МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ

ЗАХІДНОУКРАЇНСЬКИЙ НАЦІОНАЛЬНИЙ УНІВЕРСИТЕТ

Методичні вказівки для самостійної роботи студентів з курсу

«ТЕОРІЯ ІМОВІРНОСТЕЙ І МАТЕМАТИЧНА СТАТИСТИК

Тернопіль — 2022

Укладачі: ***Мартинюк Олеся Миронівна,***
кандидат фіз. -мат. наук, доцент
Єрьоменко Валерій Олександрович,
кандидат фіз. -мат. наук, доцент
Шинкарик Микола Іванович,
кандидат фіз. -мат. наук, доцент
Березька Катерина Миколаївна,
кандидат технічних наук, доцент
Руська руслана Василівна,
кандидат економічних наук, доцент
Пласконь Світлана Андріївна,
кандидат економічних наук, доцент

Затверджено на засіданні кафедри прикладної математики,
протокол № 1 від 26.08.2022 р.

Березька К. М., Мартинюк О. М., Пласконь С. А., Руська Р. В., Шинкарик М. І. Методичні рекомендації для самостійної роботи студентів з навчальної дисципліни «ТІМС» для студентів усіх спеціальностей денної та заочної форм навчання першого (бакалаврського) рівня вищої освіти. Тернопіль, ТНЕУ, 2022. 48 с.

Для студентів особливе значення має самостійна робота над дисципліною. Вивчення дисципліни ТІМС вимагає від студенів систематичної клопіткої праці над навчально-методичною і науковою літературою, приділення додаткового часу розв’язуванню задач.

Самостійна робота має на меті глибше й повніше засвоєння теоретичного матеріалу математичного інструментарію методів теорії ймовірності та математичної статистики та оволодіння практичними навиками при розв’язуванні задач.

Кожний студент самостійно знаходить у запропонованому переліку рекомендованої літератури матеріал, який стосується проблемних питань до кожної теми і опрацьовує його.

Після завершення вивчення окремих тем та курсу загалом студент перед практичними заняттями та здачею заліку самостійно перевіряє грунтовність засвоєння теоретичного матеріалу з допомогою відповідей на запитання приведених у навчальних посібниках після відповідних тем та тестових завдань.

Крім цього, для перевірки ґрунтовності і закріплення отриманих знань на лекціях, практичних заняттях та в результаті самостійного вивчення матеріалу, а також для набуття навиків застосування математичного апарату методів ТІМС студент самостійно розв'язує запропоновані задачі. При самостійному розв'язуванні задач у розрізі основних тем курсу студент керується поданими до кожної теми методичними вказівками про послідовність дій та алгоритму їх розв'язання. Подання учбових завдань відповідає послідовності вивчення теоретичного матеріалу описаного в робочій програмі з курсу

Тематика самостійної роботи студентів

1. Класичне означення ймовірності, її властивості.
2. Класифікація подій. Випадкові події, їх класифікація.
3. Залежні і незалежні події. Умовна ймовірність. Теорема множення ймовірностей.
4. Теорема додавання ймовірностей. Наслідки з неї.
5. Повна група подій, протилежні події, їх властивості.
6. Формула повної ймовірності.
7. Формули Байєса.
8. Повторні незалежні випробовування. Формула Бернуллі.
9. Локальна формула Лапласа. Функція Гауса, її властивості.
10. Інтегральна формула Лапласа. Функція Лапласа, її властивості.
11. Формула Пуассона.
12. Найімовірніше число настання події в повторних незалежних випробовуваннях.
13. Ймовірність відхилення відносної частоти від сталої ймовірності в повторних незалежних випробовуваннях.
14. Види випадкових величин. Числові характеристики дискретної величини.
15. Математичне сподівання випадкової дискретної величини, його властивості.
16. Дисперсія випадкової величини, її властивості.
17. Функція розподілу ймовірності випадкової величини, її властивості.
18. Числові характеристики неперервних випадкових величин.
19. Інтегральна функція розподілу ймовірності випадкової величини та її властивості
20. Нормальний закон розподілу, ймовірностний зміст його параметрів. Крива нормального розподілу.
21. Ймовірність попадання нормально-розподіленої величини в заданий інтервал.
22. Ймовірність відхилення нормально-розподіленої величини від свого математичного сподівання.
23. Знаходження числових характеристик у загальному випадку для цілочисельних дискретних випадкових величин (рівномірний, пуассонівський, геометричний розподіли).
24. Закон розподілу імовірностей двовимірної дискретної випадкової величини.
25. Функція розподілу двовимірної випадкової величини та її випадковості.
26. Густина розподілу імовірностей двовимірної випадкової величини та її властивості.
27. Умовні закони розподілу.
28. Залежні та незалежні випадкові величини.
29. Умовне математичне сподівання. Рівняння регресії.

30. Числові характеристики системи двох випадкових величин.
31. Кореляційний момент. Коефіцієнт кореляції.
32. Система довільного скінченого числа випадкових величин.
33. Кореляційна матриця.
34. Нормальний закон розподілу двовимірної випадкової величини.
35. Функція одного випадкового аргументу та її математичне сподівання.
36. Логарифмічний нормальний закон та χ^2-розподіл.
37. Функції двох випадкових величин.
38. Розподіл Ст'юдента, розподіл Фішера-Снедокора .
39. Нерівність Чебишева.
40. Теорема Чебишева.
41. Закон великих чисел. Теорема Бернуллі.
42. Числові характеристики вибірки.
43. Функціональна і кореляційна залежність між величинами. Умовна середня. Рівняння регресії.
44. Побудова прямої лінії регресії за незгрупованими даними методом найменших квадратів.
45. Доведення теорем про оцінювання середньої генеральної та генеральної частки для повторної та без повторної вибірки.
46. Теореми про оцінювання дисперсії генеральної та без повторної вибірки.
47. Використання критерію узгодженості Колмогорова для перевірки гіпотези про нормальний розподіл генеральної сукупності.
48. Оцінка достовірності емпіричних коефіцієнтів кореляції і регресії за даними вибірки.

Змістовний модуль 1. *Теорія ймовірностей.*

Тема 1. Основні поняття теорії ймовірності

Події та їх види. Класичне означення ймовірності випадкової події. Властивості ймовірностей. Елементи комбінаторики теорії ймовірностей. Відносна частота випадкової події. Статистична ймовірність. Операції над подіями (алгебра подій). Діаграми В'єна. Геометрична ймовірність. [1], С. 5–16; [2], С. 17–30; [5], 4–24.

Тема 2. Теореми множення і додавання ймовірностей та їх наслідки

Умовна ймовірність. Теорема множення ймовірностей. Теореми додавання ймовірностей. Основна властивість подій, які утворюють повну групу. Алгоритми розв'язування задач з використанням теорем додавання та множення ймовірностей. Ймовірність появи хоча б однієї події. Ймовірність відбуття тільки однієї події. Формула повної ймовірності. Формула Байєса. Алгоритм розв'язування задач з використанням формул повної ймовірності та Байєса. [1], С. 29–39; [2], С. 31–53; [4], 4–23; [5], 30–54.

Тема 3. Повторні незалежні випробування

Формула Бернуллі. Найімовірніше число появи події. Локальна формула Лапласа. Формула Пуассона. Інтегральна формула Лапласа. Ймовірність відхилення відносної частоти події від її постійної ймовірності. Алгоритм розв'язування задач для повторних незалежних випробувань. [1], С. 56–68; [2], С. 55–63; [4], С. 47–55; [5], 68–83.

Тема 4. Дискретні випадкові величини та їх числові характеристики

Випадкові величини та їх види. Закон розподілу ймовірностей дискретної випадкової величини. Основні розподіли дискретних (цілочисельних) випадкових величин: рівномірний, біноміальний, Пуассонівський, геометричний, гіпергеометричний. Найпростіший потік подій. Дії над випадковими величинами. Числові характеристики дискретних випадкових величин та їх властивості (математичне сподівання, дисперсія, середньоквадратичне відхилення, початковий та центральний момент). Числові характеристики біноміального розподілу. [1], 76–126; [2], С. 64–100; [4], С. 68–90; [5], 90–115.

Тема 5. Неперервні випадкові величини та їх числові характеристики

Функція розподілу ймовірностей і її властивості. Густина розподілу ймовірностей та її властивості. Числові характеристики неперервних випадкових величин. [1], 99–116; [2], С. 111–127; [4], С. 68–90; [5], 122–136.

Тема 6. Основні закони неперервних випадкових величин

Нормальний закон: імовірнісний зміст параметрів розподілу; нормальна крива та вплив параметрів розподілу на її форму; ймовірність попадання у заданий інтервал; знаходження ймовірності заданого відхилення; правило трьох сигм. Закон рівномірного розподілу. Показників закон. Гамма-розподіл та розподіл Ерланга. Розподіл хі-квадрат. [1], 116–126; [2], 127–155; [4], С. 80–90; [5], 140–151.

Тема 7. Системи випадкових величин

Закон розподілу ймовірностей двовимірної дискретної випадкової величини. Функція розподілу двовимірної випадкової величини та її властивості. Густина розподілу ймовірностей двовимірної випадкової величини та її властивості. Умовні закони розподілу. Залежні і не залежні випадкові величини. Умовне математичне сподівання. Рівняння регресії. Числові характеристики системи двох випадкових величин. Кореляційний момент. Коефіцієнт кореляції. Система довільного скінченого числа випадкових величин. Кореляційна матриця. Нормальний закон розподілу двовимірної випадкової величини. [1], 130–152; [2], 155–185

Тема 8. Функція випадкових величин

Функція одного випадкового аргументу та її математичне сподівання. Логарифмічний нормальний закон та хі-розподіл. Функції двох випадкових величин. Розподіл С'юдента, розподіл Фішера-Снедокора. [1], 154–160 [2], 143–147

Тема 9. Закон великих чисел

Лема та нерівність Чебишева. Теорема Чебишева (стійкість середніх). Теорема Чебишева (стійкість середніх). Теорема Бернуллі (кількість відносних частот). Центральна гранична теорема Ляпунова. [1], 162–169, [2], 101–111; [3], С. 99–102, 129–139; [4], С. 134–137.

Змістовний модуль 2. *Математична статистика*

Тема 10. 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. Числові характеристики сукупностей, що складається із груп. [2] С. 185–196, [3] С. 5–48; [4], С. 141–145; [5], С. 171–195.

Тема 11. Статистичне оцінювання

Точкові статистичні оцінки параметрів розподілу та їхні властивості. Оцінка середньої генеральної для простої вибірки (повторної та безповторної). Оцінка генеральної частки для простої вибірки. Середні квадратичні помилки простої вибірки. Виправлена дисперсія вибіркова. Інтервальні статистичні оцінки. Довірчі інтервали для оцінок $\bar{x}_г$ та p для немалих і малих вибірок. Знаходження мінімального обсягу вибірки. Довірчі інтервали для $D_Г, \sigma_Г$ у випадку малої вибірки. [2] С. 197–252, [3] С. 48–95; [4], С. 141–145; [5], С. 202–218.

Тема 12. Перевірка статистичних гіпотез

Статистичні гіпотези та їхні види. Статистичний критерій перевірки основної гіпотези. Потужність критерію. Параметричні статистичні гіпотези. Перевірка гіпотез про рівність дисперсій нормальних генеральних сукупностуй. Критерій узгодженості Пірсона та Колмогорова (на прикладі перевірки гіпотези про нормальний закон розподілу). Критерій однорідності двох виборок (критерій Смирнова). [2] 281–346, [3] С. 95–149; [4], С. 141–145; [5], С. 223–249.

Тема 13. Елементи кореляційного і регресійного аналізу

Поняття стохастичності та стохастичної залежності, кореляції та регресії. Основні задачі кореляційного та регресійного аналізу. Лінійні емпіричні рівняння парної кореляції. Вибірковий коефіцієнт лінійної кореляції та його властивості. Оцінка достовірності емпіричних коефіцієнтів кореляції та регресії за даними вибірки. Нелінійна парна кореляція. Вибіркове кореляційне відношення та його властивості. Регресійний аналіз: парна та множинна лінійна регресія. [2] С. 253–280, [3] С. 153–220; [4], С. 141–145; [5], С. 250–300.

Змістовий модуль 1. ТЕОРІЯ ЙМОВІРНОСТІ

§1. Випадкові події та операції над ними. Означення ймовірності

Випадкові події

Забезпечення певного комплексу умов називають *випробуванням* або *дослідом*, а можливий результат випробування — *подією*. Наприклад, підкидання монети є випробуванням, а випадання «герба» або «номіналу» — подіями. Події позначатимемо великими латинськими літерами: A, B, C.

Подію називають *випадковою*, якщо вона може відбутися або не відбутися в даному випробуванні.

Достовірною називають подію, яка обов'язково відбудеться в даному випробуванні.

Неможливою називають подію, яка точно не відбудеться в даному випробуванні.

Зауважимо, що будь-яка подія пов'язана з певним випробуванням.

Дві події називають *сумісними*, якщо поява однієї з них не виключає появи іншої в одному й тому самому випробуванні.

Дві події називають *несумісними*, якщо вони не можуть відбутися одночасно в одному й тому самому випробуванні.

Попарно несумісні випадкові події A_1, A_2, ..., A_n утворюють *повну групу подій*, якщо внаслідок випробування одна з них обов'язково відбудеться. Наприклад, події «виграш», «програш» і «нічия» (для певного гравця) утворюють повну групу подій у випробуванні — грі в шахи двох суперників.

Елементарними подіями (*наслідками*) у певному випробуванні називають усі можливі результати цього випробування, які не можна розкласти на простіші. Множину всіх можливих елементарних подій називають *простором елементарних подій*, який позначають Ω. Наприклад, при підкиданні грального кубика простір елементарних подій утворюють події ω_i = {випаде i очок}, $i = 1, 2, 3, 4, 5, 6$.

Елементарні події, при появі яких відбувається певна подія, називають *сприятливими* для цієї події. Наприклад, при підкиданні грального кубика для події A = {випаде непарне число очок} сприятливими є елементарні події ω_1, ω_3, ω_5.

Кожну подію можна розглядати як деяку підмножину простору елементарних подій у даному випробуванні. Зокрема, подія $A = \Omega$ є достовірною, а подія $B = \varnothing$ — неможливою.

Операції над подіями

Сумою двох випадкових подій A і *B* називають таку подію, яка полягає в появі хоча б однієї з подій A або B, і позначають $A + B$.

Сумою n випадкових подій A_1, A_2, ..., A_n називають таку подію, яка полягає в появі принаймні однієї з цих подій.

Добутком двох випадкових подій A і *B* називають таку подію, яка полягає в одночасній появі обох подій A і B, і позначають $A \cdot B$.

Добутком n випадкових подій A_1, A_2, ..., A_n називають таку подію, яка полягає в одночасній появі всіх цих подій.

Подію $\overline{A}$ називають *протилежною* до події A в даному випробуванні, якщо вона відбувається тоді, коли не відбувається подія A. Очевидно, протилежні події несумісні й утворюють повну групу подій.

Приклад 1. У ящику містяться кульки білого та чорного кольору. Навмання з нього виймають одну кульку. Подія A = {вийнято кульку білого кольору}, подія B = {вийнято кульку чорного кольору}. Сумісні чи несумісні ці події?

Розв'язання. Ці події несумісні, тому що поява події A виключає можливість появи події B, і навпаки. У даному випробуванні події A і B є протилежними:

$$A = \overline{B}, B = \overline{A}. \bullet$$

Класичне означення ймовірності

Імовірність події A дорівнює відношенню кількості елементарних наслідків, сприятливих цій події, до кількості всіх рівноможливих елементарних наслідків у даному випробуванні.

Імовірність події A позначають $P(A)$, тому за означенням

$$P(A) = \frac{m}{n}, \qquad (1)$$

де m — кількість елементарних наслідків, сприятливих події A; n — кількість усіх елементарних наслідків у даному випробуванні. З класичного означення ймовірності випливає, що

$$0 \le P(A) \le 1, ,$$

причому $P(A) = 0$, якщо $A = \varnothing$ — неможлива подія, і $P(A) = 1$, якщо $A = \Omega$ — достовірна подія.

Приклад 2. Знайти ймовірність того, що вибране випадковим чином двозначне число ділиться на: **а)** 3; **б)** 5.

Розв'язання. У даному разі випробування полягає в тому, що вибирається випадковим чином двозначне число. Наслідком такого випробування є одне з чисел від 10 до 99. Оскільки таких чисел 90, то $n = 90$.

а) Нехай подія A = {вибране двозначне число ділиться на 3}. Оскільки кожне третє з 90 двозначних чисел ділиться на 3, то сприятливими для події A є 30 наслідків, тобто $m = 30$. Тоді за формулою (1) імовірність події A

$$P(A) = \frac{30}{90} = \frac{1}{3}.$$

б) Нехай подія B = {вибране двозначне число ділиться на 5}. Загальна кількість наслідків випробування, як і в попередньому випадку, $n = 90$. Визначимо кількість чисел, які діляться на 5. Очевидно, що таких чисел буде $m = 18$ (кожне п'яте число ділиться на 5). Отже,

$$P(A)=\frac{18}{90}=\frac{1}{3}. \bullet$$

Класичне означення ймовірності передбачає, що кількість елементарних наслідків скінченна. Якщо множина всіх елементарних наслідків випробування нескінченна, застосовують геометричне означення ймовірності.

Геометричне означення ймовірності

Нехай множина всіх елементарних наслідків випробування нескінченна і утворює деяку множину Ω, усі елементарні наслідки рівноможливі, причому події A сприяють ті елементарні події, які утворюють множину $A\subseteq\Omega$. Тоді ймовірність події A дорівнює відношенню міри множини A до міри множини Ω, тобто

$$P(A)=\frac{m(A)}{m(\Omega)}. \qquad (2)$$

Мірою множини на прямій, площині, у просторі є відповідно довжина, площа, об'єм геометричної фігури, яку утворює ця множина.

Статистичне означення ймовірності

Оскільки класичне означення ймовірності передбачає, що всі елементарні наслідки випробування рівноможливі, що важко обґрунтувати, то розглядають ще й *статистичне означення ймовірності*.

Відносною частотою події A називають відношення кількості випробувань, у яких подія A відбулася, до кількості всіх проведених випробувань. Відносну частоту події A позначають $W(A)$. Тоді

$$W(A)=\frac{m}{n},$$

де m — кількість випробувань, у яких відбулася подія A; n — кількість усіх проведених випробувань.

Число, до якого прямує значення частоти події A при великій кількості випробувань, називають *ймовірністю події A*:

$$P(A)=\lim_{n\to\infty}W(A).$$

§2. Елементи комбінаторики та їх застосування при обчисленні ймовірностей

При обчисленні ймовірностей подій досить часто потрібно підраховувати кількість елементарних подій (сприятливих деякій події або всіх можливих подій). Здебільшого це зумовлює великі труднощі, подолати які допомагає комбінаторика, що вивчає способи підрахунку кількості розміщень, перестановок, комбінацій.

Перш ніж представити деталі, нагадаємо, що вираз $n!$ читається «ен-факторіал» і означає добуток усіх натуральних чисел до n:

$$n! = 1 \cdot 2 \cdot 3 \cdot \ldots \cdot n,$$

причому вважають, що $0! = 1$.

Розміщеннями із l елементів по k називають множини із k елементів, вибраних із l елементів, які можуть розрізнятися між собою як складом елементів, так і їх порядком. Наприклад, розміщеннями із трьох елементів по два будуть такі множини: $\{1;2\}$, $\{1;3\}$, $\{2;1\}$, $\{2;3\}$, $\{3;1\}$, $\{3;2\}$. Кількість усіх розміщень із l елементів по k визначають за формулою

$$A_l^k = l\cdot(l-1)\cdot(l-2)\cdot\ldots\cdot(l-k+1) = \frac{l!}{(l-k)!}. \qquad (3)$$

Перестановками із l елементів називають множини із l елементів, що відрізняються лише їх порядком. Наприклад, перестановками із трьох елементів будуть такі множини: $\{1;2;3\}$, $\{1;3;2\}$, $\{2;1;3\}$, $\{2;3;1\}$, $\{3;1;2\}$, $\{3;2;1\}$. Кількість усіх перестановок із l елементів визначають так:

$$P_l = 1\cdot 2\cdot\ldots\cdot l = l!. \qquad (4)$$

Комбінаціями із l елементів по k називають множини із k елементів, вибраних із l елементів, які розрізняються між собою тільки складом елементів. Наприклад, комбінаціями із трьох елементів по два будуть такі множини: $\{1;2\}$, $\{1;3\}$, $\{2;3\}$. Кількість усіх комбінацій із l елементів по k визначають за формулою

$$C_l^k = \frac{(l-k+1)\cdot(l-k+2)\cdot\ldots\cdot l}{k!} = \frac{l!}{k!\cdot(l-k)!}. \qquad (5)$$

Між переліченими поняттями існують такі співвідношення:

$$P_l = A_l^l = l!; \quad C_l^k = \frac{A_l^k}{P_k}; \quad C_l^k = C_l^{l-k};$$

$$C_l^0 = C_l^l = 1; \quad C_l^1 = l; \quad C_l^0 + C_l^1 + C_l^2 + \ldots + C_l^l = 2^l.$$

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.

Приклад 3. Шістнадцять варіантів контрольної роботи написані на окремих картках і розподіляються випадковим чином серед 14 студентів, які сидять в одному ряду. Кожний студент отримує одну картку. Знайти ймовірність того, що: **а)** варіанти 1 і 2 не будуть використані; **б)** варіанти 1 і 2 видадуть студентам, які сидять поруч.

Розв'язання. Маємо випробування розподілу 16 білетів серед 14 студентів. У цьому разі події відрізняються одна від одної не лише номерами варіантів, що розподіляються серед студентів, а й порядком розподілу. Тому такі сполучення називають розміщеннями, а кількість таких розміщень визначається за формулою (3):

$$n = A_{16}^{14} = \frac{16!}{2!}.$$

а) Позначимо через A подію, яка полягає в тому, що варіанти 1 і 2 залишаться нерозподіленими. Тоді інші 14 білетів розподіляться серед 14 студентів. Такі сполучення називають перестановками, а їх кількість визначається за формулою (4):

$$m = \mathrm{P}_{14} = 14!$$

Отже, застосувавши класичну формулу ймовірності (1), матимемо:

$$P(A) = \frac{m}{n} = \frac{14! \cdot 2!}{16!} = \frac{1}{15 \cdot 8} \approx 0,008.$$

б) Нехай подія B полягає в тому, що варіанти 1 і 2 видані студентам, які сидять поруч. У ряду із 14 місць є 13 пар сусідніх місць, причому в кожній парі варіанти можуть розподілятися двома способами:

$$m_1 = 13 \cdot 2 = 26.$$

Інші 14 варіантів білетів розподіляються між 12 студентами

$$m_2 = A_{14}^{12} = \frac{14!}{2!}$$

способами. Тому події В сприяють

$$m = m_1 \cdot m_2 = 26 \cdot \frac{14!}{2!}$$

наслідків.

Отже, імовірність події B

$$P(B) = \frac{26 \cdot 14! \cdot 2!}{2! \cdot 16!} = \frac{13}{120} \approx 0,108. \bullet$$

Приклад 4. У податковій адміністрації зареєстровано 6 приватних і 4 державних підприємства. Знайти ймовірність того, що серед навмання вибраних трьох підприємств приватними будуть: **а)** три; **б)** два; **в)** не більше одного.

Розв'язання. Оскільки, не ставиться умова впорядкованості підмножини із вибраних трьох підприємств, то потрібно використати комбінації.

Тоді $n = C_{10}^{3} = \frac{10!}{3! \cdot 7!} = \frac{7! \cdot 8 \cdot 9 \cdot 10}{1 \cdot 2 \cdot 3 \cdot 7!} = 120$. Для пункту **а)** одержимо: $m = C_6^3 = 20$. $P(A) = \frac{20}{120} = \frac{1}{6}$.

Пункт **б)** відрізняється від попереднього тим, що вибрана тут група із трьох підприємств включає два приватні й одне державне. За правилом множення одержимо: $m = C_6^2 \cdot C_4^1 = 15 \cdot 4 = 60$. $P(B) = \frac{60}{120} = \frac{1}{2}$.

Для виконання пункту **в)** розкриємо зміст словосполучення «не більше одного». Воно означає «одне або жодного».

Тоді $m = C_6^1 \cdot C_4^2 + C_4^3 = 6 \cdot 6 + 4 = 40$. $P(C) = \frac{40}{120} = \frac{1}{3}$. •

§3. Основні формули додавання і множення ймовірностей

Імовірність суми двох довільних випадкових подій A і B дорівнює сумі їх імовірностей без імовірності їхнього добутку, тобто

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

Якщо події A і B несумісні, то $P(A \cdot B) = 0$. Тоді

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

Якщо випадкові події $A_1, A_2, ..., A_n$ попарно несумісні, то ймовірність появи хоча б однієї з цих подій дорівнює сумі їх імовірностей:

$$P(A_1 + A_2 + ... + A_n) = P(A_1) + P(A_2) + ... + P(A_n).$$

Сума ймовірностей випадкових подій $A_1, A_2, ..., A_n$, які утворюють повну групу, дорівнює одиниці, тобто

$$P(A_1) + P(A_2) + ... + P(A_n) = 1.$$

Зокрема, для протилежних подій A і $\overline{A}$ виконується рівність

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

Випадкові події A та B називають *залежними*, якщо ймовірність появи однієї з них залежить від появи або непояви іншої. В іншому випадку події A та B називають *незалежними*.

Імовірність події B, обчислена за умови появи події A, називають умов*ною ймовірністю* події B (за умови появи події A) і позначають $P(B/A)$ або $P_A(B)$.

Якщо події A та B незалежні, то $P_A(B) = P(B)$, і навпаки, $P_B(A) = P(A)$.

Імовірність добутку двох випадкових подій A і B дорівнює добутку ймовірності однієї з цих подій на умовну ймовірність іншої події за умови, що перша подія відбулася, тобто

$$P(A \cdot B) = P(A) \cdot P(B/A) = P(B) \cdot P(A/B)$$

або

$$P(A \cdot B) = P(A) \cdot P_A(B) = P(B) \cdot P_B(A).$$

Якщо події незалежні, то

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

Нехай є n незалежних випадкових подій $A_1, A_2, ..., A_n$. Імовірність появи хоча б однієї з цих подій визначається за формулою

$$P(A_1 + A_2 + ... + A_n) = 1 - P(\overline{A_1}) \cdot P(\overline{A_2}) \cdot ... \cdot P(\overline{A_n}).$$

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).

Приклад 5. Мішень складається з трьох областей. Імовірність влучення стрільцем у першу область мішені дорівнює 0,45, у другу область — 0,35, у третю — 0,15. Яка ймовірність того, що при одному пострілі стрілець влучить: **а)** у першу або в другу область; **б)** не влучить у мішень?

Розв’язання. Уведемо позначення:

подія A_1 = {влучення в першу область};
подія A_2 = {влучення в другу область};
подія A_3 = {влучення в третю область};
подія A_4 = {невлучення в мішень}.

Тоді $P(A_1) = 0{,}45$; $P(A_2) = 0{,}35$; $P(A_3) = 0{,}15$.

а) При одному пострілі події A_1, A_2, A_3 несумісні. Тому

$$P(A_1 + A_2) = P(A_1) + P(A_2) = 0{,}45 + 0{,}35 = 0{,}8.$$

б) Події A_1, A_2, A_3, A_4 попарно несумісні та утворюють повну групу випадкових подій. Тому

$$P(A_1) + P(A_2) + P(A_3) + P(A_4) = 1.$$

Отже,

$$P(A_4) = 1 - P(A_1) - P(A_2) - P(A_3) = 1 - 0{,}45 - 0{,}35 - 0{,}15 = 0{,}05. \bullet$$

Приклад 6. Робітник обслуговує одночасно три верстати. Імовірність порушення протягом години для першого верстата дорівнює 0,1, для другого — 0,15, для третього — 0,2. Яка ймовірність того, що: **а)** усі три верстати працюватимуть протягом години; **б)** хоча б один з них не вийде з ладу?

Розв’язання. A_1 = {перший верстат працюватиме протягом години}; A_2 = {другий верстат працюватиме протягом години}; A_3 = {третій верстат працюватиме протягом години}.

а) Позначимо подію A = {усі три верстати працюватимуть протягом години}. Тоді $A = A_1 \cdot A_2 \cdot A_3$. Події A_1, A_2, A_3 є незалежними, тому $P(A) =$ $= P(A_1 \cdot A_2 \cdot A_3) = P(A_1) \cdot P(A_2) \cdot P(A_3) = (1 - 0{,}1) \cdot (1 - 0{,}15) \cdot (1 - 0{,}2) = 0{,}612.$

б) B = {хоча б один з трьох верстатів не вийде з ладу}. Тоді $B = A_1 + A_2 + A_3$. Події A_1, A_2, A_3 є сумісними, тому скористатися теоремою про суму сумісних подій ми не можемо, оскільки вона має місце лише для двох сумісних подій. Через це представимо подію B як $B=A_1\cdot\overline{A_2}\cdot\overline{A_3}+\overline{A_1}\cdot A_2\cdot\overline{A_3}+\overline{A_1}\cdot\overline{A_2}\cdot A_3+A_1\cdot A_2\cdot\overline{A_3}+$ $+\overline{A_1}\cdot A_2\cdot A_3+A_1\cdot\overline{A_2}\cdot A_3+A_1\cdot A_2\cdot A_3.$ Події, що є доданками даної суми, є вже між собою несумісними. Тому ймовірність суми несумісних подій дорівнює суми їх ймовірностей. Проте даний спосіб є нераціональним, тому запропонуємо інший. Подія $\overline{B}$ = {усі три верстати вийдуть з ладу} є протилежною до події B. Тому $P(B)=1-P(\overline{B})$. $\overline{B}=\overline{A_1}\cdot\overline{A_2}\cdot\overline{A_3}$. $P(\overline{B})=P(\overline{A_1}\cdot\overline{A_2}\cdot\overline{A_3})=P(\overline{A_1})\cdot P(\overline{A_2})\cdot P(\overline{A_3})=$ $=0,1\cdot 0,15\cdot 0,2=0,003$. •

§4. Формула повної ймовірності. Формула Байєса

Якщо випадкова подія A може відбутися лише сумісно з однією із несумісних між собою подій B_1, B_2, ..., B_n, що утворюють повну групу, то ймовірність появи події A обчислюється за формулою повної ймовірності

$$P(A)=\sum_{k=1}^{n}P(B_k)\cdot P_{B_k}(A). \qquad (6)$$

В умовах теореми невідомо, з якою із несумісних подій B_1, B_2, ..., B_n відбудеться подія A. Тому появу кожної з цих подій можна вважати гіпотезою, а $P(B_k)$ — імовірністю k-ї гіпотези.

Якщо в результаті проведеного випробування відбулася подія A, то умовна ймовірність $P_A(B_k)$ може не дорівнювати $P(B_k)$. Щоб отримати умовну ймовірність, використовують формулу Байєса:

$$P_A(B_k)=\frac{P(B_k)\cdot P_{B_k}(A)}{P(A)}=\frac{P(B_k)\cdot P_{B_k}(A)}{\sum\limits_{k=1}^{n}P(B_k)\cdot P_{B_k}(A)}. \qquad (7)$$

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.

Приклад 7. Деталі, виготовлені цехом заводу, потрапляють для перевірки їхньої стандартності до одного з двох контролерів. Імовірність того, що деталь потрапить до першого контролера, дорівнює 0,6, до другого — 0,4. Імовірність того, що придатна деталь буде визнана стандартною першим контролером, дорівнює 0,94, другим — 0,98. Придатна деталь при перевірці визнана стандартною. Знайти ймовірність того, що деталь перевіряв:

• перший контролер;

• другий контролер.

Розв'язання. Позначимо такі події:

A = {придатна деталь визнана стандартною};
B_1 = {деталь перевіряв перший контролер};
B_2 = {деталь перевіряв другий контролер}.

Тоді згідно з умовою задачі

$$P(B_1) = 0{,}6;\ P(B_2) = 0{,}4;\quad P_{B_1}(A) = 0{,}94;\quad P_{B_2}(A) = 0{,}98.$$

Згідно з формулою Байєса (7) маємо

$$P_A(B_1) = \frac{P(B_1)\cdot P_{B_1}(A)}{P(A)} = \frac{0{,}6\cdot 0{,}94}{0{,}6\cdot 0{,}94 + 0{,}4\cdot 0{,}98} = \frac{141}{239} \approx 0{,}59;$$

$$P_A(B_2) = \frac{P(B_2)\cdot P_{B_2}(A)}{P(A)} = \frac{0{,}4\cdot 0{,}98}{0{,}6\cdot 0{,}94 + 0{,}4\cdot 0{,}98} = \frac{98}{239} \approx 0{,}41.\ \bullet$$

§5. Незалежні повторні випробування. Формула Бернуллі

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 $P_n(m_1 \le m \le m_2)$ зовсім не обов'язково (а деколи й помилково) використовувати інтегральну формулу Лапласа.

1. Обчислення $P_n(m)$.

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

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

в) Якщо ж n велике, а p дуже мале (значно менше 0,1) і $\lambda = np \le 9$, то застосовується формула Пуассона. При великому n, дуже малому q ($q << 0{,}1$) і при виконанні нерівності $\lambda' = nq \le 9$ 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 $npq > 9$ 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 $\lambda' = nq \le 9$, 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.

Формула Бернуллі

Нехай проводиться п незалежних випробувань, причому ймовірність появи події A в кожному випробуванні одна й та сама, тобто не залежить від її появи в інших випробуваннях. Таку серію повторних незалежних випробувань називають *схемою Бернуллі*. Прикладом повторних незалежних випробувань є підкидання монети певну кількість разів, стрільба по мішені тощо.

Нехай випадкова подія A може відбутися в кожному випробуванні з однаковою ймовірністю $P(A) = p$ або не відбутися з імовірністю $q = 1 - p$. Імовірність того, що при n випробуваннях подія A відбудеться m разів, визначається за *формулою Бернуллі*:

$$P_n(m) = C_n^m \cdot p^m \cdot q^{n-m}. \qquad (8)$$

Імовірність появи події A в n випробуваннях m разів, де число m перебуває між числами k_1 і k_2, $0 \le k_1 \le k_2 \le n$, знаходиться за формулою

$$P_n(k_1 \le m \le k_2) = P_n(k_1) + P_n(k_1 + 1) + ... + P_n(k_2).$$

Імовірність появи події A в n випробуваннях хоча б один раз

$$P_n(1 \le m \le n) = 1 - q^n.$$

Найімовірніша кількість m_0 появи події A в n випробуваннях визначається з нерівностей

$$n \cdot p - q \le m_0 \le n \cdot p + p.$$

Якщо ймовірність появи події A в кожному випробуванні дорівнює p, то кількість n випробувань, які необхідно здійснити, щоб з ймовірністю P можна було стверджувати, що подія A відбудеться хоча б один раз, обчислюється за формулою

$$n > \frac{\ln(1-P)}{\ln(1-p)}.$$

Приклад 8. Прилад складено з 10 блоків, надійність кожного з яких 0,8. Блоки можуть виходити з ладу незалежно один від одного. Знайти ймовірність того, що:

а) відмовлять два блоки;

б) відмовить хоча б один блок;

в) відмовлять не менше двох блоків.

Знайти найімовірнішу кількість блоків, що вийдуть з ладу.

Розв'язання. Позначимо подію A= {відмова блока}. Тоді

$$P(A) = p = 1 - 0{,}8 = 0{,}2.$$

Тому $q = 0{,}8$. За умовою задачі $n = 10$. Застосувавши формулу Бернуллі (8) та наслідки з неї, матимемо:

а) $P_{10}(2) = C_{10}^2\ p^2\ q^8 = 45 \cdot (0{,}2)^2 \cdot (0{,}8)^8 \approx 0{,}302;$

б) $P_{10}(1 \le m \le 10) = 1 - q^{10} = 1 - (0,8)^{10} \approx 0,893;$

в) $P_{10}(2 \le m \le 10) = 1 - \left[P_{10}(0) + P_{10}(1)\right] = 1 - \left[C_{10}^{0} \cdot (0,2)^{0} \cdot (0,8)^{10} + C_{10}^{1} \cdot (0,2)^{1} \cdot (0,8)^{1}\right] \approx 0,624.$

Найімовірнішу кількість блоків, що вийдуть з ладу, знайдемо з нерівностей

$$n \cdot p - q \le m_0 \le n \cdot p + p,$$

тобто

$$10 \cdot 0,2 - 0,8 \le m_0 \le 10 \cdot 0,2 + 0,2.$$

Отже,

$$m_0 = 2. \bullet$$

Граничні теореми у схемі Бернуллі

Для наближеного обчислення ймовірності появи події A у n незалежних випробуваннях m разів схеми Бернуллі при великих n і малих p таких, що $np < 10$, доцільно використовувати *формулу Пуассона*:

$$P_n(m) = \frac{\lambda^m}{m!} e^{-\lambda}, \quad \lambda = np.$$

Приклад 9. Підручник надруковано тиражем 90000 примірників. Імовірність неправильного брошурування підручника дорівнює 0,0001. Знайти ймовірність того, що тираж має 5 бракованих підручників.

Розв'язання. Брошурування підручника можна розглядати як випробування, що вкладається в схему Бернуллі. Кількість випробувань n велика, а ймовірність кожного випробування p незначна. Тому в цьому разі доцільно застосувати формулу Пуассона. Згідно з умовою задачі маємо

$$n = 90000,\ p = 0,0001.$$

Отже, при $\lambda = np = 9$ маємо

$$P_{90000}(5) = \frac{9^5}{5!} e^{-9} \approx 0,0607. \bullet$$

Локальна теорема Муавра — Лапласа. Якщо у схемі Бернуллі кількість випробувань n достатньо велика, а ймовірність p появи події A в усіх випробуваннях однакова та відмінна від нуля й одиниці, то ймовірність появи m разів події A можна знайти наближено за формулою

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

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

Інтегральна теорема Муавра — Лапласа. Якщо у схемі Бернуллі кількість випробувань n достатньо велика, а ймовірність p появи події A в усіх випробуваннях однакова та відмінна від нуля й одиниці, то ймовірність появи події A не менше m_1 і не більше m_2 разів можна знайти за наближеною формулою

$$P_n(m_1 \le m \le m_2) = \Phi(x_2) - \Phi(x_1),$$

де $\Phi(x)=\frac{1}{\sqrt{2\pi}}\int\limits_0^x e^{-\frac{t^2}{2}}dt$ — інтегральна функція Лапласа, $x_1=\frac{m_1-np}{\sqrt{npq}}$, $x_2=\frac{m_2-np}{\sqrt{npq}}$.

Зазначимо, що для функцій Гаусса та Лапласа складені спеціальні таблиці (див. додаток).

Приклад 10. Відомо, що 30 лампочок зі 100 на даному виробництві є бракованими. Партія лампочок у кількості 500 штук була одержана магазином для реалізації. Знайти ймовірність того, що з них бракованими виявиться:

а) 155 штук;

б) не менше 100 і не більше 200 штук.

Розв'язання. Оскільки $n=500$ є достатньо великим числом, то використовувати формулу Бернуллі нераціонально. Зазначимо, що в даній задачі має місце схема Бернуллі, оскільки випуск бракованої лампочки (ймовірність дорівнює $\frac{30}{100}$) не залежить від того, чи були бракованими попередні.

а) Оскільки $m=155$, то використаємо локальну формулу Муавра — Лапласа $P_n(m)=\frac{1}{\sqrt{npq}}\varphi(x)$.

У нашому випадку $x=\frac{155-500\cdot 0,3}{\sqrt{500\cdot 0,3\cdot 0,7}}=0{,}49$, $\varphi(0{,}49)=0{,}3538$. Отже, $P_{500}(155)=\frac{1}{\sqrt{500\cdot 0,3\cdot 0,7}}\cdot 0,3538=0,033.$

б) Оскільки $100\le m\le 200$, то використаємо інтегральну формулу Лапласа $P_n(m_1\le m\le m_2)=\Phi(x_2)-\Phi(x_1)$.

У нашому випадку $x_2=\frac{200-500\cdot 0,3}{\sqrt{500\cdot 0,3\cdot 0,7}}=4{,}88$, $x_1=\frac{100-500\cdot 0,3}{\sqrt{500\cdot 0,3\cdot 0,7}}=-4{,}88$. Оскільки функція Лапласа є непарною, тобто $\Phi(-x)=-\Phi(x)$, то $\Phi(-4{,}88)=$ $=-\Phi(4{,}88)=0{,}4999$.

Отже, $P_{500}(100\le m\le 200)=\Phi(4{,}88)-\Phi(-4{,}88)=0,49+0,49=0,98.$ •

Теорема Бернуллі. Якщо в n незалежних випробуваннях імовірність p появи події A однакова й подія A відбудеться m разів, то для будь-якого додатного числа $\varepsilon>0$ виконується рівність

$$\lim_{n\to\infty}P\left(\left|\frac{m}{n}-p\right|>\varepsilon\right)=0,$$

тобто подія, для якої відхилення визначається формулою

$$\left|\frac{m}{n}-p\right|>\varepsilon$$

при великих значеннях n майже неможлива. Тому протилежна подія

$$\left|\frac{m}{n}-p\right|\le\varepsilon$$

майже достовірна:

$$\lim_{n\to\infty}P\left(\left|\frac{m}{n}-p\right|\le\varepsilon\right)=1.$$

Дану формулу можна записати так:

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

Ці формули доцільно застосовувати за умови $n > 100$, $npq > 20$.

Приклад 11. Імовірність появи події в кожному із 625 незалежних випробувань дорівнює 0,8. Знайти ймовірність того, що відносна частота появи події відхиляється від імовірності за абсолютною величиною не більше ніж на 0,04.

Розв'язання. За умовою задачі

$$n = 625;\quad p = 0{,}8;\quad q = 0{,}2;\quad \varepsilon = 0{,}04.$$

Потрібно знайти $P\left(\left|\frac{m}{n}-0{,}8\right|\le 0{,}04\right)$.

За теоремою Бернуллі маємо

$$P\left(\left|\frac{m}{n}-0{,}8\right|\le 0{,}04\right)=2\Phi\left(0{,}04\sqrt{\frac{625}{0{,}8\cdot 0{,}2}}\right)=2\Phi(2{,}5)=2\cdot 0{,}4938=0{,}9876.\ \bullet$$

Приклад 12. Імовірність появи деякої події в кожному з незалежних випробувань дорівнює 0,5. Знайти кількість випробувань n, при якій з імовірністю 0,7698 можна очікувати, що відносна частота появи події відхилиться від її ймовірності за абсолютною величиною не більше ніж на 0,02.

Розв'язання. За умовою задачі

$$p = 0{,}5;\ q = 0{,}5;\ \varepsilon = 0{,}02.$$

Потрібно знайти кількість випробувань n, для якої

$$P\left(\left|\frac{m}{n}-0{,}8\right|\le 0{,}02\right)=0{,}7698.$$

За теоремою Бернуллі маємо

$$P\left(\left|\frac{m}{n}-0{,}8\right|\le 0{,}02\right)=2\Phi\left(0{,}02\sqrt{\frac{n}{0{,}5\cdot 0{,}5}}\right)=0{,}7698.$$

Звідси

$$\Phi\left(0{,}04\sqrt{n}\right)=0{,}3849.$$

Визначивши за додатком аргумент інтегральної функції Лапласа, при якому ця функція дорівнює 0,3849, отримаємо рівняння

$$0{,}04\sqrt{n}=1{,}2.$$

Отже,

$$n=\left(\frac{1{,}2}{0{,}04}\right)^2=900.\ \bullet$$

§6. Випадкові величини

Дискретні випадкові величини

Випадковою величиною, пов'язаною з даним імовірнісним експериментом, називають величину, яка при кожному проведенні цього експерименту набуває певного числового значення, причому заздалегідь невідомо, якого саме.

Випадкову величину називають *дискретною*, якщо множина її можливих значень є скінченною або зліченною множиною.

Законом розподілу випадкової величини називають відповідність, що встановлює зв'язок між можливими значеннями випадкової величини та відповідними їм імовірностями.

У випадку дискретної випадкової величини закон розподілу найзручніше описувати за допомогою ряду розподілу — таблиці, де наведено всі можливі значення цієї випадкової величини та відповідні їм імовірності:

X	x_1	x_2	...	x_n
P	p_1	p_2	...	p_n

причому $p_i = P\{X = x_i\}$, $\sum_{i=1}^{n} p_i = 1.$

Ламану з вершинами в точках (x_i, p_i) називають многокутником розподілу (рис. 1).

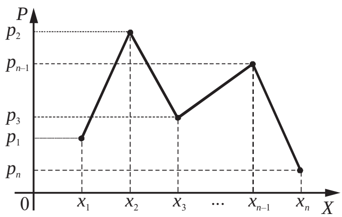

Рис. 1

Інший спосіб задання розподілу випадкової величини — аналітичний — зазначення її функції розподілу.

Функцією розподілу випадкової величини X називають функцію $F(x)$, значення якої дорівнює ймовірності того, що випадкова величина X набуде значення, яке менше x, тобто

$$F(x) = P(X < x).$$

Найважливішими є такі типи дискретних розподілів: рівномірний, біноміальний, показниковий, геометричний, гіпергеометричний.

Біноміальний розподіл (*розподіл Бернуллі*) — це розподіл випадкової величини, яка набуває значень $i \in \{0, 1, ..., n\}$ з імовірностями

$$p_i = C_n^i \, p^i \, (1-p)^{n-i},$$

де $p \in (0; 1)$. Біноміальний розподіл має випадкова величина, яка дорівнює кількості «успіхів» у серії з n незалежних випробувань, у кожному з яких «успіх» відбувається з імовірністю p.

Рівномірний розподіл — це розподіл випадкової величини, яка набуває n різних значень з однаковими ймовірностями

$$p_i = \frac{1}{n}.$$

Показниковий розподіл (*розподіл Пуассона*) — це розподіл випадкової величини, яка набуває значень $k \in \{0, 1, 2, ..., n\}$ з імовірностями

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

де $\lambda = np$, $n \to \infty$, $p \to 0$.

Геометричний розподіл — це розподіл випадкової величини, яка набуває значень $k \in \{1, 2, ...\}$ з імовірностями

$$p_k = pq^{k-1},$$

де $q = 1 - p$. Геометричний розподіл має випадкова величина, яка дорівнює кількості спроб до першого «успіху» в серії незалежних випробувань, у кожному з яких імовірність успіху дорівнює p.

Гіпергеометричний розподіл — це розподіл випадкової величини, яка набуває значень $k \in \{0, 1, ..., m\}$ з імовірностями

$$p_k = \frac{C_n^k \, C_{N-n}^{m-k}}{C_N^m},$$

де $n \geq m$, $N \geq n$.

Гіпергеометричний розподіл має така випадкова величина. Нехай у ящику міститься N однакових за фізичними властивостями кульок, серед яких S білих і $(N - S)$ чорних. З ящика навмання виймається m кульок. Тоді випадкова величина X, яка дорівнює кількості білих кульок серед m вийнятих, має гіпергеометричний розподіл.

Приклад 13. Нехай випадкова величина X дорівнює кількості номерів, угаданих гравцем у лотереї «6 із 39». Побудувати ряд розподілу випадкової величини X. Знайти значення функції розподілу в точці $x = 3$.

Розв'язання. Випадкова величина X має гіпергеометричний розподіл. Справді, процес розіграшу можна змоделювати так: у лототроні (ящику) міститься $N = 39$ однакових за фізичними властивостями кульок, серед яких $S = 6$ «білих» (так можна називати кульки, які в результаті розіграшу випадуть) і $N - S = 33$ «чорних». Вважатимемо, що кульки добре перемішані. Тому можна вважати, що з 39 кульок навмання виймається $m = 6$ шт. Тоді X — це кількість «білих» кульок серед шести вибраних (це номери, які загадані гравцем).

Очевидно, випадкова величина X набуває значень від 0 до 6, причому

$$P(X = i) = \frac{C_6^i \cdot C_{33}^{6-i}}{C_{39}^6}.$$

Виконавши підрахунки для $i = 0, 1, 2, 3, 4, 5, 6$, отримаємо:

X	0	1	2	3	4	5	6
P	$\frac{1107568}{3262623}$	$\frac{1424016}{3262623}$	$\frac{613800}{3262623}$	$\frac{109120}{3262623}$	$\frac{7920}{3262623}$	$\frac{198}{3262623}$	$\frac{1}{3262623}$

Значення функції розподілу $P(x)$ випадкової величини X у точці $x = 3$

$$F(3) = P(X < 3) = P(X = 0 \text{ або } X = 1 \text{ або } X = 2) =$$
$$= P(X = 0) + P(X = 1) + P(X = 2).$$

Іншими словами, імовірність нічого не виграти, заповнивши один лотерейний квиток, дорівнює

$$\frac{1107568}{3262623} + \frac{1424016}{3262623} + \frac{613800}{3262623} = \frac{3145384}{3262623} \approx 0{,}9641. \bullet$$

Числовими характеристиками випадкової величин є математичне сподівання, дисперсія, середнє квадратичне відхилення.

Математичним сподіванням дискретної випадкової величини називають число, яке дорівнює сумі всеможливих добутків значень випадкової величини та їх імовірностей:

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

Математичне сподівання має такі властивості.

Властивість 1. $M(C) = C$, де $C = \mathrm{const}$.

Властивість 2. $M(kX) = kM(X)$.

Властивість 3. $M(X + Y) = M(X) + M(Y)$.

Властивість 4. $M(XY) = M(X) \cdot M(Y)$, де X і Y — незалежні випадкові величини.

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

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

Дана формула еквівалентна формулі

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

Дисперсія має такі властивості:

Властивість 1. $D(X) \geq 0$.

Властивість 2. $D(C) = 0$, $C = \mathrm{const}$.

Властивість 3. $D(kX) = k^2 \cdot D(X)$.

Властивість 4. $D(X + Y) = D(X) + D(Y)$, де X і Y — незалежні випадкові величини.

Для дискретної випадкової величини дані формули можна записати так:

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

Середнім квадратичним відхиленням випадкової величини називають число $\sigma(X) = \sqrt{D(X)}$.

Дисперсія та середнє квадратичне відхилення характеризують ступінь розсіювання значень випадкової величини навколо її математичного сподівання.

Приклад 14. Норма прибутку акцій є випадковою величиною, закон розподілу якої є таким:

X	59	42	6	–6	–12
P	0,1	0,3	0,3	0,2	0,1

Знайти сподівану норму прибутку, а також середнє квадратичне відхилення даної випадкової величини.

Розв'язання.

$$M(X)=\sum_{i=1}^{n} x_i\ p_i = 59\cdot 0,1+42\cdot 0,3+6\cdot 0,3+(-6)\cdot 0,2+(-12)\cdot 0,1=17,9.$$

$$D(X)=\sum_{i=1}^{n} x_i^2\cdot p_i-(M(X))^2=59^2\cdot 0,1+42^2\cdot 0,3+6^2\cdot 0,3+(-6)^2\cdot 0,2+(-12)^2\cdot 0,1-$$

$$-17,9^2=589,29.$$

$$\sigma(X)=\sqrt{D(X)}=\sqrt{559,29}=24,28.\ \bullet$$

Неперервні випадкові величини

Неперервною випадковою величиною називають величину, функція розподілу якої є неперервною.

Найпростішим і найпоширенішим способом задання розподілу неперервної випадкової величини є задання її функції розподілу $F(x)$. Крім того, можна задати випадкову величину і через густину розподілу випадкової величини (диференціальною функцією розподілу) $f(x)$:

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

Очевидно, що:

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

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

3) $P(X\in(\alpha;\beta))=P(\alpha<X<\beta)=F(\beta)-F(\alpha)=\int_{\alpha}^{\beta} f(x)dx.$

Якщо відома диференціальна функція, то формула для обчислення інтегральної функції $F(x)$ має вигляд:

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

Найважливішими є такі типи неперервних розподілів: рівномірний, нормальний та експоненціальний.

Розподіл випадкової величини називають *рівномірним* на $[a; b]$, якщо диференціальна функція має вигляд:

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

Функція розподілу має вигляд:

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

Розподіл випадкової величини називають *експоненціальним* (*показниковим*) з параметром $\lambda>0$, якщо його диференціальна функція має вигляд:

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

Функція розподілу має вигляд:

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

Розподіл випадкової величини називають *нормальним* з параметрами a та σ, якщо його густина розподілу має вигляд:

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

Функція $y=f(x)$ швидко спадає, якщо $x\to\pm\infty$. Площа фігури, обмеженої графіком функції $y=f(x)$ і віссю Ox, дорівнює 1.

Імовірність того, що нормально розподілена величина набуде значення, яке належить відрізку $[\alpha;\ \beta]$, обчислюється так:

$$P(\alpha\le X\le\beta)=\Phi\left(\frac{\beta-a}{\sigma}\right)-\Phi\left(\frac{\alpha-a}{\sigma}\right),$$

де $\Phi(x)$ — інтегральна функція Лапласа.

Зазначимо, що ймовірностний зміст параметрів a та σ полягає втому, що

$$a=M(X),\quad \sigma=\sqrt{D(X)}.$$

Імовірність того, що відхилення нормально розподіленої величини від свого математичного сподівання за абсолютною величиною не перевищить деяке число $\varepsilon\ge 0$, обчислюють так:

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

Приклад 15. Випадкова величина X розподілена за нормальним законом, математичне сподівання якого $a=4$. Знайти ймовірність того, що випадкова величина відхилиться від свого математичного сподівання на величину, що не перевищує потроєного середнього квадратичного відхилення.

Розв'язання.

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

$$P(|X-4|\le 3\sigma)=2\Phi\left(\frac{3\sigma}{\sigma}\right);$$

$$P(|X-4|\le 3\sigma)=2\Phi(3);$$

$$P(|X-4|\le 3\sigma)=2\cdot 0,49865=0,9973.\ \bullet$$

Даний приклад дає можливість сформулювати правило «трьох σ»: нормально розподілена випадкова величина набуває своїх значень з інтервалу $[a-3\sigma;\ a+3\sigma]$, якщо ж ця випадкова величина набуває значень за межами даного інтервалу, то вони є малоймовірними.

Математичним сподіванням неперервної випадкової величини називають число

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

Дисперсію неперервної випадкової величини обчислюють як

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

або за другою розрахунковою формулою,

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

Приклад 16. Випадкову величину X задано густиною розподілу:

$$f(x)=\begin{cases}0, & \text{якщо } x\le 0;\\ \dfrac{1}{2\sqrt{x}}, & \text{якщо } 0<x\le \dfrac{1}{4};\\ 16x, & \text{якщо } \dfrac{1}{4}<x\le \dfrac{\sqrt{2}}{4};\\ 0, & \text{якщо } x>\dfrac{\sqrt{2}}{4}.\end{cases}$$

Знайти числові характеристики даної випадкової величини.

Розв'язання. Математичне сподівання випадкової величини X обчислимо за формулою $M(X)=\int\limits_{-\infty}^{+\infty}x\,f(x)\,dx$:

$$M(X)=\int_{-\infty}^{0}x\cdot 0dx+\int_{0}^{\frac{1}{4}}x\cdot\frac{1}{2\sqrt{x}}dx+\int_{\frac{1}{4}}^{\frac{\sqrt{2}}{4}}x\cdot 16xdx+\int_{\frac{\sqrt{2}}{4}}^{+\infty}x\cdot 0dx=\frac{1}{2}\int_{0}^{\frac{1}{4}}\sqrt{x}dx+$$

$$+16\int_{\frac{1}{4}}^{\frac{\sqrt{2}}{4}}x^2dx=\frac{1}{2}\cdot\left.\frac{x^{\frac{3}{2}}}{\frac{3}{2}}\right|_0^{\frac{1}{4}}+16\cdot\left.\frac{x^3}{3}\right|_{\frac{1}{4}}^{\frac{\sqrt{2}}{4}}=\frac{\sqrt{2}}{6}-\frac{1}{24}\approx 0{,}1940.$$

За формулою $D(X)=\int\limits_{-\infty}^{+\infty}x^2 f(x)\,dx-\left(M(X)\right)^2$ знайдемо дисперсію випадкової величини X:

$$D(X)=\int_{-\infty}^{0}x^2\cdot 0dx+\int_{0}^{\frac{1}{4}}x^2\cdot\frac{1}{2\sqrt{x}}dx+\int_{\frac{1}{4}}^{\frac{\sqrt{2}}{4}}x^2\cdot 16xdx+\int_{\frac{\sqrt{2}}{4}}^{+\infty}x^2\cdot 0dx-\left(\frac{\sqrt{2}}{6}-\frac{1}{24}\right)^2=$$

$$\frac{1}{2}\int_{0}^{\frac{1}{4}}x\sqrt{x}dx+16\int_{\frac{1}{4}}^{\frac{\sqrt{2}}{4}}x^3dx=\frac{1}{2}\cdot\left.\frac{x^{\frac{5}{2}}}{\frac{5}{2}}\right|_0^{\frac{1}{4}}+16\cdot\left.\frac{x^4}{4}\right|_{\frac{1}{4}}^{\frac{\sqrt{2}}{4}}-\frac{11}{192}+\frac{\sqrt{2}}{72}=\frac{\sqrt{2}}{72}-\frac{1}{240}\approx 0{,}0155.$$

Середнє квадратичне відхилення випадкової величини X

$$\sigma(X)=\sqrt{D(X)}=\sqrt{\frac{\sqrt{2}}{72}-\frac{1}{240}}\approx 0{,}1244. \bullet$$

§7. Закон великих чисел

За певних умов сумарна поведінка великої кількості випадкових величин є передбачуваною. Це випливає з теорем, які носять загальну назву — закон великих чисел.

Закони великих чисел описують стійкість середніх значень масових випадкових явищ. Історично першим законом великих чисел була теорема Бернуллі.

Теорема Бернуллі. Нехай m — кількість появ події A в серії n незалежних випробовувань, у кожному з яких подія A відбувається з імовірністю p. Тоді для довільного $\varepsilon > 0$ справджується рівність

$$\lim_{n\to\infty} P\left(\left|\frac{m}{n} - p\right| \le \varepsilon\right) = 1.$$

Лема Чебишева. Якщо всі можливі значення випадкової величини X є невід'ємними, тоді ймовірність того, що вона при випробовуванні набере значення, більшого від певного додатного числа b, не більше від дробу $\frac{M(X)}{b}$:

$$P(X > b) \le \frac{M(X)}{b}.$$

Нерівність Чебишева. Імовірність того, що відхилення випадкової величини X від її математичного сподівання за абсолютною величиною не перевищить деяке число $\varepsilon > 0$, не менше від:

$$P(|X - M(X)| \le \varepsilon) \ge 1 - \frac{D(X)}{\varepsilon^2}.$$

Теорема Чебишева. Нехай X_i — попарно незалежні випадкові величини, рівномірно обмежені за дисперсіями ($D(X_i) \le C$, $i = 1, 2, ..., n$). Тоді для довільного $\varepsilon > 0$

$$\lim_{n\to\infty} P\left(\left|\frac{1}{n}\sum_{i=1}^{n} X_i - \frac{1}{n}\sum_{i=1}^{n} M(X_i)\right| < \varepsilon\right) = 1.$$

Проте зручніше записати дану теорему у вигляді

$$P\left(\left|\frac{1}{n}\sum_{i=1}^{n} X_i - \frac{1}{n}\sum_{i=1}^{n} M(X_i)\right| < \varepsilon\right) \ge 1 - \frac{C}{n\varepsilon^2}.$$

Центральна гранична теорема Ляпунова. Якщо X_i, $i = 1, 2, ..., n$ — попарно незалежні однаково розподілені випадкові величини з математичним сподіванням $M(X_i) = a$ і $D(X_i) = \sigma^2$, то за великого n розподіл суми $X_1 + X_2 + ... + X_n = X$ близький до нормального закону.

Як наслідок зі сказаного вище, можна отримати формулу, яка має велике практичне застосування:

$$P\left(\frac{m}{n} - \varepsilon \le p \le \frac{m}{n} + \varepsilon\right) \approx 2\Phi\left(\frac{\varepsilon}{\sqrt{D(X)}}\right) = 2\Phi\left(\varepsilon\sqrt{\frac{n}{p(1-p)}}\right).$$

Приклад 17. Використовуючи нерівність Чебишева, знайти ймовірність події A, яка полягає в тому, що випадкова величина X набуде значення, яке відрізнятиметься від математичного сподівання $M(X)$ на величину, що не перевищує потроєного середнього квадратичного відхилення. Чи зміниться відповідь, якщо відомо, що випадкова величина X має нормальний розподіл?

Розв'язання. За нерівністю Чебишева

$$P(A) = P(|X - M(X)| \le 3\sigma(X)) \ge 1 - \frac{D(X)}{(3\sigma(X))^2} = 1 - \frac{D(X)}{9D(X)} = \frac{8}{9}.$$

Отже,

$$P(A) \ge \frac{8}{9} \approx 0{,}89.$$

Якщо X має нормальний розподіл (див. приклад 15), то

$$P(A)=P\left(\left|X-M(X)\right|\le 3\sigma(X)\right)=2\Phi\left(\frac{3\sigma(X)}{\sigma(X)}\right)=2\Phi(3)=0,9973.\ \bullet$$

Приклад 18. При виробництві дискет брак становить 1%. Скільки дискет потрібно відібрати для перевірки якості, щоб з імовірністю 0,95 можна було стверджувати, що у випадковій вибірці дискет відсоток бракованих відрізняється від 1% не більш як на 0,5%?

Розв'язання. Кількість бракованих дискет є випадковою величиною

$$X = X_1 + X_2 + ... + X_n,$$

де X_1, X_2, ..., X_n — незалежні однаково розподілені випадкові величини; X_i — випадкова величина, яка дорівнює кількості бракованих дискет при виготовленні однієї дискети, тобто X_i може набувати значення або 0, або 1 з імовірністю відповідно 0,99 і 0,01. Якщо n — досить велике число, то за центральною граничною теоремою розподіл випадкової величини X близький до нормального. Тому

$$P\left(\left|\frac{m}{n}-p\right|\le \varepsilon\right)\approx 2\Phi\left(\frac{\varepsilon}{\sqrt{D(X)}}\right)=2\Phi\left(\varepsilon\sqrt{\frac{n}{p(1-p)}}\right),$$

де $\frac{m}{n}$ — частота браку; $p = 0,01$; $\varepsilon = 0,005$; n — невідома кількість дискет. Число n потрібно вибрати таким, щоб

$$P\left(\left|\frac{m}{n}-0,01\right|\le 0,005\right)\approx 2\Phi\left(0,005\sqrt{\frac{n}{0,01\cdot 0,99}}\right)=0,95.$$

Іншими словами,

$$\Phi\left(0,5\sqrt{\frac{n}{99}}\right)=0,475.$$

За таблицею значень функції Лапласа (дод. 2) знаходимо значення аргументу x таке, що $\Phi(x) = 0,475$:

$$x = 1,96.$$

Розв'язавши рівняння $0,5\sqrt{\frac{n}{99}}=1,96,$ отримаємо: $n\ge 99\cdot\left(\frac{1,96}{0,5}\right)^2\ge 1522.$ $\bullet$

Змістовий модуль 2. МАТЕМАТИЧНА СТАТИСТИКА

§1. Елементи математичної статистики. Вибірковий метод

Нехай для вивчення кількісної (дискретної чи неперервної) ознаки X із генеральної сукупності отримано вибірку $x_1, x_2, \ldots, x_n$ обсягом n.

Якщо записати елементи вибірки в порядку зростання, отримаємо *варіаційний ряд*.

Спостережувані різні значення x_i ознаки X називають *варіантами*, кількість значень однієї варіанти у вибірці — її частотою n_i (сума частот усіх варіант дорівнює обсягу вибірки), відношення частоти до обсягу вибірки — *відносною частотою* або *емпіричною ймовірністю* $\omega_i = \frac{n_i}{n}$ (сума відносних частот усіх варіант дорівнює одиниці).

Статистичним розподілом вибірки називають перелік варіант варіаційного ряду з відповідними їм частотами і/або відносними частотами. Статистичний розподіл також задають у вигляді послідовності замкнених справа напівінтервалів і відповідних їм частот і/або відносних частот (частотою інтервалу вважають суму частот усіх варіант з даного інтервалу).

Щоб підкреслити зазначені відмінності в першому випадку говорять про *точковий*, а в другому — про *інтервальний* статистичний розподіл вибірки.

Приклад 1. Під час дослідження кількісної ознаки X із генеральної сукупності було отримано вибірку

$$4, 3, 6, 4, 7, 2, 5, 1, 2, 5, 4, 4, 3, 5, 6, 3, 4, 1, 3, 4.$$

Знайти обсяг вибірки, побудувати варіаційний ряд вибірки та її статистичний розподіл.

Розв'язання. Оскільки вибірка складається з 20 значень, то обсяг вибірки $n = 20$.

Побудуємо варіаційний ряд вибірки:

$$1, 1, 2, 2, 3, 3, 3, 3, 4, 4, 4, 4, 4, 4, 5, 5, 5, 6, 6, 7.$$

У даній вибірці всього сім різних значень, тобто варіант:

$$1, 2, 3, 4, 5, 6, 7.$$

Знайдемо їх частоти:

$$n_1 =2;\ n_2 = 2;\ n_3 = 4;\ n_4 = 6;\ n_5 = 3;\ n_6 = 2;\ n_7 = 1.$$

Запишемо шуканий статистичний розподіл:

x_i	1	2	3	4	5	6	7
n_i	2	2	4	6	3	2	1

•

Емпіричною функцією розподілу називають функцію

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

де n_x — кількість елементів вибірки, менших від x (тобто сума частот усіх варіант, менших від x); n — обсяг вибірки.

Емпірична функція розподілу має такі властивості.

Властивість 1. Емпірична функція розподілу $F^*(x)$ невід'ємна і не перевищує одиниці при будь-яких значеннях x:

$$F^*(x)\in[0;\ 1].$$

Властивість 2. Емпірична функція розподілу $F^*(x)$ неспадна:

$$x < y \Rightarrow F^*(x) \le F^*(y).$$

Властивість 3. Якщо x_1 — найменша варіанта, а x_n — найбільша, то емпірична функція розподілу $F^*(x) = 0$ при $x \le x_1$ і $F^*(x) = 1$ при $x > x_n$.

Приклад 2. Знайти емпіричну функцію розподілу за даним розподілом вибірки:

x_i	3	5	7	10	15
n_i	2	4	7	4	3

Розв'язання. Знайдемо обсяг вибірки:

$$n = 2 + 4 + 7 + 4 + 3 = 20.$$

Оскільки найменша варіанта дорівнює трьом,

$$F^*(x) = 0,$$

якщо $x \le 3$.

Значення $X < 5$, а саме значення $x_1 = 3$ спостерігалося двічі, тому

$$F^*(x) = \frac{2}{20} = 0{,}1,$$

якщо $3 < x \le 5$.

Значення $X < 7$, а саме $x_1 = 3$ і $x_2 = 5$ спостерігалися $2 + 4 = 6$ разів, тому

$$F^*(x) = \frac{6}{20} = 0{,}3,$$

якщо $5 < x \le 7$.

Значення $X < 10$, а саме $x_1 = 3$, $x_2 = 5$ і $x_3 = 7$ спостерігалися $2 + 4 + 7 = 13$ разів, тому

$$F^*(x) = \frac{13}{20} = 0{,}65,$$

якщо $7 < x \le 10$.

Значення $X < 15$, а саме $x_1 = 3$, $x_2 = 5$, $x_3 = 7$ і $x_4 = 10$ спостерігалися $2 + 4 + 7 + 4 = 17$ разів, тому

$$F^*(x) = \frac{17}{20} = 0{,}85,$$

якщо $10 < x \le 15$.

Оскільки $x_5 = 15$ — найбільша варіанта, то

$$F^*(x) = 1$$

якщо $x > 15$.

Отже, запишемо шукану емпіричну функцію:

$$F^*(x)=\begin{cases}0, & \text{якщо } x\le 3;\\ 0{,}1, & \text{якщо } 3<x\le 5;\\ 0{,}3, & \text{якщо } 5<x\le 7;\\ 0{,}65, & \text{якщо } 7<x\le 10;\\ 0{,}85, & \text{якщо } 10<x\le 15;\\ 1, & \text{якщо } x>15.\end{cases}$$

Графік цієї функції зображено на рис. 1.

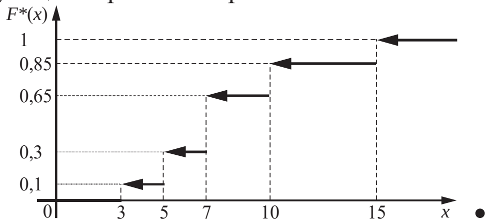

Рис. 1

Полігоном частот називають ламану, відрізки якої з'єднують точки

$$(x_1, n_1), (x_2, n_2), ..., (x_k, n_k),$$

де x_i — варіанти вибірки; n_i — відповідні частоти, $i = 1, 2, ...$.

Полігоном відносних частот називають ламану, відрізки якої з'єднують точки

$$(x_1, \omega_1), (x_2, \omega_2), ..., (x_k, \omega_k).$$

де x_i — варіанти вибірки; ω_i — відповідні емпіричні ймовірності, $i = 1, 2, ...$.

Гістограмою частот називають ступінчасту фігуру, що складається з прямокутників, основами яких є частинні інтервали довжиною Δ, а висоти дорівнюють відношенню $h_i = \frac{n_i}{\Delta}$ (*густина, щільність частоти*). Площа i-го частинного прямокутника

$$\Delta \cdot \frac{n_i}{\Delta} = n_i,$$

тобто сумі частот варіант, що потрапили в i-й частинний інтервал. Площа гістограми частот дорівнює сумі всіх частот, тобто обсягу вибірки n.

Гістограмою відносних частот називають ступінчасту фігуру, що складається з прямокутників, основами яких є частинні інтервали довжиною Δ, а висоти дорівнюють відношенню $h_i = \frac{\omega_i}{\Delta}$ (*густина, щільність відносної частоти*). Площа i-го частинного прямокутника

$$\Delta \cdot \frac{\omega_i}{\Delta} = \omega_i,$$

тобто сумі відносних частот варіант, що потрапили в i-й інтервал. Площа гістограми відносних частот дорівнює сумі всіх відносних частот, тобто одиниці.

Приклад 3. Побудувати полігон і гістограму частот за даним розподілом вибірки:

x_i	1	2	3	4	5	6	7
n_i	5	1	8	3	2	2	6

Розв'язання. Відкладемо на осі абсцис значення варіант x_i, а на осі ординат — значення відповідних їм частот n_i. Послідовно з'єднуючи між собою точки $(x_i; n_i)$ відрізками, отримуємо полігон частот (рис. 2).

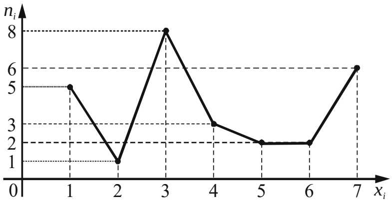

Рис. 2

Побудуємо гістограму для даного прикладу. Для цього знайдемо довжину інтервалу Δ: $\Delta = \frac{x_{\max} - x_{\min}}{k}$, де k — кількість прямокутників (вибирають довільним чином).

Нехай $k = 4$, тоді $\Delta = \frac{7-1}{4} = 1,5.$

Тоді висоти відповідних прямокутників (густина частот) h_i можна знайти з таких міркувань: до першого інтервалу [1; 2,5] попадають дві варіанти — $x_1 = 1$ і $x_2 = 2$, частоти яких $n_1 = 5$ і $n_2 = 1$. Тому $h_1 = \frac{n_1 + n_2}{\Delta} = \frac{5+1}{1,5} = 4.$ До другого інтервалу (2,5; 4] попадуть $x_3 = 3$ і $x_4 = 4$, частоти яких $n_3 = 8$ і $n_4 = 3$. Тому $h_2 = \frac{n_3 + n_4}{\Delta} = \frac{8+3}{1,5} \approx 7,33.$ До інтервалу (4; 5,5] потрапила лише $x_5 = 5$, частота якої $n_5 = 2$. Тому $h_3 = \frac{n_5}{\Delta} = \frac{2}{1,5} \approx 1,33.$ До четвертого інтервалу (5,5; 7] потраплять варіанти $x_6 = 6$ і $x_7 = 7$, частоти яких $n_6 = 2$ і $n_7 = 6$. Тому $h_4 = \frac{n_6 + n_7}{\Delta} = \frac{2+6}{1,5} \approx 5,33.$

Отримаємо гістограму частот, зображену на рисунку 3.

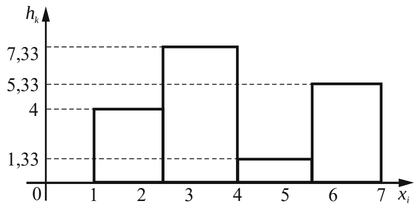

Рис. 3

Зазначимо, що сума площ прямокутників дорівнює об'єму вибірки

$$n = 5 + 1 + 8 + 3 + 2 + 2 + 6 = 27. \bullet$$

§2. Числові характеристики вибірки

Середнє арифметичне значення вибірки називають *вибірковим середнім* $\overline{x_{в}}$ і обчислюють за формулами

$$\overline{x_{в}} = \frac{\sum_{i=1}^{k} x_i n_i}{n} \quad \text{або} \quad \overline{x_{в}} = \sum_{i=1}^{k} x_i \omega_i ,$$

де x_i — значення i-ї варіанти; n_i — частота i-ї варіанти; n — обсяг вибірки; k — кількість варіант у вибірці; ω_i — відносна частота i-ї варіанти.

Середнє значення квадрата відхилення значень елементів вибірки від вибіркового середнього називають *вибірковою дисперсією* $D_{в}$ і обчислюють за формулами

$$D_{в} = \frac{\sum_{i=1}^{k} \left(x_i - \overline{x_{в}} \right)^2 n_i}{n} \quad \text{або} \quad D_{в} = \sum_{i=1}^{k} \left(x_i - \overline{x_{в}} \right)^2 \omega_i .$$

Після перетворень формули для знаходження вибіркової дисперсії дещо спрощуються:

$$D_{в} = \frac{\sum_{i=1}^{k} x_i^2}{n} - \overline{x}_{в}^2 \quad \text{або} \quad D_{в} = \sum_{i=1}^{k} x_i^2 \omega_i - \overline{x}_{в}^2 .$$

тобто вибіркова дисперсія дорівнює різниці середнього квадрата елементів вибірки й квадрата вибіркового середнього.

Квадратний корінь з вибіркової дисперсії $\sigma_{в} = \sqrt{D_{в}}$ називають *середнім квадратичним відхиленням вибірки*.

Приклад 4. Задано статистичний розподіл вибірки:

x_i	1	3	4	7	10	12	15
n_i	5	2	12	7	4	3	2

Знайти вибіркове середнє, вибіркову дисперсію та середнє квадратичне відхилення вибірки.

Розв'язання. Обчислимо обсяг вибірки:

$$n = 5 + 2 + 12 + 7 + 4 + 3 + 2 = 35.$$

Знайдемо відповідно вибіркове середнє, вибіркову дисперсію та середнє квадратичне відхилення вибірки:

$$\overline{x_{в}} = \frac{\sum_{i=1}^{7} x_i n_i}{n} = \frac{1 \cdot 5 + 3 \cdot 2 + 4 \cdot 12 + 7 \cdot 7 + 10 \cdot 4 + 12 \cdot 3 + 15 \cdot 2}{35} = \frac{214}{35} \approx 6,1;$$

$$D_{в} = \frac{\sum_{i=1}^{7} \left(x_i - \overline{x_{в}} \right)^2 n_i}{n} = \frac{\sum_{i=1}^{7} x_i^2 n_i}{n} - \overline{x_{в}}^2 = \frac{1^2 \cdot 5 + 3^2 \cdot 2 + 4^2 \cdot 12 + 7^2 \cdot 7 + 10^2 \cdot 4 + 12^2 \cdot 3 + 15^2 \cdot 2}{35} -$$

$$- \left(\frac{214}{35} \right)^2 = \frac{18604}{1225} \approx 15,2;$$

$$\sigma_{в} = \sqrt{D_{в}} = \sqrt{\frac{18604}{1225}} \approx 3,9. \ \bullet$$

Медіаною (*Me*) називають значення середнього елемента варіаційного ряду. Якщо обсяг вибірки $n = 2m + 1$ непарний, то медіаною буде значення елемента варіаційного ряду з номером $m + 1$:

$$Me = x_{m+1}.$$

Якщо обсяг вибірки $n = 2m$ парний, то медіаною буде середнє значення елементів варіаційного ряду з номерами m і $m + 1$:

$$Me = \frac{x_m + x_{m+1}}{2}.$$

Якщо задано інтервальний статистичний розподіл вибірки, то спочатку знаходять *медіанний частинний інтервал*, тобто перший частинний інтервал, для якого сума частот усіх попередніх частинних інтервалів з даним включно перевищує половину обсягу вибірки. У цьому разі медіану знаходять за формулою

$$Me = x_{Me}^{\min} + \Delta_{Me} \frac{\frac{1}{2}n - \sum_{i=1}^{Me-1} n_i}{n_{Me}},$$

де $x_{Me}^{\min}$ — початок медіанного частинного інтервалу; Δ_{Me} — довжина медіанного частинного інтервалу; n — обсяг вибірки; n_{Me} — частота медіанного частинного інтервалу; $Me-1$ — номер попереднього до медіанного частинного інтервалу.

Приклад 5. На одному з відрізків залізниці планується створити зупинку пасажирського поїзда. Розподіл населених пунктів з чисельністю їх населення наведено в таблиці.

На якому кілометрі залізниці розташований населений пункт, км	10	12	15	25	28	30	33
Чисельність населення, тис. осіб	5	2	3	10	1	4	6

На якому кілометрі залізниці потрібно розташувати цю зупинку, щоб сумарна відстань, яку покриватимуть потенційні пасажири до цієї зупинки, була найменшою.

Розв'язання. Оскільки медіана має властивість, що сума абсолютних величин відхилень елементів вибірки від медіани менша, ніж від будь-якої іншої величини, то для розв'язання прикладу потрібно знайти медіану.

Спочатку визначимо обсяг вибірки:

$$n = 5 + 2 + 3 + 10 + 1 + 4 + 6 = 31.$$

Отже, серединою (середнім членом) варіаційного ряду буде елемент із номером 16: $Me = x_{16}$. Оскільки варіаційний ряд можна записати у вигляді

$$\underbrace{10, ..., 10}_{5 \text{ разів}}, 12, 12, 15, 15, 15, \underbrace{25, ..., 25}_{10 \text{ разів}}, 28, \underbrace{30, ..., 30}_{4 \text{ рази}}, \underbrace{33, ..., 33}_{6 \text{ разів}};$$

легко бачити, що $x_{16} = 25$, тобто зупинку слід розташувати на 25-му кілометрі залізниці. •

Модою (Mo) називають варіанту з найбільшою частотою. У випадку інтервального статистичного розподілу вибірки з однаковими за довжиною частинними інтервалами *модальний частинний інтервал* визначається за найбільшою частотою, а при різних за довжиною частинних інтервалах — за найбільшою густиною $\max\limits_i \frac{n_i}{\Delta_i}$, де n_i, Δ_i — відповідно частота та довжина i-го частинного інтервалу. У цьому разі моду знаходять за формулою

$$Mo = x_{Mo}^{\min} + \Delta_{Mo} \frac{n_{Mo} - n_{Mo-1}}{2n_{Mo} - n_{Mo-1} - n_{Mo+1}},$$

де $x_{Mo}^{\min}$ — початок модального частинного інтервалу; Δ_{Mo} — довжина модального частинного інтервалу; n_{Mo} — частота модального частинного інтервалу; n_{Mo-1} — частота попереднього до модального частинного інтервалу; n_{Mo+1} — частота наступного за модальним частинного інтервалу.

Варіаційним розмахом називають різницю між найбільшим і найменшим значеннями вибірки:

$$R = x_{\max} - x_{\min}.$$

§3. Статистичні оцінки параметрів розподілу

Точкові оцінки

Статистичною оцінкою Θ^* невідомого параметра Θ теоретичного розподілу називають функцію $f(X_1, X_2, ..., X_n)$ від спостережуваних випадкових величин $X_1, X_2, ..., X_n$.

Точковою називають статистичну оцінку, яка визначається одним єдиним числом $\Theta^* = f(x_1, x_2, ..., x_n)$, де $x_1, x_2, ..., x_n$ — результати n спостережень над кількісною ознакою випадкової величини X (вибірка).

Незміщеною називають точкову оцінку Θ^*, математичне сподівання якої дорівнює оцінюваному параметру Θ за будь-якого обсягу вибірки:

$$M(\Theta^*) = \Theta.$$

Зміщеною називають точкову оцінку Θ^*, математичне сподівання якої відмінне від оцінюваного параметра Θ:

$$M(\Theta^*) \neq \Theta.$$

Незміщеною оцінкою генерального середнього (*математичного сподівання*) є вибіркове середнє

$$\overline{x}_{\text{в}} = \frac{\sum\limits_{i=1}^{k} n_i x_i}{n},$$

де n_i — частота варіанти x_i; x_i — варіанта вибірки; $n = \sum\limits_{i=1}^{k} n_i$ — обсяг вибірки.

Зміщеною оцінкою генеральної дисперсії є вибіркова дисперсія

$$D_{\text{в}} = \frac{\sum\limits_{i=1}^{k} n_i \left(x_i - \overline{x}_{\text{в}}\right)^2}{n}.$$

Ця оцінка зміщена, оскільки

$$M\left(D_{\text{в}}\right)=\frac{n-1}{n}D_{\text{г}}.$$

Незміщеною оцінкою генеральної дисперсії є виправлена вибіркова дисперсія

$$s=\sqrt{\frac{n}{n-1}D_{\text{в}}}.$$

Для обчислення вибіркової дисперсії можна скористатися більш зручною формулою

$$D_{\text{в}}=\overline{x^2}_{\text{в}}-\overline{x}_{\text{в}}^2=\frac{\sum_{i=1}^{k}n_i x_i^2}{n}-\frac{\left(\sum_{i=1}^{k}n_i x_i\right)^2}{n^2}.$$

Інтервальні оцінки

Інтервальною називають оцінку, яка визначається числовим інтервалом. *Довірчим* називають інтервал (Θ_1; Θ_2), у який із заданою надійністю γ (імовірністю, близькою до 1) потрапляє оцінюваний параметр Θ:

$$P(\Theta_1<\Theta<\Theta_2)=\gamma,\quad \gamma\to 1.$$

Для повторної вибірки обсягом n середня вибіркова $\overline{x}_{\text{в}}$ є незміщеною і спроможною оцінкою середньої генеральної $\overline{x}_{\text{г}}$. Якщо n доволі велике, тоді $\overline{x}_{\text{в}}$ з достатнім ступенем точності розподілена за нормальним законом із параметрами

$$a=\overline{x}_{\text{г}},\qquad \sigma=\sqrt{\frac{D_{\text{г}}}{n}}.$$

Для безповторної вибірки параметри відповідно рівні

$$a=\overline{x}_{\text{г}},\qquad \sigma=\sqrt{\frac{D_{\text{г}}}{n}\cdot\frac{N-n}{N-1}},$$

де N — обсяг генеральної сукупності.

Для повторної вибірки обсягом n середня вибіркова частка ω є незміщеною і спроможною точковою оцінкою невідомої генеральної частки p. Якщо n доволі велике, тоді ω з достатнім ступенем точності розподілена за нормальним законом із параметрами

$$a=p,\qquad \sigma=\sqrt{\frac{p\left(1-p\right)}{n}}.$$

Для безповторної вибірки параметри відповідно рівні

$$a=p,\qquad \sigma=\sqrt{\frac{p\left(1-p\right)}{n}\cdot\frac{N-n}{N-1}}.$$

Проте N, як правило, доволі велике, тому приймають $N-1=N$.

Можливі значення, знайдені на основі даних простої вибірки, не співпадають із оцінюваними параметрами, і кожне таке неспівпадання називають *помилкою репрезентативності*, яка викликана тим, що досліджується не всі генеральна сукупність, а лише її частина. Тому виникає потреба в оцінці середніх величин таких помилок.

Середньою квадратичною помилкою при оцінюванні середньої генеральної $\bar{x}_г$ та генеральної частки p називають середнє квадратичне відхилення середньої вибіркової $\bar{x}_в$ та вибіркової частки ω.

Можна вказати такі формули для визначення середніх квадратичних помилок:

$$\bar{\sigma}_{\bar{x}} = \sqrt{\frac{D_г}{n}}, \qquad \bar{\sigma}'_{\bar{x}} = \sqrt{\frac{D_г}{n}\left(1-\frac{n}{N}\right)},$$

$$\bar{\sigma}_{\omega} = \sqrt{\frac{p(1-p)}{n}}, \qquad \bar{\sigma}'_{\omega} = \sqrt{\frac{p(1-p)}{n}\left(1-\frac{n}{N}\right)}.$$

Тоді, щоб знайти довірчий інтервал $(\bar{x}_в - \Delta; \bar{x}_в + \Delta)$ з певною надійністю γ, необхідно знайти граничну помилку Δ (найбільше відхилення середньої вибіркової або вибіркової частки від середньої генеральної чи генеральної частки відповідно, яке можливе для заданої довірчої ймовірності γ) за формулою $\Delta = t\bar{\sigma}$, де t — корінь рівняння $2\Phi(t) = \gamma$ (див. додаток 2).

Формули для обчислення граничних помилок:

- для середньої вибіркової:

  а) у випадку повторної вибірки: $\Delta = t\sqrt{\frac{D_в}{n}}$;

  б) у випадку безповторної вибірки: $\Delta = t\sqrt{\frac{D_в}{n}\left(1-\frac{n}{N}\right)}$;

- для вибіркової частки:

  а) у випадку повторної вибірки: $\Delta = t\sqrt{\frac{\omega(1-\omega)}{n}}$;

  б) у випадку безповторної вибірки: $\Delta = t\sqrt{\frac{\omega(1-\omega)}{n}\left(1-\frac{n}{N}\right)}$.

Приклад 6 У результаті статистичних досліджень випадкової величини X отримано вибірку обсягу $n = 25$ із таким статистичним розподілом:

x_i	0	1	2	3	4	5	6
n_i	1	3	4	6	5	4	2

Знайти з надійністю $\gamma = 0{,}99$ інтервальну оцінку математичного сподівання a випадкової величини X за вибірковим середнім. Вважається, що випадкова величина X нормально розподілена.

Розв'язання. Оскільки в умові задачі не вказано, яка вибірка, то вважаємо, що вона є повторною. Перш ніж знайти довірчий інтервал для математичного сподівання $\bar{x}_г = a$ випадкової величини X за формулою

$$\bar{x}_в - t_\gamma \frac{s}{\sqrt{n}} < a < \bar{x}_в + t_\gamma \frac{s}{\sqrt{n}},$$

де $s^2 = \frac{n}{n-1}D_в$ — виправлена вибіркова дисперсія, необхідно визначити вибіркове середнє $\bar{x}_в$, «виправлене» вибіркове середнє квадратичне відхилення s ($n < 30$) і t_γ за додатком 2:

$$\bar{x}_в = \frac{\sum_{i=1}^{7} n_i x_i}{n} = \frac{0\cdot 1 + 1\cdot 3 + 2\cdot 4 + 3\cdot 6 + 4\cdot 5 + 5\cdot 4 + 6\cdot 2}{25} = 3{,}24;$$

$$s=\sqrt{\frac{n}{n-1}D_{\text{в}}}=\sqrt{\frac{n}{n-1}\left(\frac{\sum\limits_{i=1}^{7}n_i x_i^2}{n}-\left[\overline{x}_{\text{в}}\right]^2\right)}=$$

$$=\left[\frac{25}{25-1}\left(\frac{0^2\cdot 1+1^2\cdot 3+2^2\cdot 4+3^2\cdot 6+4^2\cdot 5+5^2\cdot 4+6^2\cdot 2}{25}-3,24^2\right)\right]^{\frac{1}{2}}=$$

$$=\sqrt{\frac{25}{24}\cdot 2,5024}\approx 1,6145.$$

Тоді за додатком 2 знайдемо t з рівняння $\Phi(t)=\frac{\gamma}{2}$, $\Phi(t)=\frac{0,99}{2}$, $\Phi(t)=0,495$, $t=2,58$.

Шуканий довірчий інтервал

$$3,24-2,58\cdot\frac{1,6145}{\sqrt{25}}<a<3,24+2,58\cdot\frac{1,6145}{\sqrt{25}},$$

тобто

$$2,407<a<4,073.\ \bullet$$

Кореляційний і регресійний аналіз

Випадкові величини можуть бути незалежними або залежними. Вид залежності може бути функціональним, що трапляється рідко, або стохастичним, який полягає в тому, що при зміні можливих значень однієї випадкової величини відбувається зміна закону розподілу іншої. Найважливішим випадком такого зв’язку є кореляційний зв’язок.

Кореляційною залежністю називають залежність між значеннями однієї випадкової величини і умовним середнім значенням іншої. Вона має вигляд:

$$\overline{y}_x=f(x)$$

або

$$\overline{x}_y=\varphi(y).$$

Перше рівняння називають *рівнянням регресії Y на X*, друге — *рівнянням регресії X на Y*, а їхні графіки — лініями регресії *Y на X* та *X на Y*. Якщо обидві лінії регресії — прямі, то кореляцію називають *лінійною*.

Вибірковим рівнянням прямої лінії регресії Y на X називають рівняння

$$\overline{y}_x-\overline{y}=r_{\text{в}}\frac{\sigma_y}{\sigma_x}\left(x-\overline{x}\right),$$

де $\overline{y}_x$ — умовне середнє, $\overline{x},\ \overline{y}$ — вибіркові середні випадкових величин X та Y відповідно, $\sigma_x,\ \sigma_y$ — вибіркові середні квадратичні відхилення, $r_{\text{в}}$ — вибірковий коефіцієнт кореляції, який дорівнює

$$r_b=\frac{\overline{xy}-\overline{x}\,\overline{y}}{\sigma_x\sigma_y},$$

де $\overline{xy}=\dfrac{\sum\limits_{i=1}^{k}\sum\limits_{j=1}^{m}x_i y_j n_{ij}}{n}.$

У випадку кореляції X на Y рівняння регресії має вигляд:

$$\overline{x}_y-\overline{x}=r_{\text{в}}\frac{\sigma_x}{\sigma_y}\left(y-\overline{y}\right).$$

Зазначимо, що $r_в$ в обох рівняннях одне й теж. Воно характеризує тісноту лінійного зв'язку між випадковими величинами X та Y.

Дані рівняння одержані за допомогою методу найменших квадратів.

Кореляційна залежність може бути також і нелінійною, а саме, параболічною, гіперболічою, показниковою тощо. У цьому випадку рівняння регресії мають інший вигляд, функції $f(x)$ і $\varphi(y)$ набувають відповідного вигляду. Так, наприклад, у випадку параболічної залежності

$$\overline{y}_x = a_2 x^2 + a_1 x + a_0,$$

де a_0, a_1, a_2 — невідомі параметри. Щоб їх знайти, необхідно розв'язати систему нормальних рівнянь

$$\begin{cases} a_2 \sum_{i=1}^{k} n_{x_i} x_i^4 + a_1 \sum_{i=1}^{k} n_{x_i} x_i^3 + a_0 \sum_{i=1}^{k} n_{x_i} x_i^2 = \sum_{i=1}^{k} n_{x_i} \overline{y}_{x_i} x_i^2; \\ a_2 \sum_{i=1}^{k} n_{x_i} x_i^3 + a_1 \sum_{i=1}^{k} n_{x_i} x_i^2 + a_0 \sum_{i=1}^{k} n_{x_i} x_i = \sum_{i=1}^{k} n_{x_i} \overline{y}_{x_i} x_i; \\ a_2 \sum_{i=1}^{k} n_{x_i} x_i^2 + a_1 \sum_{i=1}^{k} n_{x_i} x_i + a_0 n = \sum_{i=1}^{k} n_{x_i} \overline{y}_{x_i}, \end{cases}$$

де $\overline{y}_{x_i} = \dfrac{\sum_{j=1}^{m} n_{x_i y_j} y_j}{n_{x_i}}$ — умовне середнє, m — кількість спостережуваних варіант випадкової величини Y, $n_{x_i y_j}$ — частота варіанти $(x_i; y_j)$.

Приклад 7. Дано результати статистичних досліджень факторіальної ознаки x та результативної ознаки y:

x	2	3	4	2	3	4	2	3	3	4	2	3	4	5	5	5	5	5	7	7
y	4	5	1	3	4	3	4	4	3	3	4	4	3	5	4	5	4	5	6	6

Побудувати рівняння кореляційної залежності. Дані таблиці незгруповані.

Розв'язання. Для зручності згрупуємо дані, тобто запишемо до таблиці.

X \ Y	1	3	4	5	6	n_x
2	—	1	3	—	—	4
3	—	1	3	1	—	5
4	1	—	3	—	—	4
5	—	—	2	3	—	5
7	—	—	—	—	—	2
n_y	1	2	11	4	2	$n = 20$

Визначимо характер зв'язку між X та Y.

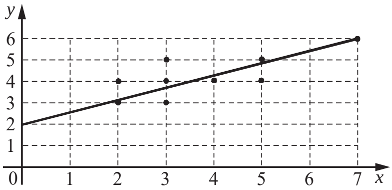

Рис. 4

Згідно графіка робимо висновок, що між x та y існує прямолінійна залежність, рівняння якої $\overline{y}_x = kx + b.$ Щоб знайти параметри k і b, запишемо канонічну систему рівнянь: $\begin{cases} k\overline{x^2} + b\overline{x} = \overline{xy}; \\ k\overline{x} + b = \overline{y}. \end{cases}$

Проте рівняння прямолінійної регресії у вигляді $\overline{y}_x - \overline{y} = r_{\text{в}} \dfrac{\sigma_y}{\sigma_x}(x - \overline{x})$ зручніше, оскільки за величиною $r_{\text{в}}$ можна зробити висновок про тісноту зв'язку між X та Y. Нагадаємо, що

$$\overline{y} = \frac{\sum_{j=1}^{m} y_j n_{y_j}}{n}; \qquad \overline{x} = \frac{\sum_{i=1}^{k} x_i n_{x_i}}{n}; \qquad \sigma_y = \sqrt{D_y}; \qquad D_y = \sigma_y^2 = \frac{\sum_{j=1}^{m} \left(y_j - y\right)^2 n_{y_j}}{n},$$

або за розрахунковою формулою

$$\sigma_y^2 = \overline{y^2} - \left(\overline{y}\right)^2 = \frac{\sum_{j=1}^{m} y_j^2 n_{y_j}}{n} - \left(\overline{y}\right)^2.$$

Аналогічно для $\sigma_x^2 = \overline{x^2} - \left(\overline{x}\right)^2 = \dfrac{\sum_{i=1}^{k} x_i^2 n_{y_i}}{n} - \left(\overline{x}\right)^2.$

Наостанок, $r_b = \dfrac{\overline{xy} - \overline{x}\,\overline{y}}{\sigma_x \sigma_y}$, де $\overline{xy} = \dfrac{\sum_{i=1}^{k}\sum_{j=1}^{m} x_i y_j n_{ij}}{n}.$

Отже,

$$\overline{x} = \frac{2 \cdot 4 + 3 \cdot 5 + 4 \cdot 4 + 5 \cdot 5 + 7 \cdot 2}{20} = \frac{78}{20} = 3{,}90;$$

$$\overline{y} = \frac{1 \cdot 1 + 3 \cdot 2 + 4 \cdot 11 + 5 \cdot 4 + 6 \cdot 2}{20} = \frac{83}{20} = 4{,}15;$$

$$\sigma_x^2 = \frac{2^2 \cdot 4 + 3^2 \cdot 5 + 4^2 \cdot 4 + 5^2 \cdot 5 + 7^2 \cdot 2}{20} - \left(\frac{78}{20}\right)^2 = 2{,}19; \quad \sigma_x = \sqrt{2{,}19} = 1{,}48;$$

$$\sigma_y^2 = \frac{1^2 \cdot 1 + 3^2 \cdot 2 + 4^2 \cdot 11 + 5^2 \cdot 4 + 6^2 \cdot 2}{20} - \left(\frac{83}{20}\right)^2 = 1{,}13; \quad \sigma_y = \sqrt{1{,}13} = 1{,}06.$$

Обчислимо спочатку $\overline{xy} - \overline{x}\,\overline{y}$:

$$\overline{xy} - \overline{x}\,\overline{y} = \frac{2 \cdot 3 \cdot 1 + 2 \cdot 4 \cdot 3 + 3 \cdot 3 \cdot 1 + 3 \cdot 4 \cdot 3 + 3 \cdot 5 \cdot 3 + 4 \cdot 1 \cdot 1 + 4 \cdot 4 \cdot 3 + 5 \cdot 4 \cdot 2 + 5 \cdot 5 \cdot 3 + 7 \cdot 6 \cdot 2}{20} -$$

$$- \frac{78}{20} \cdot \frac{83}{20} = 0{,}87.$$

Тоді вибірковий коефіцієнт кореляції дорівнюватиме:

$$r_{\text{в}} = \frac{0{,}87}{1{,}48 \cdot 1{,}06} \approx 0{,}55.$$

Таким чином, рівняння прямолінійної регресії має вигляд:

$$\overline{y}_x - 4{,}15 = 0{,}55 \cdot \frac{1{,}06}{1{,}48}(x - 3{,}90),$$

або

$$\overline{y}_x = 0{,}39x + 2{,}61.$$

Отже, між величинами *X* та *Y* існує прямолінійна кореляція, рівняння регресії якої $\overline{y}_x = 0{,}39x + 2{,}61.$ При цьому вибірковий коефіцієнт кореляції $r_{\text{в}} \approx 0{,}55$ свідчить про середній лінійний зв'язок між *X* та *Y*, до того ж, оскільки $r_{\text{в}} > 0$, то має місце додатна кореляція, тобто при зростанні *X* зростає відповідне значення результативної ознаки *Y*.

Глосарій

Безповторною називається вибірка, в процесі утворення якої відібраний об'єкт в генеральну сукупність не повертається.

Біноміальним називається закон розподілу цілочисельної випадкової величини, імовірності якого знаходяться за формулою Бернуллі.

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) до добутку середніх квадратичних відхилень вибіркових.

Вибірковою сукупністю або просто **вибіркою** називається сукупність **випадково** відібраних об'єктів із генеральної сукупності.

Вибірковою часткою називається відношення числа m об'єктів вибірки з ознакою □ до обсягу вибірки.

Випадкова — це та подія, яка при випробуванні може як відбутися, так і не відбутися.

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.

Випробування – це здійснення намічених дій і отримання результату при виконанні певного комплексу умов S.

Відносною частотою випадкової події називається відношення числа випробовувань, в яких подія відбулася, до загального числа фактично проведених випробовувань.

Внутрігруповою дисперсією називається середнє арифметичне всіх групових дисперсій, зважених по обсягах груп.

Генеральною називається вся сукупність однотипних об'єктів, яка підлягає вивченню.

Геометрична імовірність — імовірність попадання точки в область (відрізок, частину площини, частину тіла тощо).

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**).

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**).

Граничною помилкою називається найбільше відхилення середньої вибіркової (або вибіркової частки) від середньої генеральної (або генеральної частки), яке можливе для заданої довірчої імовірності.

Груповою дисперсією j-ої групи називається дисперсія розподілу значень ознаки, що складає цю сукупність, відносно групової середньої.

Густиною розподілу імовірностей неперервної випадкової величини X називається функція $f(x)$, яка дорівнює першій похідній від функції розподілу $F(x)$.

Дискретною (**перервною**) називається випадкова величина, можливі значення якої є ізольованими числами.

Дисперсією (розкидом) випадкової величини X називається математичне сподівання квадрату відхилення X від $M(X)$.

Дисперсією вибірковою статистичного розподілу (1.2) називається середнє арифметичне квадратів відхилень варіант від середньої вибіркової.

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.

Довірчим називається інтервал $(\Theta^* - \delta;\ \Theta^* + \delta)$, який із заданою надійністю γ покриває невідомий параметр Θ.

Достовірною називають подію, яка при випробуванні обов'язково відбувається.

Друга задача математичної статистики полягає в розробці методів аналізу статистичних даних в залежності від цілей дослідження.

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

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.

Загальною дисперсією називається дисперсія значень ознаки всієї сукупності відносно загальної середньої.

Законом розподілу імовірностей дискретної випадкової величини називається відповідність між можливими значеннями та імовірностями, з якими вони набираються випадковою величиною.

Зведеними характеристиками вибірки розуміють середню, дисперсію та початкові емпіричні моменти.

Інтенсивністю потоку називається середнє число подій, які вібуваються за одиницю часу.

Інтервальною називається статистична оцінка, яка визначається **двома числами** — кінцями інтервалу.

Коефіцієнтом варіації називається величина, яка дорівнює відношенню середнього квадратичного відхилення до середньої вибіркової і виражений у відсотках.

Комбінаціями називаються такі групи, які відрізняються одна від одної хоча б одним елементом.

Математичним сподіванням дискретної випадкової величини X називається сума добутків всіх її можливих значень на відповідні імовірності.

Медіаною дискретного статистичного розподілу вибірки називається таке число, яке ділить варіаційний ряд, що "породжує" цей розподіл, на дві рівні за кількістю варіант частини.

Міжгруповою дисперсією називається дисперсія групових середніх відносно загальної середньої.

Модою дискретного статистичного розподілу називається варіанта, якій відповідає найбільша частота.

Модою дискретної випадкової величини називається те її можливе значення, якому відповідає найбільша імовірність її появи.

Найпростішим називається **потік подій**, який має властивості **стаціонарності**, відсутності післядії та ординарності.

Незалежними називаються дві події, якщо імовірність однієї з них не змінюється від того, відбулася чи ні інша.

Неможливою називають подію, яка при випробуванні обов'язково не відбувається.

Неперервною називається випадкова величина, можливі значення якої заповнюють суцільно деякий скінченний або нескінченний проміжок.

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

Обсягом сукупності (генеральної або вибіркової) називається число об'єктів цієї сукупності.

Перестановками називаються групи, які відрізняються тільки порядком розташування елементів, а не самими елементами.

Перша задача математичної статистики — вказати метод відбору і групування статистичних даних, а також знаходження числа необхідних випробувань (статистичних даних).

Повторними називаються випробування, коли проводиться декілька (можливо, дуже велике число) випробувань.

Повторною називається вибірка, при утворенні якої відібраний об'єкт (перед відбором наступного) повертається в генеральну сукупність.

Подія – це результат випробування.

Полігоном відносних частот називається ламана, прямолінійні відрізки якої з’єднують сусідні точки (x_1, w_1), (x_2, w_2), …, (x_k, w_k).

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).

Потоком подій називається послідовність подій, які відбуваються у випадкові моменти часу.

Початковим емпіричним моментом *m*-го порядку розподілу називається середнє значення варіант у степені m.

Початковим моментом порядку *k* дискретної випадкової величини X називається математичне сподівання величини X^k.

Простим випадковим (**власне випадковим**) називається такий відбір, при якому об’єкти відбираються по одному випадковим чином із усієї генеральної сукупності.

Простою називається гіпотеза, яка містить тільки одне припущення.

Протилежними називаються дві події, які утворюють повну групу.

Регресійний аналіз — це дослідження односторонніх статистичних залежностей між випадковими величинами.

Рівномірно розподіленою є випадкова величина**,** якщо імовірності в законі розподілу є рівними.

Розмахом варіації називається величина, яка дорівнює різниці між найбільшою та найменшою варіантами розподілу.

Розміщеннями називаються такі групи, які відрізняються одна від іншої або хоча б одним елементом, або порядком розташування цих елементів в групі.

Середнім квадратичним відхиленням випадкової величини X називається квадратний корінь із дисперсії.

Середньою квадратичною помилкою при оцінюванні невідомих середньої генеральної $\bar{x}_г$ та генеральної частки p називається середнє квадратичне відхилення середньої вибіркової $\bar{x}_в$ та вибіркової частки w відповідно.

Складною називається гіпотеза, яка складається із скінченного або нескінченного числа простих гіпотез.

Статистичним критерієм (або просто **критерієм**, чи **статистикою**) називається випадкова великина, яка використовується для перевірки основної гіпотези і закон розподілу якої (точний або наближений) відомий.

Статистичним розподілом вибірки називається відповідність між варіантами та частотами або відносними частотами.

Статистичною називається гіпотеза про вид невідомого розподілу випадкової величини (кількісної ознаки об’єктів генеральної сукупності) або про параметри відомого розподілу.

Статистичною називається така залежність між величинами X та Y, для якої зміна спостережених значень однієї із величин зумовлює зміну умовного статистичного розподілу іншої.

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.

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.

Умовним емпіричним моментом порядку *m* називається початковий момент порядку m, обчислений для умовних варіант.

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.

Центральним емпіричним моментом *m*-го порядку розподілу (1.2) називається середня величина відхилення варіант (від середньої вибіркової) у степені *m.*
Центральним моментом порядку *k* називається математичне сподівання величини $(X - M(X))^k$.

Література

1. Єрьоменко В. О., Шинкарик М. І. Теорія ймовірностей. Навчальний посібник для студентів економічних спеціальностей. — Тернопіль: Економічна думка, 2000. 176 с.
2. Алілуйко А.М. Практикум з теорії ймовірностей та математичної статистики: навч. посібник для студентів економічних спеціальностей/ А.М.Алілуйко, Н.В.Дзюбановська, В.О.Єрьоменко, О.М.Мартинюк, М.І.Шинкарик. Тернопіль: Підручники і посібники, 2018. 352 с.
3. Єрьоменко В. О., Шинкарик М. І. Математична статистика. Навчальний посібник для студентів економічних спеціальностей. Тернопіль: Економічна думка, 2002. 248 с.
4. Жлуктенко В. І., Наконечний С. І. Теорія ймовірностей і математична статистика: Навч.методичний посібник у 2-х ч. ч. І Теорія ймовірностей . К.: КНЕУ, 2000. 304с.
5. Жлуктенко В. І., Наконечний С. І. Теорія ймовірностей і математична статистика: Навч.методичний посібник у 2-х ч. – ч. ІІ Математична статистика . К.: КНЕУ, 2003.
6. Павлова Л., Дітчук Р. Елементи комбінаторики і стохастики. Тернопіль, Підручники і посібники, 2005. 160 с. 8
7. Теорія ймовірностей та математична статистика у прикладах і задачах : навч. посіб. для студ. вищ. навч. закл. / О.Б. Жильцов; за ред. Г.О. Михаліна. К.: Київ. ун-т ім. Б. Грінченка, 2015. 336 с.
8. Зайцев Є.П. Теорія ймовірностей і математична статистика. Базовий курс з індивідуальними завданнями К., Алерта, 2017. 440 с.
9. Єрьоменко В. О., Шинкарик М.І., Мартинюк О. М., Березька К.М., Пласконь С.А., Сенів Г.В., Дзюбановська Н.В. Методичні вказівки до вивчення розділу «Теорія ймовірностей» дисципліни ТІМС для студентів всіх спеціальностей, 2017. 84 с.
10. Єрьоменко В. О., Шинкарик М.І., Мартинюк О. М., Березька К.М., Пласконь С.А., Сенів Г.В., Дзюбановська Н.В. Методичні вказівки до вивчення розділу «Математична статистика» дисципліни ТІМС для студентів всіх спеціальностей, 2017. 116 с.
11. Єрьоменко В. О., Шинкарик М.І., Мартинюк О. М., Березька К.М., Пласконь С.А., Сенів Г.В., Дзюбановська Н.В. Комплексні практичні індивідуальні завдання з теорії ймовірностей та математичної статистики для студентів всіх спеціальностей, 2017. 62 с.
12. Eremenko V.O., Plaskon S.A., Martynyuk O.M.Theory Probability and Mathematical Statistics for depth study (text of the lectures and examples for solving of the problems). Ternopil: TNEU, 2014. 192 p.
13. Теорія ймовірностей, математична статистика та імовірнісні процеси: навч. посіб. / Ю. М. Слюсарчук, Й. Я. Хром'як, Л. Л. Джавала, В. М. Цимбал ; М-во освіти і науки України, Нац. ун-т «Львів. політехніка». Львів: Вид-во Львів. політехніки, 2015. — 364с.
14. Вступ до нестандартної теорії ймовірностей: Тексти лекцій / В. Лянце, Г. Чуйко; Львів. нац. ун-т ім. І. Франка. Л., 2002. 45 с.
15. Ямненко Р.Є. Дискретна математика. К.: Четверта хвиля, 2010. – 104 с. ISBN 978-966-529- 232-6 18.
16. Карташов М.В. Л31 Імовірність, процеси, статистика : Посібник. К.: Видавничополіграфічний центр ’Київський університет’, 2008.– 494 с.
17. Турчин В.М. Теорія ймовірностей, Основні поняття, приклади, задачі. Київ, А.С.К., 2004
18. Lundberg F. Approximated framstalling av sannolikhetsfunktionen. Afterforsakring av kollektivrisker. Akad. Afhandling. Almqvist och Wiksell, Uppsala, 1903. 21. Klugmann S., Panjer H., Willmot G. Loss models. From data to desicions. John Wiley & Sons, Inc., 1998.
19. Meyn S.P. and Tweedie R.L. Markov chains and stochastic stability. Springer-Verlag, 2005.
20. Mikosch T. Non-life insurance mathematics. An introduction with stochasctic processes. SpringerVerlag, 2004.
21. Schmidli H. Lecture notes on risk theory. Aarhus, 2000.

22. Subject CT3 “Probability and Mathematical Statistics”. – Exam papers of British Institute and Faculty of Actuaries, 2005-2011 (http://www.actuaries.org.uk/students/pages/past-exam-papers).
23. Subject 106 “Actuarial mathematics 2”. Exam papers of British Institute and Faculty of Actuaries, 2000-2004 (http://www.actuaries.org.uk/students/pages/past-exam-papers).
24. Тичинська Л. М. Теорія ймовірностей. ч. 1. Історичні екскурси та основні теоретичні відомості : навчальний посібник / Л. М. Тичинська, А. А. Черепащук. Вінниця : ВНТУ, 2010. 112 с.
25. Жалдак М.І. Збірник задач і вправ з теорії ймовірностей і математичної [для студ. ф.-м. спец. педаг. універс.] / М.І. Жалдак, Н.М. Кузьміна, Г.О. Михалін. Полтава. «Довкілля-К», 2010. 728 с. Режим доступу: http://zhaldak.npu.edu.ua/drukovani-pratsi/posibnyky-tapidruchnyky
26. Ймові́рностей тео́рія // Енциклопедія сучасної України : у 30 т / ред. кол. І. М. Дзюба [та ін.] ; НАН України, НТШ, Ін-т енцикл. дослідж. НАН України. К. : Ін-т енцикл. дослідж. НАН України, 2001–2020.
27. О. Мартинюк, С. Попіна, С. Мартинюк. Імовірнісне моделювання результатів економічної діяльності як функції випадкових величин/ Вісник ТНЕУ 1 (95) 2020. С.102-112