[x_i; x_{i+1})$	[35; 38)	[38; 41)	[41; 44)	[44; 47)	[47; 50)	[50; 53)	[53; 56]
n_i	4	9	22	31	16	11	7

№ 43

$[x_i; x_{i+1})$	[27; 28,5)	[28,5; 30)	[30; 31,5)	[31,5; 33)	[33; 34,5)	[34,5; 36)	[36; 37,5]
n_i	3	12	24	35	18	6	2

№ 44

$[x_i; x_{i+1})$	[51; 55)	[55; 59)	[59; 63)	[63; 67)	[67; 71)	[71; 75)	[75; 79]
n_i	7	14	17	22	20	12	8

№ 45

$[x_i; x_{i+1})$	[42; 45,5)	[45,5; 49)	[49; 52,5)	[52,5; 56)	[56; 59,5)	[59,5; 63)	[63; 66,5]
n_i	2	14	22	34	18	15	5

№ 46

$[x_i; x_{i+1})$	[46; 49)	[49; 52)	[52; 55)	[55; 58)	[58; 61)	[61; 64)	[64; 67]
n_i	4	13	21	21	23	12	6

№ 47

$[x_i; x_{i+1})$	[26; 28,5)	[28,5; 31)	[31; 33,5)	[33,5; 36)	[36; 38,5)	[38,5; 41)	[41; 43,5]
n_i	3	8	17	44	18	6	4

№ 48

$[x_i; x_{i+1})$	[65; 69)	[69; 73)	[73; 77)	[77; 81)	[81; 85)	[85; 89)	[89; 93]
n_i	6	11	22	31	14	12	4

№ 49

$[x_i; x_{i+1})$	[40; 42)	[42; 44)	[44; 46)	[46; 48)	[48; 50)	[50; 52)	[52; 54]
n_i	7	13	19	24	21	10	6

№ 50

$[x_i; x_{i+1})$	[31; 33)	[33; 35)	[35; 37)	[37; 39)	[39; 41)	[41; 43)	[43; 45]
n_i	3	9	24	28	22	10	4

§ 2. СТАТИСТИЧНЕ ОЦІНЮВАННЯ

Точкові статистичні оцінки параметрів розподілу та їх властивості. Оцінка середньої генеральної для простої вибірки (повторної та безповторної). Оцінка генеральної частки для простої вибірки (повторної та безповторної). Середні квадратичні помилки (СКП) простої вибірки. Виправлена дисперсія вибіркова. Інтервальні статистичні оцінки (Довірчі інтервали для оцінок $\overline{x}_{\text{г}}$ та p для немалих вибірок. Знаходження мінімального обсягу вибірки. Довірчі інтервали для оцінки $\overline{x}_{\text{г}} = a$ для малої вибірки. Довірчі інтервали для $D_{\text{г}}$ та $\sigma_{\text{г}}$ у випадку малої вибірки).

КОРОТКІ ТЕОРЕТИЧНІ ВІДОМОСТІ

Точкові статистичні оцінки параметрів розподілу та їх властивості

Нехай досліджується неперервна кількісна ознака X об'єктів генеральної сукупності з метою знаходження невідомого закону розподілу. В розпорядженні дослідника є статистичні дані вибірки. В задачі 1.2 для статистичних даних знайдено коефіцієнт асиметрії $A_S^* = -0{,}0043$, значення якого дозволяє висунути гіпотезу про нормальний закон розподілу ознаки X (дослідження питання про правильність цієї гіпотези або хибність буде проведене в наступному параграфі). Оскільки нормальний розподіл повністю визначається двома параметрами a та σ, то виникає необхідність оцінити їх, тобто знайти наближені значення, використовуючи **тільки** спостережені варіанти $x_1, x_2, ..., x_n$ вибірки обсягом n. Параметр $a = M(X)$, математичне сподівання характеризує середнє арифметичне спостережених можливих значень випадкової величини X. З другого боку, $\overline{x}_{\text{в}}$ — це середнє арифметичне варіант. Тому (поки що інтуїтивно) доцільно наближати a середнім вибірковим:

$$a \approx \overline{x}_{\text{в}} = \left(\sum_{i=1}^{k} x_i n_i \right) \Big/ n. \qquad (2.1)$$

Аналогічно

$$\sigma \approx \sigma_{\text{в}} = \sqrt{\sum_{i=1}^{k} (x_i - \overline{x}_{в})^2 n_i \Big/ n}. \qquad (2.2)$$

Ще раз відмітимо, що праві частини рівностей (2.1), (2.2) є **випадковими величинами**, оскільки об'єкти у вибірку потрапляють **випадковим чином**.

Нехай тепер кількісна ознака X об'єктів генеральної сукупності є дискретною випадковою величиною. Припустимо, що статистичні розподіли генеральної та вибіркової сукупності описуються таблицями:

$$\begin{array}{c|cccc} x_i & x_1 & x_2 & \ldots & x_m \\ \hline N_i & N_1 & N_2 & \ldots & N_m \end{array}, \qquad N = \sum_{i=1}^{m} N_i, \tag{2.3}$$

$$\begin{array}{c|cccc} x_i & x_1 & x_2 & \ldots & x_m \\ \hline n_i & n_1 & n_2 & \ldots & n_m \end{array}, \qquad n = \sum_{i=1}^{m} n_i, \tag{2.4}$$

де N та n — обсяги генеральної та вибіркової сукупностей відповідно. Розподіл (2.3) є гіпотетичним — він завжди буде для нас невідомим, бо в протилежному випадку відпала б необхідність в дослідженні вибіркової сукупності. Нарешті, зауважимо, що деякі із частот розподілу (2.4) можуть дорівнювати нулю, що відповідає ситуації, коли значення кількісної ознаки об'єктів генеральної сукупності не зустрілися серед варіант вибірки (порівняйте розподіл (2.4) з (1.2)).

За аналогією із числовими характеристиками вибірки наступні формули визначають **числові характеристики генеральної сукупності**:

$$\bar{x}_{\text{г}} = \left(\sum_{i=1}^{m} x_i N_i \right) \Big/ N, \tag{2.5}$$

$$D_{\text{г}} = \left(\sum_{i=1}^{m} (x_i - \bar{x}_{\text{г}})^2 N_i \right) \Big/ N, \tag{2.6}$$

$$\sigma_{\text{г}} = \sqrt{D_{\text{г}}}. \tag{2.7}$$

Ці **числа** невідомі, і оцінки (наближення) їх дають такі рівності:

$$\bar{x}_{\text{г}} \approx \bar{x}_{\text{в}}, \qquad D_{\text{г}} \approx D_{\text{в}}, \qquad \sigma_{\text{г}} \approx \sigma_{\text{в}}. \tag{2.8}$$

Наведені приклади дозволяють зробити деякі висновки. Ліві частини **наближених** рівностей (2.1), (2.2) та (2.8) є невідомими параметрами "відомого" закону розподілу або числовими характеристиками генеральної сукупності; вони є невідомими числами для конкретної генеральної сукупності. Праві частини цих рівностей є функціями випадкових величин, які для **фіксованої** вибірки статистичних даних набирають числові значення, що можна зобразити точками. Це дозволяє назвати їх **точковими статистичними оцінками** відповідних параметрів або числових характеристик генеральної сукупності.

Позначимо узагальнено символом Θ ліві частини наближених рівностей (2.1), (2.2), (2.8) (а також багатьох інших, що можна отримати для інших законів розподілу), а символом Θ^* — праві частини цих рівностей.

Точковою статистичною оцінкою, **вибірковою функцією** або **статистикою** числового параметра Θ називається функція вибіркових значень (варіант) $\Theta^* = \Theta^*(x_1, x_2, ..., x_n)$, яка в певному статистичному сенсі є близькою до справжнього значення цього параметра.

Незміщеною називається точкова статистична оцінка Θ^*, математичне сподівання якої дорівнює оцінюваному параметру Θ при довільному обсязі вибірки, тобто

$$M(\Theta^*) = \Theta. \qquad (2.9)$$

Зміщеною називається оцінка, для якої не виконується рівність (2.9).

Нехай для оцінювання параметра Θ можуть бути використані незміщені точкові оцінки $\Theta_1^*, \Theta_2^*, ..., \Theta_k^*$. Оцінка Θ_m^*, $1 \le m \le k$, називається **ефективною**, якщо при заданому обсязі n вибірки для неї виконується рівність

$$D(\Theta_m^*) = \min_{1 \le i \le k} D(\Theta_i^*).$$

Оцінка Θ^* називається **спроможною** оцінкою параметра Θ, якщо при $n \to \infty$ вона збігається по імовірності до Θ, тобто для як завгодно малого $\varepsilon > 0$ має місце граничний перехід

$$\lim_{n \to \infty} P\left(\left| \Theta - \Theta^* \right| < \varepsilon \right) = 1.$$

Оцінки середньої генеральної та генеральної частки для простої вибірки

Теорема 2.1. **Для повторної вибірки обсягом** n **середня вибіркова** $\bar{x}_{\text{в}}$ **є незміщеною і спроможною оцінкою невідомої середньої генеральної** $\bar{x}_{\text{г}}$. **Якщо** n **досить велике, тоді** $\bar{x}_{\text{в}}$ **з достатнім ступенем точності розподілена за нормальним законом з параметрами:**

$$a = \bar{x}_{\text{г}}, \qquad \sigma = \sqrt{D_{\text{г}}/n}. \qquad (2.10)$$

Теорема 2.2. **Для безповторної вибірки обсягом** n **середня вибіркова** $\bar{x}_{\text{в}}$ **є незміщеною оцінкою невідомої середньої генеральної** $\bar{x}_{\text{г}}$. **Для досить великих** n $\bar{x}_{\text{в}}$ **з**

достатнім ступенем точності розподілена за нормальним законом з параметрами:

$$a = \overline{x}_{г}, \qquad \sigma = \sqrt{\frac{D_{г}}{n} \cdot \frac{N-n}{N-1}}, \qquad (2.11)$$

де N — обсяг генеральної сукупності.

Зауваження. Обсяг генеральної сукупності N, як правило, дуже великий. Тому заміна в знаменнику другої рівності (2.11) $N-1$ на N невідчутна для σ. В зв'язку із цим надалі будемо користуватися рівністю

$$\sigma = \sqrt{\frac{D_{г}}{n} \cdot \left(1 - \frac{n}{N}\right)}. \qquad (2.12)$$

Нехай досліджується якісна ознака об'єктів генеральної сукупності, число яких є скінченним. **Генеральною часткою** будемо називати відношення числа M об'єктів генеральної сукупності, що володіють ознакою α, до обсягу N генеральної сукупності:

$$p = M/N.$$

При такому дослідженні знову будемо виходити із положення про неможливість суцільної перевірки всієї генеральної сукупності. Тоді подія A – навмання відібраний об'єкт із генеральної сукупності має ознаку α є випадковою і $P(A)=p$.

Точковою статистичною оцінкою невідомого числа p є вибіркова частка w, отримана внаслідок дослідження об'єктів вибірки. Очевидно, що за означенням w є відносною частотою випадкової події A, тобто випадковою величиною для нефіксованої вибірки.

Наступні твердження з'ясовують властивості відносної частоти w як оцінки генеральної частки p, а також закон розподілу, якому вона підпорядковується.

Теорема 2.3. **Для повторної вибірки обсягом n середня вибіркова частка w є незміщеною і спроможною точковою оцінкою невідомої генеральної частки p. Якщо n є досить великим, тоді w з достатнім ступенем точності розподілена за нормальним законом з параметрами:**

$$a = p, \qquad \sigma = \sqrt{pq/n}, \qquad (2.13)$$

де $q = 1 - p$.

Теорема 2.4. **Для безповторної вибірки обсягом n вибіркова частка w є незміщеною точковою оцінкою невідомої генеральної частки p. Якщо n є досить великим, тоді w з достатнім ступенем точності розподілена за нормальним законом з параметрами:**

$$a = p, \qquad \sigma = \sqrt{\frac{pq}{n} \cdot \frac{N-n}{N-1}}, \tag{2.14}$$

Зауваження. Оскільки обсяг генеральної сукупності N в більшості випадків дуже великий, то заміна в знаменнику другої рівності (2.14) $N-1$ на N невідчутна для σ. З огляду на це надалі будемо користуватися рівністю

$$\sigma = \sqrt{\frac{pq}{n}\left(1 - \frac{n}{N}\right)}. \tag{2.15}$$

Середні квадратичні помилки (СКП) простої вибірки. Виправлена дисперсія вибіркова

Випадкові величини $\overline{x}_{\text{в}}$ та w є незміщеними точковими статистичними оцінками невідомих $\overline{x}_{\text{г}}$ та p відповідно. Їх реалізація або можливі значення, знайдені на основі даних простої вибірки (повторної або безповторної), не співпадають із оцінюваними параметрами. І кожне таке неспівпадання або відхилення природно називати **помилкою репрезентативності оцінки**, зумовленою тим, що досліджується не вся генеральна сукупність, а лище її частина (вибіркова сукупність). Для статистики дуже важливою є інформація про середню величину таких помилок.

Середньою квадратичною помилкою (**СКП**) при оцінюванні невідомих середньої генеральної $\overline{x}_{\text{г}}$ та генеральної частки p називається середнє квадратичне відхилення середньої вибіркової $\overline{x}_{\text{в}}$ та вибіркової частки w відповідно.

З урахуванням результатів теорем 2.1-2.4 можна вказати наступні формули для визначення середніх квадратичних помилок:

СКП середньої вибіркової повторної вибірки (див. (2.10))

$$\overline{\sigma}_{\overline{x}} = \sqrt{D_{\text{г}}/n}; \tag{2.16}$$

СКП середньої вибіркової безповторної вибірки (див. (2.12))

$$\bar{\sigma}'_{\bar{x}} = \sqrt{\frac{D_{\text{г}}}{n}\left(1-\frac{n}{N}\right)}; \tag{2.17}$$

СКП вибіркової частки повторної вибірки (див. (2.13))

$$\bar{\sigma}_w = \sqrt{pq/n}; \tag{2.18}$$

СКП вибіркової частки безповторної вибірки (див. (2.15))

$$\bar{\sigma}'_w = \sqrt{\frac{pq}{n}\left(1-\frac{n}{N}\right)}. \tag{2.19}$$

На практиці користуватися формулами (2.16)-(2.19) неможливо, оскільки для цього необхідно знати або дисперсію генеральну, або генеральну частку (проаналізуйте, наскільки обтяжливою є ця умова, згадавши початкову умову задачі оцінювання). Цю принципову трудність можна усунути за рахунок заміни у формулах СКП дисперсії генеральної $D_{\text{г}}$ на дисперсію вибіркову $D_{\text{в}}$, а генеральної частки p — на вибіркову частку w. Проте при такій заміні потрібно попередньо впевнитися, чи не з'явиться ще додаткова систематична помилка за рахунок зміщеності оцінок $D_{\text{в}}$ в обох задачах оцінювання. Якби вони виявилися зміщеними, то при заміні слід було б ввести "виправлені" оцінки невідомих дисперсій генеральних. На шляху реалізації вказаного вище підходу корисними є такі твердження.

Теорема 2.5. **Математичне сподівання дисперсії вибіркової в задачі про оцінювання невідомої середньої генеральної для повторної вибірки визначається рівністю**

$$M(D_{\text{в}}) = \frac{n-1}{n} D_{\text{г}}, \tag{2.20}$$

а для безповторної —

$$M(D_{\text{в}}) = \frac{n-1}{n} \cdot \frac{N}{N-1} D_{\text{г}}, \tag{2.21}$$

де n та N відповідно обсяги вибіркової та генеральної сукупностей.

Теорема 2.6. **Математичне сподівання дисперсії вибіркової в задачі про оцінювання невідомої генеральної частки для повторної вибірки визначається рівністю**

$$M(D_{\text{в}}) = \frac{n-1}{n} pq, \tag{2.22}$$

а для безповторної —

$$M(D_{\text{в}}) = \frac{n-1}{n} \cdot \frac{N}{N-1} pq. \qquad (2.23)$$

Зміст теорем 2.5 та 2.6 полягає в тому, що дисперсія вибіркова є зміщеною оцінкою дисперсії генеральної в задачах оцінювання невідомих $\overline{x}_{\text{г}}$ та p у випадку простої вибірки (повторної та безповторної). При цьому формули (2.20)-(2.23) вказують на заниження значень дисперсії генеральної за рахунок наявності в кожній із них множника $(n-1)/n$.

Проте цю зміщеність легко "виправити": достатньо дисперсію вибіркову помножити на дріб $n/(n-1)$.

Виправленою дисперсією називається числова характеристика S^2, яка визначається рівністю

$$S^2 = \frac{n}{n-1} D_{\text{в}}.$$

Для **немалих** вибірок (обсягом $n \geq 30$), замінивши у формулах (2.16)-(2.19) дисперсію генеральну $D_{\text{г}}$ вибірковою $D_{\text{в}}$, а генеральну частку p — вибірковою w, отримаємо використовувані на практиці формули для знаходження СКП:

СКП середньої вибіркової повторної вибірки

$$\overline{\sigma}_{\overline{x}} = \sqrt{D_{\text{в}}/n} = \sigma_{\text{в}}\big/\sqrt{n}; \qquad (2.24)$$

СКП середньої вибіркової безповторної вибірки

$$\overline{\sigma}'_{\overline{x}} = \sqrt{\frac{D_{\text{в}}}{n}\left(1-\frac{n}{N}\right)}; \qquad (2.25)$$

СКП вибіркової частки повторної вибірки

$$\overline{\sigma}_w = \sqrt{\frac{w(1-w)}{n}}; \qquad (2.26)$$

СКП вибіркової частки безповторної вибірки

$$\overline{\sigma}'_w = \sqrt{\frac{w(1-w)}{n}\left(1-\frac{n}{N}\right)}. \qquad (2.27)$$

Інтервальні статистичні оцінки

Точкова статистична оцінка Θ^* не співпадає (за виключанням рідкісних випадків) із справжнім значенням невідомого параметра Θ. Тому завжди виникає похибка при заміні невідомого параметра його оцінкою, тобто $\left|\Theta - \Theta^*\right| > 0$. Величина похибки при цьому невідома, хоча потрібно знати, до яких помилок може призвести вказана вище заміна.

Інтервальною називається статистична оцінка, яка визначається **двома числами** — кінцями інтервалу. Перевагою інтервальних оцінок є те, що вони дозволяють встановити точність і надійність оцінок.

Число $\delta > 0$, яке фігурує у нерівності

$$\left| \Theta - \Theta^{*} \right| < \delta, \qquad (2.28)$$

природно назвати **точністю** оцінки Θ^{*}. Проте говорити про виконання нерівності (2.28) можна тільки в імовірносному сенсі, бо Θ^{*} — випадкова величина. Тобто, для достатньо малого δ ця нерівність є випадковою подією.

Надійністю (довірчою імовірністю) оцінки Θ по Θ^{*} називається імовірність γ виконання нерівності (2.28):

$$P\left(\left| \Theta - \Theta^{*} \right| < \delta \right) = \gamma. \qquad (2.29)$$

Довірчим називається інтервал $(\Theta^{*} - \delta;\ \Theta^{*} + \delta)$, який із заданою надійністю γ покриває невідомий параметр Θ.

Довірчі інтервали для оцінок $\overline{x}_{г}$ та p для немалих вибірок

Найбільше відхилення середньої вибіркової (або вибіркової частки) від середньої генеральної (або генеральної частки), яке можливе для заданої довірчої імовірності γ, називається **граничною помилкою $\varDelta$**.

Гранична помилка знаходиться за формулою

$$\varDelta = t\overline{\sigma}, \qquad (2.30)$$

де t — корінь рівняння $2\Phi(t) = \gamma$.

Підставивши в рівність (2.30) вирази СКП (2.24)-(2.27), отримаємо **придатні для практики формули граничної помилки**:

середньої вибіркової повторної вибірки

$$\Delta = t\sqrt{D_{в}/n}; \qquad (2.31)$$

середньої вибіркової безповторної вибірки

$$\Delta = t\sqrt{\frac{D_{в}}{n}\left(1 - \frac{n}{N}\right)}; \qquad (2.32)$$

частки вибіркової повторної вибірки

$$\varDelta = t\sqrt{w(1-w)/n}; \qquad (2.33)$$

частки вибіркової безповторної вибірки

$$\varDelta = t\sqrt{\frac{w(1-w)}{n}\left(1-\frac{n}{N}\right)}. \qquad (2.34)$$

В результаті $(\bar{x}_{\text{в}} - \Delta;\ \bar{x}_{\text{в}} + \Delta)$ є довірчим інтервалом, який з надійністю γ покриває невідому середню генеральну $\bar{x}_{\text{г}}$. Аналогічно $(w-\varDelta;\ w+\varDelta)$ — довірчий інтервал, який з тією ж надійністю покриває невідому генеральну частку p.

Зауваження. В деяких випадках може бути відомою числова характеристика $\sigma_{\text{г}}$ (наприклад, як інформація технологічного процесу). Тоді $D_{\text{в}}$ у формулах (2.32), (2.33) слід замінити на $\sigma_{\text{г}}^2$.

Знаходження мінімального обсягу вибірки

Перед утворенням вибіркової сукупності необхідно з'ясувати, яким повинен бути її обсяг.

Наступні формули визначають мінімальні обсяги вибірки при оцінюванні невідомих:

а) середньої генеральної

$$n = \frac{t^2 D_{\text{г}}}{\Delta^2} \qquad (2.35)$$

для повторної вибірки,

$$n' = \frac{nN}{n+N} \qquad (2.36)$$

для безповторної вибірки (n визначається (2.35));

б) генеральної частки

$$n = \frac{t^2 pq}{\varDelta^2} \qquad (2.37)$$

для повторної вибірки,

$$n' = \frac{nN}{n+N} \qquad (2.38)$$

для безповторної вибірки (n визначається (2.37));
де N – обсяг генеральної сукупності.

Довірчі інтервали для оцінки $\bar{x}_{\text{г}} = a$ для малої вибірки

Нехай про досліджувану кількісну ознаку X відомо тільки те, що вона розподілена за нормальним законом. Ставиться задача: побудувати довірчий інтервал для оцінки невідомого параметра

$a = M(X)$ (або $\bar{x}_г$ у випадку скінченності обсягу генеральної сукупності) за даними повторної вибірки малого обсягу n, заданою довірчою імовірністю γ і якщо невідомий параметр $\sigma = \sigma_г$.

У [8] доведено, що довірчий інтервал для невідомого параметра a ($\bar{x}_г$) при невідомому $\sigma(\sigma_г)$ з надійністю γ має такий вид:

$$\bar{x}_в - t(\gamma, n-1)\frac{S}{\sqrt{n}} < a < \bar{x}_в + t(\gamma, n-1)\frac{S}{\sqrt{n}}, \tag{2.39}$$

де

$$S^2 = \sum_{i=1}^{n}(x_i - \bar{x}_в)^2 \Big/ (n-1), \tag{2.40}$$

параметр $t = t(\gamma, n\text{-}1)$ – корінь рівняння

$$P\left(\left| \frac{\bar{x}_в - a}{S/\sqrt{n}} \right| < t \right) = P\left(|T| < t\right) = 2\int_0^t g_k(t)dt = \gamma, \tag{2.41}$$

який можна знайти за табл.4 додатків в залежності від заданої довірчої імовірності γ і числа ступенів вільності $k = n - 1$; $g_k(t)$ – густина розподілу Ст'юдента.

Довірчі інтервали для $D_г$ та $\sigma_г$ у випадку малої вибірки

При знаходженні мінімального обсягу вибірки необхідною є інформація про дисперсію генеральну (середнє квадратичне відхилення генеральне). Часто в розпорядженні дослідника є тільки вибірка малого обсягу.

Можна довести (див. [8]), що довірчий інтервал для оцінки невідомої дисперсії генеральної $D_г = \sigma^2$ з надійністю γ має такий вид:

$$\frac{(n-1)S^2}{\chi_2^2} < \sigma^2 < \frac{(n-1)S^2}{\chi_1^2}, \tag{2.42}$$

де S визначається рівністю (2.40), значення χ_1^2 і χ_2^2 знаходяться за табл.6 додатків з використанням рівнянь

$$P\left(\chi^2(k) > \chi_1^2(p;k)\right) = p, \quad p = \frac{1+\gamma}{2}; \tag{2.43}$$

$$P\left(\chi^2(k) > \chi_2^2(p;k)\right) = p, \quad p = \frac{1-\gamma}{2}; \tag{2.44}$$

$k = n - 1$ – число ступенів вільності закону розподілу χ^2.

Із подвійної нерівності (2.42) отримаємо рівносильну подвійну нерівність

$$\frac{S\sqrt{n-1}}{\chi_2} < \sigma < \frac{S\sqrt{n-1}}{\chi_1}, \qquad (2.45)$$

яка визначає довірчий інтервал для оцінки невідомої $\sigma(\sigma_{г})$.

Зауваження. В табл. 6 додатків наведені значення $\chi^2(p;k)$, що задовольняють рівняння $P\left(\chi^2(k) > \chi^2(p;k)\right) = p$ тільки для $k = n - 1 \le 30$, а також для $k = 40, 50, 100$. Це зумовлено тим, що при зростанні k закон розподілу χ^2 наближається до нормального.

РОЗВ'ЯЗУВАННЯ ТИПОВИХ ЗАДАЧ

Задача 2.1. Результати випробувань на міцність стальних дротів однакової довжини наведені в табл. 2.1 інтервальним розподілом.

Таблиця 2.1

Розривне зусилля, (кг/мм^2)	[30; 32)	[32; 34)	[34; 36)	[36; 38)	[38; 40)	[40; 42)	[42; 44]
Кількість дротів	6	12	17	29	19	13	4

1) Знайти довірчий інтервал, який з надійністю $\gamma = 0,95$ покриває середнє розривне всієї партії 4000 дротів.
2) Знайти довірчу імовірність того, що середня вибіркова відхилиться від середньої генеральної за абсолютною величиною не більше, ніж на 0,5 кг/мм^2.

○ 1) Утворена (немала, бо $n = 100 > 30$) вибірка є безповторною, оскільки після перевірки об'єкт (дріт) не може бути повернутий в генеральну сукупність (дріт розривається внаслідок перевірки його міцності). Для знаходження меж довірчого інтервалу граничну помилку $\varDelta$ знайдемо за формулою (2.32), попередньо знайшовши числові характеристики вибірки $\bar{x}_{в}$ та $D_{в}$ за допомогою методу добутків, використаного до дискретного розподілу, варіанти x_i якого є серединами частинних інтервалів.

Результати допоміжних обчислень помістимо в табл. 2.2, де $h = 2$, $C = 37$, $u_i = (x_0 - C)/h$.

Таблиця 2.2

x_i	n_i	u_i	$n_i u_i$	$n_i u_i^2$	$n_i (u_i + 1)^2$
31	6	–3	–18	54	24
33	12	–2	–24	48	12
35	17	–1	–17	17	0
37	29	0	0	0	29
39	19	1	19	19	76
41	13	2	26	52	117
43	4	3	12	36	64
$\sum$	100		–2	226	322

Контроль:

$$n + 2\sum n_i u_i + \sum n_i u_i^2 = 100 - 4 + 226 = 322 = \sum n_i (u_i + 1)^2 .$$

Тоді $M_1^* = \sum n_i u_i / n = -0{,}02$, $M_2^* = \sum n_i u_i^2 / n = 2{,}26$ і за формулами (1.25), (1.26)

$$\overline{x}_{\text{в}} = M_1^* h + C = -0{,}02 \cdot 2 + 37 = 36{,}96,$$

$$D_{\text{в}} = \left[M_2^* - (M_1^*)^2 \right] h^2 = \left[2{,}26 - (-0{,}02)^2 \right] 2^2 = 9{,}0384.$$

За таблицями значень функції Лапласа (табл. 3 додатків) знайдемо значення t з рівняння $\Phi(t) = \gamma/2 = 0{,}95/2 = 0{,}475$, t = 1,96. За формулою (2.32), де N = 4000,

$$\varDelta = 1{,}96\sqrt{\frac{9{,}0384}{100}\left(1 - \frac{100}{4000}\right)} \approx 0{,}5818,$$

тоді ліва межа довірчого інтервалу

$$\overline{x}_{\text{в}} - \Delta = 36{,}96 - 0{,}5818 = 36{,}3782,$$

права –

$$\overline{x}_{\text{в}} + \Delta = 36{,}96 + 0{,}5818 = 37{,}5418.$$

Остаточно шуканий довірчий інтервал має такий вид (36,3782; 37,5418).

2) Шуканою є імовірність $P\left(\left| \overline{x}_{\text{в}} - \overline{x}_{\text{г}} \right| \le 0{,}5 \right)$, яку можна знайти за формулою

$$P\left(\left| \overline{x}_{\text{в}} - \overline{x}_{\text{г}} \right| < \varepsilon \right) = 2\Phi\left(\frac{\varepsilon}{\overline{\sigma}} \right),$$

оскільки $\overline{x}_{\text{в}}$ є нормально розподіленою випадковою величиною (згідно з теоремою 2.2), $M(\overline{x}_{\text{в}}) = \overline{x}_{\text{г}}$. За формулою (2.25)

$$\overline{\sigma} = \overline{\sigma}'_{\bar{x}} = \sqrt{\frac{D_{\text{в}}}{n}\left(1-\frac{n}{N}\right)} = \sqrt{\frac{9{,}0384}{100}\left(1-\frac{100}{4000}\right)} \approx 0{,}2968.$$

Тоді $\varepsilon/\overline{\sigma} = 0{,}5/0{,}2968 = 1{,}69$ і за табл. 3 додатків

$$P\left(\left|\bar{x}_{\text{в}} - \bar{x}_{\text{г}}\right| \le 0{,}5\right) = 2\Phi\left(0{,}5/0{,}2968\right) =$$
$$= 2\Phi\left(1{,}69\right) = 2\cdot 0{,}45449 = 0{,}90898. \quad \bullet$$

Задача 2.2. Із партії 9000 однотипних деталей перевірено 400 деталей. Серед них виявилося 360 першосортних деталей. Знайти межі, в яких з імовірністю 0,9542 міститься частка деталей першого сорту всієї партії, якщо вибірка: а) повторна; б) безповторна.

○ За табл. 3 додатків знаходимо, що коренем рівняння $2\Phi(t) = 0{,}9542$ є $t = 2$. Частка вибіркова $w = 360/400 = 0{,}9$. Граничну помилку повторної вибірки знайдемо за формулою (2.33) при $t = 2$, $w = 0{,}9$ і $n = 400$:

$$\Delta = 2\cdot\sqrt{\frac{0{,}9(1-0{,}9)}{400}} = 0{,}03.$$

Тоді для повторної вибірки довірчим є інтервал (0,9 – 0,03; 0,9 + 0,03) = (0,87; 0,93).

Для безповторної вибірки граничну помилку знайдемо за формулою (2.34) (при тих самих значеннях t, w та n):

$$\Delta = 2\sqrt{\frac{0{,}9\cdot 0{,}1}{400}\left(1-\frac{400}{9000}\right)} \approx 0{,}029.$$

В результаті отримаємо для безповторної вибірки такий довірчий інтервал (0,871; 0,929). ●

Задача 2.3. Партія виробів в кількості 10000 шт. перевірена на відповідність стандарту. Для цього відібрано 10% виробів, серед яких виявилось 94% стандартних. Знайти довірчу імовірність того, що відсоток таких виробів у всій партії відрізняється від відсотку їх у вибірці не більше, ніж на одиницю за абсолютною величиною, якщо вибірка: а) повторна; б) безповторна.

○ За умовою $w = 0{,}94$ (або 94%), $n = 1000$ (10% від 10000), $N = 10000$. СКП вибіркової частки у випадку повторної вибірки обчислимо за формулою (2.26)

$$\overline{\sigma}_w = \sqrt{\frac{w(1-w)}{n}} = \sqrt{\frac{0{,}94 \cdot 0{,}06}{1000}} \approx 0{,}0075,$$

а у випадку безповторної — за формулою (2.27)

$$\overline{\sigma}'_w = \sqrt{\frac{w(1-w)}{n}\left(1-\frac{n}{N}\right)} = \sqrt{\frac{0{,}94 \cdot 0{,}06}{1000}\left(1-\frac{1000}{10000}\right)} \approx 0{,}0071.$$

Довірчу імовірність знайдемо за формулою $P\left(\left|\Theta^* - \Theta\right| < \Delta\right) = 2\Phi\left(\frac{\Delta}{\overline{\sigma}}\right)$, де $\Theta^* = w$, $\Theta = p$. Для повторної вибірки при $\Delta = 0{,}01$ (або 1%) і $\overline{\sigma}_w = 0{,}0075$ одержимо

$$P\left(\left|0{,}94 - p\right| \le 0{,}01\right) = 2\Phi\left(\frac{0{,}01}{0{,}0075}\right) = 2\Phi(1{,}333) \approx 0{,}8164,$$

а для безповторної вибірки за тією ж формулою при $\overline{\sigma}'_w = 0{,}0071$ знайдемо:

$$P\left(\left|0{,}94 - p\right| \le 0{,}01\right) = 2\Phi\left(\frac{0{,}01}{0{,}0071}\right) = 2\Phi(1{,}408) \approx 0{,}8414. \bullet$$

Задача 2.4. Із партії однотипних деталей здійснено просту вибірку обсягом 100, при цьому з'ясовано, що 90 деталей виявилися першосортними. Знайти довірчий інтервал, в яких з імовірністю 0,9542 міститься частка деталей першого сорту всієї партії, користуючись "практичною" формулою граничної помилки (2.33).

○ За умовою задачі обсяг N генеральної сукупності невідомий, тому слід вважати, що вибірка є повторною. За умовою $n = 100$, $w = 90/100 = 0{,}9$, $\gamma = 0{,}9542$.

Знайдемо t із рівняння $\Phi(t) = \gamma/2 = 0{,}9542/2 = 0{,}4771$; за таблицями значень функції Лапласа $t = 2$.

Використання формули (2.33) дає:

$$\Delta = t\sqrt{\frac{w(1-w)}{n}} = 2\sqrt{\frac{0{,}9 \cdot 0{,}1}{100}} = 0{,}06,$$

звідки довірчий інтервал для цього випадку має такий вид: (0,9 – 0,06; 0,9 + 0,06) або остаточно (0,84; 0,96). ●

Задача 2.5. Визначити, якими повинні бути обсяги повторної і безповторної вибірок, щоб з імовірністю 0,95 частка деталей другого сорту в партії із 10000 деталей відрізнялась від частки у вибірці не більше, ніж на 0,03 за абсолютною величиною, якщо: а) про частку деталей другого сорту в усій партії деталей немає даних; б) деталей другого сорту не більше 5%.

○ Довірча імовірність 0,95 визначає значення параметра $t = 1{,}96$.

а) Якщо про частку деталей другого сорту в усій партії немає даних, тоді в якості добутку pq слід взяти його найбільше значення 0,25. Тому відповідно до формули (2.37) при $t = 1{,}96$, $\Delta = 0{,}03$ необхідний обсяг повторної вибірки

$$n = \frac{(1{,}96)^2 \cdot 0{,}25}{(0{,}03)^2} \approx 1068.$$

Якщо вибірка безповторна, тоді за формулою (2.38) при $n = 1068$, $N = 10000$ одержимо, що необхідний її обсяг

$$n' = \frac{1068 \cdot 10000}{1068 + 10000} \approx 965.$$

Зауваження. При визначенні n і n' заокруглення результатів до цілих чисел робимо з надлишком.

б) Необхідний обсяг повторної вибірки знайдемо за формулою (2.37) при попередніх значеннях t, Δ, а $p = 0{,}05$ (або 5%) і $q = 1 - p = 0{,}95$:

$$n = \frac{(1{,}96)^2 \cdot 0{,}05 \cdot 0{,}95}{(0{,}03)^3} \approx 203.$$

Якщо вибірка безповторна, то необхідний її обсяг знайдемо за формулою (2.38) при $n = 203$, $N = 10000$:

$$n' = \frac{203 \cdot 10000}{203 + 10000} \approx 199.$$

Висновок: обсяг повторної вибірки для випадку б) в 5,26 менший в порівнянні із випадком а). Це зумовлено тим, що у випадку б) число $pq = 0{,}0475$ і в 5,26 разів менше від максимального значення добутку pq. Тим самим визначається роль апріорної хоча б приблизної інформації про генеральну сукупність. ●

Задача 2.6. Знайти необхідні обсяги повторної і безповторної вибірок, щоб при визначенні середньої тривалості безперервної роботи блоків в партії із 6000 блоків з імовірністю 0,99 відхилення середньої генеральної від середньої вибіркової не перевищувало за абсолютною величиною 30 год. Середнє квадратичне відхилення генеральне вважати рівним 160 год.

○ Довірча імовірність 0,99 визначає $t = 2{,}58$ (за табл. 3 додатків з рівняння $\Phi(t) = 0{,}99/2 = 0{,}495$). За формулою (2.35) при $\Delta = 30$, $D_г = \sigma_г^2 = 160^2$ знайдемо необхідний обсяг повторної вибірки:

$$n = \frac{(2{,}58)^2 \cdot (160)^2}{(30)^2} \approx 189{,}34,$$

тобто вибірка повинна складатися з n = 190 електронних блоків.

Якщо вибірка безповторна, то обсяг її згідно формули (2.36) при n = 190, N = 6000 повинен складати

$$n' = \frac{190 \cdot 6000}{190 + 6000} \approx 185. \quad \bullet$$

Задача 2.7. Для визначення врожайності гречки зробили вибірку, до якої ввійшло вісім ділянок. Результати вибіркових спостережень за урожайністю (ц/га) наведені в таблиці:

Номер ділянки	1	2	3	4	5	6	7	8
Урожайність	18,2	15,1	16,9	17,8	19,1	15,4	20,5	16,3

Знайти довірчий інтервал, в якому з надійністю 0,95 перебуватиме середня врожайність ($\overline{x}_г$) гречки всього поля.

○ Знайдемо точкові незміщені статистичні оцінки для $\overline{x}_г$ та $D_г$:

$$\overline{x}_в = \frac{\sum x_i}{n} = \frac{18{,}2 + 15{,}1 + 16{,}9 + 17{,}8 + 19{,}1 + 15{,}4 + 20{,}5 + 16{,}3}{8} = 17{,}4125;$$

$$\overline{x^2} = \sum x_i^2 \big/ n = (331{,}24 + 228{,}01 + 285{,}61 + 316{,}84 + 364{,}81 + 237{,}16 +$$
$$+ 420{,}25 + 265{,}69)/8 = 306{,}20125;$$

$$D_в = \overline{x^2} - (\overline{x}_в)^2 = 306{,}21125 - 303{,}19515 = 3{,}0614;$$

$$S^2 = \frac{n}{n-1} D_в = \frac{8}{7} \cdot 3{,}0061 = 3{,}4355; \quad S = 1{,}8535.$$

Будемо вважати, що врожайність усього поля розподілена за нормальним законом. Тоді за даною надійністю $\gamma = 0{,}95$ та числом ступенів вільності $k = n - 1 = 8 - 1 = 7$, користуючись табл. 4 додатків, знайдемо значення $t(\gamma, n-1) = 2{,}365$. Обчислимо межі довірчого інтервалу (2.39):

$\overline{x}_{в}-t(\gamma,n-1)\,S\big/\sqrt{n-1}=17{,}4125-2{,}265\cdot 1{,}8535/7=$
$=17{,}4125-0{,}6263=16{,}7862;$

$\overline{x}_{в}+t(\gamma,n-1)\,S\big/\sqrt{n-1}=17{,}4125+2{,}265\cdot 1{,}8535/7=$
$=17{,}4125+0{,}6263=18{,}0388;$

Отже, довірчий інтервал для середньої врожайності гречки всього поля має такий вид:

$$16{,}7862<\overline{x}_{г}<18{,}0388.$$ ●

Задача 2.8. Кількісна ознака X об'єктів генеральної сукупності має нормальний закон розподілу. Із надійністю $\gamma = 0{,}98$ знайти довірчі інтервали для $D_{г}$ та $\sigma_{г}$, якщо відомо, що виправлена дисперсія вибіркова $S^2 = 4{,}6$, а обсяг вибірки $n = 26$.

○ Із рівнянь (2.43), (2.44) при $\alpha = 1-\gamma = 1-0{,}98 = 0{,}02$ отримаємо:

$$P\left(\chi^2>\chi_1^2\right)=\frac{1+\gamma}{2}=\frac{1+0{,}98}{2}=0{,}99,$$

$$P\left(\chi^2>\chi_2^2\right)=\frac{1-\gamma}{2}=\frac{1-0{,}98}{2}=0{,}01.$$

Число ступенів вільності $k = n - 1 = 26 - 1 = 25$.
За табл. 6 додатків знаходимо

$$\chi_1^2(0{,}99;k=25)=11{,}52;\qquad \chi_2^2(0{,}01;k=25)=44{,}31.$$

Тоді згідно із (2.42) при $S^2 = 4{,}8$ отримаємо:

$$\frac{(n-1)S^2}{\chi_2^2}=\frac{(26-1)\cdot 4{,}6}{44{,}31}=2{,}6;\qquad \frac{(n-1)S^2}{\chi_1^2}=\frac{25\cdot 4{,}6}{11{,}52}=9{,}98.$$

Звідки шукані довірчі інтервали:

$$2{,}6 < D_{г} < 9{,}89;\ 1{,}61 < \sigma_{г} < 3{,}15.$$ ●

ЗАДАЧІ ДЛЯ САМОСТІЙНОГО РОЗВ'ЯЗУВАННЯ

Завдання № 3

Вхідною інформацією кожної із задач цього завдання є інтервальні статистичні розподіли, наведені у відповідному номері задачі **завдання № 2**.

Задачі №№ 1-10.

1) Знайти довірчу імовірність того, що середнє розривне зусилля ниток усієї партії 40000 шт. відрізняється від вибіркової середньої не більше, ніж на 0,1 кг (за абсолютною величиною).

2) Знайти довірчі інтервали, у які з імовірностями: а) $\gamma = 0{,}95$; б) $\gamma = 0{,}99$ потрапить середнє розривне зусилля всіх ниток партії.

Задачі №№ 11-20.

1) Знайти довірчі інтервали, що з імовірностями: а) $\gamma = 0{,}95$; б) $\gamma = 0{,}98$ покривають середній час обробки однієї деталі, якщо за зміну виготовлено 1500 таких деталей.

2) Знайти довірчу імовірність того, що середній час обробки однієї деталі з партії деталей, виготовлених за зміну, відхилиться від середньої вибіркової не більше, ніж на 0,3 хв. за абсолютною величиною.

Задачі №№ 21-30.

1) Знайти довірчу імовірність того, що середня тривалість роботи електричних лампочок усієї партії чисельністю 30000 шт. відхилиться від середньої вибіркової не більше, ніж на 0,6 тис. год. за абсолютною величиною.

2) Знайти довірчі інтервали, котрі з імовірностями: а) $\gamma = 0{,}95$; б) $\gamma = 0{,}99$ покривають середню тривалість роботи однієї лампочки з усієї партії.

Задачі №№ 31-40.

1) Знайти довірчі інтервали, в які з імовірностями: а) $\gamma = 0{,}95$; б) $\gamma = 0{,}98$ потрапляє середній діаметр валиків з усієї партії чисельністю 2500 шт.

2) Знайти довірчу імовірність того, що середній діаметр валиків усієї партії відхилиться за абсолютною величиною від середньої вибіркової не більше, ніж на 0,06 мм.

Задачі №№ 41-50.

1) Знайти довірчі інтервали, в які з імовірностями: а) $\gamma = 0{,}95$; б) $\gamma = 0{,}98$ потрапляє середнє значення розривних зусиль стальних дротів із партії чисельністю 4000 шт.

2) Знайти довірчу імовірність того, що середнє значення розривних зусиль стальних дротів усієї партії відхилиться за абсолютною величиною від середньої вибіркової не більше, ніж на 0,1 кг/мм^2.

Завдання № 4

Задачі №№ 1-50.

I. Партія виробів в кількості *a* шт. перевіряється на відповідність стандарту. В ході перевірки відібрано *b*% виробів, серед яких виявилося *c*% стандартних.

1) Знайти довірчу імовірність того, що відсоток таких виробів у всій партії відрізняється від відсотка їх у вибірці не більне, ніж на *d*% за абсолютною величиною, якщо вибірка: а) повторна; б) безповторна.

2) Знайти довірчий інтервал, який з надійністю 0,95 покриває частку стандартних виробів усієї партії.

II. Планується здійснення перевірки партії нових виробів чисельністю *e* штук. Визначити, якими повинні бути обсяги повторної і безповторної вибірок, щоб з імовірністю 0,95 частка стандартних виробів у всій партії відрізнялася від частки у вибірці не більше, ніж на *f* (за абсолютною величиною).

Числові дані наведені в табл. 2.3.

Таблиця 2.3.

Номер задачі	*a*	*b*	*c*	*d*	*e*	*f*
1	2	3	4	5	6	7
1	10000	5	92	2	800	0,04
2	8000	10	95	1	1000	0,03
3	9000	6	91	3	500	0,01
4	15000	2	93	2	600	0,02
5	20000	1	89	3	700	0,03
6	12000	3	96	1	900	0,04
7	10000	4	91	2	700	0,03
8	9000	5	92	1	500	0,02
9	4000	3	90	4	400	0,01
10	5000	4	95	2	600	0,04
11	6000	2	89	3	500	0,02
12	8000	3	92	2	700	0,03
13	9000	4	94	1	300	0,04
14	7500	4	90	0,5	800	0,01
15	10000	1	89	2	1000	0,02
16	8000	2	88	3	300	0,03
17	9000	3	92	1	400	0,01
18	15000	1	91	4	500	0,02
19	12000	0,5	95	3	900	0,04

Продовження табл. 2.3

20	20000	0,4	93	2	600	0,03
21	6000	3	96	1	1000	0,01
22	5000	2	94	0,5	700	0,02
23	16000	0,5	91	4	800	0,04
24	12000	2	92	3	600	0,03
25	11000	2	95	2	300	0,01
26	9000	1	93	1	800	0,03
27	8000	5	92	0,5	400	0,02
28	7000	3	89	4	900	0,03
29	10000	2	88	2	500	0,05
30	3000	5	91	4	800	0,01
31	8000	6	96	3	600	0,02
32	11000	2	93	1	700	0,03
33	12000	1	94	2	1000	0,04
34	14000	0,5	89	1	500	0,01
35	10000	3	90	0,5	900	0,02
36	8000	4	88	4	600	0,04
37	6000	5	93	3	800	0,03
38	12000	5	95	1	700	0,01
39	14000	2	94	2	500	0,02
40	15000	1	93	2	600	0,03
41	18000	4	95	5	500	0,03
42	21000	3	97	2	800	0,04
43	13000	2	89	3	1100	0,06
44	17000	1,5	92	1,5	400	0,02
45	13000	2	94	1	800	0,01
46	6000	7	87	2	700	0,03
47	9000	2,5	97	3,5	500	0,02
48	14000	3	93	2	900	0,04
49	18000	4	96	0,5	700	0,01
50	16000	2	91	2,5	800	0,05

Завдання № 5

Задачі №№ 1-50.

При дослідженні часу (в год.) безвідмовної роботи електронних блоків в умовах постійних перепадів напруги отримано результати, наведені в табл. 2.4. В якості вибіркової сукупності для задачі № k, де $k = \overline{1,50}$, відібрати підряд 10 варіант, починаючи із k-ої від початку. Вважається, що досліджувана ознака X в усій партії виробів розподілена за нормальним законом.

I. Для отриманої вибірки (**малого** ! обсягу):

1) Знайти довірчу імовірність, з якою середня тривалість безвідмовної роботи блоків у всій партії відрізняється (за абсолютною величиною) від отриманої у вибірці, не більше, ніж на 0,4 год.

2) Знайти довірчі інтервали, які з надійністю 0,99 покривають: а) середню тривалість безвідмовної роботи блоків у всій партії $\left(M(X) = \overline{x}_{\text{г}}\right)$; б) середнє квадратичне відхилення випадкової величини X.

II. Виконати всі розрахунки частини I **без врахування малості вибірки**, співставити отримані результати, а також зробити висновки.

Таблиця 2.4.

10,75	8,26	7,1	11,28	9,82	10,11	5,90	8,56	7,55	10,12
9,36	7,18	6,58	10,39	10,52	11,20	6,12	9,42	9,49	11,5
10,24	8,32	7,93	11,42	9,12	10,38	7,05	8,49	8,52	10,36
8,26	7,93	11,45	9,52	10,26	5,92	8,9	7,65	10,27	9,36
7,48	6,17	10,48	11,41	6,36	9,66	8,49	10,05	8,29	11,12
9,56	7,85	11,32	7,53	9,62	10,14	7,17	8,36	9,81	8,67

§ 3. СТАТИСТИЧНА ПЕРЕВІРКА СТАТИСТИЧНИХ ГІПОТЕЗ

Статистичні гіпотези та їх види. Статистичний критерій перевірки основної гіпотези. Параметричні статистичні гіпотези (Порівняння середньої вибіркової із гіпотетичною середньою генеральною нормальної сукупності. Порівняння виправленої дисперсії вибіркової з гіпотетичною дисперсією генеральною нормальної сукупності). Критерій узгодженості Пірсона $\left(\chi^2\right)$. Перевірка гіпотези про нормальний розподіл генеральної сукупності. Критерій узгодженості Колмогорова.

КОРОТКІ ТЕОРЕТИЧНІ ВІДОМОСТІ

Статистичною називається гіпотеза про вид невідомого розподілу випадкової величини (кількісної ознаки об'єктів генеральної сукупності) або про параметри відомого розподілу.

Поряд із висунутою гіпотезою розглядають і гіпотезу, яка суперечить їй. Тому надалі будемо припускати, що у нас є дві гіпотези: H_0 та H_1, які не перетинаються. Гіпотезу H_0 будемо називати **основною** або **нульовою**, а гіпотезу H_1 — **конкуруючою** або **альтернативною**.

Гіпотези розрізняються за числом припущень. **Простою** називається гіпотеза, яка містить тільки одне припущення. **Складною** називається гіпотеза, яка складається із скінченного або нескінченного числа простих гіпотез.

Висунута статистична гіпотеза може бути правильною або хибною. Для перевірки її правильності використовують статистичні дані і статистичні методи, тому перевірку називають **статистичною**.
Для перевірки основної гіпотези потрібно мати критерій правильності цієї гіпотези. Як вже зазначалося, в розпорядженні дослідника є тільки вибірка $X_1, X_2, ..., X_n$, де X_i — значення кількісної ознаки i-ого об'єкта вибірки $(i=\overline{1,n})$, або значення вибіркової частки ω у випадку вивчення якісної ознаки об'єктів генеральної сукупності.

Статистичним критерієм (або просто **критерієм**, чи **статистикою**) називається випадкова великина K, яка використовується для перевірки основної гіпотези і закон розподілу якої (точний або наближений) відомий. Для кожного конкретного випадку величина K спеціально підбирається і може позначатися різними літерами: U або Z, якщо вона нормально розподілена, F або v^2 — по закону Фішера-

Снедокора, T — по закону Ст'юдента, χ^2 — по закону "хі-квадрат", K — по закону Колмогорова і т. д.

Можливі значення випадкової величини (критерію) K розбиваються на дві непорожні множини Q та $\overline{Q}$ $\left(Q \cap \overline{Q} = \varnothing \right)$ такі, що Q складається із значень критерію, при яких H_0 приймається, а $\overline{Q}$ — із тих значень критерію, при яких H_0 відхиляється (а отже, приймається H_1). Множину $\overline{Q}$ називають **критичною областю**, а множину Q — **областю прийняття гіпотези**, або **областю допустимих значень**.

Для конкретної вибірки обчислюється значення критерію як функції варіант. Отримане значення позначається $K_{сп}$ і називається **спостереженим значенням критерію**.

Сформулюємо **основний принцип статистичної перевірки статистичної гіпотези**: якщо спостережене значення критерію належить критичній області $\left(K_{сп} \in \overline{Q} \right)$, тоді гіпотезу H_0 відхиляють; якщо спостережене значення критерію належить області допустимих значень $\left(K_{сп} \in Q \right)$, тоді гіпотезу H_0 приймають.

Зауваження. Припустимо, що для конкретних гіпотез H_0 і H_1 множини Q і $\overline{Q}$ **визначені**. В ряді посібників по математичній статистиці під статистичним критерієм розуміється правило, яке реалізує основний принцип статистичної перевірки статистичної гіпотези.

Оскільки висновок про правильність гіпотези робиться за результатами скінченної вибірки, а вона може бути "невдалою", то завжди існує ризик прийняти хибне рішення. При цьому можуть бути допущені помилки двох родів.

Якщо буде відхилена гіпотеза H_0 (і прийнята H_1), в той час як насправді правильною є H_0, тоді це є **помилка першого роду**; її імовірність позначають α:

$$\alpha = P\left(K_{сп} \in \overline{Q} / H_0 \right) = P\left(H_1 / H_0 \right),$$

де $P\left(H_1 / H_0 \right)$ — імовірність того, що буде прийнята гіпотеза H_1, якщо насправді для генеральної сукупності правильною є гіпотеза H_0. Число α називають **рівнем значущості**.

Якщо буде прийнята гіпотеза H_0, в той час як насправді правильною є H_1, тоді буде допущена **помилка другого роду**, її імовірність позначають β:

$$\beta = P\left(K_{сп} \in Q / H_1 \right) = P\left(H_0 / H_1 \right).$$

В цьому параграфі розглянемо один із методів перевірки статистичних гіпотез, який передбачає виконання таких кроків:

1) формулювання основної і конкуруючої гіпотез;

2) вибір відповідного рівня значущості α;

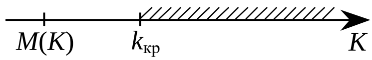
3) вибір статистичного критерія *К* перевірки гіпотези;

4) знаходження критичної області $\overline{Q}$ і області Q прийняття гіпотези;

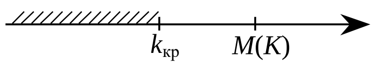
5) формулювання правила перевірки гіпотези: гіпотеза H_0 приймається при заданому рівні значущості α, якщо $K_{cn} \in Q$; гіпотеза H_0 відкидається, якщо $K_{cn} \in \overline{Q}$.

Розглянемо деякі особливості структури критичної області $\overline{Q}$. Якщо $\overline{Q}$ розташована зліва **і** справа від математичного сподівання випадкової величини *К*, то критична область називається **двосторонньою**, а критерій *К* — **двостороннім критерієм значущості**. Якщо ж $\overline{Q}$ розташована зліва **або** справа від математичного сподівання випадкової величини *К*, то критична область називається **односторонньою**, а критерій — **одностороннім**.

Реалізація основного принципу перевірки конкретної основної гіпотези H_0 передбачає знаходження множин Q та $\overline{Q}$. Оскільки ці множини числової осі не перетинаються, то існують точки, які їх розділяють. Такі точки називають **критичними точками** і позначаються $k_{\text{кр}}$. **Правосторонньою** називається критична область, яка визначається нерівністю $K > k_{\text{кр}}$, де $k_{\text{кр}} > M(K)$:



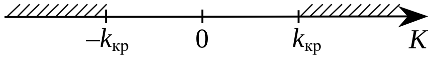
Лівосторонньою називається критична область, яка визначається нерівністю $K < k_{\text{кр}}$, де $k_{\text{кр}} < M(K)$:



Двосторонньою називається критична область, яка визначається сукупністю нерівностей

$$\left[\begin{array}{l} K < k_{\text{кр}}^{(1)}, \\ K > k_{\text{кр}}^{(2)}, \end{array}\right.$$

де $k_{кр}^{(1)} < M(K) < k_{кр}^{(2)}$. Зокрема, якщо критичні точки $k_{кр}^{(1)}$ та $k_{кр}^{(2)}$ відрізняються тільки знаком, тоді двостороння критична область визначається нерівністю $|K| > k_{кр}$:



Інтуїтивно зрозуміло, що множини Q та $\overline{Q}$ залежать від вибраного рівня значущості α, а тому слід очікувати залежності критичної точки $k_{кр}$ від α. Переконаємося в цьому, розглянувши підхід до використання основного принципу статистичної перевірки статистичної гіпотези у випадку правосторонньої критичної області. Аналіз решти випадків проводиться аналогічно. Для вибраного рівня значущості α $\overline{Q} = \{K > k_{кр}\}$. Критична точка $k_{кр}$ при умові правильності основної гіпотези H_0 задовольняє рівність

$$P(\ K > k_{кр}) = \alpha, \qquad (3.1)$$

яка трактується таким чином: при правильності H_0 малоімовірно (з врахуванням малості α), що спостережене значення критерію K виявиться більшим від $k_{кр}$.

Для кожного практично важливого критерію складено таблиці, за допомогою яких знаходиться $k_{кр}$, що задовольняє рівність (3.1). Наступний крок — обчислення спостереженого значення $K_{сп}$ критерію за даною вибіркою. А далі в силу вступає **правило**: якщо $K_{сп} > k_{кр}$, то гіпотезу H_0 відхиляють (вважається, що вона є хибною), якщо ж $K_{сп} < k_{кр}$, тоді H_0 приймають (вважається, що H_0 — правильна).

Якщо основна гіпотеза H_0 містить твердження про параметри розподілу, вид якого відомий, тоді статистичний критерій перевірки цієї гіпотези називається **параметричним**. Якщо ж у гіпотезі H_0 мова ведеться про невідомий розподіл, тоді відповідний критерій називається **критерієм узгодженості**.

Параметричні статистичні гіпотези

Розгляд статистичних гіпотез розпочнемо із параметричних гіпотез. Це зумовлено відносною простотою досліджень таких гіпотез в порівнянні із критеріями узгодженості.

Нижче будуть наведені тільки підсумки аналізу, який в повному обсязі можна знайти, зокрема, у посібнику [8].

Порівняння середньої вибіркової із гіпотетичною середньою генеральною нормальної сукупності

Нехай кількісна ознака X генеральної сукупності розподілена нормально з параметрами a та σ. При цьому середня генеральна a хоча і невідома, але є підстави припускати, що вона дорівнює гіпотетичному значенню a_0. Для перевірки гіпотези $a = a_0$ потрібно знайти середню вибіркову $\overline{x}_в$ і з'ясувати, істотно чи ні $\overline{x}_в$ відхиляється від a_0.

Розв'язування окресленої задачі суттєво залежить від того, чи відомий параметр σ.

А. Дисперсія генеральної сукупності відома

Нехай із генеральної сукупності, кількісна ознака якої нормально розподілена, організована вибірка обсягом n і знайдена середня вибіркова $\overline{x}_в$, при цьому $D_г = \sigma^2$ — відома. Потрібно при заданому рівні значущості α перевірити основну гіпотезу про рівність середньої генеральної a гіпотетичному значенню a_0, тобто $H_0 : a = a_0$.

Оскільки середня вибіркова $\overline{X}$ (вибірка поки не фіксована) є незміщеною оцінкою середньої генеральної, тобто $M(\overline{X}) = a$, то гіпотезу H_0 можна записати ще й таким чином: $M(\overline{X}) = a_0$.

Отже, потрібно перевірити, що математичне сподівання середньої вибіркової дорівнює гіпотетичній середній генеральній.

В якості критерія перевірки гіпотези H_0 використаємо випадкову величину:

$$U = (\overline{X} - a_0)\big/\sigma(\overline{X}) = (\overline{X} - a_0)\sqrt{n}\big/\sigma.$$

Ця величина розподілена нормально, причому якщо гіпотеза H_0 правильна, то $M(U) = 0,\ \ \sigma(U) = 1.$

Позначимо через $U_{спост}$ спостережене значення критерія для фіксованої вибірки:

$$U_{спост} = (\overline{x}_в - a_0)\sqrt{n}\big/\sigma. \qquad (3.2)$$

Критичну область побудуємо в залежності від виду конкуруючої гіпотези.

Правило 1. Гіпотеза $H_0 : a = a_0$ при конкуруючій гіпотезі $H_1 : a \neq a_0$ для рівня значущості α приймається, якщо $|U_{спост}| < u_{кр}$, де $u_{кр}$ — корінь рівняння

$$\Phi(u_{кр}) = (1-\alpha)\big/2\ . \qquad (3.3)$$

Якщо $|U_{\text{спост}}| > u_{\text{кр}}$, то гіпотеза H_0 відкидається.

Правило 2. При конкуруючій гіпотезі $H_1 : a > a_0$ критичну точку правосторонньої критичної області для рівня значущості α знаходять за рівністю (табл. 3 додатків):

$$\Phi(u_{\text{кр}}) = (1 - 2\alpha)/2. \tag{3.4}$$

Якщо $U_{\text{спост}} < u_{\text{кр}}$, то основна гіпотеза H_0 приймається. Якщо ж $U_{\text{спост}} > u_{\text{кр}}$ — H_0 відкидається.

Правило 3. При конкуруючій гіпотезі $H_1 : a < a_0$ знаходять корінь $u_{\text{кр}}$ рівняння (3.4) (табл. 3 додатків).

Якщо $U_{\text{спост}} > -u_{\text{кр}}$, то нема підстав відкидати гіпотезу H_0. Якщо ж $U_{\text{спост}} < -u_{\text{кр}}$, то H_0 відкидається.

Б. Дисперсія генеральної сукупності невідома

В цьому випадку в якості критерія перевірки основної гіпотези береться випадкова величина

$$T = (\overline{X} - a_0)\sqrt{n}/S, \tag{3.5}$$

де $\overline{X}$ та S — середня та "виправлене" середнє квадратичне відхилення для нефіксованої вибірки обсягу n.

Величина T розподілена за законом Ст'юдента із $k = n - 1$ ступенями вільності.

Симетричність розподілу Ст'юдента дозволяє здійснювати побудову критичних областей (в залежності від виду конкуруючої гіпотези) аналогічно тому, як це робилося у випадку відомого параметра σ.

Правило 1*. Для того, щоб при заданому рівні значущості α перевірити основну гіпотезу $H_0 : a = a_0$ про рівність невідомої середньої генеральної a (нормально розподіленої генеральної сукупності з невідомою дисперсією) гіпотетичному значенню a_0 при конкуруючій гіпотезі $H_1 : a \neq a_0$, потрібно обчислити спостережене значення критерія

$$T_{\text{спост}} = (\overline{x}_в - a_0)\sqrt{n}/S$$

і за таблицею критичних точок розподілу Ст'юдента (табл. 5 додатків) по заданому рівню значущості α, розміщеному у верхньому рядку таблиці і числу ступенів вільності $k = n - 1$ знайти критичну точку $t_{\text{двост.кр}}(\alpha, k)$.

Якщо $|T_{\text{спост}}| < t_{\text{двост.кр}}$, то нема підстав відкидати основну гіпотезу H_0.

Якщо ж $|T_{\text{спост}}| > t_{\text{двост.кр}}$, то гіпотеза H_0 відкидається на користь альтернативної гіпотези H_1.

Зауваження. Критична точка $t_{\text{двост.кр}} = t_\alpha$ є коренем рівняння

$$\int_0^{t_\alpha} g_k(t)dt = (1 - \alpha)/2,$$

де $g_k(t)$ — густина розподілу Ст'юдента, $k = n - 1$ — число ступенів вільності. Це рівняння є аналогом рівняння (3.3).

Правило 2*. При конкуруючій гіпотезі $H_1 : a > a_0$ по рівню значущості α, розміщеному в нижньому рядку табл. 5 додатків, і числу ступенів вільності $k = n - 1$ знаходимо критичну точку $t_{\text{правост.кр}}(\alpha, k)$ правосторонньої критичної області.

Якщо $T_{\text{спост}} < t_{\text{правост.кр}}$, то нема підстав відкидати основну гіпотезу H_0.

***Зауваження*.** Вибір критичної точки в табл. 5 додатків по нижньому рядку значень рівня значущості зумовлений аналогом рівності (3.4) для випадку правосторонньої критичної області, тобто $t_{\text{правост.кр}} = t_{2\alpha}$ є коренем рівняння

$$\int_0^{t_{2\alpha}} g_k(t) = (1 - 2\alpha)/2.$$

Правило 3*. При конкуруючій гіпотезі $H_1 : a < a_0$ спочатку знаходять "допоміжну" критичну точку $t_{\text{правост.кр}}(\alpha; k)$, а потім покладають межу лівосторонньої критичної області:

$$t_{\text{правост.кр}} = t_{\text{лівост.кр}}.$$

Якщо $T_{\text{спост}} > -t_{\text{правост.кр}}$, то гіпотеза H_0 приймається.

Якщо $T_{\text{спост}} < -t_{\text{правост.кр}}$, то гіпотезу H_0 відкидають.

Порівняння виправленої дисперсії вибіркової з гіпотетичною дисперсією генеральною нормальної сукупності

Нехай кількісна ознака генеральної сукупності розподілена за нормальним законом, причому дисперсія генеральна хоча і невідома, проте є підстави припускати, що вона дорівнює гіпотетичному значенню σ_0^2.

На підставі виправленої дисперсії вибіркової S^2, знайденої для вибірки обсягом n, при заданому рівні значущості α потрібно перевірити основну гіпотезу H_0, яка полягає в тому, що дисперсія генеральна дорівнює гіпотетичному значенню σ_0^2. Враховуючи незміщеність S^2 як оцінки дисперсії генеральної, основну гіпотезу можна записати таким чином:

$$H_0 : M(S^2) = \sigma_0^2. \qquad (3.6)$$

Іншими словами, потрібно з’ясувати, істотно чи неістотно відрізняються виправлена вибіркова і гіпотетична генеральна дисперсії.

В якості критерія перевірки гіпотези (3.6) візьмемо випадкову величину

$$\chi^2 = (n-1)\, S^2 \big/ \sigma_0^2, \qquad (3.7)$$

де права частина є випадковою величиною, розподіленою за законом χ^2 з $k = n-1$ ступенями вільності.

Критична область будується в залежності від виду конкуруючої гіпотези.

Правило 1. Для того, щоб при заданому рівні значущості α перевірити основну гіпотезу $H_0 : \sigma^2 = \sigma_0^2$ при конкуруючій гіпотезі $H_1 : \sigma^2 > \sigma_0^2$, потрібно обчислити спостережене значення критерія $\chi^2_{\text{спост}}$ за формулою (3.7) і за табл. 6 додатків по заданому рівню значущості α і числу ступенів вільності $k = n-1$ знайти критичну точку $\chi^2_{\text{кр}}(\alpha; k)$.

Якщо $\chi^2_{\text{спост}} < \chi^2_{\text{кр}}$, то нема підстав відкидати основну гіпотезу. Якщо $\chi^2_{\text{спост}} > \chi^2_{\text{кр}}$, то гіпотеза H_0 відкидається на користь гіпотези H_1.

Правило 2. Для того, щоб при заданому рівні значущості α перевірити гіпотезу $H_0 : \sigma^2 = \sigma_0^2$ при конкуруючій гіпотезі $H_1 : \sigma^2 \neq \sigma_0^2$, потрібно обчислити спостережене значення критерія $\chi^2_{\text{спост}}$ за формулою (3.7) і за табл. 6 додатків знайти ліву критичну точку $\chi^2_{\text{кр}}(1-\alpha/2;k)$ і праву критичну точку $\chi^2_{\text{кр}}(\alpha/2;k)$.

Якщо $\chi^2_{\text{лівост.кр}} < \chi^2_{\text{спост}} < \chi^2_{\text{правост.кр}}$, то нема підстав відкидати гіпотезу H_0.

Якщо $\chi^2_{\text{спост}} < \chi^2_{\text{лівост.кр}}$ або $\chi^2_{\text{спост}} > \chi^2_{\text{правост.кр}}$, то основна гіпотеза H_0 відкидається.

Правило 3. При конкуруючій гіпотезі $H_1 : \sigma^2 < \sigma_0^2$ знаходиться критична точка $\chi^2_{\text{кр}}(1-\alpha/2;k)$ лівосторонньої критичної області.

Якщо $\chi^2_{\text{спост}} > \chi^2_{\text{кр}}(1-\alpha/2;k)$, то нема підстав відкидати основну гіпотезу.

Якщо $\chi^2_{\text{спост}} < \chi^2_{\text{кр}}(1-\alpha/2;k)$, то гіпотеза H_0 відкидається.

Критерій узгодженості Пірсона (критерій χ^2).

Одна із найважливіших задач математичної статистики — знаходження невідомого закону розподілу випадкової величини X — кількісної ознаки об’єктів генеральної сукупності. В § 1 були висвітлені питання опрацювання вибірки з метою отримання інформації про вид емпіричного розподілу та його характеристик: середньої ознаки, величини розкиду, симетричності розподілу. Потім на основі цих даних підбирається той розподіл, який найкраще апроксимує дослідний розподіл випадкової величини.

Після вибору виду розподілу необхідно знайти (хоча б наближено) параметри того закону розподілу, який характеризує досліджувану випадкову величину.

Оскільки в більшості практично важливих випадків параметри теоретичного закону розподілу є або математичним сподіванням, або дисперсією випадкової великими, або виражаються через них (тобто через моменти перших порядків (§§4,5 Ч. 1)), то отримані вище висновки дають можливість знайти ці параметри за дослідними даними. Наприклад, для нормального розподілу параметри a та σ визначається рівностями $a = \overline{x}_в$, $\sigma = \sigma_в$ (або $\sigma = S$). Таким чином, можна говорити про "відомість" теоретичного розподілу досліджуваної ознаки X. При цьому лапки вказують на гіпотетичність такого знання.

Критерій узгодженості Пірсона $\left(\chi^2\right)$ ґрунтується на виборі певної міри розбіжності між теоретичним і емпіричним (дослідним) розподілами. При цьому задачу перевірки узгодженості можна сформулювати таким чином: на основі вибірки спостережених значень деякої випадкової величини X потрібно визначити, що емпіричний розподіл належить певному розподілу (нормальному, показниковому, біноміальному і т. д.) із *визначеними* параметрами — гіпотеза H_0 проти альтернативної гіпотези H_1 — розподіл не належить вибраному розподілу.

Нехай у відповідності із гіпотезою H_0 **відома функція розподілу імовірностей** $\boldsymbol{F(x)}$ **досліджуваної ознаки** $\boldsymbol{X}$. Позначимо через $S_1, S_2, ..., S_m$ множини, на які розбивається вся область можливих значень випадкової величини X; ці множини — або інтервали для **неперервної** випадкової величини, або групи окремих значень **дискретної** випадкової величини, які не мають спільних точок. Тоді можна обчислити імовірності того, що при випробуванні випадкова величина X набере значення із множини S_i, тобто

$$p_i = P(X \in S_i), \quad i = \overline{1, m}. \tag{3.8}$$

При цьому всі $p_i > 0$, $i = \overline{1, m}$, і

$$\sum_{i=1}^{m} p_i = 1.$$

Нехай $n_1, n_2, ..., n_m$ — відповідні емпіричні групові частоти, тобто суми частот тих варіант (значень випадкової величини X із вибірки), що потрапляють відповідно у множини $S_1, S_2, ..., S_m$.

Якщо гіпотеза H_0 правильна, то статистичний розподіл вибірки можна розглядати як емпіричний аналог для генерального розподілу, який визначається функцією $F(x)$. Це означає, що n_i є частотою (абсолютною) появи випадкової події $(X \in S_i)$, $i = \overline{1, m}$, в послідовності із n спостережень. Отже, в першому розподілі кожній множині S_i ставляться у відповідність відносна частота n_i / n, а в другому — імовірність p_i. Тоді за методом найменших квадратів в якості міри розходження між статистичним розподілом відносних частот вибірки і теоретичним розподілом імовірностей отримується міра розбіжності виду

$$\chi^2 = \sum_{i=1}^{m} \frac{(n_i - np_i)^2}{np_i} \tag{3.9}$$

така, що при збільшенні обсягу вибірки розподіл величини χ^2 наближається до граничного розподілу χ^2 із $k = m - r - 1$ ступенями вільності, де m — число інтервалів або груп, на які розбита вся множина спостережених даних, r — число параметрів гіпотетичного розподілу імовірностей, що оцінюється за даними вибірки.

Зауваження. Числа

$$n_i^0 = np_i, \quad i = \overline{1,m}, \tag{3.10}$$

називаються **теоретичними частотами** відбуття випадкових подій $(X \in S_i)$. Враховуючи рівності (3.10), із (3.9) отримаємо таку міру розбіжності між теоретичними і емпіричними частотами:

$$\chi^2 = \sum_{i=1}^{m} \frac{(n_i - n_i^0)^2}{n_i^0}. \tag{3.9*}$$

Випадкова величина, що визначається рівностями (3.9) або (3.9*), називається **критерієм Пірсона**.

Для перевірки гіпотези H_0 задамо рівень значущості α і за табл. 6 додатків знайдемо критичну точку $\chi^2_{\text{кр}}(\alpha; k)$, де $p = \alpha$, $k = m - r - 1$. Таким чином, правостороння критична область визначається нерівністю $\chi^2 > \chi^2_{\text{кр}}(\alpha; k)$, а область прийняття нульової гіпотези — нерівністю $\chi^2 < \chi^2_{\text{кр}}(\alpha; k)$. За результатами вибірки обчислюємо $\chi^2_{\text{сп}}$. Якщо $\chi^2_{\text{сп}} < \chi^2_{\text{кр}}$, то гіпотезу H_0 приймаємо, а у випадку $\chi^2_{\text{сп}} > \chi^2_{\text{кр}}$ H_0 відкидаємо.

Перевірка гіпотези про нормальний розподіл генеральної сукупності

Використання критерія узгодженості Пірсона проілюструємо для випадку перевірки гіпотези H_0: кількісна ознака X об'єктів генеральної сукупності розподілена за нормальним законом.

Нехай статистичний розподіл має такий вид:

$$\begin{array}{c|cccc} [\,x_i;x_{i+1}) & [\,x_1;x_2) & [\,x_2;x_3) & \ldots & [x_m;x_{m+1}] \\ \hline n_i & n_1 & n_2 & \ldots & n_m \end{array}, \tag{3.11}$$

де $x_{i+1}-x_i=h,\quad i=\overline{1,m},\quad \sum\limits_{i=1}^{k} n_i=n,$ число частинних інтервалів визначається наближеною рівністю $m\approx \log_2 n.$ Якщо ж статистичний розподіл є дискретним, тоді потрібно перейти до інтервального (див. розв'язування задачі 1.2). Використовуючи метод добутків, знайдемо для розподілу (3.11) $\overline{x}_в$ та $D_в$. Тоді покладемо, що густина розподілу досліджуваної ознаки X має такий вид:

$$f(x)=\frac{1}{\sigma_в\sqrt{2\pi}}e^{-\frac{(x-\overline{x}_в)^2}{2\sigma_в^2}}.$$

Враховуючи те, що можливі значення нормально розподіленої випадкової величини заповнюють всю дійсну вісь, в якості множин $S_1,S_2,\ldots S_m$ візьмемо інтервали

$$(-\infty;x_2),(x_2;x_3),\ldots(x_m;\infty).$$

Для знаходження імовірностей $p_i=P(x_i<X<x_{i+1}),\quad i=\overline{1,m},$ де $x_1=-\infty,\ \ x_{m+1}=\infty,$ використаємо формулу

$$P(\alpha<X<\beta)=\varPhi\left(\frac{\beta-a}{\sigma}\right)-\varPhi\left(\frac{\alpha-a}{\sigma}\right),$$

в якій a та σ — параметри нормального розподілу.

Отже, за табл. 3 додатків обчислимо

$$p_i=\Phi\left(\frac{x_{i+1}-\overline{x}_в}{\sigma_в}\right)-\Phi\left(\frac{x_i-\overline{x}_в}{\sigma_в}\right),\quad i=\overline{1,m},$$

поклавши $x_1=-\infty,\ \ x_{m+1}=\infty.$

В результаті за формулою (3.9) можна знайти для даної вибірки (розподілу (3.11)) $\chi^2_{\text{сп}}$. Для заданого рівня значущості α і ступеня вільності $k=m-3$ (число параметрів нормального розподілу $r=2$) за табл. 6 додатків знаходимо $\chi^2_{\text{кр}}(\alpha;k).$ Тоді гіпотеза H_0 приймається, якщо $\chi^2_{\text{сп}}<\chi^2_{\text{кр}}(\alpha;k),$ і відхиляється у випадку $\chi^2_{\text{сп}}>\chi^2_{\text{кр}}(\alpha;k).$

Зауваження. Вказане вище число ступенів вільності $k=m-3$ відноситься тільки до того випадку, коли обидва параметри нормального закону розподілу знаходяться за даними вибірки, тобто коли замість точних значень a і σ використовуються їх емпіричні значення $\bar{x}_в$ та $\sigma_в$. Якщо значення a точно відоме (наприклад, у випадку знаходження відхилень від еталону), то число ступенів вільності дорівнює $k = m - 2$. Якщо ж відомі обидва параметри, то число ступенів вільності $k = m - 1$. На практиці така ситуація зустрічається рідко, а тому для отримання числа ступенів вільності не менше п'яти потрібно, щоб число частинних інтервалів було не меншим восьми.

Критерій узгодженості Колмогорова

Критерій Пірсона можна використовувати як для випадку неперервної ознаки *X*, так і для дискретної. Проте недоліком його є те, що випадкова величина χ^2, як правило, залежить від групування варіаційного ряду вибірки. Тому іноді доцільним є використання інших критеріїв узгодженості.

Нехай $F(x)$ — теоретична функція розподілу імовірностей **неперервної** випадкової величини *X*, що є кількісною ознакою об'єктів генеральної сукупності. При цьому згідно із гіпотезою H_0 вид цієї функції відомий. Нехай $F^*(x)$ — емпірична функція розподілу, отримана внаслідок опрацювання вибірки обсягом n. Згідно із теоремою Гливенко-Кантеллі $F^*(x)$ є спроможною оцінкою функції $F(x)$. Тому можна порівнювати емпіричну функцію розподілу $F^*(x)$ із гіпотетичною $F(x)$ і якщо міра розходженості між ними мала, то вважати правильною гіпотезу H_0. Найбільш природною і простою із таких мір є рівномірна віддаль

$$D_n = \max_{-\infty<x<\infty} \left| F^*(x) - F(x) \right| . \tag{3.12}$$

Проте при побудові критерія Колмогорова більш зручно користуватись нормованою віддалю $D_n\sqrt{n}$.

Оскільки вибірка утворюється випадковим чином, то D_n є випадковою величиною, при цьому величина $D_n\sqrt{n}$ має граничний при $n \to \infty$ розподіл, обчислений в припущенні, що гіпотеза H_0 є правильною. При умові неперервності $F^*(x)$ цей розподіл має такий вид

$$P\left(D_n\sqrt{n} < x \right) \underset{n\to\infty}{\to} K(x) = \begin{cases} 0 & \text{при } x \le 0, \\ \sum\limits_{k=-\infty}^{\infty} (-1)^k e^{-2k^2x^2} & \text{при } x > 0. \end{cases} \tag{3.13}$$

Функція розподілу імовірностей $K(x)$ називається **розподілом Колмогорова**. Суттєвим є те, що **вона не залежить від конкретного розподілу** $F(x)$.

Критерій Колмогорова приписує прийняти гіпотезу H_0, якщо $D_n < K_{n;\alpha}$ і відхиляти у випадку виконання нерівності $D(x) \geq K_{n;\alpha}$, де α — рівень значущості, $K_{n;\alpha}$ — критичне значення критерію, яке знаходиться за табл. 8 додатків. Відмітимо, що в деяких посібниках і довідниках наводяться таблиці критичних значень для статистики $D_n\sqrt{n}$.

Описання табл. 8 додатків

В табл. 8 наведені критичні значення $K_{n;\alpha}$ для $n = \overline{1,100}$; $\alpha = 0{,}1$; 0,05; 0,02; 0,01.

Приклад для зчитування: при $n = 26$, $\alpha = 0{,}01$ знаходимо значення $K_{26;0{,}01} = 0{,}311$. Для обсягів вибірки $n > 100$ з урахуванням співвідношення (3.13) випливає асимптотичне співвідношення

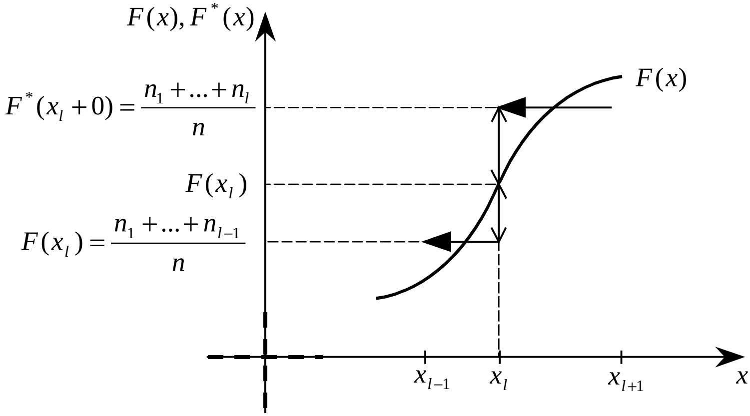
$$K_{n;\alpha} \approx k_{1-\alpha} \big/ \sqrt{n}\,, \tag{3.14}$$

де $K(k_{1-\alpha}) \approx 1 - \alpha$. Значення $k_{1-\alpha}$ знаходяться за табл. 9 додатків. Наприклад, якщо $\alpha = 0{,}05$, тоді $k_{1-0{,}05} = 1{,}358$ і за формулою (3.14) отримаємо $K_{n;0{,}05} \approx 1{,}358 \big/ \sqrt{n}$. При $n = 100$ наближене критичне значення дорівнює 0,1358, а відповідне точне значення за табл. 8. — 0,134.

Розглянемо практичну реалізацію критерія Колмогорова. Нехай статистичний розподіл вибірки задається таблицею $\begin{array}{c|cccc} x_i & x_1 & x_2 & \ldots & x_k \\ \hline n_i & n_1 & n_2 & \ldots & n_k \end{array}$, де $\sum\limits_{i=1}^{k} n_i = n$, $x_1 < x_2 < \ldots < x_k$. Тоді емпірична функція розподілу $F^*(x)$ має такий вид:

$$F^*(x) = \begin{cases} 0, & \text{якщо } x \leq x_1, \\ n_1/n, & \text{якщо } x_1 < x \leq x_2, \\ (n_1 + n_2)/n, & \text{якщо } x_2 < x \leq x_3, \\ \ldots & \ldots, \\ (n_1 + n_2 + \ldots + n_{k-1})/n & \text{якщо } x_{k-1} < x \leq x_k, \\ 1, & \text{якщо } x > x_k. \end{cases} \tag{3.15}$$

При знаходженні значення статистики D_n, що визначається за формулою (3.12), потрібно враховувати сходинковий характер графіка функції $F^*(x)$. Найбільша за абсолютною величиною різниця між $F^*(x)$ і $F(x)$ буде досягатися в одній із точок розподілу (мал. 3.1).



Мал. 3.1.

Тобто

$$D_n = \max\left\{ D_n^{(1)}; D_n^{(2)} \right\}, \tag{3.16}$$

де

$$\begin{aligned} D_n^{(1)} &= \max_{l=\overline{1,k}} \left| \frac{n_1+n_2+...+n_l}{n} - F(x_l) \right| = \max_{l=\overline{1,k}} \left| F^*(x_l+0) - F(x_l) \right|, \\ D_n^{(2)} &= \max_{l=\overline{1,k}} \left| F(x_l) - \frac{n_1+n_2+...+n_{l-1}}{n} \right| = \max_{l=\overline{1,k}} \left| F(x_l) - F^*(x_l) \right|. \end{aligned} \tag{3.17}$$

Відмітимо, що другий алгебраїчний доданок в другому співвідношенні (3.17) для $l=1$ дорівнює 0, бо із (3.15) $F^*(x_1)=0$, а перший доданок у перщому співвідношенні (3.17) при $l=k$ дорівнює 1.

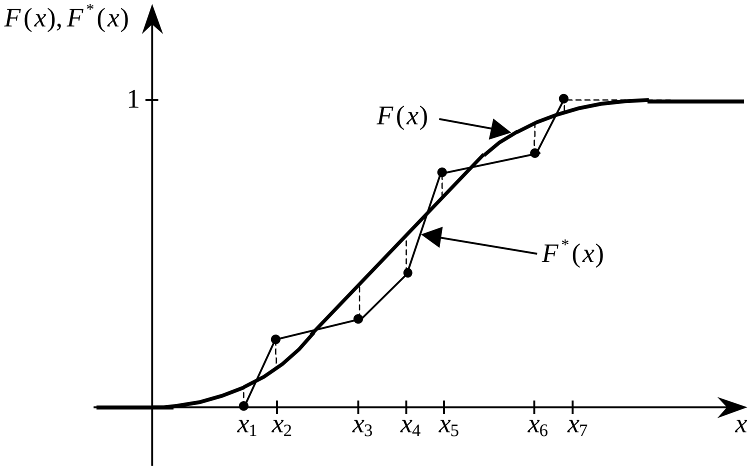
Якщо у статистичному розподілі вибірки всі частоти дорівнюють одиниці, тоді D_n знаходиться з такого співвідношення

$$D_n = \max_{1\le l\le n} \left| F(x_l) - \frac{2l-1}{2n} \right| + \frac{1}{2n}. \tag{3.18}$$

Формули (3.16), (3.17) або (3.18) дають практичну реалізацію **двостороннього** критерія Колмогорова для випадку **дискретного розподілу** частот або відносних частот вибірки. Розглянемо інтервальний статистичний розподіл вибірки

$[x_i;x_{i+1})$	$[x_1;x_2)$	$[x_2;x_3)$	...	$[x_k;x_{k+1})$
n_i	n_1	n_2	...	n_k

Ескізи графіків емпіричної та теоретичної функцій розподілу в цьому випадку (мал. 3.2) вказують на суттєву відмінність при знаходженні відхилень $\left|F^*(x)-F(x)\right|$ в порівнянні із дискретним розподілом (мал. 3.1).



Мал. 3.2.

В цьому випадку слід вибирати односторонню критичну область. Тоді статистика критерію Колмогорова задається формулою

$$D_n^+ = \max_{1\le i\le k+1}\left|F^*(x_i)-F(x_i)\right|. \tag{3.19}$$

У випадку використання одностороннього критерія на рівні значущості α гіпотеза H_0 відкидається, якщо $D_n^+ > K_{n;\alpha}^+$, де $K_{n;\alpha}^+$ — критичне значення. Оскільки $K_{n;\alpha}^+ \approx K_{n;2\alpha}$, то $K_{n;\alpha}^+$ можна знаходити за табл. 8 для $n=\overline{1,100}$ і $\alpha = 0{,}05; 0{,}025; 0{,}01; 0{,}005$.

При $n\to\infty$ має місце асимптотичне співвідношення ([17, с. 238])

$$K^{+}_{n;\alpha} \approx \sqrt{\frac{-\ln\alpha}{2n}}. \qquad (3.20)$$

Наприклад, для $n = 100$ і $\alpha = 0{,}05$ приблизне значення, отримане з допомогою цього співвідношення, дорівнює 0,1224, а відповідне точне значення — 0,121 $(K^{+}_{100;0,05} = K_{100;0,1})$.

РОЗВ'ЯЗУВАННЯ ТИПОВИХ ЗАДАЧ

Задача 3.1. Торгівельна фірма розглядає питання про відкриття в новому мікрорайоні міста філії. Відомо, що фірма буде працювати прибутково, якщо щомісячний середній дохід мешканців мікрорайону перевищує 500 у. о. Відомо також, що дисперсія доходів $\sigma^2 = 400$ у. о.

Знайти умови прийняття рішення, з допомогою якого на підставі вибірки обсягом $n = 100$ і рівня значущості $\alpha = 0{,}05$ можна встановити, що робота філії буде прибутковою.

○ Вважаємо, що середній місячний дохід навмання вибраного мешканця є нормально розподіленою випадковою величиною.

Фірма не відкриє філію, якщо середній місячний дохід мешканців не перевищить 500 у. о. Тому будемо вважати, що $H_0 : a_0 = 500$, а $H_1 : a > 500$. Оскільки дисперсія відома, то згідно із правилом 2 гіпотеза H_1 приймається, якщо $U_{\text{спост}} > u_{\text{кр}}$, де $u_{\text{кр}}$ — корінь рівняння (3.4), тобто $\Phi(u_{\text{кр}}) = (1 - 2 \cdot 0{,}05)/2 = 0{,}45.$ Звідки за табл. 3 додатків $u_{\text{кр}} = 1{,}65.$ Із врахуванням (3.2) і умов задачі

$$U_{\text{спост}} = (\bar{x}_в - a_0)\sqrt{n}\big/\sigma = (\bar{x}_в - 500)\sqrt{100}\big/2.$$

Тому H_1 приймається і, отже, філію відкривають, якщо середній місячний дохід 100 мешканців

$$\bar{x}_в > 500 + 2 \cdot 1{,}65 = 503{,}3. \quad \bullet$$

Задача 3.2. Для уточнення норм виробки на підприємстві проведено 31 незалежне вимірювання продуктивності праці робітників, виконуючих однотипну операцію. Середня продуктивність праці склала $\bar{x}_в = 7{,}2$ (одиниць продукції за 1 год), а середнє квадратичне відхилення $\sigma_в = 0{,}5$ (одиниць продукції за год). Потрібно перевірити гіпотезу, що при масовому випуску цієї продукції середня продуктивність a_0 складе 7,5 одиниць продукції за 1 год при конкуруючій гіпотезі $a < 7{,}5$ одиниць продукції за 1 год при рівні значущості $\alpha = 0{,}01$.

○ Будемо вважати, що продуктивність праці навмання взятого робітника є нормально розподіленою випадковою величиною. Згідно із умовою задачі дисперсія цієї величини невідома, тому в якості статистичного критерія візьмемо випадкову величину (3.5), спостережене значення якої

$$T_{\text{спост}} = (\overline{x}_в - a_0)\sqrt{n}\Big/S = (7{,}2 - 7{,}5)\sqrt{31}\Bigg/\left(0{,}5\sqrt{\frac{31}{31-1}}\right) = -3{,}286.$$

При обчисленні використовуємо такий зв'язок між $\sigma_в$ та S:

$$S = \sigma_в\sqrt{n/(n-1)}.$$

Оскільки $H_1 : a < a_0$, то потрібно побудувати лівосторонню критичну область. Згідно із правилом 3* знайдемо "допоміжну" критичну точку $t_{\text{правост.кр}}(0{,}01;30)$ по рівню значущості 0,01, розміщеному в нижньому рядку табл. 5 додатків: $t_{\text{правост.кр}}(0{,}01;30) = 2{,}457$.

Тоді $t_{\text{лівост.кр}} = -2{,}457$.

Так як $T_{\text{спост}} = -3{,}286 < -2{,}457 = t_{\text{лівост.кр}}$, то основну гіпотезу відкидаємо на користь альтернативної гіпотези $H_1 : a < 7{,}5$. ●

Задача 3.3. Кількісна ознака генеральної сукупності розподілена за нормальним законом. За вибіркою обсягом $n = 16$ знайдена дисперсія вибіркова $D_в = 10{,}6$. Для рівня значущості $\alpha = 0{,}02$ перевірити основну гіпотезу $H_0 : \sigma^2 = \sigma_0^2 = 12$, якщо конкуруюча гіпотеза $H_0 : \sigma^2 \neq 12$.

○ Знайдемо спочатку виправлену дисперсію

$$S^2 = \left[\, n/(n-1)\right] D_в = (16/15)10{,}6 = 11{,}307,$$

а потім спостережене значення критерія

$$\chi^2_{\text{спост}} = (n-1)\,S^2\big/\sigma_0^2 = 15\cdot 11{,}307/12 = 14{,}133.$$

Оскільки $H_1 : \sigma^2 \neq 12$, то критична область є двосторонньою. Згідно із правилом 2 за табл. 6 додатків знаходимо критичні точки: ліву — $\chi^2_{\text{кр}}(1-\alpha/2;k)=\chi^2_{\text{кр}}(1-0{,}02/2;15)=\chi^2_{\text{кр}}(0{,}99;15)=5{,}23$ і праву — $\chi^2_{\text{кр}}(\alpha/2;k)=\chi^2_{\text{кр}}(0{,}01;15)=30{,}58.$ Так як спостережене значення критерія належить області прийняття основної гіпотези $(5{,}23<14{,}133<30{,}58)$, то нема підстав її відкидати. Іншими словами, виправлена дисперсія вибіркова (11,307) неістотно відрізняється від гіпотетичної дисперсії генеральної (12). ●

Задача 3.4. За даними табл. 1.2 про контрольні виміри товщини (в міліметрах) 200 вкладишів шатунних підшипників з допомогою критерія Пірсона перевірити гіпотезу H_0 про нормальний розподіл кількісної ознаки генеральної сукупності, якщо рівень значущості дорівнює 0,05.

○ Знайдемо число m частинних інтервалів однакової довжини, які отримуються при переході від дискретного ряду варіант з табл. 1.2 до інтервального розподілу частот, використовуючи формулу $m\approx\log_2 n.$ Оскільки обсяг вибірки $n=200$, то $m\approx\log_2 200\approx 8.$ Шуканий інтервальний розподіл наведений в табл. 1.3.

Для знаходження $\bar{x}_в$ та $\sigma_в$ від цього інтервального розподілу потрібно перейти до дискретного, "нові" варіанти якого є серединами частинних інтервалів. Для отриманого розподілу в задачі 1.2 методом добутків знайдено: $\bar{x}_в=1{,}7417;$ $\sigma_в=0{,}0113269.$

Результати допоміжних розрахунків в процесі знаходження теоретичних частот $n_i^0=p_i n$ розташуємо в табл. 3.1, де $x_1=-\infty,$ $x_9=\infty.$ Відмітимо, що згідно із властивостями функції Лапласа $\Phi(-x)=-\Phi(x),$ $\Phi(-\infty)=-0{,}5,$ $\Phi(\infty)=0{,}5.$ Крім цього, при знаходженні значень $\Phi(x)$ за допомогою табл. 3 додатків для аргументів, що мають третій знак після коми, здійснимо лінійну інтерполяцію. Наприклад, при знаходженні $\Phi(1{,}916)$ використаємо табличні значення: $\Phi(1{,}91)=0{,}47193,$ $\Phi(1{,}92)=0{,}47257.$ Тоді

$$\Phi(1{,}916)=\Phi(1{,}91)+[\Phi(1{,}92)-\Phi(1{,}91)]\cdot 0{,}6=$$
$$=0{,}47193+(0{,}47257-0{,}47193)\cdot 0{,}6=0{,}472314\approx 0{,}47231.$$

Для знаходження $\chi^2_{\text{сп}}$ складемо розрахункову табл. 3.2, з якої отримаємо $\chi^2_{\text{сп}}=2{,}5393$.

Таблиця 3.1

Інтервал $(x_i; x_{i+1})$	n_i	$x_{i+1}-\overline{x}_в$	$x_i-\overline{x}_в$	$\dfrac{x_{i+1}-\overline{x}_в}{\sigma_в}$	$\dfrac{x_i-\overline{x}_в}{\sigma_в}$	$\Phi\left(\dfrac{x_{i+1}-\overline{x}_в}{\sigma_в}\right)$	$\Phi\left(\dfrac{x_i-\overline{x}_в}{\sigma_в}\right)$	$p_i=\Phi\left(\dfrac{x_{i+1}-\overline{x}_в}{\sigma_в}\right)-\Phi\left(\dfrac{x_i-\overline{x}_в}{\sigma_в}\right)$	$n_i^0=p_i n=200\cdot p_i$
(–∞; 1,720)	3	–0,0217	–∞	–1,916	–∞	–0,47231	–0,5	0,0277	5,54
[1,720; 1,727)	13	–0,0147	–0,0217	–1,298	–1,916	–0,40285	–0,47231	0,06946	13,89
[1,727; 1,734)	32	–0,0077	–0,0147	–0,680	–1,298	–0,25175	–0,40285	0,1511	30,22
[1,734; 1,741)	49	–0,0007	–0,0077	–0,062	–0,680	–0,02472	–0,25175	0,2270	45,41
[1,741; 1,748)	42	0,0063	–0,0007	0,556	–0,062	0,21089	–0,02472	0,2356	47,12
[1,748; 1,755)	35	0,0133	0,0063	1,174	0,556	0,3798	0,21089	0,1689	33,78
[1,755; 1,762)	19	0,0203	0,0133	1,792	1,174	0,46336	0,3798	0,08356	16,71
(1,762; ∞)	7	∞	0,0203	∞	1,792	0,5	0,46336	0,03664	7,33
Всього	200	—	—	—	—	—	—	—	200

Таблиця 3.2

i	n_i	n_i^0	$n_i - n_i^0$	$(n_i - n_i^0)^2$	$(n_i - n_i^0)^2 / n_i^0$	n_i^2	n_i^2 / n_i^0
1	3	5,54	–2,54	6,4516	1,1646	9	1,6246
2	13	13,89	–0,89	0,7921	0,0570	169	12,1670
3	32	30,22	1,78	3,1684	0,1048	1024	33,8848
4	49	45,41	3,59	12,8881	0,2838	2401	52,8738
5	42	47,12	–5,12	26,2144	0,5563	1764	37,4363
6	35	33,78	1,22	1,4884	0,0441	1225	36,2641
7	19	16,71	2,29	5,2441	0,3138	361	21,6038
8	7	7,33	–0,33	0,1089	0,0149	49	6,6849
$\sum$	200	200			$\chi^2_{\text{сп}} = 2{,}5393$		202,5393

За таблицею критичних точок розподілу χ^2 (табл. 6 додатків) для рівня значущості $\alpha = 0{,}05$ і числа ступенів вільності $k = 8 - 3 = 5$ знайдемо $\chi^2_{\text{кр}}(0{,}05;5) = 11{,}07$.

Оскільки $\chi^2_{\text{сп}} < \chi^2_{\text{кр}}$, то нема підстав відкидати основну гіпотезу. Отже, дані спостережень узгоджуються із гіпотезою про нормальний розподіл кількісної ознаки генеральної сукупності. ●

Задача 3.5. На основі вибірки

0,6917; 0,1794; 0,7410; 0,3094; 0,1174;
0,5424; 0,0834; 0,6288; 0,9401; 0,6606

для рівня значущості $\alpha = 0{,}05$ з допомогою критерія Колмогорова перевірити гіпотезу H_0 про те, що дана генеральна сукупність на проміжку $[0; 1]$ має рівномірний розподіл, тобто $H_0 : F(x) = x \quad (0 \le x \le 1)$.

○ В даному випадку серед варіант відсутні рівні, тому спочатку утворимо варіаційний ряд, а потім використаємо співвідношення (3.18) для обчислення D_{10}. Результати проміжних обчислень занесені в табл. 3.3.

Таблиця 3.3

l	x_l	$F(x_l)$	$\frac{2l-1}{2n}$	$F(x_l) - \frac{2l-1}{2n}$	$\left\lvert F(x_l) - \frac{2l-1}{2n} \right\rvert$
1	0,0834	0,0834	0,05	0,0334	0,0334
2	0,1174	0,1174	0,15	–0,0326	0,0326
3	0,1794	0,1794	0,25	–0,0706	0,0706
4	0,3094	0,3094	0,35	–0,0406	0,0406
5	0,5424	0,5424	0,45	0,0924	0,0924
6	0,6288	0,6288	0,55	0,0788	0,0788
7	0,6606	0,6606	0,65	0,0106	0,0106
8	0,6917	0,6917	0,75	–0,0583	0,0583
9	0,7410	0,7410	0,85	–0,1090	$\boxed{0{,}1090}$
10	0,9401	0,9401	0,95	–0,0099	0,0099

Найбільше число 0,1090 в останньому стовпчику визначає перший доданок у формулі (3.18). Тому $D_{10} = 0{,}1090 + 0{,}05 = 0{,}1590$. За табл. 8 додатків при рівні значущості $\alpha = 0{,}05$ і $n = 10$ знаходимо критичне значення

$K_{10;0,05} = 0{,}409$. Оскільки $D_{10} = 0{,}1590 < 0{,}409 = K_{10;0,05}$, то робимо висновок, що результати спостережень не суперечать гіпотезі H_0. ●

Задача 3.6. На основі інтервального статистичного розподілу вибірки, наведеного в табл. 1.3, з допомогою критерія Колмогорова здійснити перевірку гіпотези H_0 про нормальний розподіл кількісної ознаки генеральної сукупності, якщо рівень значущості $\alpha = 0{,}05$.

○ В задачі 1.2 для даного розподілу знайдені числові характеристики $\overline{x}_в = 1{,}7417$, $\sigma_в = 0{,}0113269$, які можна вважати параметрами нормального розподілу a та σ відповідно. Тоді теоретична функція розподілу $F(x)$ має такий вид:

$$F(x) = 0{,}5 + \Phi\left(\frac{x-a}{\sigma}\right) = 0{,}5 + \Phi\left(\frac{x-\overline{x}_в}{\sigma_в}\right),$$

де $\Phi\ (x)$ — функція Лапласа, значення якої наведені в табл. 3 додатків.

Враховуючи відмінність емпіричної функції, побудованої для інтервального розподілу, в порівнянні із випадком дискретного розподілу, виберемо статистику критерія Колмогорова (3.19) (для односторонньої критичної області).

Всі розрахунки для отримання D^+_{200} зведемо у табл. 3.4, при заповненні якої використаємо відповідні стовпці табл. 3.1. Зокрема, стовпці 1, 2 та 4 табл. 3.4 взяті із стовпців 1, 2 та 5 відповідно табл. 3.1. При заповненні стовпця 3 використовується означення емпіричної функції. Нарешті, при обчисленні даних стовпця 5 були використані результати стовпця 7 табл. 3.1.

Таблиця 3.4

$[x_i; x_{i+1})$	n_i	$F^*(x_{i+1})$	$\dfrac{x_{i+1}-\overline{x}_в}{\sigma_в}$	$F(x_{i+1}) = 0{,}5 + \Phi\left(\dfrac{x_{i+1}-\overline{x}_в}{\sigma_в}\right)$	$\lvert F^*(x) - F(x)\rvert$
1	2	3	4	5	6
(–∞; 1,713)	0	0	–2,534	0,00564	0,00564
[1,713; 1,720)	3	0,015	–1,916	0,02769	0,01269
[1,720; 1,727)	13	0,080	–1,298	0,09715	$\boxed{0{,}01715}$
[1,727; 1,734)	32	0,240	–0,680	0,24825	0,00825
[1,734; 1,741)	49	0,485	–0,062	0,47528	0,00972
[1,741; 1,748)	42	0,695	0,556	0,71089	0,01589

Продовження табл. 3.4

[1,748; 1,755)	35	0,870	1,174	0,8798	0,00980
[1,755; 1,762)	19	0,965	1,792	0,96336	0,00164
[1,762; 1,769)	7	1	2,410	0,99202	0,00798
[1,769; ∞)	1	1	∞	1	0

Найбільший модуль різниці між відповідними значеннями емпіричної і (гіпотетичної) теоретичної функції розподілу вибираємо із стовпця 6 табл. 3.4:

$$D_{200}^{+} = \max_{1\le i\le k+1}\left|F(x_i) - F^{*}(x_i)\right| = 0{,}01715.$$

Критичне значення $K_{200;0,01}^{+}$ знайдемо за наближеною формулою (3.20), використавши рівність

$$\lg\alpha = 0{,}4343\ln\alpha, \text{ з якої } \ln\alpha = 2{,}3026\lg\alpha.$$

Отже,

$$\ln(0{,}1) = 2{,}3026\lg 0{,}1 = -2{,}3026,$$

$$K_{200;0,1}^{+} \approx \sqrt{\frac{2{,}3026}{2\cdot 200}} = 0{,}07587.$$

Оскільки $D_{200}^{+} = 0{,}01715 < 0{,}07587 = K_{200;0,1}^{+}$, то гіпотезу H_0 приймаємо.

Відмітимо, що в задачах 3.4 та 3.6 фактично досліджувався інтервальний статистичний розподіл з табл. 1.3. Цей розподіл породжений дискретним рядом варіант з табл. 1.2 і, природно, приводить до (хай незначних) похибок. Тому при перевірці гіпотези H_0 про нормальність розподілу генеральної сукупності, із якої отримано вибірку, задану табл. 1.2, з допомогою критерія Колмогорова бажано було б використати статистику Колмогорова (3.16) безпосередньо до варіант із табл. 1.2. Проте без використання обчислювальної техніки із відповідним програмним забезпеченням така задача є дуже працеємкою, якщо врахувати великий обсяг вибірки ($n = 200$) і слабку згрупованість варіант — наявність 48 різних варіант у табл. 1.2, для кожної із яких потрібно знаходити значення функції Лапласа.

В зв'язку із цим цікаво було б дослідити таке питання: якщо інтервальний статистичний розподіл із табл. 1.3 замінити дискретним розподілом

$$\begin{array}{c|cccccccc} x_i & 1{,}7165 & 1{,}7235 & 1{,}7305 & 1{,}7375 & 1{,}7445 & 1{,}7515 & 1{,}7585 & 1{,}7655 \\ \hline n_i & 3 & 13 & 32 & 49 & 42 & 35 & 19 & 7 \end{array}, \quad (3.21)$$

то чи буде для нього правильною гіпотеза H_0 згідно із критерієм узгодженості Колмогорова? Отримання відповіді на це питання передбачає використання статистики (3.16) двостороннього критерія Колмогорова для випадку, коли частоти n_i не дорівнюють 1, як це було для даних задачі 3.5, а остаточний висновок попередить про обережність стосовно подібних переходів. ●

Задача 3.7. На основі статистичного розподілу (3.21) з допомогою критерія Колмогорова здійснити перевірку гіпотези H_0 про нормальний розподіл генеральної сукупності при рівні значущості $\alpha = 0{,}05$.

○ В задачі 1.2 для розподілу (3.21) знайдено числові характеристики $\overline{x}_в = 1{,}7417$, $\sigma_в = 0{,}0113269$, які можна вважати параметрами нормального розподілу a та σ відповідно. Тоді, як і для задачі 3.6, теоретична функція розподілу $F(x)$ має такий вид:

$$F(x) = 0{,}5 + \Phi\left(\frac{x - \overline{x}_в}{\sigma_в}\right) = 0{,}5 + \Phi\left(\frac{x - 1{,}7417}{0{,}0113269}\right).$$

Всі розрахунки для обчислення D_{200} помістимо в табл. 3.5.

Найбільший модуль різниці між відповідними значеннями емпіричної і теоретичної функцій розподілу вибираємо із останнього стовпця табл. 3.5: $D_{200} = 0{,}12894$. За формулою (3.14) критичне значення $K_{200;0,05} = k_{1-0,05} \big/ \sqrt{200}$, де за табл. 9 додатків $k_{1-0,05} = 1{,}358$. Остаточно $K_{200;0,05} = 1{,}358 \big/ \sqrt{200} = 0{,}09603$. Оскільки $D_{200} = 0{,}12894 > 0{,}09603 = K_{200;0,05}$, то робимо висновок, що гіпотезу H_0 про нормальність розподілу генеральної сукупності на підставі дискретного розподілу вибірки (3.21) потрібно відкинути. ●

Таким чином, згідно із критеріями узгодженості Пірсона та Колмогорова, використаними до інтервального розподілу вибірки із табл. 1.3, гіпотеза H_0 не відхиляється (задачі 3.4 та 3.6), а використання критерію Колмогорова до дискретного розподілу, породженого цим самим інтервальним розподілом, вже приводить до висновку про хибність цієї ж гіпотези. З’ясуємо причину такого неспівпадання висновків. Найбільші розходження між значеннями емпіричної та гіпотетичної теоретичної функцій розподілу, які перевищують критичне значення $K_{200;0,05}$, отримані для варіант 1,7375 та 1,7445 з найбільшими частотами 49 та 42 відповідно (див. табл. 3.5). Оскільки $\overline{x}_в = 1{,}7417,$ то заміна модальних інтервалів їх серединами приводить до “віддалення” отриманих варіант від $\overline{x}_в$ із збереженням великих частот, що і приводить до різкого збільшення відхилення $F^*(x)$ від $F(x)$ в цих точках.

Нарешті, відмітимо, що критерій Колмогорова рекомендується використовувати при обсязі вибірки $n \geq 30$.

ЗАДАЧІ ДЛЯ САМОСТІЙНОГО РОЗВ’ЯЗУВАННЯ

Завдання № 6

Для інтервальних статистичних розподілів, наведених в завданні № 2, за критеріями Пірсона та Колмогорова для рівня значущості $\alpha = 0{,}05$ перевірити гіпотезу H_0 — кількісна ознака генеральної сукупності розподілена за нормальним законом.

Таблиця 3.5

x_l	n_l	$F^*(x_l)$	$F^*(x_l+0)$	$\frac{x_l-\overline{x}_в}{\sigma_в}$	$F(x_l)=0{,}5+$ $+\Phi\left(\frac{x_l-\overline{x}_в}{\sigma_в}\right)$	$F^*(x_l)-F(x_l)$	$F^*(x_l+0)-F(x_l)$
1,7165	3	0	0,015	–2,225	0,01304	–0,01304	0,00196
1,7235	13	0,015	0,080	–1,608	0,05392	–0,03892	0,02608
1,7305	32	0,080	0,240	–0,989	0,16134	–0,08134	0,07866
1,7375	49	0,240	0,485	–0,371	0,35606	–0,11606	0,12894
1,7445	42	0,485	0,695	0,247	0,59755	–0,11255	0,09745
1,7515	35	0,695	0,870	0,865	0,80648	–0,11148	0,06352
1,7585	19	0,870	0,965	1,483	0,93096	–0,06096	0,03404
1,7655	7	0,965	1	2,101	0,98218	–0,01718	0,01782

§ 4. ЕЛЕМЕНТИ КОРЕЛЯЦІЙНОГО ТА РЕГРЕСІЙНОГО АНАЛІЗУ

Поняття стохастичної та статистичної залежності, кореляції і регресії. Основні задачі кореляційного і регресійного аналізу. Лінійні емпіричні рівняння парної кореляції. Вибірковий коефіцієнт лінійної кореляції та його властивості. Оцінка достовірності емпіричних коефіцієнтів кореляції і регресії за даними вибірки. Нелінійна парна кореляція. Вибіркове кореляційне відношення та його властивості. Регресійний аналіз: парна лінійна регресія.

КОРОТКІ ТЕОРЕТИЧНІ ВІДОМОСТІ

Поняття стохастичної і статистичної залежності, кореляції і регресії. Основні задачі кореляційного і регресійного аналізу

Дві випадкові величини можуть бути або незалежними, або пов'язані функціональною залежністю, або залежністю, яка називається стохастичною. Функціональна залежність, розглянута в §8 Ч. 1, досить рідко зустрічається в економічних дослідженнях. Дуже часто доводиться вивчати такі випадкові величини, для яких зміна можливих значень (при проведенні випробувань) приводить до зміни закону (умовного) розподілу іншої (див. §7 Ч. 1). Такий зв'язок між випадковими величинами називається **стохастичним**; він виникає тоді, коли на обидві величини впливають випадкові фактори, серед яких є спільні.

Найбільш важливим випадком стохастичного зв'язку є так званий **кореляційний** зв'язок, який встановлює залежність між значеннями випадкової величини X і умовним математичним сподіванням $M(Y \mid X = x)$ *) випадкової величини Y:

$$M(Y \mid X = x) = g(x), \qquad (4.1)$$

або між значенням випадкової величини Y і умовним математичним сподіванням $M(X \mid Y = y)$ *) випадкової величини X:

$$M(X \mid Y = y) = q(y). \qquad (4.2)$$

Функції $g(x)$ і $q(y)$ називаються **функціями регресії** Y на X та X на Y відповідно, а їх графіки — лініями регресії Y на X та X на Y,

---

*) Умовні математичні сподівання дискретних і неперервних випадкових величин визначаються в §7 Ч. 1.

рівняння (4.1) та (4.2) називаються **рівняннями регресії Y на X та X на Y** відповідно.

Використовуючи інформацію §§7,8 Ч. 1 можна отримати важливу для наступного викладу основну властивість регресії величини Y на величину X: якщо $g(x)$ є функцією регресії величини Y на X, то математичне сподівання квадрата відхилення величини Y від функції $g(x)$ менше, ніж від будь-якої функції $h(X) \neq g(X)$, тобто

$$M\left[\,Y - g(X)\right]^2 < M\left[\,Y - h(X)\right]^2. \qquad (4.3)$$

Аналогічною властивістю володіє і регресія величини X на величину Y.

Прикладами кореляційного зв'язку є стохастична взаємозалежність між: 1) окремими параметрами тіла людини або тварини; 2) обсягом виробництва підприємства та коефіцієнтом використання основних засобів.

На практиці сумісний закон розподілу випадкових величин X та Y (двовимірної випадкової величини (X; Y) §7 Ч. 1) невідомий. В розпорядженні дослідника є двовимірна вибірка

$$(x^{(1)}; y^{(1)}), (x^{(2)}; y^{(2)}), ..., (x^{(n)}; y^{(n)}), \qquad (4.4)$$

згрупувавши (для великих n) дані якої, можна отримати двовимірний статистичний розподіл, наведений в табл. 4.1.

Таблиця 4.1

X	Y						
	y_1	y_2	...	y_j	...	y_m	n_{x_i}
x_1	n_{11}	n_{12}	...	n_{1j}	...	n_{1m}	n_{x_1}
x_2	n_{21}	n_{22}	...	n_{2j}	...	n_{2m}	n_{x_2}
...	...	...	...	...	...	...	...
x_i	n_{i1}	n_{i2}	...	n_{ij}	...	n_{im}	n_{x_i}
...	...	...	...	...	...	...	...
x_k	n_{k1}	n_{k2}	...	n_{kj}	...	n_{km}	n_{x_k}
n_{y_j}	n_{y_1}	n_{y_2}	...	n_{y_j}	...	n_{y_m}	n

Тут n_{ij} — частота спільної появи ознак x_i, y_j (пари $(x_i; y_j)$); $\sum\limits_{i=1}^{k}\sum\limits_{j=1}^{m} n_{ij} = n$ — обсяг вибірки;

$$n_{x_i} = \sum_{j=1}^{m} n_{ij}, \ (i = \overline{1;k}); \quad n_{y_j} = \sum_{i=1}^{k} n_{ij}, \ \left(j = \overline{1;m}\right).$$

Кожному значенню X відповідає ряд значень Y, тобто зміна значень X приводить до зміни умовного статистичного розподілу $Y \mid x$. Аналогічно зміна значень Y веде до зміни умовного статистичного розподілу $X \mid y$.

Наприклад, при $X = x_1$ та $X = x_2$ відповідні умовні статистичні розподіли величини Y мають такі види:

$Y \mid x_1$	y_1	y_2	...	y_m
n_i	n_{11}	n_{12}	...	n_{1m}

;

$Y \mid x_2$	y_1	y_2	...	y_m
n_i	n_{21}	n_{22}	...	n_{2m}

.

Статистичною називається така залежність між величинами X та Y, для якої зміна спостережених значень однієї із величин зумовлює зміну умовного статистичного розподілу іншої.

Умовною середньою $\overline{y}_x$ називається середнє арифметичне спостережених значень Y, які відповідають значенню $X = x$. Наприклад, згідно із табл. 4.1

$$\overline{y}_{x_2} = \frac{y_1 n_{21} + y_2 n_{22} + \ldots + y_j n_{2j} + \ldots + y_m n_{2m}}{n_{x_2}}.$$

Умовною середньою $\overline{x}_y$ називається середнє арифметичне спостережених значень X, які відповідають значенню $Y = y$. Наприклад, у відповідності із табл. 4.1

$$\overline{x}_{y_1} = \frac{x_1 n_{11} + x_2 n_{21} + \ldots + x_k n_{k1}}{n_{y_1}}.$$

Можна довести, що умовні середні є незміщеними і спроможними точковими статистичними оцінками відповідних умовних математичних сподівань:

$$\overline{y}_x \approx M(Y \mid X = x); \quad \overline{x}_y \approx M(X \mid Y = y). \qquad (4.5)$$

Вкажемо підхід до знаходження статистичних наближень (емпіричних функцій) $\widehat{g}(x) = \widehat{g}(x; a_0, a_1, ..., a_m)$ та $\widehat{q}(y) = \widehat{q}(y; b_0, b_1, ..., b_m)$ невідомих функцій регресій $g(x)$ та $q(y)$ відповідно, а отже, з урахуванням співвідношень (4.5) і емпіричних рівнянь Y на X

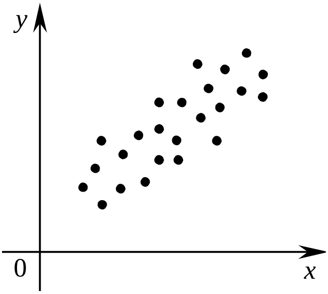
$$\overline{y}_x = \widehat{g}(x, a_0, a_1, ..., a_m) \qquad (4.6)$$

та X на Y

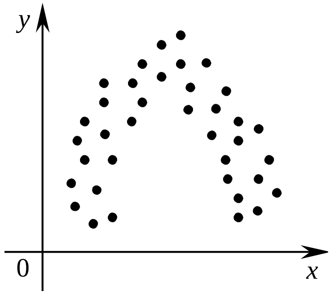
$$\overline{x}_y = \widehat{q}(y, b_0, b_1, ..., b_m) \qquad (4.7)$$

де $a_0, a_1, ..., a_m$ та $b_0, b_1, ..., b_m$ — невідомі параметри.

На першому кроці в прямокутній системі координат xOy будується послідовність пар чисел (4.4). Отримана сукупність точок називається **кореляційним полем** або **діаграмою розсіювання**. Конфігурація кореляційного поля дозволяє висунути гіпотезу про вид функції $\widehat{g}(x)$. Наприклад, у випадку кореляційного поля, зображеного на мал. 4.1, $\widehat{g}(x)$ доцільно шукати у вигляді лінійної функції $a_1x + a_0$, а у випадку мал. 4.2 у вигляді параболічної функції другого порядку $a_2x^2 + a_1x + a_0$.



Мал. 4.1.



Мал. 4.2.

На другому кроці потрібно знайти невідомі параметри. Ця задача розв’язується з допомогою так званого **методу найменших квадратів** (МНК), суть якого полягає в наступному. Оскільки невідома функція регресії Y на X $g(X)$ згідно із (4.3) мінімізує величину $M\left[Y - h(X)\right]^2$, а оцінкою $M(Y \mid x)$ є умовна середня $\overline{y}_x$, що відповідає певним спостереженим значенням $X = x$, то емпірична лінія регресії (4.6) повинна задовольняти рівності

$$F(a_0, a_1, ..., a_m) \equiv \sum_{i=1}^{n}\left[\ y^{(i)} - \widehat{g}(x^{(i)}; a_0, a_1, ..., a_m)\right]^2 \Rightarrow \min. \quad (4.8)$$

Права частина цієї рівності є сума квадратів віддалей вздовж осі $0y$ від точок $(x^{(i)}; y^{(i)})$ послідовності (4.4) до відповідних точок лінії регресії із тією ж абсцисою.

Для випадку згрупованих даних із табл. 4.1 рівність (4.8) набирає такого виду

$$F(a_0, a_1, ..., a_m) \equiv \sum_{i=1}^{k}\ \sum_{j=1}^{m}\left[\ y_j - \widehat{g}(x_i; a_0, a_1, ..., a_m)\right]^2 n_{ij} \Rightarrow \min. \quad (4.8^*)$$

Таким чином, емпірична функція $\widehat{g}(x;a_0,a_1,...,a_m)$ повинна усереднити (згладити) спостережені дані $(x^{(i)};y^{(i)})$, $i=\overline{1,n}$, з послідовності (4.4) або (x_i,y_j), $i=\overline{1,k}$, $j=\overline{1,m}$, із табл. 4.1. При цьому невідомі параметри $a_0,a_1,...,a_m$ повинні мінімізувати функцію $F(a_0,a_1,...,a_m)$; необхідною умовою цього є виконання системи рівнянь

$$\left\{\frac{\partial F(a_0,a_1,...,a_m)}{\partial a_i}=0,\quad i=\overline{0,m},\right. \qquad (4.9)$$

яка називається **системою нормальних рівнянь**.

Метод найменших квадратів дозволяє **при встановленому виді емпіричної функції регресії** $\widehat{g}(x;a_0,a_1,...,a_m)$ так знайти невідомі параметри $a_0,a_1,...,a_m$, що вона буде найкращою оцінкою функції регресії $g(x)$ в тому розумінні, що сума квадратів відхилень спостережених значень випадкової величини Y від відповідних ординат емпіричної функції буде найменшою.

Аналогічно знаходиться емпірична функція $\widehat{q}(y;b_0,b_1,...,b_m)$ регресії X на Y.

Зауваження. Враховуючи випадковий характер організації вибірки, можна зробити висновок, що для нефіксованої вибірки знайдені параметри є випадковими величинами.

Знаходження емпіричних рівнянь регресії — це тільки перший крок дослідження статистичних зв’язків між випадковими величинами. Наступним є встановлення сили або тісноти цих зв’язків. Відомо (§7 Ч. 1), що мірою лінійного зв’язку (стохастичного) між двома випадковими величинами X та Y є коефіцієнт кореляції

$$r=r_{XY}=\frac{K_{XY}}{\sigma_X\sigma_Y}=\frac{M(XY)-M(X)M(Y)}{\sigma_X\sigma_Y}, \qquad (4.10)$$

де K_{XY} — кореляційний момент або коефіцієнт коваріації (сумісної варіації), який ще часто позначають $\operatorname{cov}(X,Y)$, $\sigma_X=\sqrt{D(X)}$, $\sigma_Y=\sqrt{D(Y)}$. Зокрема, якщо величини X та Y незалежні, то коефіцієнт кореляції $r=0$; якщо $r=\pm1$, то X та Y пов’язані **лінійною** функціональною залежністю. Звідси випливає, що коефіцієнт кореляції r вимірює силу (тісноту) **лінійного** зв’язку між X та Y.

Використовуючи метод моментів, тобто замінивши числові характеристики їх статистичними оцінками, можна отримати точкову статистичну оцінку коефіцієнта кореляції — вибірковий (емпіричний) коефіцієнт кореляції.

$$r_в = r_{xy} = \frac{\overline{xy} - \overline{x}\cdot\overline{y}}{\sigma_x \sigma_y}, \qquad (4.11)$$

де для незгрупованих даних (4.4)

$$\overline{xy} = \sum_{i=1}^{n} x^{(i)} y^{(i)} \big/ n, \qquad (4.12)$$

а для згрупованих даних із табл. 4.1

$$\overline{xy} = \sum_{i=1}^{k} \sum_{j=1}^{m} n_{ij}\, x_i y_j \big/ n = \sum n_{xy}\, xy \big/ n, \qquad (4.13)$$

σ_x та σ_y — середні квадратичні відхилення вибіркові для ознак X та Y відповідно.

Зауваження. Оскільки кореляційний момент рівносильно (4.10) визначається ще й таким чином: $K_{XY} = M\left\{ \left[X - M(X) \right] \left[Y - M(Y) \right] \right\}$, то формулу (4.11) можна подати і в такому вигляді

$$r_в = \frac{\sum_{i=1}^{n} (x^{(i)} - \overline{x})(y^{(i)} - \overline{y})}{n \sigma_x \sigma_y}. \qquad (4.11^*)$$

Оскільки $r_в$ для нефіксованої вибірки є випадковою величиною, то знайдене значення $r_в$ для конкретної вибірки може суттєво відрізнятись від r. В зв’язку із цим необхідно побудувати довірчі інтервали для оцінки r, а також перевірити статистичну гіпотезу про некорельованість X та Y $(r = o)$.

У випадку нелінійності хоча б однієї із функцій регресії $g(x)$ та $q(y)$ коефіцієнт кореляції r вже не дає інформації про силу зв’язку між X та Y. Натомість ступінь концентрації розподілу поблизу лінії регресії показує **кореляційне відношення** Y на X:

$$\eta_{Y/X} = 1 - \frac{\sigma^2_{Y/X}}{\sigma^2_Y}, \qquad (4.14)$$

де

$$\sigma^2_{Y/X} = D(Y \mid X = x) = M\left\{ \left[Y - g(x) \right]^2 \mid X = x \right\}. \qquad (4.15)$$

Із означення випливає, що кореляційне відношення змінюється в межах від 0 до 1 включно; воно дорівнює 1 тоді і тільки тоді, коли

$\sigma^2_{Y/X} = 0,$ тобто весь розподіл зосереджений на лінії регресії (має місце функціональна залежність). Це відношення дорівнює нулю тоді і тільки тоді, коли лінія регресій *Y* на *X* являє собою горизонтальну пряму, що проходить через центр розподілу, тобто якщо *Y* та *X* некорельовані.

Аналогічно вводиться означення кореляційного відношення *X* на *Y* $\eta_{X/Y}$.

Можна довести, що у всіх випадках виконуються нерівності

$$r^2 \le \eta^2_{X/Y}, \quad r^2 \le \eta^2_{Y/X}. \tag{4.16}$$

За статистичними даними двовимірної вибірки можна знайти вибіркові кореляційні відношення $\hat{\eta}_{Y/X}$ та $\hat{\eta}_{X/Y}$, які є точковими статистичними оцінками $\eta_{Y/X}$ та $\eta_{X/Y}$ відповідно.

Сукупність методів оцінки кореляційних характеристик і перевірка статистичних гіпотез про них за даними вибірки називається **кореляційний аналізом**. В кореляційному аналізі використовуються такі основні методи:

1) побудова кореляційного поля і складання кореляційної таблиці;

2) знаходження вибіркового (емпіричного) коефіцієнта кореляції або кореляційного відношення;

3) перевірка статистичних гіпотез про значущість (істотність) зв’язку.

В практично важливих задачах одна із випадкових величин є “результуючою”, а інша — від якої вона залежить (факторна величина). Наприклад, залежність урожайності сільськогосподарських культур *Y* від маси внесених добрив *X*. Це зразок односторонньої залежності між випадковими величинами, яка досліджується в регресійному аналізі.

Більш того, факторна величина може бути **детермінованою** (невипадковою), тобто значення якої не тільки відоме, але й ним можна керувати. При дослідженні лінійної залежності результуючої величини *Y* від факторної величини *X* загальний вираз емпіричного рівняння регресії має такий вид

$$\bar{y} = a_0 + a_1 x. \tag{4.17}$$

В загальному випадку емпіричне рівняння регресії визначається таким чином

$$\bar{y} = \hat{g}(x; a_0, a_1, ..., a_m), \tag{4.18}$$

де $a_0, a_1, ..., a_m$ — невідомі параметри.

Тепер, коли визначені об’єкти дослідження цього параграфа (рівняння регресії (4.6), (4.7), (4.17), (4.18)), можемо сформулювати основні задачі регресійного аналізу.

Регресійний аналіз — це дослідження односторонніх статистичних залежностей між випадковими величинами. При цьому деякі із факторних величин можуть бути невипадковими величинами.

Задачі регресійного аналізу:

1) визначення форми залежностей;

2) знаходження функції регресії;

3) побудова точкових та інтервальних оцінок параметрів функції регресії;

4) знаходження точкових та інтервальних оцінок умовних математичних сподівань, необхідних для визначення меж, в яких із заданою надійністю будуть міститися середні значення досліджуваної величини, якщо інші пов’язані з нею величини набирають певних значень;

5) перевірка узгодженості знайденої емпіричної функції регресії спостереженим даним.

Основна мета регресійного аналізу — **теоретично обґрунтований і статистично надійний точковий та інтервальний прогноз** значень залежної величини або умовного математичного сподівання цієї величини.

Зауваження. Якщо рівняння регресії описує об’єкт дослідження із економічної сфери і воно обґрунтоване в теоретично-економічному сенсі, то його називають **економетричним рівнянням**.

Лінійні емпіричні рівняння парної кореляції. Вибірковий коефіцієнт лінійної кореляції та його властивості

Найбільш простим випадком є той, коли **обидві функції регресії** $g(x)$ і $q(y)$ в рівняннях регресії (4.1) та (4.2) є лінійними, тобто обидві лінії регресії є прямими лініями; вони називаються **прямими регресії**. В цьому випадку будемо говорити про лінійну кореляцію між випадковими величинами X та Y.

Можна довести, що коли сумісний розподіл імовірностей величин X і Y є **нормальним розподілом** на площині (див. [7, п. 7.8]), тоді **кореляційний зв’язок завжди є лінійним**. В зв’язку із цим слід очікувати лінійного кореляційного зв’язку між статистично залежними випадковими величинами X та Y, якщо кожну із них можна розглядати як суму великого числа незалежних або майже незалежних випадкових доданків.

Нехай конфігурація кореляційного поля, отриманого внаслідок зображення у вигляді точки в прямокутній декартовій системі координат $x0y$ кожної із пар чисел вибіркової послідовності (4.4), дозволяє висунути припущення про лінійну кореляційну залежність між X та Y. Тобто рівняння регресії (4.1) та (4.2) в цьому випадку набирають відповідно такого виду

$$M(Y \mid X = x) = \alpha_1 x + \alpha_0, \tag{4.19}$$

$$M(X \mid Y = y) = \beta_1 y + \beta_0, \tag{4.20}$$

де α_i, β_i $(i = 1,0)$ — сталі.

Тоді емпіричні рівняння регресії Y на X та X на Y будемо шукати у вигляді

$$\overline{y}_x = a_1 x + a_0, \tag{4.21}$$

$$\overline{x}_y = b_1 y + b_0, \tag{4.22}$$

де a_0, a_1 та b_0, b_1 — невідомі параметри, які є **точковими статистичними оцінками** відповідних **чисел** рівнянь (4.19) і (4.20).

Використавши метод найменших квадратів як для незгрупованих даних(4.4), так і для згрупованих даних з табл.4.1, можна отримати **систему нормальних рівнянь**

$$\begin{cases} \overline{x^2} a_1 + \overline{x} a_0 = \overline{xy}, \\ \overline{x} a_1 + a_0 = \overline{y}, \end{cases} \tag{4.23}$$

для знаходження параметрів рівняння регресії (Y на X) (4.21).

Система (4.23) має єдиний розв’язок

$$a_1 = \frac{\overline{xy} - \overline{x} \cdot \overline{y}}{\overline{x^2} - (\overline{x})^2}, \quad a_0 = \overline{y} - \overline{x} a_1. \tag{4.24}$$

Коефіцієнт a_1 рівняння (4.21) емпіричної прямої регресії Y на X називається **вибірковим (емпіричним або статистичним) коефіцієнтом регресії *Y* на *X*** і позначається $a_{Y/X}$. Оскільки $\overline{x^2} - (\overline{x})^2 = \sigma_x^2$, а

$$\overline{xy} - \overline{x} \cdot \overline{y} = K^*_{XY} (= \overline{\operatorname{cov}}(X, Y)) \tag{4.25}$$

— **вибірковий (емпіричний або статистичний) кореляційний момент** або **вибіркова коваріація**, то першу формулу (4.24) можна записати в такому виді

$$a_{Y/X} = a_1 = \frac{K^*_{XY}}{\sigma^2_X}\left(\equiv \frac{\overline{\mathrm{cov}}(X,Y)}{\sigma^2_X}\right), \qquad (4.26)$$

а шукане рівняння (4.21) з урахуванням другої рівності (4.25) набуде вигляду

$$\overline{y}_x - \overline{y} = \frac{K^*_{XY}}{\sigma^2_X}(x - \overline{x}). \qquad (4.27)$$

Це рівняння показує, що емпірична пряма регресії Y на X проходить через точку з координатами $(\overline{x}, \overline{y})$, яка називається **середньою точкою кореляційного поля**.

Аналогічно можна отримати емпіричне рівняння прямої X на Y, якщо мінімізувати сумарні квадрати відхилень точок $(\overline{x}_i, y_i)$, $i = \overline{1,m}$, від шуканої прямої, тобто прямої

$$\overline{x}_y - \overline{x} = b_{X/Y}(y - \overline{y}), \qquad (4.28)$$

де вибірковий (емпіричний або статистичний) коефіцієнт регресії X на Y визначається за формулою

$$b_{X/Y} = \frac{\overline{xy} - \overline{x}\cdot\overline{y}}{\sigma^2_y} = \frac{K^*_{XY}}{\sigma^2_y}\left(\equiv \frac{\overline{\mathrm{cov}}(X,Y)}{\sigma^2_y}\right). \qquad (4.29)$$

У формулах вибіркових коефіцієнтів регресії (4.26) і (4.29) чисельники співпадають, а знаменники завжди додатні, оскільки є дисперсіями вибірковими випадкових величин X та Y відповідно. Тому $a_{Y/X}$ і $b_{X/Y}$ мають однакові знаки. Відмітимо також, що вибірковий коефіцієнт регресії $a_{Y/X}\,(b_{X/Y})$ — це міра, яка на основі вибіркових даних в середньому вказує на вплив зміни змінної X (або Y) на змінну Y (або X).

Рівністю (4.11) **формально** був визначений вибірковий (емпіричний) коефіцієнт кореляції $r_в$ як точкова статистична оцінка коефіцієнта кореляції $r = r_{XY}$ — міри лінійного кореляційного зв'язку між випадковими величинами X та Y. Проте вид емпіричних коефіцієнтів регресії Y на X та X на Y вказує на природний зв'язок $r_в$ із $a_{Y/X}$ і $b_{X/Y}$ (за аналогією із зв'язком в теорії кореляції як розділу теорії імовірностей). А тому виникає необхідність детально вивчити властивості $r_в$ як **оцінки** сили лінійного кореляційного зв'язку між величинами X та Y. При цьому слід очікувати властивостей, аналогічних із властивостями r_{XY}.

Вибірковим (емпіричним або статистичним) коефіцієнтом кореляції $r_в = r_{XY}$ **випадкових величини** X **та** Y, **між якими припускається лінійний кореляційний зв'язок, називається відношення емпіричного кореляційного момента (коефіцієнта коваріації)** $K^*(X,Y)(=\overline{\text{cov}}(X,Y))$ **до добутку середніх квадратичних відхилень вибіркових** σ_x **та** σ_y :

$$r_в = r_{xy} = \frac{K^*_{XY}}{\sigma_x \sigma_y}\left(= \frac{\overline{\text{cov}}(X,Y)}{\sigma_x \sigma_y}\right) = \frac{\overline{xy} - \overline{x} \cdot \overline{y}}{\sigma_x \sigma_y}, \qquad (4.30)$$

де $\overline{xy}$ для незгрупованих даних (4.4) вибірки обчислюється за формулою (4.12), а для згрупованих даних із табл. 4.1 — за (4.13).

Зауваження. Для випадку, коли x_i та y_i $(i = \overline{1,k})$ є великими числами, а обсяг вибірки $n \geq 50$, то зручніше користуватися такою формулою

$$r_в = \frac{\sum_{i=1}^{n}(x_i - \overline{x})(y_i - \overline{y})}{\sqrt{\left[\sum_{i=1}^{n}(x_i - \overline{x})^2\right]\left[\sum_{i=1}^{n}(y_i - \overline{y})^2\right]}}. \qquad (4.30^*)$$

Використовуючи формули (4.26) і (4.29), отримаємо вираз через емпіричні коефіцієнти регресій для вибіркового коефіцієнта кореляції:

$$r_в = \pm\sqrt{\frac{(K^*_{XY})^2}{\sigma_x^2 \sigma_y^2}} = \pm\sqrt{\frac{K^*_{XY}}{\sigma_x^2} \cdot \frac{K^*_{XY}}{\sigma_y^2}} = \pm\sqrt{a_{Y/X}\, b_{X/Y}},$$

де знак перед коренем визначається знаком емпіричних коефіцієнтів регресій.

З другого боку, і емпіричні коефіцієнти регресій можна виразити через вибірковий коефіцієнт кореляції:

$$a_{Y/X} = \frac{K^*_{XY}}{\sigma_x^2} = \frac{K^*_{XY}}{\sigma_x \sigma_y} \cdot \frac{\sigma_y}{\sigma_x} = r_в \frac{\sigma_y}{\sigma_x},$$

$$b_{X/Y} = \frac{K^*_{XY}}{\sigma_y^2} = \frac{K^*_{XY}}{\sigma_x \sigma_y} \cdot \frac{\sigma_x}{\sigma_y} = r_в \frac{\sigma_x}{\sigma_y}.$$

Тоді емпіричне рівняння прямих регресій із врахуванням цих формул можна записати в такому вигляді:

$$\overline{y}_x - \overline{y} = r_в \frac{\sigma_y}{\sigma_x}(x - \overline{x}), \qquad (4.31)$$

$$\overline{x}_y - \overline{x} = r_в \frac{\sigma_x}{\sigma_y}(y - \overline{y}). \qquad (4.32)$$

Наведемо еквівалентні (4.30) вирази вибіркового коефіцієнта кореляції.

Нехай $\{(x_i, y_i), i = \overline{1,n}\}$ — незгруповані дані двовимірної вибірки обсягом *n*. Величину

$$D_{yx} = \frac{1}{n}\sum_{i=1}^{n}(y_i - \widehat{y}_i)^2 = \frac{1}{n}\sum_{i=1}^{n}\left[y_i - \overline{y} - r_в \frac{\sigma_y}{\sigma_x}(x_i - \overline{x})\right]^2 \qquad (4.33)$$

природно називати дисперсією спостережених значень y_i навколо емпіричної прямої регресії (4.31) *Y* на *X*, врахувавши при цьому, що геометрично різниця $y_i - \widehat{y}_i$ означає відхилення по ординаті точки (x_i, y_i) від точки $(x_i, \widehat{y}_i)$ прямої (4.31).

Можна довести, що

$$D_{yx} = \sigma_y^2(1 - r_в^2), \quad r_в = \pm\sqrt{1 - \frac{D_{yx}}{\sigma_y^2}}, \qquad (4.30^{**})$$

де знак перед коренем вибирається у відповідності із знаком коефіцієнта регресії.

Аналогічно ввівши дисперсію спостережених значень x_i навколо емпіричної прямої регресії (4.32) *X* на *Y* за формулою

$$D_{xy} = \frac{1}{n}\sum_{i=1}^{n}(x_i - \widehat{x}_i)^2 = \frac{1}{n}\sum_{i=1}^{n}\left[x_i - \overline{x} - r_в \frac{\sigma_x}{\sigma_y}(y_i - \overline{y})\right]^2,$$

можна отримати рівність

$$r_в = \pm\sqrt{1 - \frac{D_{xy}}{\sigma_x^2}}. \qquad (4.30^{***})$$

Властивості вибіркового коефіцієнта кореляції

Властивість 1. Величина $r_в$ є безрозмірною, тобто вона не залежить від вибору одиниць виміру випадкових величин *X* та *Y*.

Властивість 2. **Вибірковий коефіцієнт кореляції $r_в$ не перевищує за абсолютною величиною одиницю,** тобто $|r_в| \le 1$.

Властивість 3. **Вибірковий коефіцієнт кореляції $r_в = \pm 1$ тоді і тільки тоді, коли між випадковими величинами X та Y існує лінійний функціональний зв'язок.**

Властивість 4. **Якщо між випадковими величинами X і Y відсутній хоча б один із кореляційних зв'язків, то вибірковий коефіцієнт кореляції $r_в$ дорівнює нулю.**

Властивість 5. **Рівність $r_в = \pm 1$ є необхідною і достатньою умовою співпадань регресій Y на X і X на Y.**

Із розглянутих властивостей $r_в$ можна зробити висновок про те, що вибірковий коефіцієнт кореляції є мірою тісноти (сили) лінійного кореляційного зв'язку між випадковими величинами X і Y. Справді, якщо $|r_в| = 1,$ то між X і Y існує лінійний функціональний зв'язок, а якщо $r_в = 0,$ то лінійний кореляційний зв'язок відсутній. А із формул (4.30**) і (4.30***) випливає, що у випадку збільшення $|r_в|$ до одиниці сила кореляційного зв'язку зростає, оскільки сума квадратів відхилень спостережених значень від прямих регресій прямує до нуля. Якщо ж $|r_в| \to 0,$ то сила кореляційного зв'язку зменшується, бо сума квадратів відхилень зростає.

Спрощений метод обчислення параметрів прямих регресії за кореляційною таблицею

У випадку рівновіддаленості як варіант x_i ($i = \overline{1,k}$), так і варіант y_j ($j = \overline{1,m}$) кореляційної табл.4.1 параметри рівнянь (4.31) та (4.32) можуть бути обчислені спрощеним методом (за аналогією із обчисленням зведених характеристик вибірки методом добутків). З цією метою здійснюється перехід до умовних варіант u_i та v_j за формулами

$$u_i = \frac{x_i - C_1}{h_1} \ (i = \overline{1,k}), \quad v_j = \frac{y_j - C_2}{h_2} \ (j = \overline{1,m}), \tag{4.34}$$

де $h_1=x_{i+1}-x_i$ ($i=\overline{1,k-1}$), $h_2=y_{j+1}-y_j$ ($j=\overline{1,m-1}$), C_1 та C_2 – хибні нулі варіант x_i та y_j відповідно.Зв'язок між числовими характеристиками початкових та умовних варіант визначається такими формулами:

$$\bar{x}=\bar{u}h_1+C_1,\ \bar{y}=\bar{v}h_2+C_2,\ \sigma_x=\sigma_u h_1,\ \sigma_y=\sigma_v h_2, \tag{4.35}$$

$$r_в=\frac{\frac{1}{n}(\sum n_{uv}uv)-\bar{u}\cdot\bar{v}}{\sigma_u\sigma_v}.$$

Оцінка достовірності емпіричних коефіцієнтів кореляції і регресії за даними вибірки

Будемо вважати, що генеральна сукупність має нормальний розподіл. Тоді для вибірки великого обсягу ($n\geq 50$) довірчі інтервали з надійністю γ при оцінюванні теоретичного коефіцієнта кореляції r, а також теоретичних коефіцієнтів регресії α і β в рівняннях (4.19), (4.20) мають такий вид:

$$r_в-t\frac{1-r_в^2}{\sqrt{n}}<r<r_в+t\frac{1-r_в^2}{\sqrt{n}}, \tag{4.36}$$

$$a_{Y/X}-t\frac{\sigma_y}{\sigma_x}\cdot\frac{1-r_в^2}{\sqrt{n}}<\alpha_1<a_{Y/X}+t\frac{\sigma_y}{\sigma_x}\cdot\frac{1-r_в^2}{\sqrt{n}}, \tag{4.37}$$

$$b_{X/Y}-t\frac{\sigma_x}{\sigma_y}\cdot\frac{1-r_в^2}{\sqrt{n}}<\beta_1<b_{X/Y}+t\frac{\sigma_x}{\sigma_y}\cdot\frac{1-r_в^2}{\sqrt{n}}. \tag{4.38}$$

де t – корінь рівняння $2\Phi(t)=\gamma$, $\Phi(t)$ – функція Лапласа.

Наведеними довірчими інтервалами можна користуватись лише при великих n ($n\geq 50$). Якщо ж обсяг двовимірної вибірки невеликий ($n<30$), то розподіл вибіркового коефіцієнта кореляції суттєво відрізняється від нормального. В цьому випадку для оцінки теоретичного коефіцієнта кореляції використовується знайдене Фішером перетворення, при якому вибірковий коефіцієнт кореляції прирівнюється до гіперболічного тангенса[*)] деякої величини z: $r_в=\text{th}z$, звідки

[*)] Гіперболічний тангенс $y=\text{th}x$ визначається рівністю $y=\text{th}x=(e^x-e^{-x})\big/(e^x+e^{-x})$; відповідна обернена гіперболічна функція має вид $x=\text{arth}y=\frac{1}{2}\ln\frac{1+y}{1-y}$.

$$z = \frac{1}{2}\ln\frac{1+r_в}{1-r_в}.$$

Відомо, що випадкова величина z ($r_в$ розглядається для нефіксованої вибірки) вже при невеликих значеннях n $(n > 10)$ розподілена приблизно нормально, причому

$$M(z) \approx \frac{1}{2}\ln\frac{1+r}{1-r} + \frac{r}{2(n-1)}, \qquad D(z) \approx \frac{1}{n-3}.$$

Довірчий інтервал з надійністю γ для теоретичного коефіцієнта кореляції r двовимірної нормально розподіленої генеральної сукупності для невеликих обсягів $(n < 30)$ має такий вид

$$\left(\operatorname{th}(z - t/\sqrt{n-3});\ \operatorname{th}(z + t/\sqrt{n-3}) \right), \tag{4.39}$$

де $z = \frac{1}{2}\ln\frac{1+r_в}{1-r_в}$ і знаходиться за табл. 10 додатків перетворення Фішера (z-перетворення), t – корінь рівняння $2\Phi(t) = \gamma$, th(...) знаходиться за табл. 11 додатків оберненого перетворення Фішера.

Для коефіцієнтів регресій при невеликих n $(n < 30)$ розглянемо статистики (випадкові величини)

$$T_1 = \frac{\sigma_x\sqrt{n-2}}{\sigma_y\sqrt{1-r_в^2}}(a_{Y/X} - \alpha_1), \qquad T_2 = \frac{\sigma_y\sqrt{n-2}}{\sigma_x\sqrt{1-r_в^2}}(b_{X/Y} - \beta_1).$$

Можна довести, що у випадку нормально розподіленої вибірки кожна із них має розподіл Ст'юдента із $k = n-2$ ступенями вільності. Довірчі інтервали для теоретичних коефіцієнтів регресії α_1 та β_1 будуються стандартним чином:

$$a_{Y/X} - t_\gamma \frac{\sigma_y\sqrt{1-r_в^2}}{\sigma_x\sqrt{n-2}} < \alpha_1 < a_{Y/X} + t_\gamma \frac{\sigma_y\sqrt{1-r_в^2}}{\sigma_x\sqrt{n-2}}, \tag{4.40}$$

$$b_{X/Y} - t_\gamma \frac{\sigma_x\sqrt{1-r_в^2}}{\sigma_y\sqrt{n-2}} < \beta_1 < b_{X/Y} + t_\gamma \frac{\sigma_x\sqrt{1-r_в^2}}{\sigma_y\sqrt{n-2}}, \tag{4.41}$$

де $t_\gamma = t(\gamma, k)$ знаходиться за табл. 4 додатків розподілу Ст'юдента з числом ступенів вільності $k = n-2$.

Завершуючи розгляд статистичного оцінювання коефіцієнтів теоретичних прямих регресій, вкажемо довірчі інтервали для вільних членів цих прямих:

$$\bar{y} - t_\gamma \sigma_y \sqrt{\frac{(n-1)(1-r_в^2)}{n(n-2)}} < \alpha_0 < \bar{y} + t_\gamma \sigma_y \sqrt{\frac{(n-1)(1-r_в^2)}{n(n-2)}}, \qquad (4.42)$$

$$\bar{x} - t_\gamma \sigma_x \sqrt{\frac{(n-1)(1-r_в^2)}{n(n-2)}} < \beta_0 < \bar{x} + t_\gamma \sigma_x \sqrt{\frac{(n-1)(1-r_в^2)}{n(n-2)}}, \qquad (4.43)$$

де $t_\gamma = t(\gamma, k)$ знаходиться за табл. 4 додатків розподілу Ст'юдента з числом ступенів вільності $k = n - 2$.

Відмітимо, що довірчі інтервали (4.42), (4.43) дають добрі наближення як у випадку малих вибірок $(n < 30)$, так і великих, оскільки при зростанні n розподіл Ст'юдента наближається до нормального.

При кореляційному аналізі необхідно оцінити достовірність стохастичного зв'язку між випадковими величинами, тобто з'ясувати, чи не зумовлена величина отриманого вибіркового коефіцієнта кореляції випадковим характером утворення вибірки. Для цього оцінюється значущість або істотність $r_в$. Перевіряється гіпотеза $H_0 : r = 0$ (гіпотеза некорельованості величин) проти однієї із альтернативних гіпотез $H_1 : r \neq 0,$ або $H_1 : r < 0,$ або $H_1 : r > 0.$

Для великих n критерієм перевірки слугує вибірковий коефіцієнт кореляції (для нефіксованої вибірки!). При цьому критична область має вид

$$|r_в| > u_\alpha \frac{1 - r_в^2}{\sqrt{n}}, \qquad (4.44)$$

де u_α – значення, знайдене за табл. 3 додатків нормального розподілу для рівня значущості α. Якщо спостережене значення $r_в$ (для конкретної вибірки) потрапить в цю область, то гіпотезу H_0 відкидають.

Для невеликих n $(n < 30)$ використовується статистика

$$T = \frac{r_в \sqrt{n-2}}{\sqrt{1 - r_в^2}}, \qquad (4.45)$$

яка має розподіл Ст'юдента із $k = n - 2$ ступенів вільності. Якщо $T_{спост} > t_\alpha(n-2),$ де $t_\alpha(n-2) = t_{двост.кр}(\alpha; n-2)$ — критична точка розподілу Ст'юдента, що знаходиться за табл. 5 додатків для двосторонньої критичної області (рівень значущості α вибирається за верхнім рядком), тоді основну гіпотезу про відсутність кореляційної залежності випадкових величин слід відкинути.

Отже, **якщо основна гіпотеза $H_0 : r = 0$ відкидається, тоді для оцінки точності $r_в$ будуються довірчі інтервали**.

Для перевірки гіпотези про силу кореляційного зв'язку, тобто для перевірки основної гіпотези $H_0 : r = r_0$ при альтернативній гіпотезі $H_1 : r \neq r_0$, використовується статистика

$$U = (z - z_0)\sqrt{n-3}, \qquad (4.46)$$

де

$$z_0 = \frac{1}{2}\ln\frac{1+r_0}{1-r_0} + \frac{r_0}{2(n-1)}, \quad z = \frac{1}{2}\ln\frac{1+r_в}{1-r_в}. \qquad (4.47)$$

Якщо гіпотеза H_0 є правильною, тоді випадкова величина U розподілена асимптотично нормально з параметрами (0,1). Тому використовувати цю статистику можна для немалих обсягів вибірки.

При використанні (двостороннього) критерія на рівні значущості α гіпотеза H_0 відкидається, якщо

$$U_{\text{спост}} > u_{\text{кр}},$$

де $u_{\text{кр}}$ — корінь рівняння

$$\Phi(u_{кр}) = (1-\alpha)/2. \qquad (4.48)$$

Зауваження 1. Відмітимо, що стосовно перевірки гіпотези $H_0 : r = 0$ **зберігається умова про нормальність розподілу двовимірної генеральної сукупності**.

Зауваження 2. Побудову довірчих інтервалів для теоретичних коефіцієнтів кореляції та регресії слід здійснювати **після** того, як статистичну гіпотезу про некорельованість випадкових величин відкинуто на підставі даних вибірки.

Нелінійна парна кореляція.
Вибіркове кореляційне відношення та його властивості

Розглянемо тепер випадок, коли хоча б одна із двох послідовностей точок

$$(x_1, \bar{y}_{x_1}), (x_2, \bar{y}_{x_2}), \ldots, (x_k, \bar{y}_{x_k}), \qquad (4.49)$$

$$(y_1, \bar{x}_{y_1}), (y_2, \bar{x}_{y_2}), \ldots, (y_m, \bar{x}_{y_m}) \qquad (4.50)$$

дає підстави зробити висновок про існування нелінійного кореляційного зв'язку між випадковими величинами X та Y. На основі цих статистичних даних потрібно оцінити параметри нелінійної регресії та силу кореляційної залежності.

Нехай, наприклад, конфігурація точок (4.49) дозволяє зробити припущення про наявність параболічної кореляції другого порядку. В цьому випадку вибіркове рівняння регресії *Y* на *X* слід шукати в такому вигляді

$$\overline{y}_x = a_2 x^2 + a_1 x + a_0, \tag{4.51}$$

де $a_i \, (i = 0,1,2)$ — невідомі параметри.

Користуючись методом найменших квадратів, можна отримати систему нормальних рівнянь:

$$\begin{cases} \overline{x^4}a_2 + \overline{x^3}a_1 + \overline{x^2}a_0 = \overline{\overline{y}_x x^2}, \\ \overline{x^3}a_2 + \overline{x^2}a_1 + \overline{x}a_0 = \overline{\overline{y}_x x}, \\ \overline{x^2}a_2 + \overline{x}a_1 + a_0 = \overline{\overline{y}_x}, \end{cases} \tag{4.52}$$

де

$$\overline{x^l} = \left(\sum_{i=1}^{k} x_i^l n_{x_i} \right) \Big/ n, \quad l = \overline{1,4}, \tag{4.53}$$

$$\overline{\overline{y}_x x^l} = \left(\sum_{i=1}^{k} \overline{y}_{x_i} x_i^l n_{x_i} \right) \Big/ n, \quad l = \overline{0,2}.$$

Оскільки $\overline{\overline{y}_x} = \overline{y}$, то розв'язавши третє рівняння системи (4.52) відносно a_0 і підставивши в (4.51), після простих перетворень отримаємо

$$\overline{y}_x = \overline{y} + a_1(x - \overline{x}) + a_2(x^2 - \overline{x^2}). \tag{4.51*}$$

Знайдені із системи рівнянь (4.52) параметри a_1 та a_2 підставимо в (4.51*); в підсумку отримується шукане рівняння регресії *Y* на *X*.

У випадку параболічної регресії (другого порядку) *X* на *Y*

$$\overline{x}_y = b_2 y^2 + b_1 y + b_0$$

невідомі параметри b_2, b_1, b_0 знаходяться як розв'язок такої системи нормальних рівнянь

$$\begin{cases} \overline{y^4}b_2 + \overline{y^3}b_1 + \overline{y^2}b_0 = \overline{\overline{x}_y y^2}, \\ \overline{y^3}b_2 + \overline{y^2}b_1 + \overline{y}b_0 = \overline{\overline{x}_y y}, \\ \overline{y^2}b_2 + \overline{y}b_1 + b_0 = \overline{x}. \end{cases}$$

Вище було розглянуто вибірковий коефіцієнт кореляції $r_в$, який характеризує тісноту **лінійного** кореляційного зв'язку між випадковими величинами *X* та *Y*.

Якщо ж криві регресії відрізняються від прямолінійної форми, то в якості міри зв'язку за даними двовимірної вибірки використовується вибіркове (статистичне або емпіричне) кореляційне відношення, запропоноване Пірсоном.

Нехай дані спостережень над кількісними ознаками *X* та *Y* зведені в кореляційну табл. 4.1. Тим самим можна вважати спостережені значення *Y* розбитими на групи, при цьому кожна група містить ті значення *Y*, які відповідають конкретним значенням *X*.

Для регресії *Y* на *X* можна ввести відповідності

$$\begin{array}{ccc} x_1 & \overline{y}_{x_1} & n_{x_1}, \\ x_2 & \overline{y}_{x_2} & n_{x_2}, \\ \dots & \dots & \dots \\ x_k & \overline{y}_{x_k} & n_{x_k}, \end{array}$$

а для регресії *X* на *Y* відповідності

$$\begin{array}{ccc} y_1 & \overline{x}_{y_1} & n_{y_1}, \\ y_2 & \overline{x}_{y_2} & n_{y_2}, \\ \dots & \dots & \dots \\ y_m & \overline{x}_{y_m} & n_{y_m}, \end{array}$$

де $\sum_{i=1}^{k} n_{x_i} = n, \quad \sum_{i=1}^{m} n_{y_j} = n.$

Оскільки всі значення ознаки $Y(X)$ розбиті на групи, то дисперсію (загальну) ознаки $Y(X)$ можна представити у вигляді суми внутрігрупової і міжгрупової дисперсій (див. теорему 1.2):

$$D_{y\ \text{заг}} = D_{y\ \text{внгр}} + D_{y\ \text{міжгр}}, \tag{4.54}$$

$$D_{x\ \text{заг}} = D_{x\ \text{внгр}} + D_{x\ \text{міжгр}}. \tag{4.55}$$

Вибірковим (статистичним чи емпіричним) кореляційним відношенням *Y* до *X* називається відношення міжгрупового середнього квадратичного відхилення до загального середнього квадратичного відхилення ознаки *Y*:

$$\eta_{yx} = \sigma_{y\ \text{міжгр}} / \sigma_{y\ \text{заг}} \tag{4.56}$$

або

$$\eta_{yx} = \sigma_{\bar{y}_x} / \sigma_y, \qquad (4.56^*)$$

де

$$\sigma_{\bar{y}_x} = \sqrt{D_{y \text{ міжгр}}} = \sqrt{\sum_i n_{x_i} (\bar{y}_{x_i} - \bar{y})^2 / n}, \qquad (4.57)$$

$$\sigma_y = \sqrt{D_{y \text{ заг}}} = \sqrt{\sum_j n_{y_j} (y_j - \bar{y})^2 / n},$$

де n —обсяг вибірки, n_{x_i} — частота значення x_i ознаки *X*, n_{y_i} — частота значення y_j ознаки *Y*, $\bar{y}$ — загальна середня ознаки *Y*, $\bar{y}_{x_i}$ — умовна середня ознаки *Y*, що відповідає значенню x_i ознаки *X*.

Аналогічно визначається вибіркове кореляційне відношення *X* до *Y*:

$$\eta_{xy} = \sigma_{\bar{x}_y} / \sigma_x .$$

Зміст вибіркових кореляційних відношень визначається такими властивостями.

Властивість 1. Вибіркові кореляційні відношення невід'ємні і не перевищують одиниці:

$$0 \le \eta_{yx} \le 1, \quad 0 \le \eta_{xy} \le 1.$$

Властивість 2. Рівність нулю вибіркового кореляційного відношення є необхідною і достатньою умовою відсутності кореляційного зв'язку між випадковими величинами *X* та *Y*.

Властивість 3. Рівність одиниці вибіркового кореляційного відношення є необхідною і достатньою умовою функціональної залежності між випадковими величинами *X* та *Y*.

Властивість 4. Для однієї і тієї ж двовимірної вибірки випадкові кореляційні відношення не менші від абсолютної величини вибіркового коефіцієнта кореляції:

$$\eta_{yx} \ge |r_в|, \quad \eta_{xy} \ge |r_в|.$$

Властивість 5. Для того, щоб регресія була точно лінійною, необхідно і достатньо, щоб вибіркові кореляційні

відношення дорівнювали вибірковому коефіцієнту кореляції:

$$\eta_{yx} = |r_в|, \quad \eta_{xy} = |r_в|.$$

Розглянуті властивості розкривають зміст вибіркових кореляційних відношень, які є мірою тісноти (сили) відповідного кореляційного зв'язку між випадковими величинами *X* та *Y*, причому форма зв'язку може бути довільною. Справді, якщо $\eta_{yx} \to 1$ або $\eta_{xy} \to 1,$ то відповідний зв'язок випадкових величин стає більш тісним, оскільки міжгрупова дисперсія (дисперсія умовних середніх) прямує до загальної дисперсії ознаки *Y* або *X*, тобто точки $(x_i, \overline{y}_{x_i}),\quad i = \overline{1,k},$ $\left((\overline{x}_{y_j}, y_j)_{j=\overline{1,m}} \right)$ наближаються до відповідної вибіркової кривої регресії. Відмітимо також, що **недоліком вибіркових кореляційних відношень** є те, що вони не дозволяють судити, наскільки близько розташовані спостережені точки до кривої певного виду, наприклад, до параболи, гіперболи тощо. Це пояснюється тим, що при визначенні кореляційних відношень **конкретна** форма зв'язку не бралася до уваги.

Зупинимося тепер детальніше на деяких практично важливих видах нелінійної кореляційної залежності між *X* та *Y*. При цьому обмежимося розглядом односторонньої залежності.

За методикою оцінок параметрів парні нелінійні регресії розподіляють на два види: 1) нелінійні за факторами, але лінійні за невідомими параметрами, які підлягають оцінці; 2) нелінійні за факторами і параметрами. Регресії, нелінійні за факторами, але лінійні за оцінюваними параметрами, називаються **квазілінійними**.

Парну квазілінійну регресію можна записати в такому загальному вигляді: $\overline{y}_x = a\varphi(x) + b.$ Заміною $z = \varphi(x)$ квазілійна парна регресія приводиться до лінійної парної регресії $\overline{y}_z = az + b.$ Формули для оцінок параметрів набувають вигляду

$$\widehat{a} = \frac{n\sum_{i=1}^{n} z_i y_i - \sum_{i=1}^{n} z_i \sum_{i=1}^{n} y_i}{n\sum_{i=1}^{n} z_i^2 - \left(\sum_{i=1}^{n} z_i \right)^2}, \qquad \widehat{b} = \overline{y} - \widehat{a}\overline{z}.$$

В табл. 4.2 наведено ряд парних квазілінійних регресій.

Таблиця 4.2

Квазілінійна регресія	Заміна змінної	Лінійна регресія	Формули для оцінок параметрів
1	2	3	4
$\overline{y}_x = a/x + b$	$1/x = z$	$\overline{y}_z = az + b$	$\widehat{a} = \dfrac{n\sum_{i=1}^{n} y_i/x_i - \sum_{i=1}^{n} 1/x_i \sum_{i=1}^{n} y_i}{n\sum_{i=1}^{n} 1/x_i^2 - \left(\sum_{i=1}^{n} 1/x_i\right)^2},$ $\widehat{b} = \overline{y} - \widehat{a}\overline{z},$ де $\overline{z} = \left(\sum_{i=1}^{n} 1/x_i\right)\Big/n.$
$\overline{y}_x = a\ln x + b$	$\ln x = z$	$\overline{y}_z = az + b$	$\widehat{a} = \dfrac{n\sum_{i=1}^{n} y_i \ln x_i - \sum_{i=1}^{n} \ln x_i \sum_{i=1}^{n} y_i}{n\sum_{i=1}^{n} \ln^2 x_i - \left(\sum_{i=1}^{n} \ln x_i\right)^2},$ $\widehat{b} = \overline{y} - \widehat{a}\overline{z},$ де $\overline{z} = \left(\sum_{i=1}^{n} \ln x_i\right)\Big/n.$
$\overline{y}_x = ae^x + b$	$e^x = z$	$\overline{y}_z = az + b$	$\widehat{a} = \dfrac{n\sum_{i=1}^{n} y_i e^{x_i} - \sum_{i=1}^{n} e^{x_i} \sum_{i=1}^{n} y_i}{n\sum_{i=1}^{n} e^{2x_i} - \left(\sum_{i=1}^{n} e^{x_i}\right)^2},$ $\widehat{b} = \overline{y} - \widehat{a}\overline{z},$ де $\overline{z} = \left(\sum_{i=1}^{n} e^{x_i}\right)\Big/n.$
$\overline{y}_x = a\sqrt{x} + b$	$\sqrt{x} = z$	$\overline{y}_z = az + b$	$\widehat{a} = \dfrac{n\sum_{i=1}^{n} y_i \sqrt{x_i} - \sum_{i=1}^{n} \sqrt{x_i} \sum_{i=1}^{n} y_i}{n\sum_{i=1}^{n} x_i - \left(\sum_{i=1}^{n} \sqrt{x_i}\right)^2},$ $\widehat{b} = \overline{y} - \widehat{a}\left(\sum_{i=1}^{n} \sqrt{x_i}\right)\Big/n.$
$\overline{y}_x = ax^2 + b$	$x^2 = z$	$\overline{y}_z = az + b$	$\widehat{a} = \dfrac{n\sum_{i=1}^{n} x_i^2 y_i - \sum_{i=1}^{n} x_i^2 \sum_{i=1}^{n} y_i}{n\sum_{i=1}^{n} x_i^4 - \left(\sum_{i=1}^{n} x_i^2\right)^2},$

Продовження табл. 4.2

			$\hat{b} = \bar{y} - \hat{a}\left(\sum_{i=1}^{n} x_i^2\right)\Big/ n.$
$\bar{y}_x = ax^3 + b$	$x^3 = z$	$\bar{y}_z = az + b$	$\hat{a} = \dfrac{n\sum_{i=1}^{n} x_i^3 y_i - \sum_{i=1}^{n} x_i^3 \sum_{i=1}^{n} y_i}{n\sum_{i=1}^{n} x_i^6 - \left(\sum_{i=1}^{n} x_i^3\right)^2},$ $\hat{b} = \bar{y} - \hat{a}\left(\sum_{i=1}^{n} x_i^3\right)\Big/ n.$

Розглянемо тепер регресію, нелінійну за параметрами, виду $\bar{y}_x = 1\big/(ax+b)$. Заміною $y^{(1)} = 1\big/ y$ вона приводиться до лінійної регресії $\bar{y}_x^{(1)} = ax + b.$ Оцінки параметрів a та b знаходяться за формулами

$$\hat{a} = \frac{n\sum_{i=1}^{n} x_i / y_i - \sum_{i=1}^{n} x_i \sum_{i=1}^{n} 1\big/ y_i}{n\sum_{i=1}^{n} x_i^2 - \left(\sum_{i=1}^{n} x_i\right)^2}, \qquad \hat{b} = \left(\sum_{i=1}^{n} 1\big/ y_i\right)\Big/ n - \hat{a}\left(\sum_{i=1}^{n} x_i\right)\Big/ n.$$

Більш складною (з точки зору заміни змінних) є регресія виду $\bar{y}_x = 1\big/(ae^{-x} + b)$. Для приведення її до лінійної регресії потрібно зробити заміну обох величин: $y^{(1)} = 1\big/ y,\quad z = e^{-x}$. Тоді отримається лінійна регресія $\bar{y}_z^{(1)} = az + b,$ оцінки параметрів якої знаходяться за формулами

$$\hat{a} = \frac{n\sum_{i=1}^{n} 1\big/ y_i e^{x_i} - \sum_{i=1}^{n} 1\big/ e^{x_i} \sum_{i=1}^{n} 1\big/ y_i}{n\sum_{i=1}^{n} 1\big/ e^{2x_i} - \left(\sum_{i=1}^{n} 1\big/ e^{x_i}\right)^2}, \qquad \hat{b} = \left(\sum_{i=1}^{n} 1\big/ y_i\right)\Big/ n - \hat{a}\left(\sum_{i=1}^{n} 1\big/ e^{x_i}\right)\Big/ n.$$

Розглянемо регресію $\bar{y}_x = a^x b.$ Якщо $a > 0,$ $b > 0,$ то прологарифмувавши ліву і праву частини рівняння, отримаємо

$$\ln \bar{y}_x = x \ln a + \ln b.$$

Після заміни $\bar{y}_x^{(1)} = \ln \bar{y}_x,$ $a_1 = \ln a,$ $b_1 = \ln b$ одержується лінійна парна регресія $\bar{y}_x^{(1)} = a_1 x + b_1.$ Для неї знаходяться оцінки параметрів a_1 і b_1 :

$$\widehat{a}_1 = \frac{n\sum_{i=1}^{n} x_i \ln y_i - \sum_{i=1}^{n} x_i \sum_{i=1}^{n} \ln y_i}{n\sum_{i=1}^{n} x_i^2 - \left(\sum_{i=1}^{n} x_i\right)^2}, \quad \widehat{b}_1 = \left(\sum_{i=1}^{n} \ln y_i\right)\Big/ n - \widehat{a}\left(\sum_{i=1}^{n} x_i\right)\Big/ n.$$

Тоді оцінки параметрів вихідної регресії матимуть такий вид: $\widehat{a} = e^{\widehat{a}_1}$, $\widehat{b} = e^{\widehat{b}_1}$.

Для регресії $\overline{y}_x = e^{a/x} b$ після логарифмування отримаємо

$$\ln \overline{y}_x = \frac{a}{x} + \ln b.$$

Внаслідок виконання заміни $y^{(1)} = \ln y$, $z = 1/x$, $b_1 = \ln b$ приходимо до парної лінійної регресії $\overline{y}_z^{(1)} = az + b_1$, оцінки параметрів якої знаходяться за формулами

$$\widehat{a} = \frac{n\sum_{i=1}^{n} \frac{\ln y_i}{x_i} - \sum_{i=1}^{n} \frac{1}{x_i} \sum_{i=1}^{n} \ln y_i}{n\sum_{i=1}^{n} \frac{1}{x_i^2} - \left(\sum_{i=1}^{n} \frac{1}{x_i}\right)^2}, \quad \widehat{b}_1 = \frac{\sum_{i=1}^{n} \ln y_i}{n} - \widehat{a}\frac{\sum_{i=1}^{n} \frac{1}{x_i}}{n}.$$

Тоді $\widehat{b} = e^{\widehat{b}_1}$.

Розглянемо регресію $\overline{y}_x = x^a b$, $b > 0$. Після логарифмування рівняння і заміни отримаємо лінійну парну регресію

$$\overline{y}_z^{(1)} = az + b_1, \text{ де } y^{(1)} = \ln y, \ z = \ln x, \ b_1 = \ln b.$$

Оцінки параметрів a і b_1 знаходяться за формулами

$$\widehat{a} = \frac{n\sum_{i=1}^{n} \ln x_i \ln y_i - \sum_{i=1}^{n} \ln x_i \sum_{i=1}^{n} \ln y_i}{n\sum_{i=1}^{n} \ln^2 x_i - \left(\sum_{i=1}^{n} \ln x_i\right)^2},$$

$$\widehat{b}_1 = \left(\sum_{i=1}^{n} \ln y_i\right)\Big/ n - \widehat{a}\left(\sum_{i=1}^{n} \ln x_i\right)\Big/ n, \quad x_i > 0, \quad y_i > 0 \quad (i = \overline{1,n}).$$

Оцінка параметра b знаходиться таким чином

$$\widehat{b} = e^{\widehat{b}_1}, \text{ де } \widehat{b}_1 > 0.$$

Для показниково-степеневої парної регресії $\overline{y}_x = x^{ax} b$ $(x > 0, b > 0)$ після логарифмування і заміни отримаємо $\overline{y}_z^{(1)} = az + b_1$, де $y^{(1)} = \ln y, \quad z = x\ln x, \quad b_1 = \ln b.$ Параметри a і b_1 оцінюються за формулами

$$\hat{a} = \frac{n\sum_{i=1}^{n} x_i \ln x_i \ln y_i - \sum_{i=1}^{n} x_i \ln x_i \sum_{i=1}^{n} \ln y_i}{n\sum_{i=1}^{n} x_i^2 \ln^2 x_i - \left(\sum_{i=1}^{n} x_i \ln x_i\right)^2},$$

$$\hat{b}_1 = \left(\sum_{i=1}^{n} \ln y_i\right)\Big/ n - \hat{a}\left(\sum_{i=1}^{n} x_i \ln x_i\right)\Big/ n, \quad x_i > 0, \quad y_i > 0 \quad (i = \overline{1,n}).$$

Тоді оцінка параметра b визначається за формулою $\hat{b} = e^{\hat{b}_1}$.

Нарешті розглянемо нелінійну регресію $\bar{y}_x = \dfrac{x}{ax+b}$. Запишемо це рівняння в такому вигляді: $\dfrac{x}{\bar{y}_x} = ax + b.$ Заміною $y^{(1)} = \dfrac{x}{y}$ приводимо парну нелінійну регресію до парної лінійної $\bar{y}_x^{(1)} = ax + b.$ Параметри a і b оцінюються за такими формулами:

$$\hat{a} = \frac{n\sum_{i=1}^{n} x_i^2 / y_i - \sum_{i=1}^{n} x_i \sum_{i=1}^{n} x_i / y_i}{n\sum_{i=1}^{n} x_i^2 - \left(\sum_{i=1}^{n} x_i\right)^2},$$

$$\hat{b} = \left(\sum_{i=1}^{n} x_i / y_i\right)\Big/ n - \hat{a}\left(\sum_{i=1}^{n} x_i\right)\Big/ n, \quad y_i > 0 \quad (i = \overline{1,n}).$$

Регресійний аналіз: парна лінійна регресія

Для описання, аналізу і прогнозування явищ і процесів в економіці використовують математичні моделі у вигляді рівнянь або функцій. Модель економічного об’єкта або виробничого процесу, відображаючи основні його властивості і абстрагуючись від другорядних, дозволяє передбачати його поведінку в певних умовах.

При використанні регресійних моделей результат дії економічної системи або об’єкта у вигляді одного або кількох результуючих показників представляється як функція факторів, що впливають на нього. Як правило, суттєвих факторів небагато, а несуттєвих — достатньо велике число, а тому останніми повністю нехтувати не можна.

В процесі виробництва і розподілу продукції спостерігають масові події, що багаторазово повторюються. Наприклад, серійне виготовлення деталей, вузлів, виробів, багаторазово повторювані акти продажу в продовольчих та промтоварних магазинах. Незважаючи на розвиток економіки і її окремих частин, протягом відносно невеликих часових відрізків і в межах окремих економічних підсистем спостерігається стабільність в умовах відбуття масових подій. По крайній мірі, особливо при прогнозуванні, передбачається можливість багаторазового повторення виробничої ситуації.

Об'єктом дослідження даного пункту є найпростіша регресійна модель — парна лінійна регресія, яку схематично можна зобразити таким чином:

$$Y = \alpha_1 x + \alpha_0 + \varDelta, \tag{4.58}$$

де x — детермінований (невипадковий) фактор, перші два доданки — детермінована складова; α_1 — теоретичний коефіцієнт регресії, який показує, наскільки в середньому зміниться детермінована складова, якщо фактор зміниться на одиницю; $\varDelta$ — випадкова складова, що уособлює сумарну дію різних випадкових факторів. Складова $\varDelta$ називається **похибкою** (залишком, флуктуацією або збуренням).

Нехай для значень $x_1, x_2, ..., x_n$ (нефіксованих) спостерігаються або вимірюються відповідні моделі (4.58) значення *Y*: $y_1, y_2, ..., y_n$. В результаті отримуються *n* рівнянь

$$y_i = \alpha_1 x_i + \alpha_0 + \varDelta_i, \quad i = \overline{1, n}. \tag{4.59}$$

Відносно випадкових величин $\varDelta_1, \varDelta_2, ..., \varDelta_n$ введемо такі припущення (в літературі по економетрії вони називаються **специфікацією** моделі):

1) похибка $\varDelta_i (i = \overline{1, n})$ є нормально розподілена випадкова величин з параметрами

$$M(\varDelta_i) = 0, \quad D(\varDelta_i) = \sigma^2, \quad i = \overline{1, n};$$

2) похибки $\varDelta_i$ та $\varDelta_j$ є незалежними для $i \neq j,\ i = \overline{1, n},\ \ j = \overline{1, n}.$

Мета цього пункту полягає в розв'язуванні таких задач:

1) знаходження оцінок невідомих параметрів α_0 і α_1, вивчення їх властивостей і побудова для них довірчих інтервалів;

2) знаходження оцінки прогнозного значення $y^*(x)$ регресії $y(x)$ для довільного значення *x* і побудова для $y^*(x)$ довірчих інтервалів;

3) вказати критерій того, наскільки добре регресія описує дану систему спостережень.

Статистичні оцінки $\widehat{\alpha}_0, \widehat{\alpha}_1$ параметрів регресії α_0, α_1 вибираються таким чином, щоб емпіричні значення детермінованої складової

$\widehat{y}_i = \widehat{a}_0 + \widehat{a}_1 x_i$ якомога ближче наближались до фактичних значень результуючої ознаки y_i. Оскільки випадкова складова моделі має нормальний розподіл, то оцінки, що володіють найкращими властивостями, отримуються з допомогою МНК.

Нехай δ_i — реалізація (можливе значення конкретної вибірки) випадкової величини $\Delta_i, i = \overline{1,n}$. Згідно із МНК мінімізуємо функцію

$$Q(\widehat{\alpha}_0, \widehat{\alpha}_1) = \sum_{i=1}^{n} \delta_i^2 = \sum_{i=1}^{n} (y_i - \widehat{y}_i)^2 = \sum_{i=1}^{n} (y_i - \widehat{\alpha}_1 x_i - \widehat{\alpha}_0)^2.$$

В результаті отримаємо систему нормальних рівнянь відносно невідомих $\widehat{\alpha}_0, \widehat{\alpha}_1$:

$$\begin{cases} \left(\sum_{i=1}^{n} x_i^2\right)\widehat{\alpha}_1 + \left(\sum_{i=1}^{n} x_i\right)\widehat{\alpha}_0 = \sum_{i=1}^{n} x_i y_i, \\ \left(\sum_{i=1}^{n} x_i\right)\widehat{\alpha}_1 + n\widehat{\alpha}_0 = \sum_{i=1}^{n} y_i. \end{cases}$$

За формулою Крамера

$$\widehat{\alpha}_1 = \frac{n\sum_{i=1}^{n} x_i y_i - \sum_{i=1}^{n} x_i \sum_{i=1}^{n} y_i}{n\sum_{i=1}^{n} x_i^2 - \left(\sum_{i=1}^{n} x_i\right)^2}, \tag{4.60}$$

а з другого рівняння системи

$$\widehat{\alpha}_0 = \left(\sum_{i=1}^{n} y_i - \widehat{a}_1 \sum_{i=1}^{n} x_i\right) \Big/ n$$

або

$$\widehat{\alpha}_0 = \overline{y} - \widehat{\alpha}_1 \overline{x}. \tag{4.61}$$

Незміщена оцінка дисперсії випадкової регресійної моделі (4.58) має такий вид:

$$\widehat{\sigma}^2 = \frac{1}{n-2} \sum_{i=1}^{n} \delta_i^2 = \frac{1}{n-2} \sum_{i=1}^{n} (y_i - \widehat{y}_i)^2. \tag{4.62}$$

де

$$\widehat{y}_i = \overline{y} + \widehat{\alpha}_1 (x_i - \overline{x}).$$

Емпіричні оцінки $\sigma_{\widehat{\alpha}_1}$ та $\sigma_{\widehat{\alpha}_0}$ визначаються формулами

$$\widehat{\sigma}_{\widehat{\alpha}_1} = \widehat{\sigma} \Big/ \sqrt{\sum_{i=1}^{n} (x_i - \overline{x}_i)^2}, \tag{4.63}$$

$$\widehat{\sigma}_{\widehat{\alpha}_0} = \widehat{\sigma}\sqrt{\sum_{i=1}^{n} x_i^2 \Big/ \left[\ n\sum_{i=1}^{n}(x_i - \overline{x})^2\right]}\ , \qquad (4.64)$$

а довірчі інтервали для оцінок невідомих параметрів α_1 та α_0:

$$\widehat{\alpha}_1 - t(\gamma; n-2)\widehat{\sigma}_{\widehat{\alpha}_1} < \alpha_1 < \widehat{\alpha}_1 + t(\gamma; n-2)\widehat{\sigma}_{\widehat{\alpha}_1}\ , \qquad (4.65)$$

$$\widehat{\alpha}_0 - t(\gamma; n-2)\widehat{\sigma}_{\widehat{\alpha}_0} < \alpha_0 < \widehat{\alpha}_0 + t(\gamma; n-2)\widehat{\sigma}_{\widehat{\alpha}_0} \qquad (4.66)$$

де $t(\gamma;\ k)$ знаходиться за табл. 4 додатків.

В дійсності може виявитися, що фактор x не впливає на результуючу ознаку Y, що еквівалентно рівності $\alpha_1 = 0,$ однак при цьому вибірковий коефіцієнт $\widehat{\alpha}_1,$ взагалі кажучи, відмінний від нуля. Для перевірки істотності відхилення $\widehat{\alpha}_1$ від нуля використовується критерій значущості. Розглядається статистична гіпотеза $H_0 : \alpha_1 = 0$ і конкуруюча гіпотеза $H_1 : \alpha_1 \neq 0.$ В якості статистичного критерію береться випадкова величина

$$t = \frac{\widehat{\alpha}_1}{\widehat{\sigma}_{\widehat{\alpha}_1}} = \frac{\sum\limits_{i=1}^{n}(x_i - \overline{x})y_i}{\widehat{\sigma}\Big/\sum\limits_{i=1}^{n}(x_i - \overline{x})^2}, \qquad (4.67)$$

яка має розподіл Ст'юдента із $k = n - 2$ ступенями вільності. Для фіксованої вибірки спостережене значення критерію $t_{\text{спост}},$ знайдене за формулою (4.67), порівнюється із критичною точкою $t_\alpha = t_{\text{двост.кр}}(\alpha; k),$ знайденою за табл. 5 додатків (рівень значущості α вибирається за верхнім рядком, тобто для двосторонньої критичної області). Якщо $|t_{\text{спост}}| < t_\alpha,$ то з імовірністю помилки α приймається гіпотеза $H_0,$ в протилежному випадку, тобто при $|t_{\text{спост}}| > t_\alpha,$ приймається гіпотеза $H_1.$ У випадку прийняття гіпотези H_0 фактор x виключається із моделі і тим самим приймається, що найкраще описання системи спостережень здійснюється з допомогою середньої арифметичної $\widehat{y} = \overline{y}.$

Знайдемо **довірчу зону рівняння регресії**. Теоретичне (підібране за емпіричними даними) лінійне рівняння регресії можна записати у вигляді $\widehat{y} = \overline{y} + \widehat{\alpha}_1(x - \overline{x}),$ де $\overline{y}$ і $\widehat{\alpha}_1$ зазнають впливу помилок, тобто є випадковими величинами для нефіксованої вибірки. Виявляється, що $\overline{y}$ і $\widehat{\alpha}_1$ є некорельованими випадковими величинами.

Теоретичне значення $\hat{y}_k$ при заданому $x_k \, (k = \overline{1,n})$ із врахуванням рівності (4.61) має такий вид:

$$\hat{y}_k = \hat{\alpha}_0 + \hat{\alpha}_1 x_k = \bar{y} + \hat{\alpha}_1 (x_k - \bar{x}).$$

При цьому

$$\hat{\sigma}_{\hat{y}_k} = \hat{\sigma}\left\{ \tfrac{1}{n} + (x_k - \bar{x})^2 \Big/ \sum_{i=1}^{n} (x_i - \bar{x})^2 \right\}^{1/2}, \quad k = \overline{1,n}, \tag{4.68}$$

де $\hat{\sigma}^2$ — точкова незміщена оцінка σ^2, визначена формулою (4.62).

Випадкова величина

$$t = \hat{y}_k \big/ \hat{\sigma}_{\hat{y}_k}$$

має розподіл Ст'юдента із $k = n - 2$ ступенями вільності, тому за даною надійністю γ

$$P\left(-\hat{\sigma}_{\hat{y}_k} t(\gamma; k) < \hat{y}_k < \hat{\sigma}_{\hat{y}_k} t(\gamma; k) \right) = \gamma.$$

Знаючи надійність γ та число ступенів вільності $k = n - 2,$ за табл. 4 додатків відшукаємо значення $t(\gamma; k).$

Отже, довірчий інтервал для так званих **базисних даних** має такий вигляд

$$\left(\hat{y}_k - \hat{\sigma}_{\hat{y}_k} t(\gamma; k); \hat{y}_k + \hat{\sigma}_{\hat{y}_k} t(\gamma; k) \right). \tag{4.69}$$

Якщо для всіх $k = \overline{1,n},$ знайти значення лівих і правих кінців інтервала (4.69), а потім прямолінійними відрізками сполучити кожні дві сусідні точки, то в результаті отримається **довірча зона для базисних точок**. При цьому вважається, що послідовність x_i розташована в порядку зростання.

Розрахунки і перевірка достовірності отриманих оцінок коефіцієнтів регресії не є самоціллю, це тільки необхідний проміжний етап. Основне — це використання моделі для аналізу і прогнозу поведінки досліджуваного економічного явища. Прогноз здійснюється підстановкою фактора *x* в оцінку детермінованої складової:

$$\hat{y}_\text{п} = \bar{y} + \alpha_1 (x_\text{п} - \bar{x}). \tag{4.70}$$

Це точкова оцінка детермінованої складової в прогнозній точці $x_\text{п}$. Можна довести, що вираз (4.70) є найбільш точним (в середньоквадратичному) прогнозом. Для того, щоб виміряти точність цієї оцінки і побудувати довірчий інтервал для $\hat{y}_\text{п},$ необхідно знайти дисперсію точкової оцінки $\hat{y}_\text{п}(x_\text{п}).$ При цьому потрібно врахувати ще одну невизначеність — розсіювання навколо прямої регресії. Оскільки $D(\varDelta_i)=\sigma^2$, то рівнянню $y_{п}=\widehat{\alpha}_1 x_{п}+\widehat{\alpha}_0+\varDelta_{п}$ відповідає дисперсія

$$D(y_{п_i})=\sigma^2(x_{п}-\overline{x})^2\Big/\sum_{i=1}^{n}(x_i-\overline{x})^2+\sigma^2\big/n+\sigma^2.$$

Замінюючи σ його точковою оцінкою $\widehat{\sigma}$, визначеною формулою (4.62), отримаємо за аналогією із вище розглянутим випадком (для базисних даних) такий довірчий інтервал

$$\left(\,\widehat{y}_{п}-t(\gamma;k)\;\widehat{\sigma}_{\widehat{y}_{п}};\,\widehat{y}_{п}+t(\gamma;k)\;\widehat{\sigma}_{\widehat{y}_{п}}\right),\tag{4.71}$$

де $t(\gamma;k)$ знаходиться за табл. 4 додатків для заданої надійності γ і числа ступенів вільності $k=n-2$,

$$\widehat{\sigma}_{\widehat{y}_{п}}=\widehat{\sigma}\left[(x_{п}-\overline{x})^2\Big/\sum_{i=1}^{n}(x_i-\overline{x})^2+1/n+1\right]^{1/2}.\tag{4.72}$$

Залишається ще розглянути питання: наскільки добре рівняння регресії описує дану систему спостережень $\left\{\,(x_i,y_i),i=\overline{1,n}\right\}$. В якості міри якості такого узгодження служить **коефіцієнт детермінації**; при цьому за базу порівняння береться описання з допомогою середнього арифметичного.

Складемо наступні суми квадратів відхилень:

$S^2=\sum_{i=1}^{n}(y_i-\overline{y})^2$ — спостережених значень від їх середньої арифметичної;

$\widehat{S}^2=\sum_{i=1}^{n}(\widehat{y}_i-\overline{y})^2$ — розрахункових (за моделлю) від середньої арифметичної спостережених значень;

$S_R^2=\sum_{i=1}^{n}(y_i-\widehat{y}_i)^2$ — спостережених від розрахункових значень.

Співвідношення між ними має такий вид:

$$S^2=\widehat{S}^2+S_R^2.\tag{4.73}$$

Суму $\widehat{S}^2$ природно назвати поясненою частиною варіації, оскільки значення $\widehat{y}_i$ знаходиться за рівнянням регресії, а $\widehat{S}_R$ — непоясненою (згідно із моделлю) частиною. Тому відповідно до рівності (4.73) чим "менше" значення S_R^2, тим точніше рівняння регресії описує спостережені дані.

Вибірковим коефіцієнтом детермінації називається відношення поясненої частини варіації до усієї варіації в цілому: $d = \widehat{S}^2 / S^2$. Враховуючи рівність (4.77), отримаємо:

$$d = \widehat{S}^2 / S^2 = (S^2 - S_R^2) / S^2 = 1 - S_R^2 / S^2 . \tag{4.74}$$

Отже, чим "ближчий" коефіцієнт d до одиниці, тим кращим є опис рівнянням регресії емпіричних даних, якщо при цьому, зрозуміло, модель є методично правильною.

Статистичну значущість вибіркового коефіцієнта детермінації можна перевірити за допомогою F-статистики:

$$F = \frac{1-d}{d} \cdot \frac{m-1}{n-m}, \tag{4.75}$$

де d — вибірковий коефіцієнт детермінації, n — число спостережень, m — число параметрів рівняння регресії.

Випадкова величина F розподілена за законом Фішера-Снедокора із параметрами $k_1 = m-1,\ k_2 = n-m.$

Оскільки в даному випадку $m = 2,$ то спостережене значення критерію обчислюється за формулою

$$F_{\text{спост}} = \frac{1-d}{d} \cdot \frac{1}{n-2}. \tag{4.76}$$

Для заданої надійності γ знаходиться рівень значущості $\alpha = 1-\gamma$ і за табл. 7 додатків шукається критична точка $F_{\text{кр}}(\alpha; k_1, k_2)$. Якщо $F_{\text{спост}} < F_{\text{кр}}(\alpha; k_1, k_2),$ тоді з надійністю γ можна вважати, що вибірковий коефіцієнт детермінації статистично значущий і включений у регресію фактор достатньо пояснює стохастичну залежність показника.

ПРИКЛАДИ РОЗВ'ЯЗУВАННЯ ЗАДАЧ

Задача 4.1. Отримані статистичні дані десяти однотипних підприємств стосовно кофіцієнта використання основних засобів X і добового обсягу виробництва Y(тис. грн.):

x_i	0,4	0,45	0,5	0,55	0,6	0,65	0,7	0,75	0,8	0,9 .
y_i	2,3	2,4	3,2	3,8	4,5	4,7	5,1	5,6	7,2	8,4

1) скласти систему нормальних рівнянь і знайти коефіцієнти рівняння $\overline{y}_x = \alpha_1 x + a_0$ прямої регресії Y на X;

2) обчислити вибірковий коефіцієнт кореляції $r_{\text{в}}$;

3) перевірити правильність статистичної гіпотези $H_0: r = 0$ для рівня значущості $\alpha = 0{,}05$;

4) якщо гіпотеза з 3) відхилена, тоді для рівня значущості $\alpha = 0{,}05$ перевірити гіпотезу $H_0: r = r_в$ при конкуруючій гіпотезі $H_1: r \neq r_в$;

5) побудувати довірчі інтервали для α_1, α_0, r для надійності $\gamma = 0{,}95$.

○ 1) Для знаходження коефіцієнтів системи нормальних рівнянь (4.23)

$$\begin{cases} \overline{x^2}a_1 + \overline{x}a_0 = \overline{xy}, \\ \overline{x}a_1 + a_0 = \overline{y}, \end{cases}$$

складемо розрахункову табл. 4.3,

Таблиця 4.3

i	x_i	y_i	x_i^2	x_iy_i	y_i^2
1	0,4	2,3	0,16	0,92	5,29
2	0,45	2,4	0,2025	1,08	5,76
3	0,5	3,2	0,25	1,6	10,24
4	0,55	3,8	0,3025	2,09	14,44
5	0,6	4,5	0,36	2,7	20,25
6	0,65	4,7	0,4225	3,055	22,09
7	0,7	5,1	0,49	3,57	26,01
8	0,75	5,6	0,5625	4,2	31,36
9	0,8	7,2	0,64	5,76	54,76
10	0,9	8,4	0,81	7,56	70,56
Σ	6,3	47,2	4,2	32,535	260,76

останній стовпець якої потрібний для обчислення σ_y. Використовуючи нижній рядок табл. 4.3, отримаємо (обсяг вибірки n=10):

$$\overline{x} = \sum_{i=1}^{10} x_i / n = 6{,}3/10 = 0{,}63\,; \quad \overline{y} = \sum_{i=1}^{10} y_i / n = 47{,}2/10 = 4{,}72\,;$$

$$\overline{x^2} = \sum_{i=1}^{10} x_i^2 / n = 4{,}2/10 = 0{,}42\,; \quad \overline{xy} = \sum_{i=1}^{10} x_i y_i / n = 32{,}535/10 = 3{,}2535\,;$$

$$\overline{y^2} = \sum_{i=1}^{10} y_i^2 / n = 260{,}76/10 = 26{,}076;$$

$$\begin{cases} 0{,}42a_1 + 0{,}63a_0 = 3{,}2535, \\ 0{,}63a_1 + a_0 = 4{,}72. \end{cases}$$

Знайдемо єдиний розв'язок системи нормальних рівнянь згідно із формулами (4.24):

$$a_1 = \frac{\overline{xy} - \overline{x} \cdot \overline{y}}{\overline{x^2} - (\overline{x})^2} = \frac{3{,}2535 - 0{,}63 \cdot 4{,}72}{0{,}42 - (0{,}63)^2} = \frac{0{,}2799}{0{,}0231} \approx 12{,}117 \,,$$

$$a_0 = \overline{y} - \overline{x}a_1 = 4{,}72 - 0{,}63 \cdot 12{,}117 = -2{,}914 \,.$$

Отже, емпіричне рівняння прямої регресії Y на X має такий вид:

$$\overline{y}_x = 12{,}117x - 2{,}914 \,.$$

2) Вибірковий коефіцієнт кореляції $r_в$ знайдемо за формулою (4.30):

$$r_в = \frac{\overline{xy} - \overline{x} \cdot \overline{y}}{\sigma_x \sigma_y} = \frac{\overline{xy} - \overline{x} \cdot \overline{y}}{\sqrt{\overline{x^2} - (\overline{x})^2}\sqrt{\overline{y^2} - (\overline{y})^2}} =$$

$$= \frac{3{,}2535 - 0{,}63 \cdot 4{,}72}{\sqrt{0{,}42 - (0{,}63)^2}\sqrt{26{,}076 - (4{,}72)^2}} \approx 0{,}945.$$

3) Оскільки обсяг вибірки n=10 – малий, то використаємо критерій (статистику) (4.45). Для даної вибірки

$$T_{спост.} = \frac{r_в \sqrt{n-2}}{\sqrt{1 - r_в^2}} = \frac{0{,}945\sqrt{8}}{\sqrt{1 - (0{,}945)^2}} = 8{,}172 \,.$$

В даному випадку критична область двостороння. Тому критичну точку $t_{двост.кр.}(\alpha;\, n\text{-}2)$ = $t_\alpha(n\text{-}2)$ знайдемо за верхнім рядком табл. 5 додатків при α=0,05, k=n-2=8: $t_{0,05}(8)$=2,306.

Так як $t_{спост.}$=8,172> $t_{0,05}(8)$=2,306, то гіпотезу H_0 слід відкинути на користь конкуруючої гіпотези $H_1 : r \neq 0$, тобто можна зробити висновок про існування кореляційного зв'язку між X та Y.

4) Перевіримо гіпотезу $H_0 : r = r_0 = 0{,}94$ при конкуруючій гіпотезі $H_1 : r \neq r_0 = 0{,}94$.

Для обчислення спостереженого значення критерія (4.46) знайдемо спочатку значення z_0 та z за формулами (4.47), використавши табл. 10 додатків (перетворення Фішера) і здійснивши лінійну інтерполяцію:

$$z = \frac{1}{2}\ln\frac{1 + r_в}{1 - r_в} = \frac{1}{2}\ln\frac{1 + 0{,}945}{1 - 0{,}945} = (1{,}7380 + 1{,}8318)/2 = 1{,}7849 \,,$$

$$z_0=\frac{1}{2}\ln\frac{1+r_0}{1-r_0}+\frac{r_0}{2(n-1)}=\frac{1}{2}\ln\frac{1+0{,}94}{1-0{,}95}+\frac{0{,}94}{2(10-1)}=$$

$$=1{,}7380+0{,}0522=1{,}7902.$$

Тоді

$$U_{спост.}=(z-z_0)\sqrt{n-3}=(1{,}7849-1{,}7902)\sqrt{7}=-0{,}0141.$$

Коренем рівняння (4.48) $\Phi(u_{кр})=(1-0{,}05)/2=0{,}475$ (за табл.3 додатків) є $u_{кр.}$=1,96.

Оскільки $|U_{спост.}|=|-0{,}0141|<u_{кр.}=1{,}96$, то нема підстав відкидати основну гіпотезу $H_0: r=0{,}94$.

5) Знову врахуємо малість обсягу вибірки. Тоді довірчі інтервали з надійністю γ=0,95 для оцінок теоретичного коефіцієнта регресії α_1 та вільного члена α_0 використаємо (4.40) та (4.42):

$$a_1-t_\gamma\frac{\sigma_y\sqrt{1-r_в^2}}{\sigma_x\sqrt{n-2}}<\alpha_1<a_1+t_\gamma\frac{\sigma_y\sqrt{1-r_в^2}}{\sigma_x\sqrt{n-2}},$$

$$\bar{y}-t_\gamma\sigma_y\sqrt{\frac{(n-1)(1-r_в^2)}{n(n-2)}}<\alpha_0<\bar{y}+t_\gamma\sigma_y\sqrt{\frac{(n-1)(1-r_в^2)}{n(n-2)}},$$

де $t_\gamma=t(\gamma,k)$ знаходиться за табл. 4 додатків розподілу Ст'юдента з числом ступенів вільності $k=n-2$.

Згідно із 1) та 2)

$$\sigma_x=\sqrt{\overline{x^2}-(\bar{x})^2}=\sqrt{0{,}42-(0{,}63)^2}=0{,}152,$$

$$\sigma_y=\sqrt{\overline{y^2}-(\bar{y})^2}=\sqrt{26{,}076-(4{,}72)^2}=1{,}949,$$

а за табл. 4 додатків $t_{0,95}$(0,95; 8)=2,306.

Підставивши дані, отримаємо

$$12{,}117-2{,}306\frac{1{,}949\sqrt{1-(0{,}945)^2}}{0{,}152\sqrt{8}}<\alpha_1<12{,}117+2{,}306\frac{1{,}949\sqrt{1-(0{,}945)^2}}{0{,}152\sqrt{8}},$$

$$4{,}72-2{,}306\cdot1{,}949\sqrt{\frac{9[1-(0{,}945)^2]}{8}}<\alpha_0<4{,}72+{,}306\cdot1{,}949\sqrt{\frac{9[1-(0{,}945)^2]}{8}},$$

або остаточно

$$8{,}705 < \alpha_1 < 15{,}529,$$
$$3{,}161 < \alpha_0 < 6{,}279.$$

Для знаходження довірчого інтервалу з надійністю γ=0,95 для оцінки теоретичного коефіцієнта кореляції *r* використаємо (4.39):

$$\left(\operatorname{th}(z - t/\sqrt{n-3});\ \operatorname{th}(z + t/\sqrt{n-3}) \right),$$

де $z = \frac{1}{2}\ln\frac{1+r_в}{1-r_в}$ і знаходиться за табл. 10 додатків перетворення Фішера (*z*-перетворення), t — корінь рівняння $2\Phi(t) = \gamma$, th(...) знаходиться за табл. 11 додатків оберненого перетворення Фішера.

Для даного випадку *t*=1,96, за табл. 10 додатків

$$z = \frac{1}{2}\ln\left(\frac{1+0{,}945}{1-0{,}945}\right) = (1{,}7380 + 1{,}8318)/2 = 1{,}7849,$$

$$z - t/\sqrt{n-3} = 1{,}7849 - 1{,}96/\sqrt{7} = 1{,}044,$$

$$z + t/\sqrt{n-3} = 1{,}7849 + 1{,}96/\sqrt{7} = 2{,}526,$$

за табл. 11 додатків з використанням інтерполяції

$$\operatorname{th}(1{,}044) = 0{,}7616 + (0{,}7658 - 0{,}7616)\cdot 0{,}44 = 0{,}7635,$$

$$\operatorname{th}(2{,}526) = 0{,}9874.$$

Отже, шуканий довірчий інтервал має вид:

0,7635<*r*<0,9874 ●

Зробимо зауваження відносно кореляційної табл. 4.1. Досі вважалося, що вона отримується внаслідок групування спостережених пар чисел (x_i, y_i) таких, для яких є однаковими або x_i, або y_i, або і x_i і y_i. Така ситуація є характерною для великих обсягів вибірки і дискретних величин *X* та *Y*. Проте для неперервних величин навіть для дуже великих обсягів вибірки вище згадане співпадання значень є рідкісним. Тим не менше, формування кореляційної таблиці і в таких випадках є корисним, оскільки внаслідок цього зменшується обсяг обчислень, а геометрична інтерпретація кореляційного поля стає більш наочною. Проте, як і у випадку одновимірної вибірки, “платою” за такі зручності є похибка числових характеристик вибірки, значення якої стають достатньо малими для великих *n* і вдалому виборі “нових” варіант.

Ідея побудови кореляційної таблиці у випадку великого обсягу вибірки досліджуваних **неперервних** ознак полягає в розбитті основного прямокутника, в якому знаходиться все кореляційне поле, на сукупність частинних прямокутників з однаковими довжинами сторін із наступним обчисленням частот попадання спостережених точок (пар чисел) в кожний із них. Ілюструє цю ідею, зокрема, наступна задача.

Задача 4.2. В табл. 4.4 наведені значення ваги (кг) x_i, $i = \overline{1{,}50}$, та росту (см) y_i, $i = \overline{1{,}50}$, випадково відібраних 50 чоловіків.
Потрібно: 1) скласти кореляційну таблицю спостережених даних; 2) знайти вибіркові рівняння регресії *Y* на *X* та *X* на *Y* за отриманою кореляційною таблицею.

Таблиця 4.4

x_i	69,80	71,46	82,41	77,08	63,77	66,78	88,15	73,60	77,54	56,84
y_i	165,8	175,7	176,0	170,5	163,0	170,4	172,1	181,0	180,2	157,4
x_i	72,00	73,03	73,98	73,70	69,90	72,33	78,63	83,46	80,46	82,73
y_i	171,8	171,7	169,2	166,2	161,7	162,7	167,7	170,5	171,7	184,5
x_i	81,85	71,01	76,69	68,97	65,41	75,28	83,33	75,73	70,43	73,01
y_i	173,8	175,5	179,8	167,8	163,7	177,4	176,4	178,1	170,2	179,2
x_i	77,24	64,48	77,93	72,51	72,08	70,02	79,91	67,22	68,42	76,20
y_i	168,1	159,1	171,3	179,6	166,9	162,9	182,8	166,3	174,4	177,8
x_i	68,41	69,92	58,18	79,16	73,34	71,66	84,04	77,41	73,36	72,94
y_i	171,1	170,6	153,9	175,3	170,6	173,8	181,4	174,2	170,4	163,0

○ 1) Спостережені ознаки *x* та *y* знаходяться в проміжках [56,84; 88,15] та [153,9; 184,5]. Основним прямокутником, в якому знаходиться все кореляційне поле, будемо вважати прямокутник $\{56 \le x \le 89; 153 \le y \le 185]$. Проміжок [56; 89] довжиною 33 розіб'ємо на 11 частинних інтервалів, а проміжок [153; 185] довжиною 32 — на 8 інтервалів. В результаті отримується 88 частинних прямокутників. Побудову кореляційної таблиці здійнимо в два етапи: на першому кроці підрахуємо число попадань спостереженої пари чисел у частинний прямокутник, помічаючи кожне попадання штрихом всередині відповідного прямокутника кореляційної таблиці; на другому кроці здійснимо перехід до "нових" варіант — середин кожного із частинних інтервалів. Результатом першого етапу є табл. 4.5, а другого — табл. 4.6.

Таблиця 4.5

X	Y								
	[153; 157)	[157; 161)	[161; 165)	[165; 169)	[169; 173)	[173; 177)	[177; 181)	[181; 185]	n_{x_i}
[56;59)	′	′							2
[59;62)									0
[62;65)		′	′						2
[65;68)			′	′	′				3
[68;71)			″	″	‴	′			8
[71;74)			″	″	′′′′′	‴	‴		15
[74;77)							′′′′		4
[77;80)				″	″	″	′	′	8
[80;83)					′	″		′	4
[83;86)					′	′		′	3
[86;89]					′				1
n_{y_j}	1	2	6	7	14	9	7	4	50

Таблиця 4.6

X	Y								
	155	159	163	167	171	175	179	183	n_{x_i}
57,5	1	1							2
60,5									0
63,5		1	1						2
66,5			1	1	1				3
69,5			2	2	3	1			8
72,5			2	2	5	3	3		15
75,5							4		4
78,5				2	2	2	1	1	8
81,5					1	2		1	4
84,5					1	1		1	3
87,5					1				1
n_{y_i}	1	2	6	7	14	9	8	3	50

2) Для знаходження вибіркових рівнянь регресії Y на X та X на Y використаємо спрощений метод обчислення параметрів за кореляційною таблицею. В якості хибних нулів візьмемо моди варіант x та y: C_1=72,5, C_2=171. Варіанти x_{i} та y_{j} є рівновіддаленими

із кроками h_1=3 та h_2=4 відповідно. За формулами (4.34) перейдемо до умовних варіант. В результаті отримаємо кореляційну табл. 4.7

Таблиця 4.7

U	V -4	-3	-2	-1	0	1	2	3	n_u
-5	1	1							2
-4									0
-3		1	1						2
-2			1	1	1				3
-1			2	2	3	1			8
0			2	2	5	3	3		15
1							4		4
2				2	2	2	1	1	8
3					1	2		1	4
4					1	1		1	3
5					1				1
n_v	1	2	6	7	14	9	8	3	50

Величини $\bar{u}$, $\bar{v}$, σ_u, σ_v можна обчислити методом добутків, проте оскільки числа u_i, v_j малі, то знайдемо їх за означенням:

$$\bar{u} = \left(\sum n_u u\right)/n = [(-5)\cdot 2 + (-4)\cdot 0 + (-3)\cdot 2 + (-2)\cdot 3 + (-1)\cdot 8 + 1\cdot 4 +$$

$$+ 2\cdot 8 + 3\cdot 4 + 4\cdot 3 + 5\cdot 1]/50 = 0{,}38;$$

$$\overline{u^2} = \left(\sum n_u u^2\right)/n = [(-5)^2\cdot 2 + (-4)^2\cdot 0 + (-3)^2\cdot 2 + (-2)^2\cdot 3 + (-1)^2\cdot 8 +$$

$$+ 1^2\cdot 4 + 2^2\cdot 8 + 3^2\cdot 4 + 4^2\cdot 3 + 5^2\cdot 1]/50 = 4{,}68;$$

$$\sigma_u = \sqrt{\overline{u^2} - (\bar{u})^2} = \sqrt{4{,}68 - (0{,}38)^2} = 2{,}130\ ;$$

$$\bar{v} = \left(\sum n_v v\right)/n = [(-4)\cdot 1 + (-3)\cdot 2 + (-2)\cdot 6 + (-1)\cdot 7 + 1\cdot 9 + 2\cdot 8 +$$

$$+ 3\cdot 3]/50 = 0{,}1;$$

$$\overline{v^2} = \left(\sum n_v v^2\right)/n = [(-4)^2\cdot 1 + (-3)^2\cdot 2 + (-2)^2\cdot 6 + (-1)^2\cdot 7 + 1^2\cdot 9 +$$

$$+ 2^2\cdot 8 + 3^2\cdot 3]/50 = 2{,}66;$$

$$\sigma_v = \sqrt{\overline{v^2} - (\bar{v})^2} = \sqrt{2{,}66 - (0{,}1)^2} = 1{,}628\ .$$

Для знаходження $r_в$ обчислимо $\left(\sum n_{uv} uv\right)/n$, врахувавши нульові значення варіант u та v:

$$\left(\sum n_{uv}uv\right)/n = [(-5)\cdot(-4)\cdot 1 + (-5)\cdot(-3)\cdot 1 + (-3)\cdot(-3)\cdot 1 +$$
$$+(-3)\cdot(-2)\cdot 1 + (-2)\cdot(-2)\cdot 1 + (-2)\cdot(-1)\cdot 1 + (-1)\cdot(-2)\cdot 2 +$$
$$+(-1)\cdot(-1)\cdot 2 + (-1)\cdot 1\cdot 1 + 1\cdot 2\cdot 4 + 2\cdot(-1)\cdot 2 + 2\cdot 1\cdot 2 + 2\cdot 2\cdot 1 +$$
$$+2\cdot 3\cdot 1 + 3\cdot 1\cdot 2 + 3\cdot 3\cdot 1 + 4\cdot 1\cdot 1 + 4\cdot 3\cdot 1]/50 = 2{,}2.$$

Тоді за формулами (4.35)

$$\bar{x} = \bar{u}h_1 + C_1 = 0{,}38\cdot 3 + 72{,}5 = 73{,}64\,;$$
$$\bar{y} = \bar{v}h_2 + C_2 = 0{,}1\cdot 4 + 171 = 171{,}4\,;$$
$$\sigma_x = \sigma_u\cdot h_1 = 2{,}13\cdot 3 = 6{,}39\,;$$
$$\sigma_y = \sigma_v\cdot h_2 = 1{,}628\cdot 4 = 6{,}512\,;$$
$$r_в = \frac{\frac{1}{n}\left(\sum n_{uv}uv\right) - \bar{u}\cdot\bar{v}}{\sigma_u\sigma_v} = \frac{2{,}2 - 0{,}38\cdot 0{,}1}{2{,}13\cdot 1{,}628} = 0{,}624\,.$$

Емпіричні рівняння (4.31) та (4.32) прямих регресій із врахуванням отриманих даних матимуть такий вид

$$\bar{y}_x - 171{,}4 = 0{,}624\frac{6{,}512}{6{,}39}(x - 73{,}64)\,,$$
$$\bar{x}_y - 73{,}64 = 0{,}624\frac{6{,}39}{6{,}512}(y - 171{,}4)$$

або остаточно

$$\bar{y}_x = 0{,}636x + 124{,}57\,,$$
$$\bar{x}_y = 0{,}612y - 31{,}31\,. \bullet$$

Задача 4.3. 1) За даними кореляційної табл. 4.6 задачі 4.2 з надійністю $\gamma = 0{,}95$ знайти довірчі інтервали для теоретичних коефіцієнтів кореляції і регресії Y на X та X на Y, а також вільних членів прямих регресій. 2) Ігноруючи немалість обсягу вибірки, знайти довірчий інтервал для коефіцієнта кореляції, використавши формули для малих вибірок. Порівняти результати.

○ 1) Для знаходження інтервальних оцінок теоретичного коефіцієнта кореляції та параметрів теоретичних прямих регресії використаємо числові характеристики двовимірної вибірки, знайдені в задачі 4.2:

$$r_в = 0{,}624,\quad \sigma_x = 6{,}39,\quad \sigma_y = 6{,}512,$$
$$a_{Y/X} = 0{,}636,\quad b_{X/Y} = 0{,}612,\quad \bar{x} = 73{,}64,\quad \bar{y} = 171{,}4.$$

Оскільки обсяг даної вибірки $n = 50$ достатньо великий, то 95%-ові довірчі інтервали для r, α_1 та β_1 будемо шукати за формулами (4.36)-(4.38), де $t = 1{,}96$ — корінь рівняння $2\Phi(t) = 0{,}95,$ знайдений за табл. 3 додатків.

Тобто

$$0{,}624 - 1{,}96 \cdot \frac{1-(0{,}624)^2}{\sqrt{50}} < r < 0{,}624 + 1{,}96 \cdot \frac{1-(0{,}624)^2}{\sqrt{50}},$$

$$0{,}636 - 1{,}96 \cdot \frac{6{,}512}{6{,}31} \cdot \frac{1-(0{,}624)^2}{\sqrt{50}} < \alpha_1 < 0{,}636 + 1{,}96 \cdot \frac{6{,}512}{6{,}39} \cdot \frac{1-(0{,}624)^2}{\sqrt{50}},$$

$$0{,}612 - 1{,}96 \cdot \frac{6{,}39}{6{,}512} \cdot \frac{1-(0{,}624)^2}{\sqrt{50}} < \beta_1 < 0{,}612 + 1{,}96 \cdot \frac{6{,}39}{6{,}512} \cdot \frac{1-(0{,}624)^2}{\sqrt{50}}.$$

Виконавши обчислення, отримаємо

$$0{,}455 < r < 0{,}793; \quad 0{,}464 < \alpha_1 < 0{,}808; \quad 0{,}446 < \beta_1 < 0{,}778.$$

Для знаходження довірчих інтервалів для α_0 та β_0 використаємо формули (4.42), (4.43), знайшовши попередньо $t_{0,95} = t(0{,}95;47)$ за табл. 4 додатків розподілу Ст'юдента: $t = 2{,}025$ (при цьому здійснюємо лінійну інтерполяцію для значень $k = 40$ та $k = 50$).

Тоді

$$171{,}4 - 2{,}025 \cdot 6{,}512\sqrt{\frac{49(1-0{,}624^2)}{50 \cdot 48}} < \alpha_0 < 171{,}4 + 2{,}025 \cdot 6{,}512\sqrt{\frac{49(1-0{,}624^2)}{50 \cdot 48}},$$

$$73{,}64 - 2{,}025 \cdot 6{,}39\sqrt{\frac{49(1-0{,}624^2)}{50 \cdot 48}} < \beta_0 < 73{,}64 + 2{,}025 \cdot 6{,}39\sqrt{\frac{49(1-0{,}624^2)}{50 \cdot 48}},$$

або остаточно $169{,}928 < \alpha_0 < 172{,}872, \quad 72{,}195 < \beta_0 < 75{,}085.$

2) Для інтервального оцінювання теоретичного коефіцієнта кореляції r використаємо тепер (4.39). Для $r_в = 0{,}624$ за табл. 8 додатків перетворення Фішера знайдемо $z = 0{,}725 + (0{,}74/4 - 0{,}725) \cdot 0{,}4 = 0{,}7316$, а для $\gamma = 0{,}95$ за табл. 3 додатків знайдемо корінь рівняння $2\Phi(t) = 0{,}95,$ тобто $t = 1{,}96.$ Тоді інтервал (4.39) набере такого виду

$$\left(\mathrm{th}(0{,}7316 - 1{,}96\big/\sqrt{47});\mathrm{th}(0{,}7341 + 1{,}96\big/\sqrt{47})\right)$$

або $(\text{th}\,0{,}45; \text{th}\,1{,}02)$. Знайшовши за табл. 11 додатків межі останнього інтервалу, остаточно отримаємо шуканий довірчий інтервал $0{,}422 < r < 0{,}770$. Порівняння цього інтервалу із довірчим інтервалом, знайденим в першій частині, дозволяє зробити висновок про незначну похибку, яка виникає внаслідок використання інтервалу (4.39) при оцінюванні *r* для немалих вибірок. ●

Задача 4.4. Перевірити основну гіпотезу $H_0 : r = 0$ при конкуруючій гіпотезі $H_1 : r \neq 0$ за даними кореляційної табл.4.6 задачі 4.2 при рівні значущості $\alpha = 0{,}05$. При цьому використати критерії як для великих *n*, так і для малих *n*.

○ Оскільки альтернативна гіпотеза $H_1 : r \neq 0$, то критична область — двостороння. Тому u_α в (4.44) є коренем рівняння (4.48), тобто $u_\alpha = 1{,}96$ (за табл. 3 додатків). Враховуючи, що $r_в = 0{,}624$, $n = 50$, обчислимо в даному випадку праву частину нерівності (4.48):

$$u_\alpha \frac{1 - r_в^2}{\sqrt{n}} = 1{,}96 \cdot \frac{1 - (0{,}624)^2}{\sqrt{50}} = 0{,}169.$$

Але $r_в = 0{,}624 > 0{,}169$, тому основну гіпотезу про некорельованість випадкових величин X та Y відкидаємо, допускаючи помилку лише у 5% випадків.

Здійснимо тепер перевірку H_0, використавши статистику (4.45) для невеликих обсягів вибірки. Для даної вибірки

$$T_{\text{спост}} = \frac{r_в \sqrt{n-2}}{\sqrt{1 - r_в^2}} = \frac{0{,}624\sqrt{48}}{\sqrt{1 - (0{,}624)^2}} = 5{,}533.$$

В даному випадку критична область також двостороння. Тому критичну точку $t_\alpha(n-2)$ знайдемо за верхнім рядком табл. 5 додатків при $\alpha = 0{,}05$, $k = n - 2 = 48$, здійснивши лінійну інтерполяцію:

$$t_{0,05}(48) = t_{0,05}(40) + [t_{0,05}(60) - t_{0,05}(40)] \cdot \frac{8}{20} =$$

$$= 2{,}021 + (2{,}000 - 2{,}021) \cdot \frac{8}{20} = 2{,}021 - 0{,}0084 = 2{,}0126.$$

Оскільки $T_{\text{спост}} = 5{,}533 > t_{0{,}05}(48) = 2{,}0126,$ то робимо висновок про існування кореляційного зв'язку між X та Y, тобто приймаємо конкуруючу гіпотезу. ●

Задача 4.5. Для рівня значущості $\alpha = 0{,}05$ перевірити гіпотезу $H_0 : r = 0{,}62$ при конкуруючій гіпотезі $H_1 : r \neq 0{,}62$ за даними кореляційної табл. 4.6 задачі 4.2.

○ Для обчислення спостереженого значення критерія (4.46) знайдемо спочатку значення z_0 та z (формули (4.47)), використавши табл. 10 додатків перетворення Фішера і лінійну інтерполяцію:

$$z = \frac{1}{2}\ln\frac{1+r_в}{1-r_в} = \frac{1}{2}\ln\frac{1+0{,}624}{1-0{,}624} = 0{,}725 + (0{,}74/4 - 0{,}725)\cdot 0{,}4 = 0{,}7316,$$

$$z_0 = \frac{1}{2}\ln\frac{1+r_0}{1-r_0} + \frac{r_0}{2(n-1)} = \frac{1}{2}\ln\frac{1+0{,}62}{1-0{,}62} + \frac{0{,}62}{2\cdot 49} = 0{,}725 + 0{,}0063 = 0{,}7313.$$

Тоді

$$U_{\text{спост}} = (z - z_0)\sqrt{n-3} = (0{,}7313 - 0{,}7313)\sqrt{47} = 0{,}0021.$$

Коренем рівняння $\Phi(u_{\text{кр}}) = (1-0{,}05)/2$ є $u_{\text{кр}} = 1.96.$

Оскільки $U_{\text{спост}} = 0{,}0021 < u_{\text{кр}} = 1{,}96,$ то нема підстав відкидати основну гіпотезу про те, що теоретичний коефіцієнт кореляції $r = 0{,}62.$ ●

Задача 4.6. Знайти емпіричне рівняння регресії Y на X за даними кореляційної табл. 4.8.

Таблиця 4.8

X	Y				
	[5; 7)	[7; 9)	[9; 11)	[11; 13]	n_x
[0,9; 1,1)	—	—	1	8	9
[1,1; 1,3)	2	—	6	—	8
[1,3; 1,5)	6	1	—	—	7
[1,5; 1,7)	1	9	1	—	11
[1,7; 1,9)	—	2	5	—	7
[1,9; 2,1]	—	—	1	7	8
n_y	9	12	14	15	n = 50

○ Для знаходження послідовності точок виду (4.49) перейдемо від кореляційної табл. 4.8 до табл. 4.9, для якої "новими" варіантами є середини відповідних частинних інтервалів, а останній стовпчик заповнимо, використовуючи означення умовної середньої:

$$\overline{y}_{x_i} = \left(\sum y_j n_{ij} \right) / n_{x_i}$$

(Наприклад, $\overline{y}_{x_1} = (10 \cdot 1 + 12 \cdot 8)/9 \approx 11{,}78$; $\overline{y}_{x_2} = (6 \cdot 2 + 10 \cdot 6)/8 = 9$).

Таблиця 4.9

X	Y					
	6	8	10	12	n_x	$\overline{y}_x$
1	—	—	1	8	9	11,78
1,2	2	—	6	—	8	9
1,4	6	1	—	—	7	6,29
1,6	1	9	1	—	11	8
1,8	—	2	5	—	7	9,43
2	—	—	1	7	8	11,75
n_y	9	12	14	15	$n = 50$	

Із табл. 4.9 отримуємо послідовність точок

(1; 11,78), (1,2; 9), (1,4; 6,29), (1,6; 8), (1,8; 9,43), (2; 11,75),

розташування яких в прямокутній декартовій системі координат дозволяє висунути припущення про наявність параболічної кореляції другого порядку.

Для знаходження коефіцієнтів системи нормальних рівнянь (4.52) складемо розрахункову табл. 4.10.

Таблиця 4.10

x_i	n_{x_i}	$\overline{y}_{x_i}$	$n_{x_i} x_i$	$n_{x_i} x_i^2$	$n_{x_i} x_i^3$	$n_{x_i} x_i^4$	$n_{x_i} \overline{y}_{x_i}$	$n_{x_i} \overline{y}_{x_i} x_i$	$n_{x_i} \overline{y}_{x_i} x_i^2$
1	9	11,78	9	9	9	9	106,02	106,02	106,02
1,2	8	9	9,6	11,52	13,824	16,589	72	86,4	103,68
1,4	7	6,29	9,8	13,72	19,208	26,891	44,03	61,642	86,3
1,6	11	8	17,6	28,16	45,056	72,090	88	140,8	225,28
1,8	7	9,43	12,6	22,68	40,824	73,483	66,01	118,818	213,872
2	8	11,75	16	32	64	128	94	188	376
$\sum$	50	—	74,6	117,08	191,912	326,053	470,06	701,68	1111,152

За формулами (4.53) із останнього рядка табл. 4.10 отримаємо:

$\overline{x} = 74{,}6/50 = 1{,}492$; $\overline{x^2} = 117{,}08/50 = 2{,}342$; $\overline{x^3} = 119{,}912/50 = 3{,}838$;

$\overline{x^4} = 326{,}053/50 = 6{,}521$; $\overline{y} = \overline{\overline{y}_x} = 470{,}06/50 = 9{,}401$;

$\overline{\overline{y}_x x} = 701{,}68/50 = 14{,}034$; $\overline{\overline{y}_x x^2} = 1111{,}152/50 = 22{,}223$.

Тоді система рівнянь (4.52) набере такого виду

$$\begin{cases} 6{,}521a_2 + 3{,}838a_1 + 2{,}342a_0 = 22{,}223, \\ 3{,}838a_2 + 2{,}342a_1 + 1{,}492a_0 = 14{,}034, \\ 2{,}342a_2 + 1{,}492a_1 + a_0 = 9{,}401. \end{cases} \qquad (4.77)$$

Розв'язавши її, отримаємо

$$a_2 = 11{,}492, \quad a_1 = -34{,}010, \quad a_0 = 33{,}230.$$

Отже, шукане емпіричне рівняння регресії Y на X має такий вид

$$\overline{y}_x = 11{,}492x^2 - 34{,}01x + 33{,}23. \qquad (4.78)$$ ●

Зауваження. При розв'язуванні системи нормальних рівнянь з'ясовується, що визначник матриці, складеної із коефіцієнтів при невідомих, близький до нуля. Для таких систем характерним є те, що навіть дуже малі зміни коефіцієнтів приводять до **значних** змін розв'язків. Для ілюстрації цього положення розглянемо систему рівнянь

$$\begin{cases} 6{,}52a_2 + 3{,}84a_1 + 2{,}34a_0 = 22{,}22, \\ 3{,}84a_2 + 2{,}34a_1 + 1{,}49a_0 = 14{,}03, \\ 2{,}34a_2 + 1{,}49a_1 + a_0 = 9{,}4, \end{cases}$$

яка отримується із (4.77) внаслідок заокруглення коефіцієнтів до сотих. Її розв'язок

$$a_2 = 55{,}359, \quad a_1 = -162{,}958, \quad a_0 = 122{,}666$$

суттєво відрізняється від розв'язку системи (4.77) і визначає емпіричне рівняння регресії в такому вигляді:

$$\overline{y}_x = 55{,}359x^2 - 162{,}958x + 122{,}666. \qquad (4.79)$$

Табл. 4.11 характеризує "якість" отриманих емпіричних рівнянь регресії в плані порівняння розрахункових значень $\bar{y}_x^{(4.78)}$ і $\bar{y}_x^{(4.79)}$, отриманих із формул (4.78) та (4.79) відповідно, із спостереженими значеннями $\bar{y}_x^{(\text{спост})}$ (дані останнього стовпця табл. 4.9).

Таблиця 4.11

x_i	1	1,2	1,4	1,6	1,8	2
$\bar{y}_x^{(4.78)}$	10,71	8,97	8,14	8,24	9,25	11,18
$\bar{y}_x^{(4.79)}$	15,07	6,84	3,03	3,66	8,71	18,19
$\bar{y}_x^{(\text{спост})}$	11,78	9	6,29	8	9,43	11,75

На підставі цієї таблиці можна зробити висновок, що рівняння регресії (4.78) є кращим в порівнянні із (4.79). Разом з тим слід очікувати покращення результатів внаслідок збільшення числа знаків після коми у коефіцієнтів нормальної системи рівнянь. Іншими словами, **досягнення потрібної точності результатів без використання обчислювальної техніки є дуже трудоємкою задачею**.

Задача 4.7. Обчислити вибіркове кореляційне відношення за даними задачі 4.6.

○ За означенням (4.56*) $\eta_{yx} = \sigma_{\bar{y}_x} / \sigma_y$, тому потрібно знайти

$$\sigma_{\bar{y}_x} = \sqrt{\sum_i n_{x_i} (\bar{y}_{x_i} - \bar{y})^2 / n}, \quad \sigma_y = \sqrt{\sum_j (y_j - \bar{y})^2 n_{y_j} / n},$$

де за даними розв'язування задачі 4.6 $\bar{y} = 9{,}401$, $n = 50$, n_{x_i} та $\bar{y}_{x_i}$ наведені в двох останніх стовпцях табл. 4.9, а n_{y_j} — в останньому рядку цієї ж таблиці.

Підставивши вказані значення, отримаємо:

$$\sigma_{\bar{y}_x} = \left\{ \left[9(11{,}87 - 9{,}401)^2 + 8(9 - 9{,}401)^2 + 7(6{,}29 - 9{,}401)^2 + 11(8 - 9{,}401)^2 + \right.\right.$$

$$\left.\left. + 7(9{,}43 - 9{,}401)^2 + 8(11{,}75 - 9{,}401)^2 \right] / 50 \right\}^{1/2} = 1{,}927;$$

$$\sigma_y = \left\{ \left[9(6 - 9{,}401)^2 + 12(8 - 9{,}401)^2 + 14(10 - 9{,}401)^2 + \right.\right.$$

$$\left.\left. + 15(12 - 9{,}401)^2 \right] / 50 \right\}^{1/2} = 2{,}169.$$

Тоді $\eta_{yx} = 1{,}927 / 2{,}169 = 0{,}888$ і, отже, між випадковими величинами X та Y існує достатньо тісна кореляційна залежність. ●

Задача 4.8. Торгівельне підприємств має велику кількість філій, і керівництво цього підприємства вивчає питання про залежність y (річний товарообіг однієї філії, млн. грн.) від x (торгівельної площі, тис. м^2). Для дванадцяти філій за певний рік зафіксовані такі значення показників y та x:

i	1	2	3	4	5	6	7	8	9	10	11	12
y_i	2,93	5,27	6,85	7,01	7,02	8,35	4,33	5,77	7,68	3,16	1,52	3,15
x_i	0,31	0,98	1,21	1,29	1,12	1,49	0,78	0,94	1,29	0,48	0,24	0,55

На обсяг товарообігу впливають такі фактори: середньоденна інтенсивність потоку покупців, об'єм основних фондів, їх структура, середньосписочна чисельність працівників, площа підсобних приміщень тощо. Припускається, що в досліджуваній групі філій значення цих факторів приблизно однакові, тому вплив відмінностей їх значень на зміну обсягу товарообігу є незначним.

Потрібно:

1) знайти статистичні оцінки параметрів лінійного рівняння регресії;
2) обчислити вибірковий коефіцієнт детермінації;
3) для рівня значущості $\alpha = 0{,}05$ перевірити адекватність лінійної моделі емпіричним даним (з'ясувати статистичну значущість вибіркового коефіцієнта детермінації);
4) перевірити правильність статистичної гіпотези $H_0 : \hat{\alpha}_1 = 0$ для рівня значущості $\alpha = 0{,}05$;
5) побудувати довірчі інтервали для параметрів рівняння регресії з надійністю $\gamma = 0{,}95$;
6) побудувати довірчу зону для базисних даних із надійністю $\gamma = 0{,}95$;
7) знайти прогнозне значення річного товарообігу для нової філії, торгівельна площа якої складає 1,82 тис. м^2, а також із надійністю $\gamma = 0{,}95$ побудувати довірчий інтервал для цього прогнозного значення.

○ Для спрощення розрахунків при розв'язуванні задачі складемо табл. 4.12, заповнивши спочатку перших 5 стовпців.

1) Обсяг вибірки $n = 12$. Сума другого-п'ятого стовпців визначають коефіцієнти системи нормальних рівнянь. Тому згідно із формулою (4.60)

$$\hat{a}_1 = \frac{n\sum_{i=1}^{n} x_i y_i - \sum_{i=1}^{n} x_i \sum_{i=1}^{n} y_i}{n\sum_{i=1}^{n} x_i^2 - \left(\sum_{i=1}^{n} x_i\right)^2} = \frac{12 \cdot 66{,}0307 - 10{,}68 \cdot 63{,}04}{12 \cdot 11{,}4058 - (10{,}68)^2} = 5{,}2221.$$

Із останнього рядка табл. 4.12 (2-го і 3-го стовпців) знаходимо

$$\bar{x} = \sum_{i=1}^{n} x_i / n = 10{,}68/12 = 0{,}89; \quad \bar{y} = \sum_{i=1}^{n} y_i / n = 63{,}04/12 = 5{,}2533,$$

тоді за формулою (4.61)

$$\hat{\alpha}_0 = \bar{y} - \hat{\alpha}_1 \bar{x} = 5{,}2533 - 5{,}2221 \cdot 0{,}89 = 0{,}6056.$$

Таблиця 4.12

№ п/п філії	x_i	y_i	x_i^2	$x_i y_i$	$x_i - \overline{x}$	$(x_i - \overline{x})^2$	$(x_i - \overline{x})y_i$	$(y_i - \overline{y})^2$	$\widehat{y}_i$	$(y_i - \widehat{y}_i)^2$
1	2	3	4	5	6	7	8	9	10	11
1	0,31	2,93	0,0961	0,9083	-0,58	0,3364	-1,6994	5,3977	2,2245	0,4977
2	0,98	5,27	0,9604	5,1646	0,09	0,0081	0,4743	0,0003	5,7233	0,2055
3	1,21	6,85	1,4641	8,2885	0,32	0,1024	2,1920	2,5495	6,9243	0,0743
4	1,29	7,01	1,6641	9,0429	0,40	0,1600	2,8040	3,0860	7,3421	0,1103
5	1,12	7,02	1,2544	7,8624	0,23	0,0529	1,6146	3,1212	6,4544	0,3199
6	1,49	8,35	2,2201	12,4415	0,60	0,3600	5,0100	9,5896	8,3865	0,0013
7	0,78	4,33	0,6084	3,3774	-0,11	0,0121	-0,4763	0,8525	4,6779	0,1217
8	0,94	5,77	0,8836	5,4238	0,05	0,0025	0,2885	0,2670	5,5144	0,0653
9	1,29	7,68	1,6641	9,9072	0,40	0,1600	3,0720	5,8889	7,3421	0,1142
10	0,48	3,16	0,2304	1,5168	-0,41	0,1681	-1,2956	4,3819	3,1122	0,0023
11	0,24	1,52	0,0576	0,3648	-0,65	0,4225	-0,9880	13,9375	1,8589	0,1149
12	0,55	3,15	0,3025	1,7325	-0,34	0,1156	-1,071	4,4239	3,4778	0,1074
$\sum$	10,68	63,04	11,4058	66,0307		1,9006	9,9251	53,496		1,7348

Отже, емпіричне рівняння регресії має такий вид:

$$\widehat{y} = 5{,}2221x + 0{,}6056. \qquad (4.80)$$

2) Для обчислення вибіркового коефіцієнта детермінації d спочатку заповнимо 9-й, 10-й та 11-й стовпці, при цьому $\widehat{y}_i$ знайдемо із рівняння (4.80): $\widehat{y}_i = 5{,}2221x_i + 0{,}6056$. Підсумок 9-го та 11-го стовпців дасть відповідно $S^2 = 53{,}4960$ та $S_R^2 = 1{,}7348$. Тоді за формулою (4.74)

$$d = 1 - S_R^2 / S^2 = 1 - 1{,}7348/53{,}4960 = 0{,}9676.$$

3) Статистичну значущість отриманого вибіркового коефіцієнта детермінації перевіримо за критерієм (4.75). Згідно із формулою (4.74)

$$F_{\text{спост}} = \frac{1-d}{d} \cdot \frac{1}{n-2} = \frac{1-0{,}9676}{0{,}9676} \cdot \frac{1}{12-2} = 0{,}0033.$$

За табл. 7 додатків критичних точок розподілу Фішера-Снедокора для рівня значущості $\alpha = 0{,}05$ та ступенів вільності $k_1 = m - 1 = 1$, $k_2 = n - m = 12 - 2 = 10$ ($m = 2$ — число параметрів рівняння регресії) знайдемо

$$F_{\text{кр}}(0{,}05;1{,}10) = 4{,}96.$$

Оскільки $F_{\text{спост}} = 0{,}0033 < 4{,}96 = F_{\text{кр}}$, то з надійністю $\gamma = 1 - \alpha = 0{,}95$ можна стверджувати, що вибірковий коефіцієнт детермінації статистично значущий, а тому рівняння (4.80) адекватно описує емпіричні дані.

Отже, на підставі значення d можна вважати, що в загальній величині дисперсії залежної змінної y частка поясненої моделлю дисперсії регресії складає 96,76% і тільки 3,24% припадає на неврахованi фактори.

4) Заповнимо стовпці з номерами 6, 7 і 8. Підсумки 7-го і 8-го стовпців дадуть

$$\sum_{i=1}^{n} (x_i - \overline{x})^2 = 1{,}9006, \quad \sum_{i=1}^{n} (x_i - \overline{x}) y_i = 9{,}9251.$$

Використавши формулу (4.62) і підсумок 11-го стовпця, отримаємо

$$\widehat{\sigma}^2 = \frac{1}{n-2} \sum_{i=1}^{n} (y_i - \widehat{y}_i)^2 = \frac{1}{12-2} \cdot 1{,}7348 = 0{,}1735,$$

$$\widehat{\sigma} = \sqrt{0{,}1735} = 0{,}4165.$$

За формулою (4.67) розраховуємо спостережене значення критерію перевірки значущості (відмінності від нуля α_1):

$$t=\frac{\sum_{i=1}^{n}(x_i-\overline{x})y_i}{\widehat{\sigma}\Big/\sum_{i=1}^{n}(x_i-\overline{x})^2}=\frac{9{,}9251}{0{,}4165/1{,}9006}=45{,}291.$$

За табл. 5 додатків критичних точок розподілу Ст'юдента для $\alpha=0{,}05$ (по верхньому рядку) і числу ступенів вільності $k=n-1=12-1=11$ знаходимо $t_{0,05}=t_{\text{äâî ñò.êð}}(0,05;11)=2,201.$ Оскільки $t_{\text{спост}}=45{,}291>>2{,}201=t_{0,05}$, то гіпотеза $H_0:\alpha_1=0$ відкидається і з імовірністю 0,95 приймається гіпотеза $H_1:\alpha_1\neq 0.$

5) Для побудови довірчого інтервалу для параметра α_0 знайдемо спочатку $\widehat{\sigma}_{\widehat{\alpha}_0}$, використавши формулу (4.64) і підсумки 4-го та 7-го стовпців табл. 4.12:

$$\widehat{\sigma}_{\widehat{\alpha}_0}=\widehat{\sigma}\sqrt{\sum_{i=1}^{n}x_i^2\Big/\left[\;n\sum_{i=1}^{n}(x_i-\overline{x})^2\right]}=0{,}4165\sqrt{11{,}4058/(12\cdot 1{,}9006)}=0{,}2945.$$

За табл. 4 додатків для $\gamma=0{,}92$ і $n-2=10$ знайдемо $t(0{,}95;10)=2{,}228.$ Тоді використавши співвідношення (4.66) і врахувавши те, що $\widehat{\alpha}_0=0{,}6056,$ отримаємо

$$0{,}6056-2{,}228\cdot 0{,}2945<\alpha_0<0{,}6056+2{,}228\cdot 0{,}2945$$

або остаточно

$$-0{,}0506<\alpha_0<1{,}2618.$$

Аналогічно за формулою (4.63)

$$\widehat{\sigma}_{\widehat{\alpha}_1}=\widehat{\sigma}\Big/\sqrt{\sum_{i=1}^{n}(x_i-\widehat{x}_i)^2}=0{,}4165\Big/\sqrt{1{,}9006}=0{,}3021$$

і співвідношенням (4.65) одержимо

$$5{,}2221-2{,}228\cdot 0{,}3021<\alpha_1<5{,}2221+2{,}228\cdot 0{,}3021$$

або остаточно

$$4{,}5490<\alpha_1<5{,}8952.$$

6) Для побудови довірчої зони для базисних даних обчислимо спочатку значення $\widehat{\sigma}_{\widehat{y}_k}$, використавши формулу (4.68)

$$\widehat{\sigma}_{\widehat{y}_k} = \widehat{\sigma}\left\{ 1\big/ n + (x_k - \overline{x})^2 \Big/ \sum_{i=1}^{n} (x_i - \overline{x})^2 \right\}^{1/2}, \quad k = \overline{1, n},$$

і 7-й стовпець табл. 4.12:

$$\widehat{\sigma}_{\widehat{y}_1} = 0{,}4165\{1/12 + 0{,}3364/1{,}9006\}^{1/2} = 0{,}2125,$$

$$\widehat{\sigma}_{\widehat{y}_2} = 0{,}4165\{1/12 + 0{,}0081/1{,}9006\}^{1/2} = 0{,}1233,$$

$\widehat{\sigma}_{\widehat{y}_3} = 0{,}1543,$ $\widehat{\sigma}_{\widehat{y}_4} = 0{,}1705,$ $\widehat{\sigma}_{\widehat{y}_5} = 0{,}1389,$

$\widehat{\sigma}_{\widehat{y}_6} = 0{,}2175,$ $\widehat{\sigma}_{\widehat{y}_7} = 0{,}1247,$ $\widehat{\sigma}_{\widehat{y}_8} = 0{,}1212,$

$\widehat{\sigma}_{\widehat{y}_9} = 0{,}1705,$ $\widehat{\sigma}_{\widehat{y}_{10}} = 0{,}1726,$ $\widehat{\sigma}_{\widehat{y}_{11}} = 0{,}2303,$

$\widehat{\sigma}_{\widehat{y}_{12}} = 0{,}1581.$

Довірчі інтервали для базисних даних визначаються формулою (4.69)

$$\left(\widehat{y}_k - \widehat{\sigma}_{\widehat{y}_k} t(\gamma; n-2); \widehat{y}_k + \widehat{\sigma}_{\widehat{y}_k} t(\gamma; n-2) \right),$$

де $t(\gamma; n-2) = t(0{,}95,\ 10) = 2{,}228.$

Використавши дані 10-го стовпця табл. 4.12, послідовно отримаємо довірчі інтервали:

$k = 1$ $(2{,}2245 - 0{,}2125 \cdot 2{,}228; 2{,}2245 + 0{,}2125 \cdot 2{,}228) \leftrightarrow (1{,}7511; 2{,}2730),$

$k = 2$ — (5,4486; 5,9980), $k = 3$ — (6,5805; 7,2681),
$k = 4$ — (6,9623; 7,7219), $k = 5$ — (6,1449; 6,7639),
$k = 6$ — (7,9019; 8,8711), $k = 7$ — (4,4001; 4,9557),
$k = 8$ — (5,2443; 5,7844), $k = 9$ — (6,9622; 7,7229),
$k = 10$ — (2,7277; 3,468), $k = 11$ — (1,3458; 2,3720),
$k = 12$ — (3,1256; 3,8300).

Для графічної побудови довірчої зони для базисних даних попередньо потрібно розсортувати значення $x_i (i = \overline{1, n})$ в порядку зростання і кожному значенню x_i поставити у відповідність дві точки з координатами $(x_i; y_i^*)$, $(x_i; y_i^{**})$, де y_i^* та y_i^{**} — лівий та правий кінець i-ого побудованого довірчого інтервалу (для базисних даних). Кожні дві сусідні точки $\left((x_i; y_i^*), (x_{i+1}; y_{i+1}^*) \right)$ та $\left((x_i; y_i^{**}), (x_{i+1}; y_{i+1}^{**}) \right)$ сполучаються прямолінійними відрізками.

7) Знайдемо прогноз значення річного товарообігу для нової філії, торгівельна площа якої складає 1,82 тис. м^2. Для цього в рівняння (4.80) підставимо $x_{\text{п}} = 1{,}82,$ тоді $\widehat{y}_{\text{п}} = 5{,}2221 \cdot 1{,}82 + 0{,}6056 = 10{,}1098.$ За формулою (4.72)

$$\widehat{\sigma}_{\widehat{y}_\text{п}}=\widehat{\sigma}\left[(x_\text{п}-\overline{x})^2\Big/\sum_{i=1}^{n}(x_i-\overline{x})^2+1/n+1\right]^{1/2}=$$

$$=0{,}4165\left[(1{,}82-0{,}89)^2/1{,}9006+1/12+1\right]^{1/2}=0{,}5166,$$

а за співвідношенням (4.71) довірчий інтервал для цього прогнозного значення з надійністю $\gamma=0{,}95$ має вид

$$(10{,}1098-2{,}228\cdot 0{,}5166;10{,}1098+2{,}228\cdot 0{,}5166)$$

або остаточно

$$8{,}9588<y_\text{п}<11{,}2608.\qquad\bullet$$

ЗАДАЧІ ДЛЯ САМОСТІЙНОГО РОЗВ'ЯЗУВАННЯ

Завдання № 7

Задачі №№ 1-50.

Результати спостережень ознак *X* та *Y* об'єктів генеральної сукупності дали результати, наведені в табл. 4.13.

Таблиця 4.13

i	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
x_i	1	2,1	2,9	3,8	5,2	7,3	9,2	12,5	13,8	15,2	2,2	3,1	3,9	5,2	6,2
y_i	1	2,4	3,6	4,2	6,3	7,5	8,1	9,3	10,4	12,3	2,6	3,8	4,3	6,4	7,3
i	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30
x_i	7,4	9,3	12,6	13,9	15,4	2,4	1,5	2,5	3	4,1	5,5	6,4	7	7,5	7,8
y_i	7,7	8,3	9,5	10,5	12,8	2,8	2,2	3,1	4,2	5,4	6,2	7,3	8,2	8,9	9,3
i	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45
x_i	8,2	9,3	9,8	10,4	11	11,5	13,2	5,2	7,4	9,3	12,7	2,2	0,5	1	1,3
y_i	9,8	10,5	11,1	12,2	13,1	10,4	14,3	6,5	8,2	8,9	14,2	2,9	1,4	2,1	2,4
i	46	47	48	49	50	51	52	53	54	55	56	57	58	59	60
x_i	1,5	2	2,5	3,1	5,2	6,4	7,3	8,5	4,7	6,9	5,7	4,1	3,3	2,5	2,1
y_i	2,8	3,1	3,9	4,4	7,1	8,5	9,4	10,6	7,9	6,3	5,7	4,6	2,7	1,6	1,3

Для задачі № k $(k=\overline{1{,}50})$ відібрати десять пар чисел (x_i, y_i), починаючи із $i=k$. На підставі цих даних:

1) скласти систему нормальних рівнянь і знайти коефіцієнти рівняння $\overline{y}_x=\alpha_1 x+a_0$ прямої регресії *Y* на *X*;

2) обчислити вибірковий коефіцієнт кореляції $r_\text{в}$;

3) перевірити правильність статистичної гіпотези $H_0: r=0$ для рівня значущості $\alpha=0{,}05$;

4) якщо гіпотеза з 3) відхилена, тоді для рівня значущості $\alpha = 0{,}05$ перевірити гіпотезу $H_0 : r = r_в$ при конкуруючій гіпотезі $H_1 : r \neq r_в$;

5) побудувати довірчі інтервали для α_1, α_0, r для надійності $\gamma = 0{,}95$.

ДОДАТКИ

Значення функції Гаусса $\varphi(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$

Таблиця 1

x	Соті долі x									
	0	1	2	3	4	5	6	7	8	9
0,0	0,39894	39892	39886	39876	39862	39844	39822	39797	39767	39733
0,1	39695	39654	39608	39559	39505	39448	39387	39322	39253	39181
0,2	39104	39024	38940	38853	38762	38667	38568	38466	38361	38251
0,3	38139	38023	37903	37780	37654	37524	37391	37255	37115	36973
0,4	36827	36678	36526	36371	36213	36053	35889	35723	35553	35381
0,5	35207	35029	34849	34667	34482	34294	34105	33912	33718	33521
0,6	33322	33121	32918	32713	32506	32297	32086	31874	31659	31443
0,7	31225	31006	30785	30563	30339	30114	29887	29659	29431	29200
0,8	28969	28737	28504	28269	28034	27798	27562	27324	27086	26848
0,9	26609	26369	26129	25888	25647	25406	25164	24923	24681	24439
1,0	24197	23955	23713	23471	23230	22988	22747	22506	22265	22025
1,1	21785	21546	21307	21069	20831	20594	20357	20121	19886	19652
1,2	19419	19186	18954	18724	18494	18265	18037	17810	17585	17360
1,3	17137	16915	16694	16474	16256	16038	15822	15608	15395	15183
1,4	14973	14764	14556	14350	14146	13943	13742	13542	13344	13147
1,5	12952	12758	12566	12376	12188	12001	11816	11632	11450	11270
1,6	11092	10915	10741	10567	10396	10226	10059	09893	09728	09566
1,7	09405	09246	09089	08933	08780	08628	08478	08329	08183	08038
1,8	07895	07754	07614	07477	07341	07206	07074	06943	06814	06687
1,9	06562	06438	06316	06195	06077	05959	05844	05730	05618	05508
2,0	05399	05292	05186	05082	04980	04879	04780	04682	04586	04491
2,1	04398	04307	04217	04128	04041	03955	03871	03788	03706	03626
2,2	03547	03470	03394	03319	03246	03174	03103	03034	02965	02898
2,3	02833	02768	02705	02643	02582	02522	02463	02406	02349	02294
2,4	02239	02186	02134	02083	02033	01984	01936	01888	01842	01797
2,5	01753	01709	01667	01625	01585	01545	01506	01468	01431	01394
2,6	01358	01323	01289	01256	01223	01191	01160	01130	01100	01071
2,7	01042	01014	00987	00961	00935	00909	00885	00861	00837	00814
2,8	00792	00770	00748	00727	00707	00687	00668	00649	00631	00613
2,9	00595	00578	00562	00545	00530	00514	00499	00485	00470	00457
3,0	00443	00430	00417	00405	00393	00381	00370	00358	00348	00337
3,1	00327	00317	00307	00298	00288	00279	00271	00262	00254	00246
3,2	00238	00231	00224	00216	00210	00203	00196	00190	00184	00178
3,3	00172	00167	00161	00156	00151	00146	00141	00136	00132	00127

Продовження табл. 1

x	Соті долі x									
	0	1	2	3	4	5	6	7	8	9
3,4	00123	00119	00115	00111	00107	00104	00100	00097	00094	00090
3,5	00087	00084	00081	00079	00076	00073	00071	00068	00066	00063
3,6	00061	00059	00057	00055	00053	00051	00049	00047	00046	00044
3,7	00042	00041	00039	00038	0037	00035	00034	00033	00031	00030
3,8	00029	00028	00027	00026	00025	00024	00023	00022	00021	00021
3,9	00020	00019	00018	00018	00017	00016	00016	00015	00014	00014

x	Десяті долі x				
	0	2	4	6	8
4,	0,0001338	0000589	0000249	0000101	0000040
5,	0000015				

Значення функції $P(m;\lambda) = \dfrac{\lambda^m e^{-\lambda}}{m!}$

Таблиця 2

m	λ									
	0,1	0,2	0,3	0,4	0,5	0,6	0,7	0,8	0,9	1,0
0	0,90484	81873	74082	67032	60653	54881	49659	44933	40657	36788
1	09048	16375	22225	26813	30327	32929	34761	35946	36591	36788
2	00452	01637	03334	05363	07582	09879	12166	14379	16466	18394
3	00015	00109	00333	00715	01264	01976	02839	03834	04940	06131
4		00005	00025	00072	00158	00296	00497	00767	01111	01533
5			00002	00006	00016	00036	00070	00123	00200	00307
6					00001	00004	00008	00016	00030	00051
7							00001	00002	00004	00007
8										00001

Продовження табл. 2

m	λ									
	1,5	2,0	2,5	3,0	3,5	4,0	4,5	5,0	5,5	6,0
0	0,22313	13534	08208	04979	03020	01832	01111	00674	00409	00248
1	33470	27067	20521	14936	10569	07326	04999	03369	02248	01487
2	25102	27067	25652	22404	18496	14653	11248	08422	06181	04462
3	12551	18045	21376	22404	21579	19537	16872	14037	11332	08924
4	04707	09022	13360	16803	18881	19537	18981	17547	15582	13385
5	01412	03609	06680	10082	13217	15629	17083	17547	17140	16062
6	00353	01203	02783	05041	07710	10420	12812	14622	15712	16062
7	00076	00344	00994	02160	03855	05954	08236	10444	12345	13768
8	00014	00086	00311	00810	01687	02977	04633	06528	08487	10326
9	00002	00019	00086	00270	00656	01323	02316	03627	05187	06884
10		00004	00022	00081	00230	00529	01042	01813	02853	04130
11		00001	00005	00022	00073	00192	00426	00824	01426	02253
12			00001	00006	00021	00064	00160	00343	00654	01126
13				00001	00006	00020	00055	00132	00277	00520
14					00001	00006	00018	00047	00109	00223
15						00002	00005	00016	00040	00089
16							00002	00005	00014	00033
17								00001	00004	00012
18									00001	00004
19										00001

Значення функції Лапласа $\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_0^x e^{\frac{t^2}{2}} dt$

Таблиця 3

x	Соті долі *x*									
	0	1	2	3	4	5	6	7	8	9
0,0	0,00000	00399	00798	01197	01595	01994	02392	02790	03188	03586
0,1	03983	04380	04776	05172	05567	05962	06356	06749	07142	07535
0,2	07926	08317	08706	09095	09483	09871	10257	10642	11026	11409
0,3	11791	12172	12552	12930	13307	13683	14058	14431	14803	15173
0,4	15542	15910	16276	16640	17003	17364	17724	18082	18439	18793
0,5	19146	19497	19847	20194	20540	20884	21226	21566	21904	22240
0,6	22575	22907	23237	23565	23891	24215	24537	24857	25175	25490
0,7	25804	26115	26424	26730	27035	27337	27637	27935	28230	28524
0,8	28814	29103	29389	29673	29955	30234	30511	30785	31057	31327
0,9	31594	31859	32121	32381	32639	32894	33147	33398	33646	33891

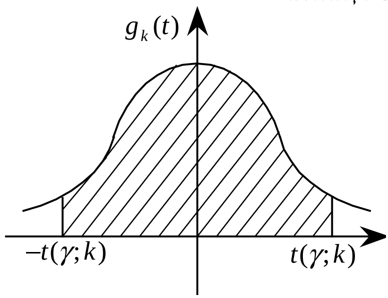
Продовження табл. 3

x	Соті долі *x*									
	0	1	2	3	4	5	6	7	8	9
1,0	34134	34375	34614	34850	35083	35314	35543	35769	35993	36214
1,1	36433	36650	36864	37076	37286	37493	37698	37900	38100	38298
1,2	38493	38686	38877	39065	39251	39435	39617	39796	39973	40147
1,3	40320	40490	40658	40824	40988	41149	41308	41466	41621	41774
1,4	41924	42073	42220	42364	42507	42647	42786	42922	43056	43189
1,5	43319	43448	43574	43699	43822	43943	44062	44179	44295	44408
1,6	44520	44630	44738	44845	44950	45053	45154	45254	45352	45449
1,7	45543	45637	45728	45818	45907	45994	46080	46164	46246	46327
1,8	46407	46485	46562	46638	46712	46784	46856	46926	46995	47062
1,9	47128	47193	47257	47320	47381	47441	47500	47558	47615	47670
2,0	47725	47778	47831	47882	47932	47982	48030	48077	48124	48169
2,1	48214	48257	48300	48341	48382	48422	48461	48500	48537	48574
2,2	48610	48645	48679	48713	48745	48778	48809	48840	48870	48899
2,3	48928	48956	48983	49010	49036	49061	49086	49111	49134	49158
2,4	49180	49202	49224	49245	49266	49286	49305	49324	49343	49361
2,5	49379	49396	49413	49430	49446	49461	49477	49492	49506	49520
2,6	49534	49547	49560	49573	49585	49598	49609	49621	49632	49643
2,7	49653	49664	49674	49683	49693	49702	49711	49720	49728	49736
2,8	49744	49752	49760	49767	49774	49781	49788	49795	49801	49807
2,9	49813	49819	49825	49831	49836	49841	49846	49851	49856	49861
3,0	49865	49869	49874	49878	49882	49886	49889	49893	49897	49900
3,1	49903	49906	49910	49913	49916	49918	49921	49924	49926	49929
3,2	49931	49934	49936	49938	49940	49942	49944	49946	49948	49950
3,3	49952	49953	49955	49957	49958	49960	49961	49962	49964	49965
3,4	49966	49968	49969	49970	49971	49972	49973	49974	49975	49976
3,5	49977	49978	49978	49979	49980	49981	49981	49982	49983	49983
3,6	49984	49985	49985	49986	49986	49987	49987	49988	49988	49989
3,7	49989	49990	49990	49990	49991	49991	49992	49992	49992	49992
3,8	49993	49993	49993	49994	49994	49994	49994	49995	49995	49995
3,9	49995	49995	49996	49996	49996	49996	49996	49996	49997	49997

x	Десяті долі *x*				
	0	2	4	6	8
4,	0,4999683	4999867	4999946	4999979	4999992
5,	4999997				

Таблиця 4

Значення $t = t(\gamma; k)$,
що задовільняють рівнянню
$$P\left(\,|T| < t\right) = 2\int_0^t g_k(t)dt = \gamma,$$
де $g_k(t)$ — густина **розподілу Ст'юдента (*t*-розподілу)**, $k = n-1$ — число ступенів вільності



$k = n-1$	γ				
	0,9	0,95	0,98	0,99	0,999
1	6,314	12,706	31,821	63,657	636,619
2	2,920	4,303	6,965	9,925	31,599
3	2,353	3,182	4,541	5,841	12,924
4	2, 132	2,776	3,747	4,604	8,610
5	2,015	2,571	3,365	4,032	6,869
6	1,943	2,447	3,143	3,707	5,969
7	1,895	2,365	2,998	3,499	5,408
8	1,860	2,306	2,896	3,355	5,041
9	1,833	2,262	2,821	3,250	4,781
10	1,812	2,228	2,764	3,169	4,587
11	1,796	2,201	2,718	3,106	4,437
12	1,782	2,179	2,681	3,055	4,318
13	1,771	2,160	2,650	3,012	4,221
14	1,761	2,145	2,624	2,977	4,140
15	1,753	2,131	2,602	2,947	4,073
16	1,746	2,120	2,583	2,921	4,015
17	1,740	2,110	2,567	2,898	3,965
18	1,734	2,101	2,552	2,878	3,922
19	1,729	2,093	2,539	2,861	3,883
20	1,725	2,086	2,528	2,845	3,850
25	1,708	2,060	2,485	2,785	3,725
30	1,697	2,042	2,457	2,750	3,646
40	1,684	2,021	2,423	2,704	3,551
50	1,676	2,009	2,403	2,678	3,496
60	1,671	2,000	2,390	2,660	3,460
70	1,667	1,994	2,381	2,648	3,435

Продовження табл. 4

80	1,664	1,990	2,374	2,639	3,416
90	1,662	1,987	2,368	2,632	3,402
100	1,660	1,984	2,364	2,626	3,390
120	1,658	1,980	2,358	2,617	3,373
∞	1,645	1,960	2,326	2,576	3,291

Таблиця 5

Критичні точки розподілу Ст'юдента (*t*-розподілу)

Для двосторонньої критичної області критична точка $t_{\text{двост.кр}}(\alpha; k) = t_\alpha$ є коренем рівняння $\int_0^{t_\alpha} g_k(t)dt = (1-\alpha)/2$; для односторонньої (правосторонньої) критичної області точка $t_{\text{правост.кр}}(\alpha; k) = t_{2\alpha}$ є коренем рівняння $\int_{t_{2\alpha}}^{\infty} g_k(t)dt = (1-2\alpha)/2$, де $g_k(t)$ — густина розподілу Ст'юдента, $k = n-1$ — число ступенів вільності. Для лівосторонньої критичної області $t_{\text{лівост.кр}}(\alpha; k) = -t_{2\alpha}$.

Число ступенів вільності $k = n-1$	Рівень значущості α (**двостороння** критична область)					
	0,10	0,05	0,02	0,01	0,002	0,001
1	6,314	12,71	31,82	63,66	318,3	637,0
2	2,920	4,303	6,965	9,925	22,33	31,60
3	2,353	3,182	4,541	5,841	10,22	12,94
4	2,132	2,776	3,747	4,604	7,173	8,610
5	2,015	2,571	3,365	4,032	5,893	6,859
6	1,943	2,447	3,143	3,707	5,208	5,959
7	1,895	2,365	2,998	3,499	4,785	5,405
8	1,860	2,306	2,896	3,355	4,501	5,041
9	1,833	2,262	2,821	3,250	4,297	4,781
10	1,812	2,228	2,764	3,169	4,144	4,587
11	1,796	2,201	2,718	3,106	4,025	4,437
12	1,782	2,179	2,681	3,055	3,930	4,318
13	1,771	2,160	2,650	3,012	3,852	4,221
14	1,761	2,145	2,624	2,977	3,787	4,140

Продовження табл.5

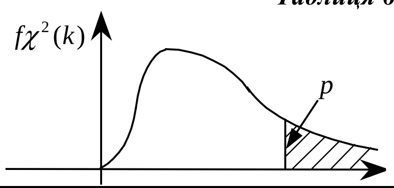
15	1,753	2,131	2,602	2,947	3,733	4,073
16	1,746	2,120	2,583	2,921	3,686	4,015
17	1,740	2,110	2,567	2,898	3,646	3,965
18	1,734	2,101	2,552	2,878	3,611	3,922
19	1,729	2,093	2,539	2,861	3,579	3,883
20	1,725	2,086	2,528	2,845	3,562	3,850
21	1,721	2,080	2,518	2,831	3,527	3,819
22	1,717	2,074	2,508	2,819	3,505	3,792
23	1,714	2,069	2,500	2,807	3,485	3,767
24	1,711	2,064	2,492	2,797	3,467	3,745
25	1,708	2,060	2,485	2,787	3,450	3,725
26	1,706	2,056	2,479	2,779	3,435	3,707
27	1,703	2,052	2,473	2,771	3,421	3,690
28	1,701	2,048	2,467	2,763	3,408	3,674
29	1,699	2,045	2,462	2,756	3,396	3,659
30	1,697	2,042	2,457	2,750	3,385	3,646
40	1,684	2,021	2,423	2,704	3,307	3,551
60	1,671	2,000	2,390	2,660	3,232	3,460
120	1,658	1,981	2,362	2,624	3,172	3,374
∞	1,645	1,960	2,326	2,576	3,090	3,291
Число ступенів вільності $k = n - 1$	0,05	0,025	0,01	0,005	0,001	0,0005
	Рівень значущості α (**одностороння** критична область)					

Таблиця 6

Значення

$P\left(\chi^2(k) > \chi^2(p;k) \right) = p,$

де k — число ступенів вільності



k	p							
	0,999	0,99	0,95	0,90	0,10	0,05	0,01	0,001
1	$0{,}157 \cdot 10^{-5}$	0,0002	0,004	0,02	2,71	3,84	6,63	10,83
2	0,002	0,02	0,10	0,21	4,61	5,99	9,21	13,82

Продовження табл. 6

3	0,02	0,12	0,35	0,58	6,25	7,82	11,34	16,27
4	0,09	0,30	0,71	1,06	7,78	9,49	13,28	18,47
5	0,21	0,55	1,15	1,61	9,24	11,07	15,08	20,51
6	0,38	0,87	1,64	2,20	10,65	12,59	16,81	22,46
7	0,60	1,24	2,17	2,83	12,02	14,06	18,48	24,32
8	0,86	1,65	2,73	3,49	13,36	15,51	20,09	26,12
9	1,15	2,09	3,33	4,17	14,68	16,92	21,67	27,88
10	1,48	2,56	3,94	4,87	15,99	18,31	23,21	29,59
11	1,83	3,05	4,58	5,58	17,28	19,68	24,72	31,26
12	2,21	3,57	5,23	6,30	18,55	21,03	26,22	32,91
13	2,62	4,11	5,89	7,04	19,81	22,36	27,69	34,53
14	3,04	4,66	6,57	7,79	21,06	23,69	29,14	36,12
15	3,48	5,23	7,26	8,55	22,31	25,00	30,58	37,70
16	3,94	5,81	7,96	9,31	23,54	26,30	32,00	39,25
17	4,42	6,41	8,67	10,09	24,77	27,59	33,41	40,79
18	4,90	7,02	9,39	10,86	25,99	28,87	34,81	42,31
19	5,41	7,63	10,12	11,65	27,20	30,14	36,19	43,82
20	5,92	8,26	10,85	12,44	28,41	31,41	37,57	45,31
21	6,45	8,90	11,59	13,24	29,62	32,67	38,93	46,80
22	6,98	9,54	12,34	14,04	30,81	33,92	40,29	48,27
23	7,53	10,20	13,20	14,85	32,01	35,17	41,64	49,73
24	8,08	10,86	13,85	15,66	33,19	36,42	43,98	51,18
25	8,65	11,52	14,61	16,47	34,38	37,65	44,31	52,62
26	9,22	12,20	15,37	17,29	35,56	38,89	45,64	54,05
27	9,80	12,88	16,15	18,11	36,74	40,11	46,96	55,48
28	10,39	13,56	16,93	18,94	37,92	41,34	48,28	56,89
29	10,99	14,26	17,71	19,77	39,09	42,56	49,59	58,30
30	11,59	14,95	18,49	20,60	40,26	43,77	50,89	59,70
40	17,92	22,16	26,51	29,05	51,81	55,76	63,69	73,40
50	24,67	29,71	34,76	37,69	63,17	67,51	76,15	86,66
100	61,92	70,07	77,93	82,36	118,50	124,34	135,81	149,45

Таблиця 7

Критичні точки $F_{кр}(\alpha; k_1, k_2)$ розподілу Фішера-Снедокора, що задовільняють рівнянню $P\left[\, F > F_{кр}(\alpha; k_1, k_2)\right] = \alpha$ при $\alpha = 0{,}05$

k_2	k_1									
	1	2	3	4	5	6	8	12	24	∞
1	161,45	199,50	215,71	224,58	230,16	233,99	238,88	243,91	249,05	254,32
2	18,51	19,00	19,16	19,25	19,30	19,33	19,37	19,41	19,45	19,50
3	10,13	9,55	9,28	9,12	9,01	8,94	8,84	8,74	8,64	8,53
4	7,71	6,94	6,59	6,39	6,26	6,16	6,04	5,91	5,77	5,63
5	6,61	5,79	5,41	5,19	5,05	4,95	4,82	4,68	4,53	4,36
6	5,99	5,14	4,76	4,53	4,39	4,28	4,15	4,00	3,84	3,67
7	5,59	4,74	4,35	4,12	3,97	3,87	3,73	3,57	3,41	3,23
8	5,32	4,46	4,07	3,84	3,69	3,58	3,44	3,28	3,12	2,93
9	5,12	4,26	3,86	3,63	3,48	3,37	3,23	3,07	2,90	2,71
10	4,96	4,10	3,71	3,48	3,33	3,22	3,07	2,91	2,74	2,54
11	4,84	3,98	3,59	3,36	3,20	3,09	2,95	2,79	2,61	2,40
12	4,75	3,88	3,49	3,26	3,11	3,00	2,85	2,69	2,50	2,30
13	4,67	3,80	3,41	3,18	3,02	2,92	2,77	2,60	2,42	2,21
14	4,60	3,74	3,34	3,11	2,96	2,85	2,70	2,53	2,35	2,13
15	4,54	3,68	3,29	3,06	2,90	2,79	2,64	2,48	2,29	2,07
16	4,49	3,63	3,24	3,01	2,85	2,74	2,59	2,42	2,24	2,01
17	4,45	3,59	3,20	2,96	2,81	2,70	2,55	2,38	2,19	1,96
18	4,41	3,55	3,16	2,93	2,77	2,66	2,51	2,34	2,15	1,92
19	4,38	3.52	3,13	2,90	2,74	2,63	2,48	2,31	2,11	1,88
20	4,35	3,49	3,10	2,87	2,71	2,60	2,45	2,28	2,08	1,84
25	4,24	3,38	2,99	2,76	2,60	2.49	2,34	2.16	1,96	1,71
30	4,17	3,32	2,92	2,69	2,53	2,42	2,27	2,09	1,89	1,62
40	4,08	3,23	2,84	2,61	2,45	2,34	2,18	2,00	1,79	1,51
60	4,00	3,15	2,76	2,52	2,37	2,25	2,10	1,92	1,70	1,39
120	3,92	3,07	2,68	2,45	2,29	2,17	2,02	1,83	1,61	1,25
∞	3,84	2,99	2,60	2,37	2,21	2,09	1,94	1,75	1,52	1,00

Таблиця 8

Критичні значення $K_{n;\alpha}$ для стаитистики критерія Колмогорова

n	α				n	α			
	0,10	0,05	0,02	0,01		0,10	0,05	0,02	0,01
1	0,950	0,975	0,990	0,995	51	0,168	0,187	0,208	0,224
2	776	842	900	929	52	166	185	207	222
3	636	708	789	829	53	165	183	205	220
4	565	624	689	734	54	163	181	203	218
5	510	563	627	669	55	162	180	201	216
6	468	519	577	617	56	160	178	199	214
7	436	483	538	576	57	159	177	198	212
8	410	454	507	542	58	158	175	196	210
9	388	430	480	513	59	156	174	194	208
10	369	409	457	489	60	155	172	193	207
11	352	391	437	468	61	154	171	191	205
12	338	375	419	449	62	153	170	190	203
13	326	361	404	433	63	151	168	188	202
14	314	349	390	418	64	150	167	187	200
15	304	338	377	404	65	149	166	185	199
16	295	327	366	392	66	148	164	184	197
17	286	318	355	381	67	147	163	183	196
18	279	309	346	371	68	146	162	180	194
19	271	301	337	361	69	145	161	181	193
20	265	294	329	352	70	144	160	179	192
21	259	287	321	344	71	143	159	177	190
22	253	281	314	337	72	142	158	176	189
23	248	275	307	330	73	141	157	175	188
24	242	269	301	323	74	140	155	174	187
25	238	264	295	317	75	139	154	173	185
26	233	259	290	311	76	138	153	172	184
27	229	254	284	305	77	137	152	171	183
28	225	250	279	300	78	136	152	169	182
29	221	246	275	295	79	136	151	168	181
30	218	242	270	290	80	135	150	167	180
31	215	238	266	285	81	134	149	166	178
32	211	234	262	281	82	133	148	165	177
33	208	231	258	277	83	132	147	164	176
34	205	227	254	273	84	132	146	163	175

Продовження табл. 8

35	202	224	251	269	85	131	145	162	174
36	199	221	247	265	86	130	144	161	173
37	196	218	244	262	87	129	144	161	172
38	194	215	241	258	88	129	143	160	171
39	192	213	238	255	89	128	142	159	170
40	189	210	235	252	90	127	141	158	169
41	187	208	232	249	91	126	140	157	169
42	185	205	229	246	92	126	140	156	168
43	183	203	227	243	93	125	139	155	167
44	181	201	224	241	94	124	138	155	166
45	179	198	222	238	95	124	138	154	165
46	177	196	219	235	96	123	137	153	164
47	175	194	217	233	97	123	136	152	163
48	173	192	215	231	98	122	135	151	162
49	171	190	213	228	99	121	135	151	162
50	170	188	211	226	100	121	134	150	161

Таблиця 9

Асимптотичні критичні значення $n_{1-\alpha}$
для статистики Колмогорова і Смирнова

α	0,20	0,10	0,05	0,02	0,01
$n_{1-\alpha}$	1,073	1,224	1,358	1,517	1,628

Таблиця 10

Перетворення Фішера (Z-перетворення)

$$Z = \frac{1}{2}\ln\frac{1+r_в}{1-r_в}.$$

$r_в$	Соті долі $r_в$									
	0	1	2	3	4	5	6	7	8	9
0,0	0,0000	0,0100	0,0200	0,0300	0,0400	0,0501	0,0601	0,0701	0,0802	0,0902
0,1	0,1003	0,1105	0,1206	0,1308	0,1409	0,1511	0,1614	0,1717	0,1820	0,1923
0,2	0,2027	0,2132	0,2237	0,2342	0,2448	0,2561	0,2661	0,2769	9,2877	0,2986
0,3	0,3095	0,3206	0,3317	0,3428	0,3541	0,3654	0,3769	0,3884	0,4001	0,4118
0,4	0,4236	0,4356	0,4477	0,4599	0,4722	0,4847	0,4973	0,5101	0,5230	0,5361
0,5	0,5493	0,5627	0,5763	0,5901	0,6042	0,6184	0,6328	0,6475	0,6625	0,6777
0,6	0,6931	0,7089	0,7250	0,7414	0,7582	0,7753	0,7928	0,8107	0,8291	0,8480
0,7	0,8673	0,8872	0,9076	0,9287	0,9505	0,9730	0,9962	1,0203	1,0454	1,0714
0,8	1,0986	1,1270	1,1568	1,1881	1,2212	1,2562	1,2933	1,3331	1,3758	1,4219
0,9	1,4722	1,5275	1,5890	1,6584	1,7380	1,8318	1,9459	2,0923	2,2976	2,6467

Таблиця 11

Обернене перетворення Фішера $r_в = \text{th}Z$.

Z	Соті долі Z									
	0	1	2	3	4	5	6	7	8	9
0,0	0,0000	0,0100	0,0200	0,0300	0,0400	0,0500	0,0599	0,0699	0,0799	0,0898
0,1	0997	1096	1194	1293	1391	1489	1586	1684	1781	1877
0,2	1974	2070	2165	2260	2355	2449	2543	2636	2729	2821
0,3	2913	3004	3095	3185	3275	3364	3452	3540	3627	3714
0,4	3800	3885	3969	4053	4136	4219	4301	4382	4462	4542
0,5	4621	4699	4777	4854	4930	5005	5080	5154	5227	5299
0,6	5370	5441	5511	5580	5649	5717	5784	5850	5915	5980
0,7	6044	6107	6169	6231	6291	6351	6411	6469	6527	6584
0,8	6640	6696	6751	6805	6858	6911	6963	7014	7064	7114
0,9	7163	7211	7259	7306	7352	7398	7443	7487	7531	7574
1,0	7616	7658	7699	7739	7779	7818	7857	7895	7932	7969
1,1	8005	8041	8076	8110	8144	8178	8210	8243	8275	8306
1,2	8337	8367	8397	8426	8455	8483	8511	8538	8565	8591
1,3	8617	8643	8668	8692	8717	8741	8764	8787	8810	8832
1,4	8854	8875	8896	8917	8937	8957	8977	8996	9015	9033
1,5	9051	9069	9087	9104	9121	9138	9154	9170	9186	9201
1,6	9217	9232	9246	9261	9275	9289	9302	9316	9329	9341
1,7	9354	9366	9379	9391	9402	9414	9425	9436	9447	9458
1,8	9468	9478	9488	9498	9508	9518	9527	9536	9545	9554
1,9	9562	9571	9579	9587	9595	9603	9611	9618	9626	9633
2,0	9640	9647	9654	9661	9668	9674	9680	9686	9693	9699
2,1	9704	9710	9716	9722	9727	9732	9738	9743	9748	9753
2,2	9757	9762	9767	9771	9776	9780	9785	9789	9793	9797
2,3	9801	9805	9809	9812	9816	9820	9823	9827	9830	9834
2,4	9837	9840	9843	9846	9849	9852	9855	9858	9861	9864
2,5	9866	9869	9871	9874	9876	9879	9881	9884	9886	9888
2,6	9890	9892	9894	9897	9899	9901	9903	9904	9906	9908
2,7	9910	9912	9914	9915	9917	9919	9920	9922	9923	9925
2,8	9926	9928	9929	9931	9932	9933	9935	9936	9937	9938
2,9	9940	9941	9942	9943	9944	9945	9946	9947	9948	9949

ЛІТЕРАТУРА

1. *Бочаров П. П., Печенкин А. В.* **Теория вероятностей. Математическая статистика**. — М.: Гардарика, 1998. — 328 с.
2. *Бугір М. К.* **Практикум з теорії імовірностей та математичної статистики.** — Тернопіль: ЦМДС, 1998. — 171 с.
3. *Булдик Г. М.* **Теория вероятностей и математичская статистика.** — Минск: Вышэйшая школа, 1989. — 285 с.
4. *Гмурман В. Е.* **Руководство к решению задач по теории вероятностей и математической статистике.** — М.: Высш. шк., 1979. — 400 с.
5. *Гмурман В. Е.* **Теория вероятностей и математическая статистика.** — М.: Высш. шк., 2004. — 479 с.
6. *Грубер Й.* **Эконометрия.** — Том 1. **Введение в эконометрию.** — К.: Астарта, 1996. — 397 с.
7. *Єрьоменко В. О., Шинкарик М. І.* **Теорія імовірностей.** — Тернопіль: Економічна думка, 2000. — 176 с.
8. *Єрьоменко В. О., Шинкарик М. І.* **Математична статистика.** — Тернопіль: Економічна думка, 2002. — 247 с.
9. *Єрьоменко В. О., Шинкарик М. І., Бабій Р. М., Процик А. І.* **Практикум з теорії імовірностей та математичної статистики** / Навчальний посібник для студентів економічних спеціальностей. — Тернопіль: Економічна думка, 2006. — 320 с.
10. *Жлуктенко В. І., Наконечний С. І.* **Теорія ймовірностей і математична статистика:** Навч. метод. посібник. У 2ч. — ч.І. Теорія ймовірностей. — К.: КНЕУ, 2000. — 304 с.
11. *Карасев А. И.* **Теория вероятностей и математическая статистика.** — М.: Статистика, 1977. — 279 с.
12. *Кибзун А.И., Горяинов Е.Р. Наумов А.В., Сиротин А.Н.* **Теория вероятностей и математическая статистика. Базовый курс с примерами и задачами** / Учебн. пособие. М.: ФИЗМАТЛИТ, 2002. —224с.
13. *Коваленко И. Н., Гнеденко Б. В.* **Теория вероятностей.** — К.: Вища школа, 1990. — 328 с.

14. *Колемаев В. А., Калинина В. Н.* **Теория вероятностей и математическая статистика:** Учебник / Под ред. В. А. Колемаева. — М.: ИНФРА-М, 2000. — 302 с.

15. *Колемаев В. А., Староверов О. В., Турундаевский В. Б.* **Теория вероятностей и математическая статистика.** — М.: Высш. шк., 1991. — 400 с.

16. *Кремер М.Ш.* **Теория вероятностей и математическая статистика.** — М.: ЮНИТИ-ДАНА, 2002. — 543 с.

17. *Мармоза А.Т.* **Практикум з математичної статистики:** Навч. посібник. —К.: Кондор, 2004. —264с.

18. *Мюллер П., Найман П., Шторм Р.* **Таблицы по математической статистике.** — М.: Финансы и статистика, 1982. — 278 с.

19. *Румшиский Л. З.* **Математическая обработка результатов эксперимента.** — М.: Наука, 1971. — 192 с.

20. **Статистика підприємництва:** Навч. посібник. — Вашків П. Г., Пастер П. І., Сторожук В. П., Ткач Є. І. — К.: Слобожанщина, 1999. — 600 с.

21. *Черняк І.О., Обушна О.М., Ставицький А.В.* **Теорія ймовірностей та математична статистика:** Збірник задач: Навч. посібник. — К.: Т-во “Знання”, КОО, 2001. — 199 с. — (Вища освіта XXI століття).

ЗМІСТ

Передмова.. 3

ЧАСТИНА ПЕРША. ТЕОРІЯ ІМОВІРНОСТЕЙ

§ 1. Визначення імовірності 4
Завдання №1 24
§ 2. Теореми множення і додавання імовірностей та їх наслідки.. 30
Завдання №2.. 54
Завдання №3.. 59
§ 3. Повторні незалежні випробування.................................. 67
Завдання №4.. 82
§ 4. Дискретні випадкові величини та їх числові характеристики.. 89
Завдання №5.. 114
§ 5. Неперервні випадкові величини та їх числові характеристики.. 120
Завдання №6.. 134
§ 6. Основні закони неперервних випадкових величин.......... 138
Завдання №7.. 150
§ 7. Системи випадкових величин (багатовимірні випадкові величини)…………………… 156
Завдання №8.. 176
§ 8. Функція випадкових величин……………………… 183
Завдання №9.. 189
§ 9. Закон великих чисел.. 191
Завдання №10.. 200

ЧАСТИНА ДРУГА. МАТЕМАТИЧНА СТАТИСТИКА

§ 1. Вступ в математичну статистику. Вибірковий метод 205
Завдання №1.. 229
Завдання №2.. 230
§ 2. Статистичне оцінювання.. 236
Завдання №3.. 252
Завдання №4.. 254
Завдання №5.. 255
§ 3. Статистична перевірка статистичних гіпотез.................. 257
Завдання №6.. 283
§ 4. Елементи кореляційного та регресійного аналізу........... 284
Завдання №7.. 334
Додатки.. 336
Література.. 349

Навчальне видання

***Алілуйко** Андрій Миколайович*
***Дзюбановська** Наталія Володимирівна*
***Єрьоменко** Валерій Олександрович*
***Мартинюк** Олеся Миронівна*
***Шинкарик** Микола Іванович*

ПРАКТИКУМ З ТЕОРІЇ ІМОВІРНОСТЕЙ ТА МАТЕМАТИЧНОЇ СТАТИСТИКИ

Навчальний посібник
для студентів економічних спеціальностей

Виготовлено згідно із СОУ 22.2-02477019-07:2012

Формат 60×84/16. 20,46 ум. др. арк., 19,88 обл.-вид. арк. Тираж 300. Замовлення № 18-945.
Видавець і виготовлювач Редакція газети «Підручники і посібники».
46000, м. Тернопіль, вул. Поліська, 6а. Тел.: (0352) 43-15-15; 43-10-21.
Збут: pip.ternopil@ukr.net Редакція: editoria@i.ua
www.pp-books.com.ua
Свідоцтво про внесення суб'єкта видавничої справи
до Державного реєстру видавців, виготівників і розповсюджувачів видавничої продукції
серія ДК № 4678 від 21.01.2014 р.
Книга-поштою: а/с 376, Тернопіль, 46011.
Тел.: (0352) 42-43-76; 097-50-35-376
pip.bookpost@gmail.com