Л. А. Вотякова

О. С. Туржанська

Теорія ймовірностей і математична статистика

Частина II

Навчально-методична розробка

УДК 519.21
ББК 22.171

Автори:
Вотякова Леся Андріївна, кандидат фізико-математичних наук, доцент кафедри алгебри і методики навчання математики Вінницького державного педагогічного університету імені Михайла Коцюбинського;

Туржанська Оксана Степанівна, кандидат педагогічних наук, доцент кафедри математики та інформатики Вінницького державного педагогічного університету імені Михайла Коцюбинського.

Рецензенти:
І. В. Хом'юк, доктор педагогічних наук, професор кафедри вищої математики Вінницького національного технічного університету;
О. Б. Панасенко, кандидат фізико-математичних наук, старший викладач кафедри алгебри і методики навчання математики Вінницького державного педагогічного університету імені Михайла Коцюбинського.

Рекомендовано до друку вченою радою Вінницького державного педагогічного університету імені Михайла Коцюбинського
(протокол № від)

Вотякова Л. А., Туржанська О. С. . – Вінниця : ТОВ фірма «Планер», 2020. – 104 с.

У посібнику подано теоретичні відомості та завдання для самостійної роботи з теорії ймовірностей та елементів математичної статистики.

Для студентів фізико-математичних спеціальностей вищих навчальних закладів.

ЗМІСТ

Розділ 1. Випадкові величини
1.1. Випадкові величини загального вигляду. Функція розподілу випадкової величини та її властивості……………………………6
1.2. Дискретні випадкові величини. Їх числові характеристики………………………8
Задачі для самостійного розв'язування………   12
1.3. Основні закони розподілу дискретної випадкової величини…………………………….14
Задачі для самостійного розв'язування………...18
1.4. Неперервні числові величини. Їх числові характеристики……………………….23
1.5. Основні розподіли неперервної випадкової величини………………………………………..25
Задачі для самостійного розв'язування………….30
1.6. Двовимірні дискретні випадкові величини. Їх числові характеристики…………………........34
1.7. Двовимірні неперервні випадкові величини. Їх числові характеристики……………………..36
1.8. Функції від випадкових величин………………...39
1.9. Закон великих чисел і центральна гранична теорема……………………………….41
Розділ 2. Елементи математичної статистики
2.1. Основні поняття математичної статистики……. 45
2.2. Числові характеристики статистичних розподілів вибірки………………………………………..49
2.3. Точкові оцінки параметрів розподілу……………52
2.4. Інтервальні оцінки параметрів розподілу……….62
Задачі для самостійного розв'язування…………68
2.5. Означення статистичної гіпотези і задача про її статистичну перевірку…………………………72
2.6. Перевірка гіпотези про порівняння середнього значення ознаки генеральної сукупності………75

2.7. Перевірка гіпотези про рівність математичних сподівань двох нормально розподілених випадкових величин……………………………….82
2.8. Перевірка гіпотези про рівність дисперсій двох незалежних випадкових величин…………84
2.9. Критерій статистичної перевірки гіпотези……...92
2.10. Критерій узгодження Пірсона…………………95
Задачі для самостійного розв'язування………..99

ПЕРЕДМОВА

Навчально-методична розробка написана відповідно до програми курсу «Теорія ймовірностей і математична статистика» для фізико-математичних спеціальностей вищих навчальних закладів і складається з двох розділів.

У першому розділі «Випадкові величини» розглядаються одновимірні та багатовимірні випадкові величини, їх функції розподілу, числові характеристики, найважливіші закони розподілу, які знаходять практичне застосування, закон великих чисел та центральна гранична теорема.

Другий розділ «Елементи математичної статистики» присвячений основним поняттям математичної статистики, точковим та інтервальним оцінкам параметрів розподілу, статистичній перевірці гіпотез.

До кожного розділу пропонуються задачі для самостійної роботи.

Розділ 1. Випадкові величини

1.1. Випадкові величини загального вигляду. Функція розподілу випадкової величини та її властивості

Нехай (Ω, S, P) − деякий ймовірнісний простір (Ω − простір елементарних подій, S − σ-алгебра підмножин Ω, P − ймовірність, тобто числова функція визначена на S, яка задовольняє аксіомам Колмогорова). Визначимо на Ω яку-небудь числову функцію $\xi = \xi(\omega)$.

Означення 1.1. Кажуть, що функція ξ є випадкова величина, якщо ця функція вимірна відносно ймовірності P.

Останнє означає, що $\forall\, x \in R^1 \equiv (-\infty; +\infty)$

$$\{\xi < x\} = \{\omega \in \Omega, \xi(\omega) < x\} \in S.$$

На змістовному рівні під випадковою величиною можна розуміти змінну величину, яка в даному випробуванні може приймати деякі числові значення з певними ймовірностями.

Приклад 1.1. Експеримент − підкидання двох монет. Випадкова величина ξ − число гербів, яке може випасти при киданні двох монет. В цьому випадку $\Omega = \{\omega_1, \omega_2, \omega_3, \omega_4\}$, де

$$\omega_1 = \{ц_1, ц_2\}, \omega_2 = \{ц_1, г_2\}, \omega_3 = \{г_1, ц_2\}, \omega_4 = \{г_1, г_2\}.$$

$$P(\xi = 0) = \frac{1}{4},\ P(\xi = 1) = \frac{1}{4} + \frac{1}{4} = \frac{1}{2},\ P(\xi = 2) = \frac{1}{4}.$$

Приклад 1.2. Експеримент − підкидання монети до першого випадання герба. Випадкова величина η − число підкидань. Простір елементарних подій

$$\Omega = \{\omega_1, \omega_2, ...\} = \{г, цг, ццг...\}, \eta(\omega_i) = i ,$$

$$P\{\eta = i\} = \left(\frac{1}{2}\right)^i, i = 1, 2, \ldots .$$

З наведених прикладів видно, що випадкова величина може набувати як скінченну, так і нескінченну множину значень.

Означення 1.2. Функцією розподілу випадковові величини ξ називають функцію $F_\xi(x)$, яка визначається рівністю

$$F_\xi(x) = P\{\xi < x\}, x \in R^1. \qquad (1.1)$$

Надалі позначатимемо просто $F(x)$, опускаючи індекс ξ, якщо з контексту зрозуміло, про яку випадкову величину йде мова.

Приклад 1.3. Знайти ймовірність події $\{x_1 < \xi < x_2\}$.

Оскільки $\{\xi < x_2\} = \{\xi < x_1\} \cup \{x_1 < \xi < x_2\}$, то згідно з теоремою додавання для несумісних подій $P\{\xi < x_2\} = P\{\xi < x_1\} + P\{x_1 \le \xi < x_2\}$. Звідси знаходимо, що

$$P\{x_1 < \xi < x_2\} = F(x_2) - F(x_1). \qquad (1.2)$$

Остання формула знаходить широке застосування в теорії ймовірностей і математичній статистиці.

Функція розподілу $F(x)$ будь-якої випадкової величини має такі властивості:

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

2) Якщо $x_1 < x_2$, то $F(x_1) \le F(x_2)$, тобто $F(x)$ − неспадна функція;

3) $F(x-0) = F(x)$, тобто функція $F(x)$ неперервна зліва;

4) $F(-\infty)=\lim\limits_{x\to-\infty}F(x)=0, F(+\infty)=\lim\limits_{x\to+\infty}F(x)=1$.

Справедлива наступна теорема.

Теорема 1.1. Будь-яка функція розподілу $F(x)$ є обмеженою, неспадною, неперервною зліва і задовольняє умовам: $F(-\infty)=0, F(+\infty)=1$.

Справедливе і обернене твердження.

Теорема 1.2. Будь-яку функцію, що задовольняє умови теореми 1.1, можна розглядати як функцію розподілу деякої випадкової величини.

Зауваження 1.1. Слід мати на увазі, що своєю функцією розподілу випадкова величина визначається неоднозначно, тобто різні випадкові величини можуть мати однакову функцію розподілу.

1.2. Дискретні випадкові величини. Їх числові характеристики

Означення 1.3. Випадкова величина ξ називається дискретною, якщо множина її значень не більш як зчисленна.

Для того, щоб задати таку величину, досить вказати її можливі значення x_k та їх ймовірності p_k. Відповідність між можливими значеннями та їх ймовірностями називають законом розподілу дискретної випадкової величини. Закон розподілу зручно задавати у вигляді таблиці:

ξ	x_1	x_2	$\dots$	x_k	$\dots$
P	p_1	p_2	$\dots$	p_k	$\dots$

При цьому $x_k \in R^1, p_k \ge 0, \sum\limits_k p_k = 1$. Останнє співвідношення використовують для контролю правильності побудови закону розподілу. За законом розподілу легко побудувати і функцію розподілу дискретної випадкової величини за формулою

$$F(x) = \sum_{x_k < x} p_k \,. \tag{1.3}$$

Тоді $p_k = P\{\xi = x_k\} = F(x_k + 0) - F(x_k - 0)$.

Крім графіка функції розподілу, поширеною графічною формою ілюстрації дискретного закону розподілу є так званий *многокутник розподілу*. Для його побудови відкладатимемо на осі абсцис можливі значення випадкової величини ξ: $x_1, x_2, ..., x_n, ...$, а на осі ординат − їх ймовірності $p_1, p_2, ..., p_n, ...$. Тоді кожній парі чисел (x_i, p_i) на площині відповідатиме точка $A_i(x_i, p_i)$. Сполучивши точки A_i послідовно в порядку зростання аргумента x_i, ми отримаємо ламану лінію, яку й називають многокутником розподілу дискретної випадкової величини ξ.

Найбільш повну інформацію про характер і поведінку випадкової величини дає функція розподілу. Однак, часто функція розподілу буває нам невідомою, а на практиці інколи досить знайти тільки деякі числа, які також певним чином характеризують випадкову величину. Такі числа називають числовими характеристиками випадкової величини. Найважливішими серед них є математичне сподівання (середнє значення), дисперсія (розсіювання), середнє квадратичне відхилення (стандарт), початкові і центральні моменти.

Нехай дискретна випадкова величина ξ задана законом розподілу

ξ	x_1	x_2	$\cdots$	x_k	$\cdots$

P	p_1	p_2	$\cdots$	p_k	$\cdots$

Означення 1.4. Математичним сподіванням дискретної випадкової величини ξ називають число

$$M_\xi = \sum_i x_i p_i . \qquad (1.4)$$

Зауваження 1.1. Якщо множина значень випадкової величини ξ скінченна, то M_ξ завжди існує. У випадку зчисленної множини значень для існування M_ξ потрібно вимагати абсолютної збіжності ряду, тобто $\sum_{i=1}^{\infty} |x_i| \, p_i < +\infty$.

Властивості математичного сподівання

1. $M(C) = C$, де C − стала величина.
2. $M(C\xi) = C\,M(\xi)$.
3. $M(\alpha\xi + \beta\eta) = \alpha M\xi + \beta M\eta$.
4. Якщо ξ і η незалежні випадкові величини, то математичне сподівання їх добутку дорівнює добутку їх математичних сподівань, тобто

$$M(\xi\eta) = M\xi \cdot M\eta .$$

Для характеристики відхилення значень випадкової величини від їх математичного сподівання вводиться поняття дисперсії або розсіювання.

Означення 1.5. *Дисперсією* випадкової величини ξ називають математичне сподівання квадрата відхилення цієї величини від її математичного сподівання. Позначають дисперсію ξ символом D_ξ. Згідно з означенням,

$$D_\xi = M\left(\xi - M_\xi\right)^2 . \qquad (1.5)$$

Або, $D_\xi = \sum_k \left(x_k - M_\xi\right)^2 p_k$. (1.6)

Означення 1.6. Арифметичне значення квадратного кореня з дисперсії називають *середнім квадратичним відхиленням або стандартом.*

Позначають стандарт через σ_ξ. Згідно з означенням,

$$\sigma_\xi = \sqrt{D_\xi}\,. \tag{1.7}$$

Властивості дисперсії

1. $D_\xi \geq 0$.
2. $DC = 0$.
3. $D\left(C_\xi\right) = C^2 D_\xi$.
4. $D_\xi = M\xi^2 - \left(M\xi\right)^2$.
5. Якщо ξ і η незалежні випадкові величини, то

$$D(\xi \pm \eta) = D\xi + D\eta\,.$$

Примітка. Формула

$$D_\xi = M\xi^2 - \left(M\xi\right)^2 \tag{1.8}$$

Використовується замість формули (1.5) при практичному обчисленні дисперсії випадкової величини. Формула (1.8) відповідно має вигляд:

$$D\xi = \sum_{+\infty} {x_k}^2 p_k - \left(M\xi\right)^2\,. \tag{1.9}$$

Крім математичного сподівання і дисперсії використовується ряд інших числових характеристик випадкової величини.

Нехай k − довільне натуральне число.

Означення 1.7. *Початковим моментом* k*-го порядку* випадкової величини ξ називають число

$$v_k = M\xi^k\,. \tag{1.10}$$

Означення 1.8. *Центральним моментом* k*-го порядку* випадкової величини ξ називають число

$$\mu_k = M\left(\xi - M\xi\right)^k . \qquad (1.11)$$

Для перших чотирьох моментів справедливі формули:

$$\mu_1 = 0,$$

$$\mu_2 = \nu_2 - \nu_1^{\,2},$$

$$\mu_3 = \nu_3 - 3\nu_2\nu_1 + 2\nu_1^{\,3},$$

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

Означення 1.8. *Коефіцієнтом варіації* випадкової величини ξ називають число

$$V(\xi) = \frac{\sqrt{D_\xi}}{M_\xi}(100\%) = \frac{\sigma_\xi}{M_\xi}(100\%). \qquad (1.12)$$

Означення 1.9. *Коефіцієнтом асиметрії* випадкової величини ξ називають число

$$\beta(\xi) = \frac{\mu_3}{(\sigma\xi)^3}. \qquad (1.13)$$

Означення 1.10. *Ексцесом* випадкової величини ξ називають число

$$\gamma(\xi) = \frac{\mu_4}{(\sigma\xi)^4} - 3. \qquad (1.14)$$

Коефіцієнт асиметрії та ексцес дають уявлення про форму многокутника розподілу дискретної випадкової величини.

Розглянемо ще одну характеристику, яку використовують для дискретної випадкової величини.

Означення 1.11. *Медіаною* розподілу $F(x)$ називають таке значення аргумента $x = m$, для якого виконуються нерівності

$$F(m) \le 0{,}5 \le F(m+0). \qquad (1.15)$$

Задачі для самостійного розв'язування

1. Перевірити, чи є подані нижче функції функціями розподілу:

А) $F(x)=\frac{1}{2}+\frac{1}{2}arctgx$;

Б) $F(x)=\frac{3}{4}+\frac{1}{2\pi}arctgx$;

В) $F(x)=\begin{cases}0, & x\le 0\\ 0{,}3 & 0<x\le 1\\ 0{,}5 & 1<x\le 2\\ 1 & x>2\end{cases}$

Г) $F(x)=\begin{cases}0, & x\le 0\\ 0{,}1x & 0<x\le 5\\ 0{,}4 & 5<x\le 6\\ 1 & x>6\end{cases}$

Побудувати графіки виявлених функцій розподілу.

2. Перевірити для заданих функцій характеристичних властивостей функцій розподілу і зробити висновок.

А) $F(x)=\begin{cases}0{,}5e^{x} & x\le 0\\ 0{,}8 & 0<x\le 1\\ 1 & x>1\end{cases}$

Б) $F(x)=\begin{cases}0 & x\ge 0\\ \frac{[x]}{2} & 0<x\le 2\\ 1 & x>2\end{cases}$

В) $F(x)=\begin{cases}0 & x\le 0\\ \frac{x}{1+x} & x>0\end{cases}$

Г) $F(x)=\begin{cases}0 & x\le 0\\ 1-\frac{1-e^{-x}}{x} & x>0\end{cases}$

3. За заданою функцією розподілу знайти ймовірності:

$P(\xi<2)$, $P(-4<\xi\le 2)$, $P(\xi>2)$, $P(\xi\ge 5)$.

$$F(x)=\begin{cases}0, & x\le -4\\ 0{,}1 & -4<x\le -1\\ 0{,}3 & -1<x\le 2\\ 0{,}5 & 2<x\le 5\\ 1 & x>8\end{cases}$$

4. Закон розподілу випадкової величини ξ задано функцією розподілу:

$$F(x)=\begin{cases}0 & x\le -3\\ \dfrac{(x+3)^2}{49} & -3<x\le 4\\ 1 & x>4\end{cases}$$

Побудувати графік $F(x)$ і обчислити ймовірність $P(-1<\xi<2)$.

5. При яких значення *а* і *в* функція $F(x)$ буде функцією розподілу. Знайти *а* і *в, якщо*

$$F(x)=\begin{cases}0 & x\le -2\\ ax+b & -2<x\le 5\\ 1 & x>5\end{cases}$$

Побудувати графік.

1.3. Основні закони розподілу дискретної випадкової величини

1. Вироджений розподіл

Означення 1.12. Кажуть, що дискретна випадкова величина ξ має вироджений закон розподілу, зосереджений в точці x_0, якщо вона набуває тільки цього значення з ймовірністю 1.

Закон розподілу:

ξ	x_0
p	1

Функція розподілу

$$F(x)=\begin{cases}0, x<x_0,\\ 1, x\ge x_0.\end{cases}$$

Числові характеристики: $M_\xi = x_0, D_\xi = 0$.

Вироджений закон розподілу описує фактично невипадкові величини. Існує і обернене твердження: якщо випадкова величина ξ має скінченне математичне сподівання і нульову дисперсію, то $P\{\xi = M_\xi\} = 1$.

2. Розподіл Бернуллі

Означення 1.13. Кажуть, що дискретна випадкова величина ξ розподілена за законом Бернуллі з параметром p $(0 < p < 1)$, якщо вона набуває тільки двох значень 0 та 1 з ймовірностями $1-p$ та p.

Закон розподілу:

ξ	0	1
p	$1-p$	p

Функція розподілу

$$F(x) = \begin{cases} 0, x < 0, \\ 1-p, 0 \le x < 1, \\ 1, x \ge 1. \end{cases}$$

Числові характеристики: $M_\xi = p, D_\xi = p(1-p)$.

3. Рівномірний розподіл на множині $\{1, 2, ..., n\}$

Означення 1.14. Дискретна випадкова величина ξ називається рівномірно розподіленою на множині $\{1, 2, ..., n\}$, якщо вона набуває значень з цієї множини з ймовірностями $\frac{1}{n}$, тобто $P\{\xi = m\} = \frac{1}{n}, m = \overline{1, n}$.

Закон розподілу:

ξ	1	2	…	n
p	$1/n$	$1/n$	…	$1/n$

Функція розподілу:

$$F(x)=\sum_{x_k<x} p_k=\sum_{x_k<x}\frac{1}{n}=\begin{cases}0, x\le 1,\\ \frac{1}{n}, 1<x\le 2,\\ \dots\dots\dots\dots\dots\dots\\ \frac{n-1}{n}, n-1<x\le n,\\ 1, x>n.\end{cases}$$

Числові характеристики:

$$M_\xi=\sum_{k=1}^{n} x_k p_k=\dots=\frac{n+1}{2}, D_\xi=\sum_{k=1}^{n} {x_k}^2 p_k-\left(\frac{n+1}{2}\right)^2=\dots=\frac{n^2-1}{12}.$$

4. Гіпергеометричний закон розподілу

Означення 1.15. Кажуть, що дискретна випадкова величина ξ розподілена за гіпергеометричним законом, якщо вона набуває значень $0,1,\dots,k$ з ймовірностями

$$P\{\xi=m\}=\frac{C_M^m\ C_{N-M}^{n-m}}{C_N^n}.$$

Параметри розподілу N, M, n − натуральні числа, причому $M\le N, n\le N, k=\min(M,n)$.

До гіпергеометричного закону розподілу приводять задачі, в яких ймовірність обчислюється за формулою: $p=\frac{C_M^m\ C_{N-M}^{n-m}}{C_N^n}$.

Числові характеристики:

$$M\xi=\sum_{m=0}^{k} x_m p_m=\dots=n\frac{M}{N},$$

$$D\xi = \sum_{m=0}^{k} x_m^2 \; p_m - \left(n\frac{M}{N} \right)^2 = \ldots = \frac{nM\left(N-M\right)\left(N-n\right)}{N^2\left(N-1\right)}.$$

5. Геометричний закон розподілу

Означення 1.16. Кажуть, що дискретна випадкова величина ξ розподілена за геометричним законом, якщо вона набуває значень з множини натуральних чисел з ймовірностями

$$P\{\xi = m\} = p\left(1-p\right)^{m-1}, m = 1, 2, \ldots .$$

Закон розподілу:

ξ	1	2	$\ldots$	m	$\ldots$
p	p	pq	$\ldots$	pq^{m-1}	$\ldots$

За геометричним законом розподілена, наприклад, випадкова величина, яка дорівнює числу проведених випробувань Бернуллі до настання першого «успіху». Числові характеристики розподілу:

$$M_\xi = \sum_{k=1}^{\infty} x_k p_k = \frac{1}{p}, D_\xi = \sum_{k=1}^{\infty} x_k^2 \; p_k - \frac{1}{p^2} = \frac{q}{p^2}.$$

6. Біномний розподіл

Означення 1.17. Кажуть, що дискретна випадкова величина ξ розподілена за біномним законом, якщо вона набуває значень $0, 1, \ldots, n$ з ймовірностями:

$$P\{\xi = k\} = C_n^k \; p^k q^{n-k}, k = \overline{0, n}.$$

Закон розподілу:

ξ	0	1	$\ldots$	k	$\ldots$	n
p	q^n	npq^{n-1}	$\ldots$	$C_n^k \; p^k q^{n-k}$	$\ldots$	p^n

За біномним законом розподілена, наприклад, випадкова величина, яка дорівнює числу «успіхів» при

n випробуваннях Бернуллі. Математичне сподівання і дисперсія відповідно дорівнюють:

$$M_\xi = \sum_{k=0}^{n} x_k p_k = np, D_\xi = \sum_{k=0}^{n} x_k{}^2 p_k - (np)^2 = npq .$$

7. Розподіл Пуассона

Означення 1.18. Кажуть, що дискретна випадкова величина ξ розподілена за законом Пуассона, якщо вона набуває значень $0,1,\ldots,\mathrm{n}$ з ймовірностями $(\mathrm{n} \to \infty)$:

$$P\{\xi = m\} = \frac{\lambda^m}{m!} e^{-\lambda}, \lambda - \text{параметр розподілу.}$$

Закон розподілу:

ξ	0	1	$\ldots$	m	$\ldots$	n
p	$e^{-\lambda}$	$\lambda e^{-\lambda}$	$\ldots$	$\frac{\lambda^m}{m!} e^{-\lambda}$	$\ldots$	$\frac{\lambda^n}{\mathrm{n}!} e^{-\lambda}$

Числові характеристики розподілу:

$$M_\xi = \sum_{k=0}^{\infty} x_k p_k = \lambda, D_\xi = \sum_{k=0}^{\infty} x_k^2 \; p_k - \lambda^2 = \lambda .$$

Задачі для самостійного розв’язування

1. З допомогою якого закону розподілу можна задати такі випадкові величини:

а) число осіб жіночої статі у безповторній вибірці об'єму 5 з 30 студентів, серед яких 10 осіб чоловічої статі;

б) число осіб жіночої статі у повторній вибірці об'єму 5 з 30 студентів, серед яких 10 осіб чоловічої статі;

в) число кидань правильного грального кубика до першої появи числа очок меншого п'яти;

г) число бракованих виробів у партії з 1000 виробів, якщо ймовірність того, що випущено бракований виріб рівняється 0,008;

д) число очок, що відповідає вибраний з колоди у 36 гральних карт одній карті, якщо валету відповідає 2, дамі — 3, королю — 4, тузу — 11 очок, а останнім картам — відповідно 6, 7, 8, 9, 10 очок?

2. За даним законом розподілу дискретної випадкової величини Х маємо:

$X = x_i$	-4	-1	2	5	8	10
$P(X = x_i) = p_i$	*a*	*1,5a*	*0,5a*	*3,5a*	*2,5a*	*a*

Знайти *a*. Обчислити: $P(X < 2)$, $P(-4 < X \le 8)$.
Побудувати графік розподілу ймовірностей і накреслити її графік.

3. За заданою функцією розподілу ймовірностей

$$F(x) = P(X < x) = \begin{cases} 0, & x \le -4 \\ 0{,}1 & -4 < x \le -1 \\ 0{,}3 & -1 < x \le 2 \\ 0{,}5 & 2 < x \le 5 \\ 0{,}8 & 5 < x \le 8 \\ 1 & x > 8 \end{cases}$$

Обчислити: $P(-4 < X \le 2)$, $P(X > 2)$, $P(X \ge 5)$, $P(X \le 2)$.

4. Троє складають іспит з теорії ймовірностей. Імовірність того, що перший студент складе екзамен, становить 0,9, для другого та третього студентів ця ймовірність дорівнює відповідно 0,85; 0,8. Побудувати закон розподілу ймовірностей дискретної випадкової величини X – числа студентів, які складуть іспит з теорії ймовірностей, побудувати $F(x)$ і накреслити її графік.

5. На шляху руху автомобіля стоять п’ять світлофорів, кожний із яких з ймовірністю 0,5 дозволяє або забороняє рух. Побудувати закон розподілу ймовірностей дискретної випадкової X – числа світлофорів, що їх автомобіль промине без затримки.

6. П'ять приладів потрібно перевірити на надійність. Кожний наступний прилад підлягає перевірці лише в тому разі, якщо перевірений прилад перед цим виявляється надійним. Імовірність того, що прилад витримає перевірку на надійність, для кожного з них становить 0,8. Побудувати закон розподілу дискретної випадкової величини X- числа приладів, які пройшли випробування.

7. При підкиданні трьох гральних кубиків гравець може виграти 18 грн., якщо на трьох кубиках випаде цифра 6; 1 грн. 40 коп., якщо на двох гральних кубиках випаде цифра 6, і 20 коп., якщо лише на одному кубику з трьох випаде цифра 6. Який у середньому буде виграш гравця? Яка має бути ставка за участь у грі, щоб вона була принаймні безкоштовною?

8. Знайти математичне сподівання і дисперсію кількості очок, що з'являться в результаті одного підкидання грального кубика.

9. Відомі значення $M(X) = -2,\ D(X) = 4$. Знайти $M(-4X + 5),\ D(-4X + 5)$.

10. Монета підкидається до першої появи герба. Знайти середню кількість підкидань.

11.Знайти $M(X^2)$, якщо $D(X) = 4,\ M(X) = 1$.

12. Побудувати функцію розподілу для числа білих куль серед вибраних навгад чотирьох куль з урни, у якій 4 білі і 6 чорних куль.

13. З урни, яка містить 3 білі та 2 чорні кулі, перекладено 2 кулі в урну, яка містить 4 білі і 4 чорні кулі. Побудувати функцію розподілу числа білих куль серед двох взятих навгад з другої урни після перекладання.

14. Гральний кубик виготовлено так, що ймовірність випадання певної грані прямо пропорційна числу очок на цій грані. Знайдіть ймовірність того, що при киданні такого грального кубика випаде парне число очок.

15. В урні 5 куль, занумерованих від 1 до 5. Знайдіть ймовірність того, що серед взятих навгад трьох куль максимальним буде номер 4.

16. По мішені проведено 3 постріли. Імовірність влучення у мішень першого пострілу становить 0.1, другого - 0.2, а третього - 0.3. Знайти ряд розподілу кількості влучень при трьох пострілах. Побудувати графік функції розподілу.

17. Серед 10 виробів є один бракований. Щоб його знайти, беруть навмання один виріб за іншим і кожен взятий виріб перевіряють. Побудувати ряд розподілу кількості перевірених виробів.

18. Виконується 4 незалежних постріле по мішені. Імовірність влучення при одному пострілі дорівнює 0,25. Знайти закон розподілу кількості влучань *X*, найімовірніше число влучень та його ймовірність.

19. Монету кидають 6 разів. Скласти ряд розподілу і побудувати функцію розподілу відношення частоти появ герба до числа появ надпису.

20. Абонент забув останню цифру номера телефону і набирає її навгад.

а) Побудуйте ряд розподілу числа дзвінків, які він може здійснити.

б) Яким є найімовірніше число дзвінків?

в) Знайдіть математичне сподівання, дисперсію і середнє квадратичне відхилення цієї випадкової величини.

21. Побудуйте ряд розподілу для найменшого числа очок, яке випадає при киданні трьох правильних гральних кубиків. Знайти математичне сподівання, дисперсію і середнє квадратичне відхилення цієї випадкової величини.

22. Можливими значеннями випадкової величини ξ є числа 1, 2, 3. Знайти імовірності цих значень, якщо
$M(\xi) = 2{,}3;\ \ M(\xi^2) = 5{,}9$.
Знайдіть математичне сподівання і дисперсію випадкової величини $\eta = -2\xi - 1$.

23. Випадкова величина ξ має біномний закон розподілу з параметром n = 80 і D(ξ) = 15. Знайдіть другий параметр розподілу і математичне сподівання M(ξ). Визначити імовірність того, що ξ > 20.

24. Випадкова величина ξ має біномний розподіл з $p=0{,}6$. Знайти $P(\xi=5)$, якщо відношення $P(\xi=3)$ до $P(\xi=1)$ рівне 21.

25. У круг радіуса R навгад кидається 7 точок. Нехай ξ – число точок, що попаде у квадрат, вписаний у цей круг. Знайти: M(ξ), D>(ξ), P(ξ = 3).

26. Дискретна випадкова величина ξ може набирати тільки два можливих значення ξ_1 і ξ_2. Ймовірність того, що ξ набирає значення ξ_1 рівна 0,2. Знайти закон розподілу цієї випадкової величини, якщо M(ξ) = 2,6, $\sigma(\xi) = 0{,}8$.

27. Випадкова величина ξ має геометричний розподіл. Знайти $P(\xi = 2)$, якщо $P(\xi = 1) = 0{,}16$.

28. Випадкова величина ξ може набирати значень $\xi_1 = 1$, $\xi_2 = 3$, $\xi_3 = 3$. Знайти відповідні ймовірності цих значень, якщо M(ξ) = 2,3, $M(\xi^2)$ = 5,9. Знайти дисперсію випадкової величини $\eta = -100\xi - 0{,}01$.

29. Випадкова величина ξ має розподіл Пуассона. Знайти P(ξ = 2), якщо відношення суми P(ξ = 1) і P(ξ = 2) до P(ξ = 3) рівне 3.

1.4. Неперервні числові величини. Їх числові характеристики

Означення 1.19. Випадкова величина ξ називається *неперервною*, якщо її функція розподілу є неперервною функцією.

Означення 1.20. Випадкова величина ξ називається *неперервною* (точніше, абсолютно неперервною), якщо функція розподілу $F(x)$ можна подати у вигляді

$$F(x)=\int_{-\infty}^{x} p(y)dy\,, \tag{1.16}$$

де $p(x)$ − невід'ємна функція, яка задовольняє умову нормування

$$\int_{-\infty}^{+\infty} p(x)dx=1\,. \tag{1.17}$$

Функція $p(x)$ називається *щільністю розподілу ймовірностей*.

Якщо $F(x)$ диференційовна і її похідна обмежена, то

$$p(x)=F'(x)\,. \tag{1.18}$$

Враховуючи рівності (1.16), (1.18) функцію $F(x)$ називають *інтегральною* функцією розподілу, а функцію $p(x)$ − диференціальною функцією розподілу. Очевидно, що

$$P\{x_1\le\xi<x_2\}=\int_{x_1}^{x_2} p(y)dy\,, \tag{1.19}$$

$$P\{\xi=x\}=F(x+0)-F(x)=0\,. \tag{1.20}$$

Отже, ймовірність набути окреме значення для неперервної випадкової величини завжди дорівнює нулю. Внаслідок цього

$$P\{x_1 < \xi < x_2\} = P\{x_1 \le \xi < x_2\} = P\{x_1 < \xi \le x_2\} = P\{x_1 \le \xi \le x_2\},$$

тобто ймовірність набування значень з деякого проміжку для неперервної випадкової величини не залежить від його відкритості чи замкненості.

Зауваження. Крім дискретних і неперервних випадкових величин та їх комбінацій, існують випадкові величини, які в жодному інтервалі не є ні дискретними, ні неперервними. Такі величини називають *сингулярними* випадковими величинами. Функція розподілу $F(x)$ сингулярної випадкової величини неперервна $\forall x \in R^1$, але $F'(x) = 0$ майже скрізь. Іншими словами, *сингулярна випадкова величина* − це неперервна випадкова величина, яка не має щільності розподілу ймовірностей.

Числові характеристики

Нехай ξ − неперервна випадкова величина, задана щільністю розподілу ймовірностей $p(x)$.

Означення 1.21. *Математичним сподіванням* неперервної випадкової величини ξ називають число

$$M_\xi = \int_{-\infty}^{+\infty} x \cdot p(x)\, dx. \qquad (1.21)$$

Зауваження. Для існування M_ξ потрібно вимагати абсолютної збіжності інтегралу, тобто виконання умови $\int_{-\infty}^{+\infty} |x|\, p(x)\, dx < +\infty$.

Математичне сподівання неперервної випадкової величини має такі самі властивості, що і дискретна випадкова величини (див. п. 1.2.).

Дисперсія неперервної випадковові величини та її властивості означаються так само, як і для дискретної (див. п. 1.2.).

Для неперервної випадкової величини ξ формула (1.5) набуває вигляду

$$D_\xi = \int_{-\infty}^{+\infty} \left(x - M_\xi\right)^2 p(x)dx. \qquad (1.21)$$

Числові характеристики такі, як початковий та центральний моменти k-го порядку, коефіцієнт варіації, коефіцієнт асиметрії, ексцес для неперервної випадкової величини означаються так само, як і для дискретної (див. п.1.2).

Дамо означення ще однієї числової характеристики неперервної випадкової величини ξ.

Означення 1.22. *Модою* розподілу називається кожне значення x, при якому $p(x)$ досягає максимуму.

Часто зустрічаються розподіли, що мають єдину моду; їх називають *унімодальними.* Прикладом унімодального розподілу є нормальний закон розподілу.

1.5. Основні розподіли неперервної випадкової величини

1. Рівномірний закон розподілу на відрізку

Означення 1.23. Неперервна випадкова величина ξ називається рівномірно розподіленою на відрізку $[a,b]$, якщо її щільність розподілу ймовірностей визначається формулою

$$p(x) = \begin{cases} c, x \in [a,b], \\ 0, x \notin [a,b]. \end{cases}$$

З умови нормування знаходимо, що

$$1=\int\limits_{-\infty}^{+\infty} p(x)dx=\int\limits_a^b c\,dx=c\int\limits_a^b dx=c\,x\Big|_a^b=c(b-a).$$

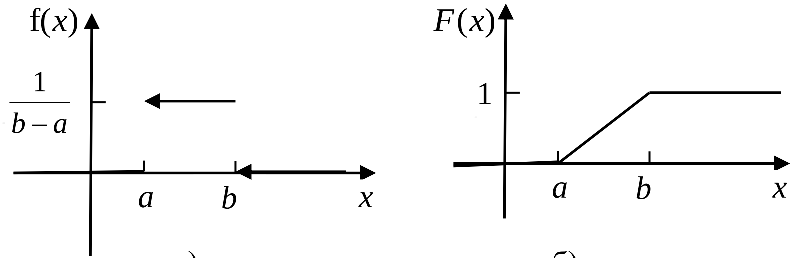
Звідси $c=\dfrac{1}{b-a}$. Таким чином, щільність розподілу ймовірностей

$$p(x)=\begin{cases}\dfrac{1}{b-a}, x\in[a,b],\\ 0, x\notin[a,b].\end{cases}$$

Функція розподілу має вигляд

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

Графіки щільності і функції розподілу мають вигляд (рис. 1.1):



а) б)

Рис. 1.1. Графік щільності і функції рівномірного розподілу

Числові характеристики рівномірного розподілу

$$M_\xi=\int\limits_{-\infty}^{+\infty} x\,p(x)dx=\ldots=\frac{a+b}{2},$$

$$D_\xi = \int\limits_{-\infty}^{+\infty} x^2\, p(x)dx - \left(\frac{a+b}{2}\right)^2 = \ldots = \frac{(b-a)^2}{12}.$$

3. Нормальний закон розподілу (розподіл Гаусса)

Означення 1.25. Неперервна випадкова величина ξ називається нормально розподіленою, якщо її щільність розподілу ймовірностей визначається формулою

$$f(x) = c\,e^{-\frac{(x-a)^2}{2\sigma^2}} \equiv c\exp\left(-\frac{(x-a)^2}{2\sigma^2}\right).$$

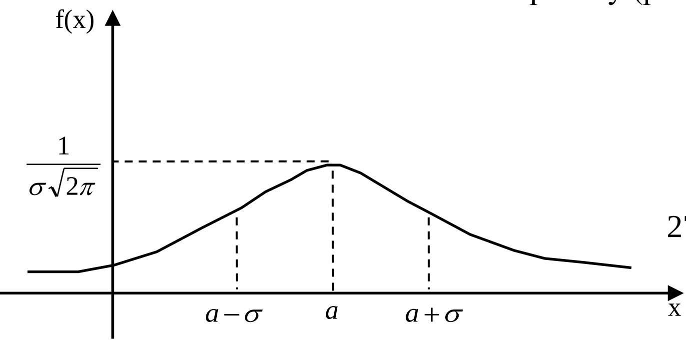
З умови нормування знаходимо, що

$$1 = \int\limits_{-\infty}^{+\infty} p(x)dx = c\int\limits_{-\infty}^{+\infty} \exp\left(-\frac{(x-a)^2}{2\sigma^2}\right)dx = c\sigma\sqrt{2\pi}.$$

Звідси $c = \dfrac{1}{\sigma\sqrt{2\pi}}$. Таким чином, щільність розподілу ймовірностей нормального закону

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

Параметри розподілу: $a \in R^1, \sigma > 0$. Функція $f(x)$ досягає максимуму при $x = a$, графік її симетричний відносно прямої $x = a$ і має вісь OX асимптотою. Точки $x = a \pm \sigma$ є точками перегину (рис. 1.2).



о

Рис. 1.2. Графік щільності нормального розподілу

Ймовірність того, що нормально розподілена випадкова величина набуває значень з деякого проміжку, наприклад, $[\alpha, \beta]$, обчислюється за формулою:

$$P\{\alpha \le \xi \le \beta\} = \int_{\alpha}^{\beta} p(x)dx = \Phi\left(\frac{\beta - a}{\sigma}\right) - \Phi\left(\frac{\alpha - a}{\sigma}\right),$$

де $\Phi(x)$ − функція Лапласа.

Якщо покласти $\alpha = a - 3\sigma, \beta = a + 3\sigma$, то

$$P\{a - 3\sigma \le \xi \le a + 3\sigma\} = \Phi(3) - \Phi(-3) = 2\Phi(3) \approx 0{,}997,$$

тобто подія $\{a - 3\sigma \le \xi \le a + 3\sigma\}$ є практично достовірною. Це означає, що практично всі можливі значення нормально розподіленої випадкової величини зосереджені на проміжку $[a - 3\sigma, a + 3\sigma]$. Останнє твердження називають **«правилом трьох сігм».**

Числові характеристики нормально розподіленої випадкової величини:

$$M_{\xi} = \int_{-\infty}^{+\infty} x\, p(x)dx = a, D_{\xi} = \int_{-\infty}^{+\infty} x^2\, p(x)dx - a^2 = \sigma^2.$$

Показниковий закон розподілу

Означення 1.26. Неперервна випадкова величина ξ називається розподіленою за показниковим (експоненціальним) законом, якщо її щільність розподілу ймовірностей визначається формулою

$$p(x) = \begin{cases} 0, x < 0, \\ ce^{-\lambda x}, x \ge 0. \end{cases}$$

З умови нормування знаходимо, що

$$1 = \int_{-\infty}^{+\infty} p(x)dx = c\int_{0}^{+\infty} e^{-\lambda x}\,dx = -\frac{c}{\lambda}e^{-\lambda x}\Big|_{0}^{+\infty} = \frac{c}{\lambda}.$$

Звідси $c = \lambda$. Отже, щільність показникового закону розподілу

$$p(x) = \begin{cases} 0, x < 0, \\ \lambda e^{-\lambda x}, x \ge 0, \end{cases}$$ λ − параметр розподілу.

Функція розподілу

$$F(x) = \int_{-\infty}^{x} p(t)dt = \begin{cases} 0, x < 0, \\ 1 - e^{-\lambda x}, x \ge 0. \end{cases}$$

Графіки щільності розподілу і функції розподілу мають вигляд:

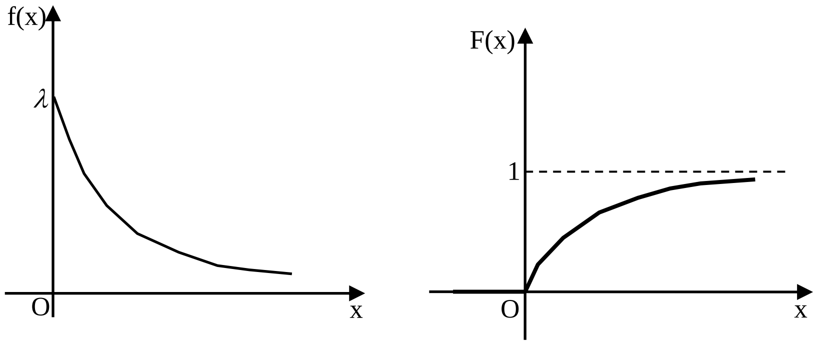


Рис. 1.3. Графік щільності і функції показникового розподілу

Числові характеристики:

$$M_\xi = \int_{-\infty}^{+\infty} x\, p(x)dx = \frac{1}{\lambda}, D_\xi = \int_{-\infty}^{+\infty} x^2\, p(x)dx - \frac{1}{\lambda^2} = \frac{1}{\lambda^2}.$$

Задачі для самостійного розв’язування

1. При яких значеннях параметрів a, b, c подані нижче функції будуть неперервними функціями розподілу:

а) $F(x) = \begin{cases} 0, & x \le 0, \\ \frac{ax}{1+bx}, & x > 0; \end{cases}$

б) $F(x) = \begin{cases} a, & x \le a, \\ bx + c, & a < x \le b, \\ b, & x > b; \end{cases}$

в) $F(x) = \begin{cases} 0, & x \le -c, \\ a + b\,arcsin\frac{x}{a}, & -c < x \le c, \\ b, & x > c; \end{cases}$

г) $F(x) = a + b\,arctg\frac{x}{c}$;

д)) $F(x) = \begin{cases} 0, & x \le 0, \\ a + be^{cx}, & x > 0; \end{cases}$

е) $F(x) = \begin{cases} a, & x \le 0, \\ b\sqrt{x}, & 0 < x \le \frac{1}{4}, \\ cx^2, & \frac{1}{4} < x \le \frac{\sqrt{2}}{4}, \\ 1, & x > \frac{\sqrt{2}}{4}; \end{cases}$

є) $F(x) = \begin{cases} a, & x \le 1, \\ b\,lnx, & 1 < x \le 2, \\ cx, & 2 < x \le 3, \\ 1, & x > 3; \end{cases}$

ж) $F(x) = \begin{cases} c, & x \le 0, \\ ax^2 + bx + c, & 0 < x \le 1, \\ 1, & x > 1; \end{cases}$

2. Випадкова величина Х задана інтегральною функцією

$$F(x)=\begin{cases}0, & x\le -1,\\ \frac{3}{4}x+\frac{3}{4}, & -1<x\le\frac{1}{3},\\ 1, & x>\frac{1}{3};\end{cases}$$

Знайти ймовірність того що X прийме значення з проміжку $\left(0,\frac{1}{3}\right)$.

3. Дана функція розподілу неперервної випадкової величини X

$$F(x)=\begin{cases}0, & x\le 0,\\ \sin x, & 0<x\le\frac{\pi}{2},\\ 1, & x>\frac{\pi}{2};\end{cases}$$

Знайти диференціальну функцію розподілу $f(x)$ (щільність імовірності).

4. Задана щільність ймовірності величини X

$$f(x)=\begin{cases}\frac{2}{3}\sin 3x, & x\in\left(0,\frac{\pi}{3}\right);\\ 0, & x\notin\left(0,\frac{\pi}{3}\right).\end{cases}$$

Знайти ймовірність того, що X прийме значення із $\left(\frac{\pi}{6},\frac{\pi}{4}\right)$.

5. Задана диференціальна функція розподілу неперервної випадкової величини X

$$f(x)=\begin{cases}x-\frac{1}{2}, & 1<x\le 2;\\ 0, & x\notin(1{,}2).\end{cases}$$

Знайти інтегральну функцію розподілу X.

6. Задана щільність імовірності випадкової величини X $f(x)=\frac{2C}{1+x^2}$.

Знайти постійний параметр C.

7. Щільність імовірності випадкової величини X задана виразом

$$f(x) = \begin{cases} a\cos x, & x \in \left[0, \frac{\pi}{2}\right]; \\ 0, & x \notin \left[0, \frac{\pi}{2}\right]. \end{cases}$$

Знайти коефіцієнт a, математичне сподівання та дисперсію.

8. Задана функція

$$F(x) = \begin{cases} 0, & x < 0, \\ \frac{1}{2}\sin x, & 0 \le x \le \pi, \\ 0, & x > \pi; \end{cases}$$

Показати, що вона буде диференціальною функцією розподілу деякої випадкової величини X. Знайти $M(X), D(X), \sigma(X)$.

9. За даною щільністю ймовірностей

$$f(x) = \begin{cases} 0, & x \le -1, \\ \frac{3}{64}(x+1)^2, & -1 < x \le 3, \\ 0, & x > 3; \end{cases}$$

Знайти $M(X), \sigma(X), Me$.

10. Задано

$$F(x) = \begin{cases} 0, & x \le 0, \\ \frac{1}{2}(1 - \cos x), & 0 < x \le \pi, \\ 1, & x > \pi; \end{cases}$$

Знайти $M(X), D(X), Me, Mo$.

11. Задано закон розподілу ймовірностей

$$f(x) = \begin{cases} 0, & x < -3, \\ \frac{1}{4\sqrt{x+3}}, & -3 < x \le 1, \\ 0, & x > 1; \end{cases}$$

Знайти $D(X), \sigma(X), Me$.

12. Задано

$$F(x)=\begin{cases}0, & x\le 1,\\ 1+x(\ln x-1), & 1<x\le e,\\ 1, & x>e;\end{cases}$$

Знайти $\mathrm{M}(X), \sigma(X)$.

13. Задано

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

Знайти $\mathrm{M}(X), D(X)$.

14. Знайдіть функцію розподілу рівномірно розподіленої на відрізку $[1;3]$ випадкової величини. Побудувати її графік. Знайдіть числові характеристики.

15. Доведіть, що коли випадкова величина ξ має показниковий розподіл з параметром λ, то $P(\xi < \mathrm{M}(\xi))$ не залежить від λ. Знайдіть імовірність того, що випадкова величина ξ набере значення більшого, ніж $\mathrm{M}(\xi)$. Знайдіть функцію розподілу і числові характеристики випадкової величини $\eta = -5\xi + 2$.

16. Щільність імовірності випадкової величини ξ має вигляд

$f(x) = 0$, *якщо* $x \le 0$, $\frac{x^k}{k!}e^{-x}$, *якщо* $x > 0$.

Знайдіть числові характеристики цієї випадкової величини.

17. При яких значеннях параметра a функція $f(x)=\frac{a}{1+x^2}$ буде щільністю розподілу ймовірностей? Знайти функцію розподілу і $\mathrm{P}(-1 < \xi < 1)$.

18. При яких значеннях параметра a функція $f(x)=\frac{a}{\sqrt{1-x^2}}$, *якщо* $-1 < x < 1$,0, *в усіх інших випадках* буде щільністю розподілу ймовірностей? Побудувати графіки щільності розподілу ймовірностей і функції розподілу. Знайти *M* (ξ) і *D*(ξ).

19. При яких значеннях параметра *a* функція $f(x) = a\cos^2 x$, *якщо* $-\frac{\pi}{2} < x < \frac{\pi}{2}$, 0, *в усіх інших випадках* буде щільністю розподілу ймовірностей? Побудувати графіки щільності розподілу ймовірностей і функції розподілу. Знайти M(ξ) і *D*(ξ).

20. Випадкова величина ξ має нормальний закон розподілу з параметрами 3 і 3. Записати щільність розподілу і побудувати її графік. Записати функцію розподілу. знайти ймовірність того, що випадкова величина ξ набере значення не меншого 1 і не більшого 3.

21. Випадкова величина ξ рівномірно розподілена на відрізку $[-1;5]$. Записати щільність розподілу і побудувати її графік. Записати функцію розподілу і побудувати її графік. Знайти ймовірність того, що випадкова величина ξ набере значення меншого 3.

22. Щільність нормально розподіленої випадкової величини ξ у точці -2 досягає максимального значення $\frac{1}{4\sqrt{\pi}}$ Знайти дисперсію випадкової величини $\eta = -2\xi + 120$. Знайти P(ξ > 0).

23. Випадкова величина ξ має показниковий розподіл з параметром 1,5. Записати щільність розподілу і побудувати її графік. Записати функцію розподілу і побудувати її графік. Знайти ймовірність того, що випадкова величина ξ набере значення не меншого 0,2 і не більшого 1.

24. Випадкова величина ξ має нормальний закон розподілу з $\mathrm{M}(\xi) = -2$. Знайти D(ξ), якщо відомо, що одна із точок перегину щільності розподілу є точка -5. Яким має бути ε, щоб $P(|\xi - M(\xi)| < \varepsilon) = 0{,}2.$

25. Випадкова величина ξ має нормальний закон розподілу з параметрами -2 і 5. Записати щільність розподілу і побудувати її графік. Записати функцію розподілу. знайти ймовірність того, що випадкова величина ξ набере значення не меншого 0 і не більшого 3.

26. Випадкова величина ξ має показниковий розподіл з математичним сподіванням M (ξ) = 2. Знайти D(ξ). Записати закон розподілу ξ. Яка з цих подій $(\xi < M\xi)$ чи $(\xi > M\xi)$ має більшу ймовірність?

27. Випадкова величина ξ рівномірно розподілена на відрізку $[-2;3]$. Записати щільність розподілу і побудувати її графік. Записати функцію розподілу і побудувати її графік. Знайти ймовірність того, що випадкова величина ξ набере значення більшого 0.

28. Випадкова величина ξ має нормальний закон розподілу з параметрами 2 і 3. Знайти дисперсію випадкової величини $\eta = 2\xi + 1000$. Яка з цих подій $(\xi < M\xi)$ чи $(\xi > M\xi)$ має більшу ймовірність?

29. Математичне сподівання і середнє квадратичне відхилення нормально розподіленої випадкової величини ξ дорівнює 10 і 5. Знайти довжину інтервалу, симетричного відносно математичного сподівання, в який з ймовірністю 0,9973 попаде ξ в результаті випробування.

30. Випадкова величина ξ рівномірно розподілена на відрізку $[a;b]$. Знайти b, якщо $a=2$ і $D(\xi)=3$. Записати щільність, побудувати її графік.

31. Випадкова величина ξ має геометричний закон розподілу з параметром $p = \frac{2}{3}$. Побудувати ряд розподілу випадкової величини ξ, знайти числові характеристики.

1.6. Двовимірні дискретні випадкові величини. Їх числові характеристики

Означення 1.30. Випадкову величину $\xi = (\xi_1, \xi_2, \dots, \xi_n): \Omega \to R^n$ називають багатовимірною випадковою величиною.

Означення 1.31. Двовимірну випадкову величину (ξ, η) називають дискретною, якщо множина її значень не більш як зчисленна.

Для задання такої величини досить вказати всі її можливі значення (x_i, y_k) та ймовірності цих значень $p_{ik} = P\{\xi = x_i, \eta = y_k\}$. Це зручно роботи з допомогою закону розподілу, який можна задати у вигляді таблиці з двома входами.

ξ	x_1	x_2	$\cdots$	x_i	$\cdots$	
y_1	p_{11}	p_{21}	$\cdots$	p_{i1}	$\cdots$	q_1
y_2	p_{12}	p_{22}	$\cdots$	p_{i2}	$\cdots$	q_2
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
y_k	p_{1k}	p_{2k}	$\cdots$	p_{ik}	$\cdots$	q_k
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
η	p_1	p_2		p_i	$\cdots$	1

Причому $p_i = \sum_k p_{ik}$, $q_k = \sum_i p_{ik}$ або

$$p_i = \sum_k P\{\xi = x_i, \eta = y_k\} = P\{\xi = x_i\},$$
$$q_k = \sum_i P\{\xi = x_i, \eta = y_k\} = P\{\eta = y_k\}.$$

Таким чином, ймовірності $\{p_i\}$ визначають розподіл величини ξ, а ймовірності $\{q_k\}$ − розподіл випадкової величини η. При цьому

$$\sum_i p_i = \sum_k q_k = \sum_{i,k} p_{ik} = 1.$$

Числові характеристики

Нехай задано n-вимірну випадкову величину $\xi = (\xi_1, \xi_2, ..., \xi_n)$, для кожної компоненти якої існує скінченне математичне сподівання.

Означення 1.32. *Математичним сподіванням вектора* ξ називають вектор

$$M_\xi = \left(M_{\xi_1}, M_{\xi_2}, ..., M_{\xi_n}\right).$$

У випадку дискретної випадкової величини ξ, заданої законом розподілу $P\{\xi = x_s, \eta = y_k\} = p_{ik}$, можна записати формулу

$$M_\xi = \left(\sum_{i,k} x_k p_{ik}, \sum_{i,k} y_k\ p_{ik}\right).$$

Означення 1.33. *Дисперсією (дисперсійною матрицею)* вектора ξ називається сукупність n^2 чисел d_{ik}, які визначаються формулами

$$d_{ik} = M\left[\left(\xi_i - M_{\xi_i}\right)\left(\xi_k - M_{\xi_k}\right)\right]; i, k = \overline{1, n} \qquad (1.22)$$

З формули (1.22) випливає, що дисперсійна матриця D_ξ симетрична, тобто $d_{ik} = d_{ki}$.

У випадку двовимірної випадкової величини $\xi = (\xi, \eta)$ дисперсійна матриця є матрицею 2-го порядку:

$$D_\xi = \begin{pmatrix} d_{11} & d_{12} \\ d_{21} & d_{22} \end{pmatrix},$$

у якій $d_{11} = M\left(\xi - M_\xi\right)^2 = D_\xi, d_{22} = M\left(\eta - M_\eta\right)^2 = D_\eta,$ $d_{12} = d_{21} = M\left[\left(\xi - M_\xi\right)\left(\eta - M_\eta\right)\right].$ Число $d_{12} = d_{21}$ називається *коваріацією або другим змішаним центарльним моментом* випадкових величин ξ і η і позначається символом $\operatorname{cov}(\xi, \eta)$. Очевидно, що

$$\operatorname{cov}(\xi, \eta) = M(\xi\eta) - M_\xi \cdot M_\eta. \qquad (1.23)$$

Таким чином, дисперсійна матриця вектора ξ має структуру

$$D_\xi = \begin{pmatrix} D_\xi & \operatorname{cov}(\xi, \eta) \\ \operatorname{cov}(\xi, \eta) & D_\eta \end{pmatrix}. \qquad (1.24)$$

Якщо випадкові величини ξ і η незалежні, то $\operatorname{cov}(\xi, \eta) = 0$ і матриця (1.24) є діагональною матрицею

$$D_\xi = \begin{pmatrix} D_\xi & 0 \\ 0 & D_\eta \end{pmatrix}.$$

1.7. Двовимірні неперервні випадкові величини. Їх числові характеристики

Означення 1.34. Двовимірну випадкову величину (ξ, η) називають *неперервною* (точніше *абсолютно неперервною*), якщо її функцію розподілу $F(x, y)$ можна зобразити у вигляді

$$F(x, y) = \int_{-\infty}^{x} \int_{-\infty}^{y} p(\alpha, \beta)\, d\alpha\, d\beta, \qquad (1.25)$$

де $p(x,y)$ − невід'ємна функція, яка задоввольняє умову нормування $\int\limits_{-\infty}^{x}\int\limits_{-\infty}^{y} p(x,y)dx\,dy = 1$.

Функція $p(x,y)$, як і в одновимірному випадку, називається *щільністю розподілу ймовірностей.*

Знаючи щільність розподілу $p(x,y)$ випадкового вектора (ξ,η), легко знайти щільності розподілу його компонент: $p_\xi(x) = \int\limits_{-\infty}^{+\infty} p(x,\beta)d\beta$, $p_\eta(x) = \int\limits_{-\infty}^{+\infty} p(\alpha,y)d\alpha$.

Розглянемо двовимірну випадкову величину $\varsigma = (\xi,\eta)$ зі щільністю розподілу ймовірностей $p(x,y)$. Математичні сподівання координат:

$$M_\xi = \int\limits_{-\infty}^{+\infty} x\, p_\xi(x)dx = \int\limits_{-\infty}^{+\infty} x \int\limits_{-\infty}^{+\infty} p(x,y)dydx = \int\limits_{-\infty}^{+\infty}\int\limits_{-\infty}^{+\infty} x\, p(x,y)dxdy,$$

$$M_\eta = \int\limits_{-\infty}^{+\infty} y\, p_\eta(y)dy = \int\limits_{-\infty}^{+\infty} y \int\limits_{-\infty}^{+\infty} p(x,y)dydx = \int\limits_{-\infty}^{+\infty}\int\limits_{-\infty}^{+\infty} y\, p(x,y)dxdy.$$

Отже, $M_\varsigma = \left(\iint\limits_{R^2} x\, p(x,y)dxdy, \iint\limits_{R^2} y\, p(x,y)dx\,dy \right)$.

Поняття дисперсії і формула обчислення дисперсії визначені у пункті 1.6.

1.8. Функції від випадкових величин

Нехай ξ − випадкова величина, а $y = f(x)$ − деяка функція, визначена на множині значень цієї величини. Утворимо функцію випадкового аргументу $\eta = f(\xi)$ і будемо вважати, що η також є випадковою величиною. Відповідно до означення випадкової величини, це означає, що множина $\{\omega : \eta(\omega) < y\} \in S$ для будь-якого $y \in R^1$, тобто є випадковою подією.

1. Якщо випадкова величина ξ − дискретна, то випадкова величина η також є дискретною випадковою величиною. За відомим рядом розподілу випадкової величини ξ можна побудувати ряд розподілу величини η.

а) Нехай $y = f(x)$ − монотонна функція. Тоді дискретуна випадкова величина $\eta = f(\xi)$ має ряд розподілу $P\{\eta = y_i\} = P\{\eta = f(x_i)\} = p_i$.

б) Якщо $y = f(x)$ − немонотонна функція, то серед її значень $y_i = f(x_i)$ можуть бути однакові. В цьому випадку стовпчики з рівними значеннями об'єднуються в один стовпчик, а відповідні їм ймовірності додаються.

2. Нехай ξ − неперервна випадкова величина з функцією розподілу $F_\xi(x)$ і щільністю розподілу ймовірностей $p_\xi(x)$.

а) Якщо $y = f(x)$ − монотонно зростаюча неперервно диференційовна функція, то для неї існує обернена монотонно зростаюча неперервно диференційовна функція $x = f^{-1}(y)$. У цьому випадку інтегральна та диференціальна функція розподілу випадкової величини η визначаються формулами:

$$F_\eta(x) = F_\xi\left(f^{-1}(x)\right),\ p_\eta(x) = p_\xi\left(f^{-1}(x)\right)\cdot\left(f^{-1}(x)\right)'.$$

б) Якщо $y = f(x)$ − монотонно спадна неперервно диференційовна функція, то інтегральна та диференціальна функції розподілу випадкової величини $\eta = f(\xi)$ визначаються формулами:

$$F_{\eta}(x) = 1 - F_{\xi}\left(f^{-1}(x)\right), p_{\eta}(x) = -p_{\xi}\left(f^{-1}(x)\right)\cdot\left(f^{-1}(x)\right)'.$$

1.9. Закон великих чисел і центральна гранична теорема

Під *законом великих чисел (ЗВЧ)* розуміють твердження, в яких встановлюється, що при певних умовах деякі події є практично достовірними. Зрозуміло, що протилежні події будуть практично неможливими, тому досить обмежитись вивченням, наприклад, практично достовірних подій. З однією з найпростіших форм закону великих чисел ми вже знайомі − це теорема Бернуллі. Згідно з цією теоремою, при великому числі випробувань випадкова подія $\left|\frac{\mu}{n} - p\right| < \varepsilon$ є практично достовірною для будь-якого $\varepsilon > 0$.

Ефективним методом доведення теорем, що виражають закон великих чисел, є **нерівності Чебишева**: якщо випадкова величина ξ має скінченні математичне сподівання і дисперсію, то для будь-якого $\varepsilon > 0$ виконуються нерівності

$$P\left\{\left|\xi - M_{\xi}\right| \geq \varepsilon\right\} \leq \frac{D_{\xi}}{\varepsilon^2}, P\left\{\left|\xi - M_{\xi}\right| < \varepsilon\right\} \geq 1 - \frac{D_{\xi}}{\varepsilon^2}. \qquad (1.26)$$

Теорема Чебишева (закон великих чисел). Якщо випадкові величини $\xi_1, \xi_2, \ldots \xi_n$ незалежні з математичним сподіванням M_{ξ} і дисперсіями, які обмежені однією і тією ж константою c. Тоді для довільного $\varepsilon > 0$ виконується рівність:

$$\lim_{n\to\infty} p\left(\left|\frac{1}{n}\sum_{k=1}^{n}\xi_k - M_{\xi}\right| < \varepsilon\right) = 1. \qquad (1.27)$$

Узагальнена теорема Чебишева. Якщо випадкові величини $\xi_1, \xi_2, ...\xi_n$ незалежні, мають скінченні математичні сподівання, і дисперсії обмежені в сукупності $D(\xi_k) \le c, k = \overline{1,\infty}$, де c − стала величина, то для довільного $\varepsilon > 0$

$$\lim_{n\to\infty} p\left(\left|\frac{1}{n}\sum_{k=1}^{n}\xi_k - \frac{1}{n}\sum_{k=1}^{n}M(\xi_k)\right| < \varepsilon\right) = 1. \qquad (1.28)$$

Формула (1.28) означає, що середнє арифметичне значень випадкових величин, коли кількість доданків нескінченно зростає, збіжна за ймовірністю до середнього арифметичного їх математичних сподівань.

Зауваження. Для використання у практичній діяльності узагальненої теореми Чебишева її можна формулювати таким чином: якщо дисперсії незалежних випадкових величин обмежені, то для досить великих n з будь-якою точністю має місце наближена рівність: $\frac{1}{n}\sum_{k=1}^{n}\xi_k \approx \frac{1}{n}\sum_{k=1}^{n}M(\xi_k)$.

Теорема Бернуллі. Нехай k − кількість успіхів у n випробуваннях Бернуллі, а p − ймовірність успіху у кожному випробуванні. Тоді для довільного $\varepsilon > 0$ виконується рівність $\lim_{n\to\infty} P\left(\left|\frac{k}{n} - p\right| < \varepsilon\right) = 1$.

Теорема Бернуллі є класичним варіантом закону великих чисел.

Задачі для самостійного розв'язування

1. Імовірність появи випадкової події в одному експерименті є величиною сталою і дорівнює 0,3. Із якою імовірністю можна стверджувати, що відносна частота цієї події при 100 експериментах буде знаходитись у межах [0,2;0,4].

2. Випадкова подія A може здійснитися при одному експерименті із імовірністю p. Експеримент повторили n раз. Яка ймовірність того, що при цьому виконується нерівність

$$np - 2\sqrt{npq} < m < np + 2\sqrt{npq}\,.$$

3. Яке повинна мати значення величина ε у нерівності Чебишова, щоб $P(|X - a| < \varepsilon) \approx 0{,}99$, коли відомо, що $D(\mathrm{X}) = 4$.

4. Із якою надійністю середнє арифметичне вимірів певної величини відповідає істинному виміру цієї величини, якщо було здійснено 500 вимірювань із точністю 0,1 і при цьому дисперсії випадкових величин – результатів вимірювання – не перевищують 0,3.

5. Скільки необхідно провести вимірів діаметра втулки, щоб середнє арифметичне цих вимірів відрізнялося від істинного розміру діаметра втулки не більше як 0,05 із надійністю 90%, якщо дисперсії випадкових величин (результатів вимірів) не перевищують 0,2.

6. Імовірність того, що за час t із ладу вийде один конденсатор, дорівнює 0,2. Яка ймовірність того, що за час t із 100 конденсаторів із ладу вийде:

1) не менш як 28 конденсаторів;

2) від 14 до 26 конденсаторів?

7. При відливанні відливок, із яких потім виготовляють на верстатах деталі, одержують у середньому 20% браку. Скільки необхідно запланувати відливок, щоб із імовірністю не меншою за 0,95 була забезпечена програма випуску деталей, для виготовлення яких необхідно 50 бездефектних відливок.

8. Здійснюється вибіркове обстеження партії електроламп для визначення тривалості їх горіння. Скільки необхідно перевірити електролампочок, щоб із імовірністю не меншою за 0,9876 можна було стверджувати, що середня тривалість горіння лампочки для всіх n штук перевірених відхилялось від її середньої величини не більше ніж на 10 годин, якщо середнє квадратичне відхилення тривалості горіння лампочок дорівнює 80 годин.

9. Випадкова величина $\overline{X}$ - середнє арифметичне 10000 незалежних випадкових величин, що мають один і той самий закон розподілу, і середнє квадратичне відхилення кожної із них дорівнює 2. Яке максимальне відхилення величини $\overline{X}$ від його математичного сподівання можна очікувати із імовірністю 0,9544?

10. Верстат із програмним управлінням виготовляє за робочу зміну 900 виробів, із яких в середньому 1% складає брак. Знайти наближено ймовірність того, що за зміну буде виготовлено не менше 810 доброякісних виробів, якщо вони виявляються доброякісними незалежно один від одного.

11. Кожна із 40 незалежних випадкових величин має гамма-розподіл із значеннями параметрів

$\alpha = 2, \lambda = 10$. На підставі центральної граничної теореми теорії ймовірностей записати наближено закон розподілу для випадкової величини $X = \sum_{i=1}^{40} X_i$.

Розділ 2
Елементи математичної статистики

2.1. Основні поняття математичної статистики

Вихідними поняттями математичної статистики є генеральна сукупність і вибірка.

Нехай X − випадкова величина, функція розподілу якої $F(x)$ нам невідома. Досліджуючи цю величину, ми здійснюємо n раз той самий експеримент, в результаті чого дістанемо n значень величини $X: x_1, x_2, ..., x_n$. Експеримент полягає в тому, що ми вибираємо навмання один з елементів множини A, реєструємо деяку його характеристику X і повертаємо цей елемент до множини. Множину A називають *генеральною сукупністю,* а групу елементів, які спостерігалися при n повтореннях експерименту, − *випадковою вибіркою.*

У більшості випадків нас цікавитимуть не самі елементи множини A, а лише згадана їх характеристика X і закон її розподілу. Тому у випадку довільної випадкової величини X множину всіх її можливих значень також називають *генеральною сукупністю,* а утворені в результаті n експериментів числа $x_1, x_2, ..., x_n$ − *вибіркою* з цієї сукупності.

Отже, із теоретико-ймовірнісного погляду *генеральна сукупність* − це випадкова величина X, задана на просторі елементарних подій Ω із виділеним у ньому полем подій F, для якого вказані їх імовірності p. *Вибірка з даної генеральної сукупності* − це результати обмеженого ряду спостережень $x_1, x_2, ..., x_n$ випадкової величини X.

Означення 2.1. Число n спостережень, що утворюють вибірку, називають *обсягом вибірки*, а числа $x_1, x_2, ..., x_n$ − *варіантами вибірки.*

Означення 2.2. Послідовність варіант, записаних у порядку зростання називають *варіаційним рядом.*

Означення 2.3. Число $w_i = \dfrac{n_i}{n} \left(\sum\limits_{n=1}^{k} w_n = 1 \right)$ називається *відносною частотою вибірки.*

Означення 2.4. Якщо ряд спостережень $x_1, x_2, ..., x_n$ утворює послідовність незалежних і однаково розподілених випадкових величин, то *вибірка* називається *випадковою.*

Статистичні ряди розподілу вибірки

Нехай вивчають деяку генеральну випадкову величину X. Для цього проводять низку незалежних дослыдыв або спостережень, у кожному з яких величина X набуває того чи іншого значення. Сукупність отриманих значень величини X (де n − число дослідів) і є утворена нами вибірка. Цю сукупність значень випадкової величини X називають *статистичним рядом.*

1. *Дискретний статистичний розподіл вибірки*

Нехай X − дискретна випадкова величина.

Означення 2.5. Дискретним статистичним розподілом вибірки називається відповідність між варіантами та їх частотами або відносними частотами.

Дискретний статистичний розподіл вибірки можна подати у формі таблиць:

- дискретний статистичний розподіл частот

Таблиця 2.1

x_i	x_1	x_2	$\cdots$	x_k
n_i	n_1	n_2	$\cdots$	n_k

- дискретний статистичний розподіл відносних частот

Таблиця 2.2

x_i	x_1	x_2	$\cdots$	x_k
w_i	w_1	w_2	$\cdots$	w_k

де $w_i = \dfrac{n_i}{n}$, $\sum\limits_{i=1}^{k} w_i = w_1 + w_2 + ... + w_k = 1.$

2. *Інтеравальний статистичний розподіл вибірки*

Нехай X − неперервна випадкова величина.

Означення 2.6. Неперервним статистичним розподілом вибірки називається відповідність між варіантами та їх частотами або відносними частотами.

Інтервальний статистичний розподіл вибірки, як і дискретний, також записують у формі таблиць:

- інтервальний статистичний розподіл частот

Таблиця 2.3

$(z_{i-1}, z_i]$	$(z_0, z_1]$	$(z_1, z_2]$	$\cdots$	$(z_{m-1}, z_m]$
n_i	n_1	n_2	$\cdots$	n_m

$$\sum_{i=1}^{m} n_i = n_1 + n_2 + ... + n_m = n;$$

- інтервальний розподіл відносних частот:

$(z_{i-1}, z_i]$	$(z_0, z_1]$	$(z_1, z_2]$	$\cdots$	$(z_{m-1}, z_m]$
w_i	w_1	w_2	$\cdots$	w_m

$$\sum_{i=1}^{m} w_i = w_1 + w_2 + ... + w_m = 1.$$

Зауважимо, що інтервальний статистичний розподіл вибірки використовують також у разі, коли випадкова величина дискретна і обсяг різних варіант вибірки є досить великим.

Інтервальний статистичний розподіл вибірки за необхідності можна замінити дискретним. Для цього досить частинні проміжки $(z_{i-1}, z_i]$ замінити числами − серединами цих проміжків (тобто прийняти $x_i = \dfrac{z_{i-1} + z_i}{2}$), а відповідні їм значення частот (відносних частот) залишити без змін.

Полігони і гістограми

Статистичний розподіл вибірки можна задати графічно полігоном або гістограмою частот (відносних частот).

Полігон розподілу вибірки використовується для зображення як дискретних, так і інтервальних варіаційних рядів, а гістограма − лише для інтервальних рядів.

Означення 2.7. Полігоном частот називають ламану, відрізки якої послідовно з’єднують точки (x_1, n_1), (x_2, n_2), …, (x_k, n_k) координатної площини.

Щоб побудувати полігон частот, на осі абсцис відкладають варіанти x_i, а на осі ординат – відповідні їм частоти n_i. Точки (x_i, n_i) з’єднують відрізками прямих і отримують полігон частот

Означення 2.8. Полігоном відносних частот називають ламану, відрізки якої послідовно з’єднують точки (x_1, w_1), (x_2, w_2), …, (x_k, w_k).

Щоб побудувати полігон відносних частот, на осі абсцис відкладають варіанти x_i, а на осі ординат – відповідні їм відносні частоти частоти w_i. Точки (x_i, w_i) послідовно з’єднують відрізками прямих і отримують полігон відносних частот.

Означення 2.9. Гістограмою частот називається східчаста фігура, яка складена з прямокутників, основами яких є частинні проміжки $[z_{i-1}, z_i)$, $i = 1, 2, ..., m$, а їх висоти $h_i = \dfrac{n_i}{z_i - z_{i-1}}$.

Щоб побудувати гістограму частот, на осі абсцис відкладаються частинні проміжки $[z_{i-1}, z_i)$ і на них, як на основі, будують прямокутники з висотами h_i. Площа кожного такого прямокутника дорівнює n_i.

Означення 2.10. Гістограмою відносних частот називається східчаста фігура, яка складена з прямокутників, основами яких є частинні проміжки $[z_{i-1}, z_i)$, $i = 1, 2, ..., m$, а їх висоти $\tilde{h}_i = \dfrac{w_i}{z_i - z_{i-1}}$.

Щоб побудувати гістограму відносних астот, на осі абсцис відкладаються частинні проміжки $[z_{i-1}, z_i)$ і на них, як на основі, будують прямокутники з висотами $\tilde{h}_i$. Площа кожного такого прямокутника дорівнює w_i.

Емпірична функція розподілу

Функція розподілу $F(x)$ випадкової величини X визначається рівністю $F(x) = P(X < x)$, тобто функція розподілу $F(x)$ означає ймовірність події $X < x$. Її називають ще теоретичною функцією розподілу випадкової величини X.

Означення 2.10. Емпіричною функцією розподілу $F^*(x)$ випадкової величини X називається функція, яка для кожного дійсного числа x дорівнює відношенню $\tilde{n}(x)$ до обсягу вибірки n, тобто:

$$F^*(x) = \frac{\tilde{n}(x)}{n} \tag{2.1}$$

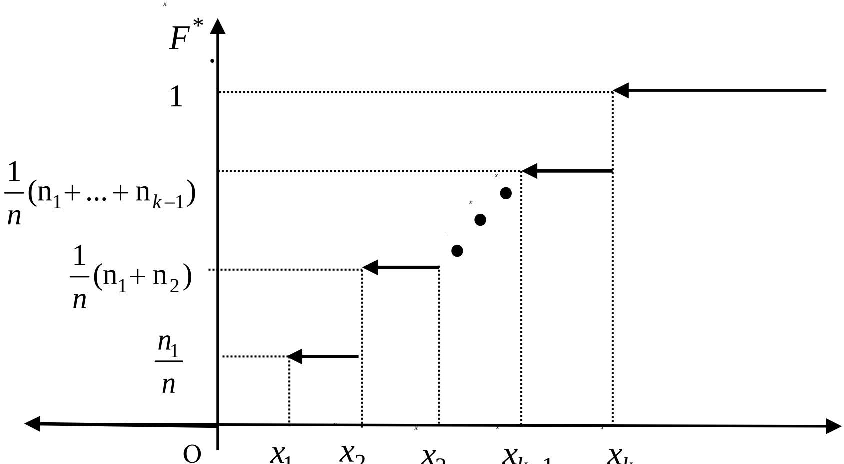
де $\tilde{n}(x)$ - накопичена частота для тих значень випадкової величини X, що менші за будь-яке дійсне число x.

Отже, емпірична функція розподілу виражає для кожного дійсного числа x відносну частоту події $X < x$, тобто $F^*(x) = W_n(X < x)$.

Якщо вихідні статистичні дані згруповані в дискретний варіаційний ряд за табл. 2.1, то емпірична функція розподілу:

$$F^*(x)=\begin{cases} 0, & x \le x_1; \\ \dfrac{n_1}{n}, & x_1 < x \le x_2; \\ \dfrac{n_1+n_2}{n}, & x_2 < x \le x_3; \\ \dots\dots\dots\dots\dots\dots\dots\dots\dots & \\ \dfrac{n_1+n_2+\dots+n_{k-1}}{n}, & x_{k-1} < x \le x_k; \\ 1, & x > x_k. \end{cases}$$

Графік емпіричної функції розподілу $F^*(x)$ зображено на рис. 2.1.



x

Рис. 2.1. Графік емпіричної функції дискретного розподілу

Якщо маємо згрупований інтервальний статистичний ряд, то можемо лише наближено побудувати емпіричну функцію розподілу $F^*(x)$. За значення x, для яких $F^*(x)$ обчислюється за формулою (2.1), природно взяти межі частинних проміжків. У всіх інших точках $F^*(x)$ визначається за допомогою лінійного інтерполювання. Це означає, що геометрично $F^*(x)$ буде зображена неперервною ламаною лінією, яка з'єднує послідовно точки $(z_i, w_i): w_0 = 0,\ w_i = w_1 + w_2 + ... + w_i, i = 1,2,...,m$ (рис. 2.2).

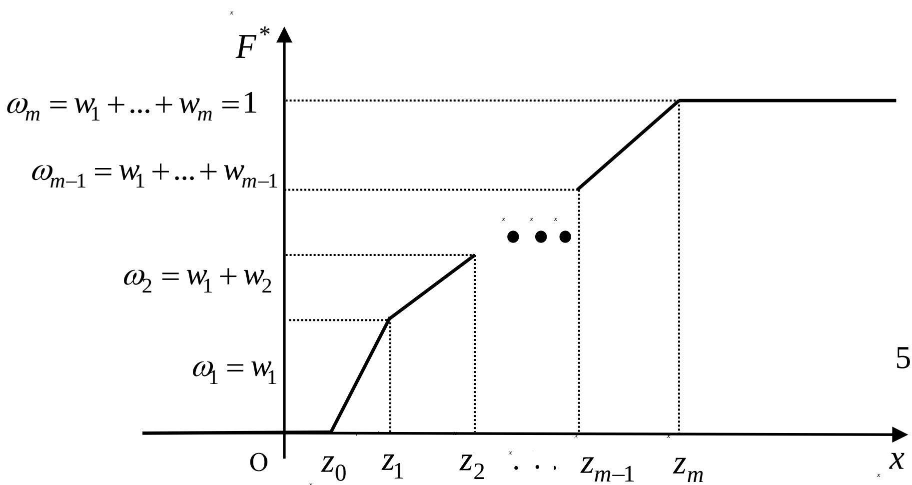


Рис. 2.2. Графік емпіричної функції інтервального розподілу

Аналітично емпірична функція розподілу інтервального статистичного ряду вибірки записується так:

$$F^*(x)=\begin{cases}0, & x\le z_0;\\ F_1^*(x), & z_0<x\le z_1;\\ F_2^*(x), & z_1<x\le z_2;\\ \dots\dots\dots\dots\dots & \\ F_m^*(x), & z_{m-1}<x\le z_m;\\ 1, & x>z_m,\end{cases}$$

де $F_i^*(x)=\dfrac{w_i}{z_i-z_{i-1}}(x-z_{i-1})+w_{i-1}, w_0=0, i=\overline{1,m}$.

2.2. Числові характеристики статистичних розподілів вибірки

У практичних задачах часто замість повного вивчення даних вибірки буває достатньо обмежетися знаходженням їх числових характеристик. Надалі, припустимо, що статистичні дані згруповано в дискретний варіаційний ряд.

Означення 2.11. Вибірковим середнім $\overline{x}$ статистичного розподілу вибірки називається середнє арифметичне значення її варіант x_i з урахуванням їх частот, тобто:

$$\overline{x} = \frac{1}{n}\sum_{i=1}^{k} n_i\, x_i = \frac{n_1x_1 + n_2x_2 + ... + n_kx_k}{n}.$$

Вибіркове середнє $\overline{x}$ є основною характеристикою статистичного рохподілу вибірки. Його узагальненням є поняття початкового емпіричного моменту.

Означення 2.12. Вибірковим початковим емпіричним моментом s-го порядку $\overline{v_s}$ статистичного розподілу вибірки називають середнє арифмитичне значення степенів порядку s варіант x_i, тобто:

$$\overline{v_s} = \overline{x^s} = \frac{1}{n}\sum_{i=1}^{k} x_i^s\, n_i .$$

Якщо $s = 1$, то $\overline{v}_1 = \overline{x}$ − вибірковому середньому.

Розглянемо основні характеристики розсіювання значень випадкової величини навколо її середнього значення.

Означення 2.13. Розмахом вибірки називають різницю між найбільшим і найменшим значеннями її варіант, тобто: $R = x_k - x_1$.

Означення 2.14. Вибірковою дисперсією $\overline{D}$ статисчного розподілу вибірки називають середнє арифметичне значення квадратів відхилень його варіант x_i від вибіркового середнього $\overline{x}$, тобто:

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

Для обчислення вибіркової дисперсії зручніше використовувати іншу формулу:

$$\overline{D} = \frac{1}{n}\sum_{i=1}^{k} x_i^2 \cdot n_i - \left(\overline{x}\right)^2 .$$

Дисперсія має вимірність, яка дорівнює квадрату вимірності значень випадкової величини, що є не зручним у дослідженнях. Щоб усунути цей недолік, за характеристику розсіювання значень випадкової величини за результатами вибірки приймають вибіркове середнє квадратичне відхилення $\overline{\sigma}$:

$$\overline{\sigma} = \sqrt{\overline{D}} = \sqrt{\frac{1}{n}\sum_{i=1}^{k}\left(x_i - \overline{x}\right)^2 n_i} \text{ або}$$

$$\overline{\sigma} = \sqrt{\overline{D}} = \sqrt{\frac{1}{n}\sum_{i=1}^{k} x_i^2\, n_i - \left(\overline{x}\right)^2} .$$

Означення 2.15. Коефіцієнтом варіації V статистичного розподілу вибірки називається відношення вибіркового середнього квадратичного відхилення $\overline{\sigma}$ до середнього вибіркового, тобто:

$$V = \frac{\overline{\sigma}}{\overline{x}} \cdot 100\% .$$

Означення 2.16. Вибірковим центарльним емпіричним моментом s-го порядку $\overline{\mu_s}$ статистичного розподілу вибірки називається середнє арифметичне значення степенів порядку s відхилень його варіант від середнього вибіркового значення $\overline{x}$, тобто:

$$\overline{\mu_s} = \frac{1}{n}\sum_{i=1}^{k}\left(x_i - \overline{x}\right)^s n_i .$$

Зокрема, $\overline{\mu_0} = 1, \overline{\mu_1} = 0, \overline{\mu_2} = \overline{D}$.

Для оцінки відхилення статистичного розподілу вибірки від нормального розподілу використовують числові характеристики − *асиметрію та ексцес.*

Означення 2.*17*. Вибірковою асиметрією $\overline{A}$ називають число, яке обчислюється за формулою:

$$\overline{A}=\frac{\overline{\mu_3}}{\overline{\overline{\sigma^3}}},$$

де $\overline{\mu_3}$ − вибірковий центральний емпіричний момент 3-го порядку, $\overline{\sigma}$ − середнє квадратичне відхилення статистичного розподілу вибірки.

Означення 2.*18*. Вибірковим ексцесом $\overline{E}$ статистичного розподілу вибірки називається число, яке обчислюється за формулою: $\overline{E}=\frac{\overline{\mu_4}}{\overline{\overline{\sigma^4}}}-3$,

де $\overline{\mu_4}$ − вибірковий центральний емпіричний момент 4-го порядку, $\overline{\sigma}$ − вибіркове середнє квадратичне відхилення статистичного розподілу вибірки.

Якщо випадкова величина X розподілена за нормальним законом, то її асиметрія та ексцес дорівнюють нулю. Тому, чим більше віддалені від нуля асиметрія та ексцес статистичного розподілу вибірки, то тим менше підстав сподіватися, що вибірка, з якої утворено варіаційний ряд, утворена з нормально розподіленої генеральної сукупності.

Задачі для самостійного розв'язування

1. Вибірку задано у вигляді розподілу частот:

а)

x_i	3	6	8
n_i	1	4	6

б)

x_i	4	8	10	12
n_i	5	3	4	12

Знайти розподіл відносних частот, емпіричну функцію розподілу та побудувати її графік.

2. За допомогою нівеліра було здійснено 15 вимірювань однієї і тієї самої відносної висоти. Результати вимірювання в сантиметрах: 89, 92, 90, 91, 88, 89, 93, 92, 94, 95, 96, 91, 93, 87, 90. Потрібно: а) побудувати дискретний статистичний розподіл; б) побудувати полігон відносних частот; в) знайти емпіричну функцію розподілу та побудувати її графік; г) обчислити $R, V, \bar{x}, \bar{\sigma}$; д) побудувати інтервальний розподіл, поділивши інтервал на 4 рівних частини.

3. Для обчислення середньої врожайності x_i озимої пшениці поле площею 3000 га було поділено на 25 рівних ділянок. Фактичний урожай на кожній ділянці наведено в таблиці, в якій n_i означає кількість ділянок із врожайністю x_i :

x_i	20	25	30	35	40	45
n_i	3	5	4	5	3	5

Для заданої вибірки виконати дії: а) побудувати полігон відносних частот та емпіричну функцію розподілу; б) обчислити $\bar{x}, \bar{\sigma}, \bar{A}, \bar{E}$.

4. Для дослідження розподілу маси птахів x_i була зібрана інформація щодо 40 качок. Ця інформація подана інтервальним статистичним розподілом:

z_{i-1}-z_i	1,0-1,5	1,5-2,0	2,0-2,5	2,5-3,0	3,0-3,5	3,5-4,0	4,0-4,5	4,5-5,0
n_i	2	4	6	11	9	4	2	2

Побудувати гістограму відносних частот.

Знайти: а) емпіричну функцію розподілу і побудувати її графік; б) побудувати гістограму частот; в) обчислити $\bar{x}, \bar{\sigma}, \bar{A}, \bar{E}$.

5. Кількість деталей x_i, виготовлених робітниками за зміну, наведено у формі інтервального статистичного розподілу:

z_{i-1}-z_i	5-15	15-25	25-35	35-45	45-55	55-65	65-75
n_i	3	4	8	17	11	5	2

Побудувати гістограму відносних частот.
Знайти: а) емпіричну функцію розподілу і побудувати її графік; в) обчислити $\bar{x}, \bar{\sigma}, \bar{A}, \bar{E}$.

6. При вивченні питання про норми виробітку ткаль спостерігалася така частота обривів пряжі на однотипних ткацьких станках на різних проміжках часу однакової довжини: 2, 1, 3, 7, 4, 1, 5, 3, 0, 5, 2, 3, 0, 2, 5, 4, 1, 0, 4, 6, 8, 3, 7, 0, 1, 5, 3, 5, 2, 3.

За наведеним даними необхідно побудувати:
а) варіаційний ряд;
б) таблицю і полігон частот;
в) таблицю і полігон відносних частот;
гістограму частот, розбивши проміжок на 4 рівних частинних інтервали;
г) гістограму відносних частот;
е) емпіричну функцію розподілу та її графік.

2.3. Точкові оцінки параметрів розподілу

Повною характеристикою випадкової величини є її закон розподілу. Для його встановлення певними дослідними методами потрібні певні затрати часу і ресурсів. Як правило вважають, що закон розподілу випадкової велични відомий (нормальний, показниковий, геометричний тощо). Але, щоб конкретизувати цей закон, потрібно знайти параметри розподілу.

Наприклад, коли відомо, що закон розподілу випадкової величини нормальний, тобто характеризується густиною $f(x)=\frac{1}{\sigma\sqrt{2\pi}}e^{-\frac{(x-a)^2}{2\sigma^2}}$, то для його конкретизаії потрібно визначити невідомі параметри a і σ.

Отже, на основі одержаної вибірки з генеральної сукупності потрібно визначити наближені числові значення невідомих параметрів розподілу. Такі наближені числові значення параметрів розподілу називають їх *точковими статистичними оцінками*, або скорочено − *точковими оцінками.*

Зауважимо, що у типових законах розподілу ймовірностей випадкової величини параметри розподілу виражаються через числові характеристики ціїє величини (математичне сподівання, дисперсію, середнє квадратичне відхилення).

2.3.1. Означення точкової оцінки параметрів розподілу

Нехай у генеральній сукупності спостерігається випадкова величина X , відомий закон розподілу якої містить невідомий параметр θ. Потрібно знайти відповідну точкову оцінку θ^* параметра θ за результатами n незалежних випробувань, у кожному з яких випадкова величина X набуває певних значень $x_1, x_2,...x_n$ (вибірка обсягу n).

Означення 2.3.1. Будь-яку однозначну функцію $\theta^* = \theta^*(x_1, x_2, ..., x_n)$, при допомозі якої знаходять наближене значення параметра θ розподілу випадкової величини X, називають точковою оцінкою цього параметра.

Зауважимо, що для різних вибірок, навіть одного обсягу, відповідні значення функції θ^* будуть, загалом кажучи, різними.

Для того, щоб оцінка θ^* була більш точною для параметра θ, тобто мала практичну цінність, вона повинна задовольняти умови незміщеності, консистентності та ефективності.

Означення 2.3.2. Точкова оцінка $\theta^* = \theta^*(x_1, x_2, ..., x_n)$ для параметра розподілу θ випадкової величини X називається незміщеною, якщо її математичне сподівання дорівнює точному значенню цього параметра, тобто $M(\theta^*) = \theta$.

У протилежному випадку точкову оцінку θ^* називають зміщеною.

Суть незміщеної оцінки θ^* параметра розподілу θ випадкової величини X полягає в тому, що вона «застрахована» від накопичення односторонніх похибок (у бік збільшення або зменшення).

Означення 2.3.3. Точкова оцінка $\theta^* = \theta^*(x_1, x_2, ..., x_n)$ параметра розподілу θ називається консистентною (змістовною), якщо θ^* збігається (необмежено наближається) за ймовірністю до оцінюваного параметра при необмеженому зростанні числа спостережень, тобто виконується така рівність: $\lim\limits_{n\to\infty} P\{|\theta^* - \theta| < \varepsilon\} = 1$, де $\varepsilon > 0$ як завгодно мале число.

Для виконання даної граничної рівності достатньо, щоб дисперсія оцінки необмежено наближалася до нуля при $n \to \infty$, тобто $\lim_{n\to\infty} D\left(\theta^*\right) = 0$ і, крім того, щоб оцінка була незміщеною.

Якщо виконана достаня умова консистентності оцінки θ^*, то ця властивість означає, що зі збільшенням обсягу вибірки n розсіювання значень точкової оцінки θ^* зменшується, тобто при достатньо великому обсязі вибірка оцінка θ^* параметра розподілу θ стає як завгодно близькою до її точного значення.

Оскільки θ^* − випадкова величина, значення якої змінюються від вибірки до вибірки, то міру її розсіювання навколо математичного сподівання θ будемо характеризувати дисперсією $D\left(\theta^*\right)$. Нехай θ^* і $\tilde{\theta}^*$ − дві незміщені оцінки параметра θ, тобто $M\left(\theta^*\right) = \theta$ і $M\left(\tilde{\theta}^*\right) = \theta$. Тоді, якщо $D\left(\theta^*\right) < D\left(\tilde{\theta}^*\right)$, то за оцінку для θ приймають θ^*.

Означення 2.3.4. Незміщена оцінка $\theta^* = \theta^*\left(x_1, x_2, ..., x_n\right)$ називається ефективною, якщо вона має найменшу дисперсію серед усіх незміщених оцінок параметра θ, обчислених за вибірками одного і того ж обсягу.

Отже, за статистичного оцінювання параметрів розподілу випадковові величини X точкова оцінка має задовольняти умови незміщеності, консистентності та ефективності. Однак на практиці не завжди вдається досягти виконання трьох зазначених вимог. Оскільки, оцінюючи θ, функція $\theta^*\left(x_1, x_2, ..., x_n\right)$ повинна бути, по можливості, не дуже складною. Через те іноді необхідно нехтувати деякими із перерахованих вище вимог до точкових оцінок.

Зазначимо, що параметри розподілу випадкової величини X у простий спосіб виражаються через її чисельні характеристики. Наприклад, якщо випадкова величина X нормально розподілена, то параметри розподілу $a = M(X), \sigma = \sigma(X)$. Надалі нас цікавить задача про оцінки основних чисельних характеристик випадкової величини X.

2.3.2. Точкова оцінка математичного сподівання

Нехай $x_1, x_2, ..., x_n$ − вибірка, отримана в результаті n незалежних спостережень над випадковою величиною X − деякою ознакою генеральної сукупності, яка має математичне сподівання $M(X) = a$.

За точкову оцінку M^* математичного сподівання $M(X) = a$ беруть вибіркове середнє (емпіричне вибіркове): $M^* = \bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$.

Якщо випадкова величина X нормально розподілена з параметрами $M(X) = a, D(X) = \sigma^2$, то оцінка $\bar{x}$ має у класі всіх незміщених оцінок для математичного сподівання a мінімальну дисперсію, яка дорівнює $\frac{\sigma^2}{n}$. Тому $\bar{x}$ є ефективною оцінкою параметра a.

2.3.3. Точкова оцінка дисперсії

Якщо випадкова вибірка складається з n незалежних спостережень $x_1, x_2, ..., x_n$ над випадковою величиною X із математичним сподіванням $M(X) = a =$ і дисперсією $D(X) = \sigma^2$, то за точкову оцінку D^* дисперсії $D(X) = \sigma^2$ беруть вибіркову дисперсію: $D^* = \overline{D} = \frac{1}{n}\sum_{i=1}^{n}(x_i - \overline{x})^2$, яка є зміщеною оцінкою для параметра $D(X) = \sigma^2$, або виправлену вибіркову дисперсію: $D^* = \tilde{D} = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \overline{x})^2$, яка є незміщеною оцінкою для параметра $D(X) = \sigma^2$.

Дріб $\frac{n}{n-1}$ називають поправкою Бесселя. За малих n поправка Бесселя значно відрізняється від одиниці. За $n > 50$ практично немає різниці між $\overline{D}$ і $\tilde{D}$.

У випадку, коли математичне сподівання a відоме і випадкова величина X нормально розподілена, то незміщеною, конзистентною та ефективною оцінкою дисперсії $D(X) = \sigma^2$ є оцінка $D^* = \overline{D} = \frac{1}{n}\sum_{i=1}^{n}(x_i - a)^2$.

2.3.4. Оцінка середнього квадратичного відхилення

Оскільки середнє квадратичне відхилення $\sigma(X) = \sqrt{D(X)}$, то за оцінку σ^* параметра σ можна вибрати один із варіантів вибіркового середнього квадратичного відхилення: $\sigma^* = \sqrt{D^*}$, де D^* − оцінка для дисперсії $D(X)$.

Зазаначимо, що всі оцінки для σ не є незміщенними, ефективними, а є лише конзистентними.

Зауваження. Усі наведені вище формули для точкових оцінок математичного сподівання і дисперсії випадвкової величини, можуть бути використані для обчислення значень цих оцінок щодо конкретних вибірок лише у випадку, коли вибірка задана вихідним статистичним рядом. Якщо дані вибірки подано у формі згрупованого варіаційного ряду (дискретного або інтервального), то для обчислення відповідних вибіркових чисельних характеристик користуємося формулами:

$$M^* = \frac{1}{n}\sum_{i=1}^{k} n_i \cdot x_i \ ,$$

$$D^* = \overline{D} = \frac{1}{n}\sum_{i=1}^{k} n_i \left(x_i - \overline{x}\right)^2 = \frac{1}{n}\sum_{i=1}^{k} n_i x_i^2 - \overline{x}^2 \ ,$$

$$D^* = \widetilde{D} = \frac{1}{n-1}\sum_{i=1}^{k} n_i \left(x_i - \overline{x}\right)^2 = \frac{1}{n-1}\sum_{i=1}^{k} n_i {x_i}^2 - \frac{n}{n-1}\overline{x}^2.$$

2.3.5. Метод моментів оцінювання параметрів розподілу

Схема проведення статистичного оцінювання параметрів $\theta_1, \theta_2, ..., \theta_m$ методом моментів:

1. Обчислюємо m теоретичних початкових моментів: $\nu_1(\theta_1, \theta_2, ..., \theta_m)$, $\nu_2(\theta_1, \theta_2, ..., \theta_m)$, …, $\nu_m(\theta_1, \theta_2, ..., \theta_m)$, де $\nu_s = M\left(X^s\right)$.

2. На підставі даних вибірки $x_1, x_2, ..., x_n$ обчислюємо m відповідних вибіркових початкових моментів:

$\bar{\nu}_s(\theta_1,...,\theta_m)=\dfrac{1}{n}\sum\limits_{i=1}^{n} x_i^s$ $, s=1,2,...,m$, якщо вибірка задана вихідним статистичним рядом, і

$\bar{\nu}_s(\theta_1,...,\theta_m)=\dfrac{1}{n}\sum\limits_{i=1}^{n} n_i\cdot x_i^s$ $,\sum\limits_{i=1}^{k} n_i=n,\ s=1,2,...,m$, якщо вибірка задана згрупованим варіаційним рядом.

3. Прирівнюючи теоретичні та відповідні їм вибіркові моменти, отримаємо систему рівнянь відносно компонент $\theta_1,\theta_2,...,\theta_m$ оцінюваного параметра

$$\begin{cases} \nu_S(\theta_1,\theta_2,...,\theta_m)=\bar{\nu}_s(\theta_1,\theta_2,...,\theta_m), \\ (s=1,2,...,m). \end{cases}$$

4. Розв'язуючи отриману систему рівнянь (точно або наближено), знаходимо шукані оцінки $(\theta_1^*,\theta_2^*,...,\theta_m^*)$.

Зазначена вище схема розглянута, виходячи з початкових (теоретичних і вибіркових) моментів. Але ця схема зберігається також у разі іншого вибору моментів: центральних або початкових і центральних у сукупності.

2.3.6. Метод максимуму правдоподібності оцінювання параметрів розподілу

Нехай вивчається випадкова величина X, розподіл якої заданий або ймовірностями її значень $f(x_i,a)$, якщо величина X дискретна, або густиною розподілу $f(x,\theta)$, якшо величина X неперервна, де $\theta=(\theta_1,\theta_2,...,\theta_m)$ − невідомий векторний параметр. Нехай $(x_1,x_2,...,x_n)$ − вибірка, отримана внаслідок n незалежних спостережень над випадковою величиною X. Ідея методу максимуму правдоподібності полягає в тому, що за оцінку θ^* приймають таке значення параметра θ, для якого ймовірність отримання вже наявної вибірки є максимальною.

Опишемо схему застосування цього методу. Для цього розгялнемо функцію $L(x_1,...,x_n,\theta)= f(x_1,\theta)\cdot f(x_2,\theta)\cdot ...\cdot f(x_n,\theta)$, яку називають функцією правдоподібності. Вона зображує сумісний розподіл випадкового вектора з незалежними компонентами, кожна з яких має той самий розподіл, що й випадкова величина X .

За оцінку невідомого параметра θ приймається таке його значення θ^*, для якого функція $L(x_1,...,x_n,\theta)$, що розглядається як функція від θ при фіксованих значеннях $x_1,x_2,...,x_n$, досягає максимуму.

Дослідження функції $L(x_1,...,x_n,\theta)$ на максимум проводиться за допомогою методів диференціального числення: знаходимо критичні точки $\theta^* = (\theta_1^*,\theta_2^*,...,\theta_m^*)$ із системи рівнянь $\frac{\partial L(x_1,...,x_n,\theta_1,...,\theta_m)}{\partial \theta_i} = 0, i=\overline{1,m}$, а далі досліджуємо критичну точку на екстремум за допомогою достатніх ознак екстремуму функції багатьох змінних.

Для спрощення обчислень зручно замість функції $L(x_1,...,x_n,\theta)$ розглядати так звану логарифмічну функцію правдоподібності $\ln L(x_1,...,x_n,\theta)$, оскільки точки екстремуму функцій L і $\ln L$ збігаються.

Методи диференціального числення дозволяють знайти критичні точки функції $\ln L$, тобто точки можливого екстремуму, а потім з’ясувати, в якій з них досягається максимум.

Для цього розглядаємо спочатку систему рівнянь:

$$\begin{cases} \dfrac{\partial \ln L(x_1,...,x_n,\theta_1,...,\theta_m)}{\partial \theta_i} = 0, \\ i = 1,2,...,m, \end{cases}$$

розв'язки якої $(\theta_1^*, \theta_2^*, ..., \theta_m^*)$ − точки можливого екстремуму. Після цього, використовуючи достатні умови існування екстремуму функції, знаходимо точку максимуму.

Задачі для самостійного розв'язування

1. Результати вибіркового спостереження за часом обробки однієї деталі робітниками наведені у таблиці, в першому рядку якої розташовані частинні інтервали часу (у хв), а в другому – число робітників, час роботи яких потрапив у відповідний інтервал.

z_{i-1}-z_i	4,0-4,4	4,4-4,8	4,8-5,2	5,2-5,6	5,6-6,0	6,0-6,4	6,4-6,8
n_i	3	8	21	31	19	14	4

Визначити точкові незміщені статистичні оцінки для M(X), D(X) та оцінку для σ(X), де випадкова величина X – інтервали часу.

2. За період між черговими переналадками обладнання здійснено контрольні виміри товщини (в міліметрах) 200 вкладишів шатунних підшипників. Результати вимірювання наведені у вигляді дискретного статистичного розподілу:

x_i	1,72	1,73	1,74	1,75	1,76	1,77	1,78	1,79
n_i	1	22	40	79	27	26	4	1

Визначити точкові незміщені статистичні оцінки для M(X), D(X) та оцінку для σ(X).

3. У п’ятьох різних аптеках ціна одного і того ж медичного препарату дорівнює (у гривнях): 86, 88, 95, 98, 102. Знайти: а) вибіркову середню ціни цього товару; б) вибіркову та виправлену дисперсії ціни; в) коефіцієнт варіації.

4. У результаті валютних торгів протягом п’яти днів було встановлено таку ціну за американський долар (у гривнях): 26,7; 26,75, 26,77, 27. Знайти: а) вибіркову середню ціну американського долара; б) вибіркову та виправлену дисперсії курсу долара; в) коефіцієнт варіації курсу долара.

5. За результати соціологічного опитування 50 фермерів визначено їх середній прибуток на місяць у тис. гривнях:

Середній прибуток	0-40	40-80	80-120	120-160	160-200	200-240
Кількість фермерів	2	3	20	15	8	2

Визначити точкові незміщені статистичні оцінки для M(X), D(X) та оцінку для σ(X), де випадкова величина X – середній прибуток фермера на місяць.

6. На основі вибірки (x_1, x_2, …, x_n) знайти методом моментів точкові оцінки параметрів a та b рівномірного розподілу випадкової величини X.

7. Знайти методом максимальної правдоподібності оцінку параметрів нормального розподілу.

8. На основі вибірки (x_1, x_2, …, x_n) знайти за методом максимальної правдоподібності точкові оцінки параметрів a та b рівномірного розподілу випадкової величини X.

2.4. Інтервальні оцінки параметрів розподілу

2.4.1. Задача про інтервальне оцінювання параметрів розподілу

Розглянемо задачу про надійність точкової оцінки θ^* параметра θ, тобто про можливе відхилення точкової оцінки θ^* від істинного значення параметра θ або про оцінку абсолютної величини різниці $|\theta - \theta^*|$.

Зрозуміло, що точкова оцінка θ^* параметра θ є тим точнішою, чим менша величина різниці $|\theta - \theta^*|$. Якщо би вдалося встановити, що $|\theta - \theta^*| < \delta$, то число $\delta > 0$ характеризувало б точність точкової оцінки θ^* для параметра θ. Однак статистичні методи не дозволяють категорично стверджувати, що $|\theta - \theta^*| < \delta$, оскільки θ^* є випадкова величина.

Можна лише говорити про ймовірність γ, з якою ця нерівність виконується.

Означення 2.4.1. Надійністю (довірчою ймовірністю) точкової оцінки θ^* параметра розподілу θ називають імовірність γ, з якою виконується нерівність $|\theta - \theta^*| < \delta$, тобто $P(|\theta - \theta^*| < \delta) = \gamma$.

На практиці надійність оцінки задається наперед, причому число γ вибирають близьким до одиниці: $\gamma = 0{,}95$; $\gamma = 0{,}99$; $\gamma = 0{,}999$.

Напишемо рівність:

$P(-\delta < \theta - \theta^* < \delta) = \gamma$, або

$$P(\theta^* - \delta < \theta < \delta + \theta^*) = \gamma. \qquad (1.3)$$

Означення. Інтервал $(\theta^* - \delta, \theta^* + \delta)$, для якого виконується рівність (1.3), називається ***довірчим (надійним) інтервалом***, а його межі $\theta^* - \delta$ і $\theta^* + \delta$ –

довірчими (надійними) межами для параметра розподілу θ.

Інакше кажучи, довірчий інтервал для параметра розподілу θ є інтервал $(\theta^* - \delta, \theta^* + \delta)$, який із ймовірністю γ “накриває» точне значення цього параметра.

Зрозуміло, що завжди бажано, щоб для заданої близької до одиниці ймовірності γ довжина довірчого інтервалу була якомога меншою. Однак практично завжди є така альтернатива: збільшення надійності γ приводить до збільшення довжини довірчого інтервалу, і навпаки.

Загальний спосіб, за допомогою якого знаходять довірчий інтервал, полягає в тому, що розв’язують рівняння (1.3) і визначають з нього число δ. А для цього потрібно обчислити ймовірність $P(\theta^* - \delta < \theta < \delta + \theta^*)$. Останнє обчислення можна зробити, коли відомий закон розподілу точкової оцінки (статистики) $\theta^*(x_1, x_2, \dots, x_n)$ або пов’язаною з нею іншої випадкової величини, бо при цьому можна використати відомі формули з теорії ймовірності:

$P(\alpha \le \theta^* < \beta) = F(\beta) - F(\alpha)$,

або $P(\alpha \le \theta^* < \beta) = \int_\alpha^\beta f(x)dx$,

де $F(x)$ – функція розподілу і $f(x)$ – густина розподілу випадкової величини θ^*.

2.4.2. Розподіл χ^2 – «хі-квадрат»

Нехай $X_1, X_2, \dots, X_n$ — незалежні і нормально розподілені випадкові величини, причому їх математичні сподівання $M(X_i) = 0$ і середньоквадратичні відхилення $\sigma(X_i) = 1$ для будь-якого $i = 1, 2, \dots, n$.. Випадкова величина

$$\chi^2 = \sum_{i=1}^{n} X_i^2$$

має розподіл χ^2 із n ступенями вільності, який характеризується густиною

$$R(x,n) = \begin{cases} 0, & x < 0; \\ A_n x^{\frac{n}{2}-1} e^{-x/2}, & x \geq 0, \end{cases}, \qquad (3.30)$$

де A_n — стала, яка визначається з умови нормування:

$$A_n \cdot \int_0^{\infty} x^{\frac{n}{2}-1} e^{-x/2}\, dx = 1.$$

Описаний закон розподілу, що характеризується густиною (3.30), у теорії ймовірностей і статистиці називають *законом «хі-квадрат».* Розподіл «хі-квадрат» залежить від одного параметра *n* і при $n \to \infty$ він наближається до нормального закону розподілу.

Таблиця щодо розподілу «хі-квадрат», яка буде використана нами далі, міститься в *додатку 5.*

2.4.3. Розподіл Ст'юдента

Нехай Z — нормально розподілена випадкова величина, причому $M(Z) = 0, \sigma(Z) = 1$, а V — незалежна від Z випадкова величина, яка розподілена за законом «хі-квадрат» із *n* ступенями вільності. Тоді випадкова величина

$$T = \frac{Z}{\sqrt{V/n}}$$

має розподіл Ст'юдента з $k = n$ ступенями вільності, який характеризується густиною:

$$S(x,n) = B_n \left(1 + \frac{x^2}{n}\right)^{-\frac{n+1}{2}}, \quad x\epsilon(-\infty, +\infty), \qquad (3.31)$$

де B_n — деяка нормуюча константа.

Розподіл Ст'юдента також залежить від одного параметра п і при п оо наближається до стандартного нормального розподілу. Таблиця щодо розподілу Стьюдента, яка буде використана нами далі, міститься в додатку 6.

2.4.4. Розподіл Фішера — Снедекора

Нехай *U* і *V* — незалежні випадкові величини, які мають χ^2 — розподіли зі ступенями вільності k_1 і k_2, відповідно. Випадкова величина $F = \dfrac{U/k_1}{V/k_2}$ залежить від двох параметрів — ступенів вільності k_1 і k_2 і задається густиною ймовірностей

$$f_F(x, k_1, k_2) = C_{k_1 k_2} \cdot x^{\frac{k_1}{2}-1}\left(1+\frac{k_1}{k_2}x\right)^{-\frac{k_1+k_2}{2}}, \quad x \geq 0, \tag{3.32}$$

де коефіцієнт $C_{k_1 k_2}$ визначається з умови нормування. Цей розподіл отримав назву *F-розподілу,* або *розподілу Фішера — Снедекора.* Зокрема, *F*-розподілу підпорядковується відношення дисперсій двох залежних вибірок обсягів *n* і *m* із двох нормально розподілених генеральних сукупностей із рівними дисперсіями. У цьому випадку $k_1 = n - 1$ і $k_2 = m - 1$.

Задачі для самостійного розв'язування

1. У 30 осіб було виміряно їх зріст x_i (см). Отримані результати наведено у вигляді інтервального статистичного розподілу:

x_{i-1}-x_i	150,5-155,5	155,5-160,5	160,5-165,5	165,5-170,5
n_i	4	8	11	7

Із надійністю $\gamma = 0{,}99$ у припущенні, що випадкова величина X – зріст особи – нормально розподілена, знайти довірчий інтервал для математичного сподівання, якщо $\sigma = 2$.

2. У лікарні випадковим чином було відібрано 20 хворих та виміряно їх кров'яний тиск x_i (в умовних одиницях). Результати вимірювання наведено у вигляді дискретного статистичного розподілу:

x_i	1,6	1,9	2,2	2,4	2,7	3,1
n_i	2	3	4	6	4	1

Із надійністю $\gamma = 0{,}999$ побудувати довірчий інтервал для математичного сподівання нормально розподіленої випадкової величини X – кров'яного тиску хворого, якщо $\sigma = 1$.

3. Відомо, що зріст 12-річної дитини є випадковою величиною X із нормальним розподілом імовірностей, причому середньоквадратичне відхилення $\sigma = 5$ см. Скільки треба виконати спостережень, щоб знайти інтервал завширшки 2 см, який з імовірністю 0,95 «накриває» невідоме математичне сподівання досліджуваної випадкової величини?

4. Для банку, що включає 1000 філій, складено випадкову вибірку з 20 філій. За цією вибіркою виявилося, що у філії банку в середньому працює 65 співробітника за середнього квадратичного відхилення $\sigma = 16$ співробітників. Користуючись 95% довірчим інтервалом, оцінити середню кількість працівників у філій банку та загальну кількість співробітників у всьому банку. Припускається, що кількість співробітників банку має нормальний розподіл.

5. Верстат штампує деталі. За вибіркою обсягом n=120 обчислено вибіркову середню радіусів виготовлених деталей. Знайти з надійністю 0,95 точність δ, з якою вибіркова середня оцінює математичне сподівання радіусів виготовлених деталей, якщо відомо, що $\sigma = 2$ мм.

6. За вибіркою обсягом n=60 обчислено вибіркове середнє прибутку підприємців деякого регіону. Знайти з надійністю 0,99 точність δ, з якою вибіркова середня оцінює математичне сподівання прибутку підприємця, якщо відомо, що $\sigma = 2{,}6$.

7. З генеральної сукупності одержано вибірку:

x_i	-0,4	-0,3	-0,1	0	0,1	0,5	0,7	1	1,1	1,4
n_i	1	2	1	1	1	1	1	1	2	1

Оцінити з надійністю 0,95 математичне сподівання нормально розподіленої ознаки генеральної сукупності за вибірковою середньою з допомогою довірчого інтервалу.

8. З партії однотипних деталей навмання було вибрано 30 штук, в кожній з яких була виміряна глибина канавки. Результати спостережень наведено у вигляді інтервального статистичного ряду:

x_{i-1} - x_i]	2,3-2,5	2,5-2,7	2,7-2,9	2,9-3,1	3,1-3,3
n_i	5	7	10	6	2

В припущенні, що випадкова величина X – нормально розподілена, знайти з надійністю 0,99 інтервал довіри для математичного сподівання X.

9. У ході перевірки двох продуктових маркетів визначили, що в одному маркеті для випадкової вибірки n=10 рахунків середнє сальдо рахунку дорівнює 1200 гривень, а у другому, за такого ж обсягу вибірки – 950 гривень. Використовуючи 95% довірчі межі, оцінити різницю середніх сальдо рахунків для двох продуктових маркетів, якщо середньоквадратичне відхилення сальдо для першого маркету $\sigma_1 = 150$ грн., а для другого - $\sigma_2 = 100$ грн. Припускається нормальний розподіл сальдо рахунку.

10. У ході перевірки квіткових магазинів виявилось, що виправлене середньоквадратичне відхилення добового виторгу цих магазинів дорівнює 1300 гривень. Знайти з надійністю 95% довірчий інтервал, що покриває генеральне середнє квадратичне відхилення, якщо виторг магазинів підпорядкований нормальному закону розподілу.

11. Автомат заповнює банки кавою. Взявши 20 банок, контролер визначив, що різниця x_i між масою банки і стандартною масою має частоту n_i, які задані таблицею:

x_i	-4	-2	-1	0	1	5
n_i	2	1	6	7	3	1

Знайти з надійністю 95% довірчий інтервал, що покриває середнє квадратичне відхилення маси банки з кавою, якщо маса банки задовольняє нормальному закону розподілу.

2.5. Означення статистичної гіпотези і задача про її статистичну перевірку

Дані вибіркових спостережень часто становлять основу для прийняття одного з кількох альтернативних рішень (продукція може бути бракованою або якісною, технологічний процес порушується або ні, точність обробки виробу в межах норми, нижча від норми або вища від неї і т. д.). Із загальнометодологічного погляду йдеться про висунення деякої гіпотези, яку відхиляють або приймають після проведення деякого експерименту. Якщо цей експеримент має статистичний (стохастичний) характер, кажуть, що гіпотеза є *статистичною.*

Означення. ***Статистичною*** *називають гіпотезу про властивості {ознаки) генеральної сукупності, що перевіряється на основі вибірки.*

Статистичними гіпотезами можуть бути, наприклад, такі твердження: розподіл імовірностей випадкової величини є нормальний; розподіл імовірностей випадкової величини є пуассонівський; у нормальному розподілі випадкової величини параметри $a = 20$ і $\sigma = 1{,}5$; у показниковому розподілі випадкової величини параметр $\lambda = 5$; випадкові величини *X* і *У* незалежні і т. п.

У математичній статистиці виділяють два основні *типи статистичних гіпотез:*

гіпотези про закон розподілу ймовірностей випадкової величини (ознаки генеральної сукупності);

гіпотези про значення параметрів розподілу випадкової величини (ознаки генеральної сукупності).

Статистичні гіпотези першого типу називають ***непараметричними,*** *а другого типу —* ***параметричними.***

Означення. ***Основною (нульовою)*** *називають висунуту гіпотезу і позначають* H_0.

Альтернативною (конкуруючою) *називають гіпотезу, яка повністю або частково логічно заперечує нульову гіпотезу, і позначають* H_1.

Наприклад, математичне сподівання нормально розподіленої випадкової величини $a = 10$ — основна гіпотеза; математичне сподівання нормально розподіленої випадкової величини $a \neq 10$ — альтернативна гіпотеза. Записують це так: $H_0\colon\ a = 10$; $H_1\colon\ a \neq 10$.

Означення. *Параметричну гіпотезу називають* ***простою,*** *якщо вона стверджує, що всі невідомі параметри мають деякі числові значення, і*

складеною, *якщо вона складається зі скінченного або нескінченного числа простих гіпотез.*

Наприклад, якщо λ — параметр показникового розподілу, то гіпотеза $H_0\colon\ \lambda = 5$ є проста, а гіпотеза $H_0\colon\ \lambda > 5$ є складена.

Задача про статистичну перевірку статистичних гіпотез формулюється так. Розглядають деяку гіпотезу про те, що розподіл імовірностей випадкової величини має той чи інший вигляд, або параметри розподілу мають ті чи інші значення.

Задача полягає в тому, щоб на основі вивчення статистичних даних (вибірки) підтвердити справедливість висунутої гіпотези чи спростувати її. При цьому вказується також імовірність того, що прийняте рішення є правильним або помилковим.

Проблема зменшення ймовірності того, що прийняте рішення є помилковим, є також однією із задач математичної статистики.

У результаті статистичної перевірки гіпотези може бути прийняте одне з двох рішень:

• гіпотеза приймається;

• гіпотеза відхиляється.

Поряд із тим у результаті статистичної перевірки статистичної гіпотези можуть бути допущені помилки (прийняті неправильні рішення) двох типів:

• гіпотеза відхиляється, але вона істинна (помилка першого роду);

• гіпотеза приймається, але вона неістинна (помилка другого роду).

Виявляється, що помилка першого роду має вагоміші наслідки, ніж помилка другого роду.

Виникає питання: як застрахувати себе від помилки першого роду або принаймні звести до мінімуму ризик її допущення? Для цього вводиться спеціальне число α, яке виражає ймовірність відкинути правильну гіпотезу.

Означення. *Імовірність допустити помилку першого роду називають* ***рівнем значущості*** *і позначають через* α**.**

Число α задають наперед і найчастіше його вибирають рівним 0,1; 0,05; 0,01. Якщо $\alpha = 0{,}05$, то це означає, що ймовірність допустити помилку першого роду є мала, а саме — ми ризикуємо її допустити у 5 випадках зі 100.

Означення. *Інформацію про випадкову величину, яка міститься в гіпотезі, називають* ***гіпотетичною,*** *або* ***теоретичною,*** *а інформацію про неї, яку отримують на основі вибірки, називають* ***статистичною,*** *або* ***емпіричною.***

2.6. Перевірка гіпотези про порівняння середнього значення (математичного сподівання) ознаки генеральної сукупності зі стандартом

У критеріях для перевірки гіпотези про значення математичного сподівання H_0: $a = a_0$,, тобто гіпотези, що математичне сподівання $a = M(X)$ досліджуваної ознаки X генеральної сукупності збігається зі стандартом a_0, використовуємо статистику $\bar{x}$ — *середнє вибіркове.* Залежно від інформації щодо генеральної сукупності, якою володіємо, розрізняємо такі моделі.

Модель А. *Гіпотеза про значення математичного сподівання нормального закону розподілу за відомої дисперсії.*

Нехай випадкова величина X нормально розподілена з невідомим математичним сподіванням $a = M(X)$, але відомою дисперсією $\sigma^2 = D(X)$.

Потрібно на основі вибірки перевірити нульову гіпотезу H_0: $a = a_0$, про рівність математичного сподівання a певному числу a_0. При цьому припускаємо, що відомі такі величини:

а) дані вибірки обсягу *n;*

б) середнє квадратичне відхилення $\sigma = \sigma(X)$**;**

в) гіпотетичне значення математичного сподівання a_0*;*

г) рівень значущості $\alpha(0 < \alpha < 1)$.

Із вивченого нами матеріалу випливає, що вибіркове середнє $\bar{x}$ *у* вибірці з нормального розподілу з параметрами (a, σ^2) має нормальний розподіл із параметрами $(a, \frac{\sigma^2}{n})$, тому в розглядуваній задачі нормоване вибіркове середнє

$$Z = \frac{(\bar{x} - a_0)\sqrt{n}}{\sigma}$$

за умови істинності гіпотези H_0 також матиме нормальний розподіл із параметрами (0, 1).

Використовуючи цей факт, можна побудувати критерій для перевірки нульової гіпотези.

Правило 1. *Якщо нульова гіпотеза* H_0: $a = a_0$, *а конкуруюча гіпотеза* H_1: $a \neq a_0$, *то перевірку гіпотези* H_0 *проводимо за такою* ***схемою:***

1) *обчислюємо емпіричне значення критерію за формулою:*

$$Z_{\text{емп}} = \frac{(\bar{x} - a_0)\sqrt{n}}{\sigma}; \qquad (3.52)$$

2) *знаходимо за таблицею значень функції Лапласа критичне значення* $z_{\text{кр}}$, *використовуючи рівняння:*

$$\Phi(z_{\text{кр}}) = \frac{1-\alpha}{2}; \qquad (3.53)$$

3) *робимо висновок про висунуту гіпотезу:*

• *якщо* $|Z_{емп}| < z_{кр}$, *то гіпотезу* H_0 *приймаємо;*

• *якщо* $|Z_{емп}| \geq z_{кр}$, *то відхиляємо гіпотезу* H_0 *на користь альтернативи* H_1.

Зауваження. Якщо нульова гіпотеза H_0: $a = a_0$, а конкуруюча гіпотеза H_1: $a > a_0$ або H_1: $a < a_0$, то перевірку цих гіпотез також проводимо за схемою *правила 1* із такими змінами:

4) *замість рівняння (3.53) для знаходження критичного значення* $z_{кр}$ *використовуємо рівняння:*

$$\Phi\left(z_{кр}\right) = \frac{1-2\alpha}{2}; \qquad (3.54)$$

5) *робимо висновки стосовно висунутої гіпотези* H_0*:*

• *якщо* $Z_{емп} < z_{кр}$, *то немає підстав відхилити гіпотезу* H_0*; якщо* $Z_{емп} \geq z_{кр}$, *то гіпотезу* H_0 *відхиляємо на користь альтернативної гіпотези* H_1: $a > a_0$*;*

• *якщо* $Z_{емп} > -z_{кр}$ *то немає підстав відхилити гіпотезу* H_0*; якщо* $Z_{емп} < -z_{кр}$, *то гіпотезу* H_0 *відхиляємо і приймаємо гіпотезу* H_1: $a < a_0$.

Модель Б. *Гіпотеза про значення математичного сподівання нормального закону розподілу при невідомій дисперсії.*

Нехай випадкова величина *X* нормально розподілена з невідомими математичним сподіванням $a = M(X)$ і дисперсією $D(X) = \sigma^2$.

Потрібно на основі вибірки перевірити нульову гіпотезу H_0: $a = a_0$ (про рівність математичного сподівання a певному числу a_0).

Припускаємо, що відомі лише такі величини:

а) дані вибірки обсягу *n;*

б) гіпотетичне значення математичного сподівання a_0;

в) рівень значущості $\alpha(0 < \alpha < 1)$.

Оскільки середнє квадратичне відхилення $\sigma = \sigma(X)$ невідоме, то для перевірки гіпотези H_0 тут ми вже не зможемо скористатися статистикою Z через те, що для неї неможливо буде обчислити спостережуване значення $Z_{емп}$. У даному випадку використовуємо статистику

$$T = \frac{(\bar{x} - a_0)\sqrt{n}}{\tilde{\sigma}},$$

де $\bar{x}$ — вибіркове середнє значення, а $\tilde{\sigma}$ — виправлене вибіркове середнє квадратичне відхилення. Можна показати, що за умови правдивості гіпотези H_0 випадкова величина T має розподіл Стьюдента з числом $k = n - 1$ ступенів вільності.

Подальша побудова критичної області для дво- та односторонніх перевірок гіпотези здійснюється аналогічно, як викладено вище, з тією лише різницею, що критичні точки (тут замість $z_{кр}$ вони будуть позначатися через $t_{кр}$) визначаються за таблицею розподілу Стьюдента, а не Лапласа. Розподіл Стьюдента симетричний, як і нормальний, проте має менший (від'ємний) ексцес, тому у «хвостах» розподілу охоплено більшу площу. Отже, за того самого рівня значущості α значення $t_{кр}$ буде більшим, аніж $z_{кр}$, тобто довірчий інтервал ширший, аніж побудований на базі нормального розподілу.

Правило 2. *Якщо нульова гіпотеза* H_0: $a = a_0$, *а конкуруюча гіпотеза* H_1: $a \neq a_0$, *то перевірку гіпотези* H_0 *проводимо за схемою:*

- *обчислюємо емпіричне значення критерію*

$$T_{емп} = \frac{(\bar{x} - a_0)\sqrt{n}}{\tilde{\sigma}}, \quad (3.55)$$

де $\bar{x}$ *— середнє вибіркове значення, а* $\tilde{\sigma}$ *— виправлене вибіркове середнє квадратичне відхилення, обчислене для конкретної вибірки;*

- *із таблиці критичних точок розподілу Стьюдента (див. додаток 6) за заданим рівнем значущості α, зосередженим у верхньому рядку таблиці, і числом ступенів вільності* $k = n - 1$ *знаходимо критичну точку* $t_{\text{кр}} = t_{\text{двостор. кр}}(\propto, k)$*;*
- *робимо висновок про гіпотезу: якщо* $|T_{\text{емп}}| < t_{\text{кр}}$ *, то приймаємо гіпотезу* H_0*; якщо* $|T_{\text{емп}}| \geq t_{\text{кр}}$ *, то гіпотезу* H_0*, відхиляємо на користь альтернативи* H_1*.*

Зауваження. *Якщо нульова гіпотеза* $H_0\colon\ a = a_0$**,** *, а конкуруюча гіпотеза* $H_1\colon\ a > a_0$ *або* $H_1\colon\ a < a_0$*,, то сформульовані гіпотези також перевіряємо* ***за*** *схемою правила 2 із такими змінами:*

- *із таблиці критичних точок розподілу Стьюдента (додаток 6) за заданим рівнем значущості α, який розміщений у нижньому рядку таблиці, і числом ступенів вільності* $k = n - 1$ *знаходимо критичну точку* $t_{\text{кр}} = t_{\text{правостор. кр}}(\propto, k)$*;*
- *робимо висновки стосовно висунутої гіпотези* H_0*:*

1) якщо $T_{\text{емп}} < t_{\text{правостор. кр}}(\propto, k)$*, то немає підстав відхилити гіпотезу* H_0*: якщо* $T_{\text{емп}} \geq t_{\text{правостор. кр}}(\propto, k)$*, то гіпотезу* H_0 *відхиляємо на користь гіпотези* $H_1\colon\ a > a_0$*;*

2) якщо $T_{\text{емп}} > -t_{\text{правостор. кр}}(\propto, k)$*, то приймаємо гіпотезу* H_0*; якщо* $T_{\text{емп}} < -t_{\text{правостор. кр}}(\propto, k)$*,, то гіпотезу* H_0 *відхиляємо і приймаємо гіпотезу* $H_1\colon\ a < a_0$*.*

Модель В. *Гіпотеза про значення математичного сподівання будь-якого закону* *розподілу за великого обсягу вибірки.* Нехай випадкова величина має будь-який закон розподілу з невідомим математичним сподіванням *M (X) = a* і скінченною, але невідомою дисперсією $D(X) = \sigma^2$.

Як і в попередніх моделях, потрібно на підставі даних випадкової вибірки перевірити нульову гіпотезу H_0: $a = a_0$ (про рівність математичного сподівання *a* певному числу a_0).

Припускаємо, що відомими є такі величини:

а) дані вибірки обсягу *n* (*n*≥100);

б) гіпотетичне значення математичного сподівання a_0;

в) рівень значущості $\alpha(0 < \alpha < 1)$.

Перевірка гіпотези H_0 у цій моделі здійснюється аналогічно, як у моделі **А**, за *правилом 1* із тією лише різницею, що для обчислення спостережуваного значення критерію за формулою (3.52) невідоме середнє квадратичне відхилення σ треба замінити на вибіркове середнє квадратичне відхилення $\bar{\sigma} = \sqrt{\bar{D}}$, знайдене за даними вибірки.

2.7. Перевірка гіпотези про рівність математичних сподівань двох нормально розподілених випадкових величин

Нехай задані дві статистичні сукупності, що характеризуються незалежними нормально розподіленими випадковими величинами *X* і Y із параметрами, відповідно, (a_x, σ_x^2) і (a_y, σ_y^2), де $a_x = M(X)$, $\sigma_x^2 = D(X)$, $a_y = M(Y)$, $\sigma_y^2 = D(Y)$. Якщо математичні сподівання a_x і a_y невідомі, то висувається гіпотеза про їх рівність, тобто H_0: $a_x = a_y$. Для перевірки гіпотези H_0 за наявності відповідної альтернативної гіпотези H_1, із кожної сукупності проводиться вибірка: з першої — обсягу *n,* унаслідок якої отримуємо вибіркове середнє $\bar{x}$, з другої — обсягу *m,* з якої отримуємо вибіркове середнє $\bar{y}$. Усі критерії перевірки гіпотези H_0 ґрунтуються на порівнянні статистик $\bar{x}$ і $\bar{y}$. Тут ми знову розглядаємо кілька моделей.

Модель А. *Дисперсії ознак X і У відомі.*

Нехай випадкові величини *X* і *У* незалежні, нормально розподілені, математичні сподівання яких a_x і a_y невідомі, а дисперсії σ_x^2 і σ_y^2 відомі. Перевіряємо гіпотезу H_0: $a_x = a_y$ (про рівність математичних сподівань випадкових величин *X* і У).

Вважаємо, що відомими є такі величини:

а) дані двох незалежних вибірок обсягів *n* і *m* значень випадкових величин *X* і *У,* відповідно;

б) дисперсії σ_x^2 і σ_y^2;

в) рівень значущості $\alpha (0 < \alpha < 1)$.

Позначимо через $(x_1, \dots, x_n)$, $(y_1, \dots, y_m)$ — вибірки значень випадкових величин *X* і У, відповідно. У припущеннях даної моделі випадкова величина

$$\bar{x} - \bar{y} = \frac{1}{n}\sum_{i=1}^{n} x_i - \frac{1}{m}\sum_{i=1}^{m} y_i$$

розподілена за нормальним законом із параметрами $\left(a_x - a_y, \ \frac{\sigma_x^2}{n} + \frac{\sigma_y^2}{m}\right)$, нормована різниця

$$Z = \frac{(\bar{x} - \bar{y}) - (a_x - a_y)}{\sqrt{\frac{\sigma_x^2}{n} + \frac{\sigma_y^2}{m}}}$$

має нормальний розподіл із параметрами (0, 1).

У припущенні, що гіпотеза H_0 — істинна, статистика Z набуває виразу:

$$Z = \frac{\bar{x} - \bar{y}}{\sqrt{\frac{\sigma_x^2}{n} + \frac{\sigma_y^2}{m}}}$$

Правило 1. Якщо нульова гіпотеза H_0: $a_x = a_y$, а конкуруюча гіпотеза H_1: $a_x \neq a_y$, то перевірка сформульованих гіпотез проводиться за схемою:

- обчислюємо емпіричне значення критерію за формулою:

$$Z_{емп} = \frac{\bar{x}-\bar{y}}{\sqrt{\frac{D(X)}{n}+\frac{D(Y)}{m}}} \qquad (3.56)$$

де $\bar{x}$, $\bar{y}$ — середні вибіркові значення вибірок для X і У відповідно;

- за таблицею функції Лапласа (див. додаток 2) знаходимо критичну точку $z_{кр}$ із рівності:

$$\Phi(z_{кр}) = \frac{1-\alpha}{2}; \qquad (3.57)$$

- *робимо висновок про гіпотезу: якщо* $|Z_{емп}| < z_{кр}$, *то приймаємо нульову гіпотезу* H_0*; якщо* $|Z_{емп}| \geq z_{кр}$, *то гіпотезу* H_0 *відхиляємо на користь альтернативної гіпотези* H_1.

Зауваження. Якщо нульова гіпотеза H_0: $a_x = a_y$, а конкуруюча гіпотеза H_1: $a_x > a_y$ або H_1: $a_x < a_y$, то перевірка гіпотези проводиться також за схемою *правила 1* із такими змінами:

- *рівняння (3.57) для знаходження критичного значення* $z_{кр}$ *замінюється рівнянням:*

$$\Phi(z_{кр}) = \frac{1-2\alpha}{2}; \qquad (3.58)$$

- *робимо такі висновки стосовно нульової гіпотези* H_0*:*

1) якщо $Z_{емп} < z_{кр}$, *то немає підстав відхилити гіпотезу* H_0*; якщо* $Z_{емп} \geq z_{кр}$, *то гіпотезу* H_0

відхиляємо на користь альтернативної гіпотези H_1: $a_x > a_y$;

2) *якщо* $Z_{\text{емп}} > -z_{\text{кр}}$, *то приймаємо гіпотезу* H_0; *якщо* $Z_{\text{емп}} < -z_{\text{кр}}$, *то гіпотезу* H_0 *відхиляємо і приймаємо альтернативну гіпотезу* H_1: $a_x < a_y$.

Модель *Б.* *Дисперсії ознак Х і У невідомі (випадок малих вибірок).*

Нехай випадкові величини *Х* і *У* незалежні, нормально розподілені, для яких як математичні сподівання a_x і a_y, так і дисперсії σ_x^2 і σ_y^2 є невідомі. Перевіряємо гіпотезу H_0: $a_x = a_y$ (про рівність математичних сподівань випадкових величин *Х* і У).

Вважаємо, що відомими є величини:

a) дані двох незалежних вибірок обсягів *n* і *m* значень випадкових величин *Х* і *У,* відповідно, які є невеликими (*n* + *m* ≤ 122);
b) рівень значущості $\alpha (0 < \alpha < 1)$.

Тепер зробимо додаткове припущення, що дисперсії обох сукупностей хоч невідомі, проте рівні, тобто σ_x^2 = σ_y^2. Зауважимо, що дане припущення дуже часто вимагає спеціальної перевірки, про яку йтиметься в наступному пункті. Якщо прийняти це припущення, то для перевірки гіпотези H_0 використовується статистика

$$T_{\text{емп}} = \frac{\bar{x}-\bar{y}}{\sqrt{(n-1)\bar{D}_x+(m-1)\bar{D}_y}} \cdot \sqrt{\frac{n\cdot m(n+m-2)}{n+m}}, \qquad (3.59)$$

яка за умови виконання гіпотези H_0 має розподіл Стьюдента з числом $k = n + m - 2$ ступенів вільності.

Правило 2. *Якщо нульова гіпотеза* H_0: $a_x = a_y$, *а конкуруюча гіпотеза* H_1: $a_x \neq a_y$, *то перевірку гіпотези* H_0 *проводимо за схемою:*

- *знаходимо емпіричне значення критерію за формулою (3.59), де* $\bar{x}$ *,* $\bar{y}$ *— вибіркові середні значення, а* $\widetilde{D}_x$*,* $\widetilde{D}_y$ *— виправлені вибіркові дисперсії, які обчислені за конкретними вибірками для випадкових величин X і У;*
- *за таблицею критичних точок розподілу Стьюдента для заданого рівня значущості α, зосередженого у верхньому рядку таблиці додатка 6, і числа ступенів вільності* $k = n + m - 2$ *знаходимо критичну точку* $t_{кр} = t_{двостор.\ кр}(\alpha, k)$*;*
- *робимо висновок про гіпотезу: якщо* $|T_{емп}| < t_{кр}$*, то приймаємо нульову гіпотезу* H_0*; якщо* $|T_{емп}| \geq t_{кр}$ *, то гіпотезу* H_0 *відхиляємо на користь альтернативи* H_1*.*

Зауваження. Якщо нульова гіпотеза $H_0\colon\ a_x = a_y$, а конкуруюча гіпотеза $H_1\colon\ a_x > a_y$ або $H_1\colon\ a_x < a_y$, то перевірку гіпотези H_0 проводимо за схемою *правила 2* із такими змінами:

- *за таблицею критичних точок розподілу Стьюдента (додаток 6) для заданого рівня значущості α, який розміщений у нижньому рядку таблиці, і числа ступенів вільності* $k = n + m - 2$ *знаходимо критичну точку* $t_{кр} = t_{правостор.\ кр}(\alpha, k)$*;*
- *робимо висновки стосовно висунутої гіпотези* H_0*:*

1) якщо $T_{емп} < t_{кр}$*, то нульову гіпотезу* H_0 *приймаємо; якщо* $T_{емп} \geq t_{кр}$*, то гіпотезу* H_0 *відхиляємо і приймаємо альтернативну гіпотезу* $H_1\colon\ a_x > a_y$*;*

2) *якщо* $T_{емп} > -t_{кр}$*, то немає підстав відхилити гіпотезу* H_0*; якщо* $T_{емп} \leq -t_{кр}$*, то гіпотезу* H_0

відхиляємо на користь альтернативної гіпотези $H_1\colon\ a_x < a_y$.

Модель *B*. *Дисперсії ознак X і У невідомі (випадок великих вибірок).* Нехай випадкові величини *X* і *У* незалежні, нормально розподілені, для яких як математичні сподівання a_x і a_y, так і дисперсії σ_x^2 і σ_y^2 є невідомі. За даними вибірок великих обсягів *n і m (n, m* $\geq$ 100) та за даним рівнем значущості $\alpha(0 < \alpha < 1)$ перевіряємо гіпотезу $C\colon\ a_x = a_y$ (про рівність математичних сподівань випадкових величин *X* і *У*) за відповідної альтернативи H_1.

За умов даної моделі критерій перевірки гіпотези H_0 будуємо аналогічно, як і в моделі А, з тією лише різницею, що для обчислення спостережуваного значення статистики за формулою (3.56) замість невідомих дисперсій D_x і D_y приймаємо вибіркові дисперсії $\overline{D_x}$, $\overline{D_y}$, які обчислюємо за допомогою даних вибірок.

2.8. Перевірка гіпотези про рівність дисперсій двох незалежних випадкових величин

Задача про перевірку гіпотези про рівність двох дисперсій виникає досить часто; наприклад, під час аналізу стабільності виробничого процесу до і після введення нової технології (коливання у випуску продукції вимірюється за допомогою квадратичного відхилення), вивчення якості вимірювальних приладів (зіставлення дисперсій показників окремих приладів), вивчення ступеня однорідності двох сукупностей щодо деякої ознаки (кваліфікації робітників, стажу персоналу і т. д.). Потреба перевірити рівність дисперсій виникає, як ми переконалися раніше, і під час порівняння середніх величин сукупностей.

Отже, нехай випадкові величини *X* і У, що характеризують дві статистичні сукупності, незалежні, нормально розподілені з невідомими дисперсіями $D(X)=\sigma_x^2$ і $D(Y)=\sigma_y^2$ відповідно. Перевіряємо гіпотезу $H_0\colon \sigma_x^2=\sigma_y^2$і (про рівність дисперсій випадкових величин *X* і *У*).

Вважаємо, що відомими є такі величини:

а) дані двох незалежних вибірок обсягів *n* і *m* для випадкових величин *X* і У, відповідно;

б) рівень значущості $\alpha(0<\alpha<1)$.

Нехай $(x_1,\dots,x_n)$, $(y_1,\dots,y_n)$ — незалежні випадкові вибірки для випадкових величин *X* і У, відповідно. Критерій перевірки гіпотези H_0 базується на зіставленні виправлених вибіркових дисперсій $\tilde{D}_x$ і $\tilde{D}_y$ обчислених за даними вибірок. Так, у припущеннях розглядуваної моделі випадкова величина

$$F=\frac{\tilde{D}_x}{\tilde{D}_y},\ \tilde{D}_x\geq\ \tilde{D}_y, \qquad (3.60)$$

за умови виконання гіпотези H_0 розподілена за законом Фішера—Снедекора з $k_1=n-1$ і $k_2=m-1$ ступенями вільності.

Правило. *Якщо нульова гіпотеза* $H_0\colon \sigma_x^2=\sigma_y^2$, *а конкуруюча* $H_1\colon \sigma_x^2>\sigma_y^2$, *то перевірку гіпотези* H_0 *здійснюємо за схемою:*

- знаходимо спостережуване значення критерію за формулою (3.60), де $\tilde{D}_x$ і $\tilde{D}_y$ — виправлені вибіркові дисперсії, які обчислені за конкретними вибірками для випадкових величини *X* і У;
- за таблицею критичних точок розподілу Фішера —Снедекора для заданого рівня значущості α і ступенів вільності $k_1=n-1$ і $k_2=m-1$

знаходимо критичну точку правосторонньої критичної області $f_{кр} = f_{кр}(\alpha, k_1, k_2)$;

- робимо висновок щодо прийняття гіпотези H_0:

якщо $F_{емп} < f_{кр}$, то гіпотезу H_0 приймаємо;

якщо $F_{емп} \geq f_{кр}$, то гіпотезу H_0 відхиляємо на користь альтернативної гіпотези H_1.

У випадку, коли $D_x < D_y$, критерій узгодження $F = D_y/D_x$ і $k_1 = m - 1$, $k_2 = n - 1$.

Зауваження. Якщо нульова гіпотеза $H_0: \sigma_x^2 = \sigma_y^2$, а конкуруюча гіпотеза $H_1: \sigma_x^2 \neq \sigma_y^2$, то перевірку гіпотези H_0 здійснюємо за

сформульованим *правилом,* в якому змінюється лише методика знаходження критичного значення $f_{кр}$, а саме: із таблиці критичних точок розподілу Фішера — Снедекора критичну точку $f_{кр} = f_{кр}(\alpha/2, k_1, k_2)$ визначаємо за рівнем значущості $\alpha/2$ (удвічі меншому від заданого) і числами ступенів вільності $k_1 = n - 1$ і $k_2 = m - 1$.

2.9. Критерій статистичної перевірки гіпотези

Нехай H_0 – нульова гіпотеза про розподіл ймовірностей випадкової величини (ознаки генеральної сукупності) або про значення параметрів цього розподілу. Далі на основі статистичних даних (вибірки) потрібно вирішити: чи гіпотезу H_0 прийняти, чи гіпотезу H_0 відхилити на користь альтернативної гіпотези H_1.

Виникає питання: як дані вибірки пов'язати зі сформульованою гіпотезою і за якими критеріями на основі кожної з різних вибірок прийняти ту чи іншу гіпотезу?

Щоб дати відповідь на це питання, для гіпотези H_0 вводять певну чисельну характеристику, яка обчислюється на основі вибірки і на підставі якої вирішують: прийняти основну гіпотезу H_0 чи альтернативну гіпотезу H_1. Зрозуміло, що вибрана чисельна характеристика для різних вибірок матиме, загалом кажучи, різні значення, які наперед невідомі, і тому вона є випадковою величиною.

Означення. ***Статистичним критерієм*** *гіпотези (або просто* ***критерієм*** *гіпотези) називається випадкова величина К, за допомогою якої проводиться перевірка гіпотези.*

Випадкову величину *К* вибирають такою, щоб закон розподілу її ймовірностей був відомий.

Означення. *Значення випадкової величини К, яке обчислене на основі певної вибірки, називають* ***емпіричним (спостережуваним)*** *значенням критерію гіпотези і позначають* $K_{емп}$ $(K_{сп})$.

Виявляється, що за одних значень $K_{емп}$ гіпотеза H_0 приймається, а за інших його значень — відхиляється.

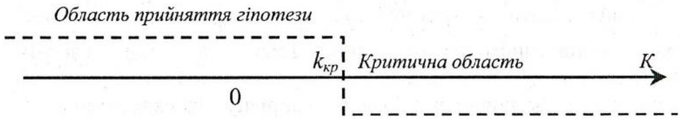
Означення. *Сукупність значень критерію К (випадкової величини К), за яких нульова гіпотеза* H_0 *відхиляється, називається* ***критичною областю,*** *а сукупність значень критерію К, за яких нульову гіпотезу* H_0 *приймають, називається* ***областю прийняття гіпотези (областю допустимих значень).***

Звідси маємо таке ***правило перевірки статистичних гіпотез:*** *якщо емпіричне (спостережуване) значення критерію* $K_{емп}$ *належить критичній області, то нульову гіпотезу відхиляють; якщо емпіричне значення критерію* $K_{емп}$ *належить* *області прийняття гіпотези, то гіпотезу H_0 приймають.*

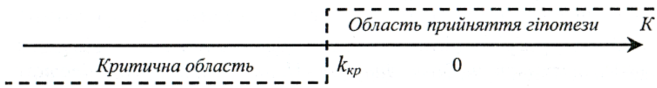
Які ж області можуть бути критичними і як їх шукати? Якщо випадкова величина K є одновимірна, то критична область, як правило, є множиною точок певних інтервалів на прямій, які відділені від області прийняття гіпотези так званими критичними точками $k_{кр}$. Тобто для знаходження критичної області достатньо визначити критичні точки.

Розрізняють три види критичних областей:

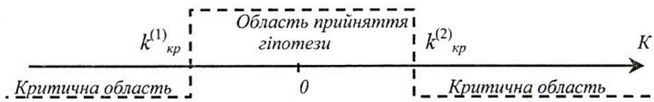
- *правостороння критична область* — це та область на числовій прямій, що визначається нерівністю $K > k_{кр}$:



- *лівостороння критична область* — це та область на числовій прямій, що визначається нерівністю $K < k_{кр}$:



- *двостороння критична область* — це та область на числовій прямій, що визначається сумою інтервалів: $K < k_{кр}^{(1)}$ і $K > k_{кр}^{(2)}$:



Для знаходження критичної області задаються рівнем значущості α і шукають критичні точки $k_{кр}$ із таких співвідношень:

а) для правосторонньої критичної області:

$$P(K < k_{кр}) = \alpha \quad (k_{кр} > 0); \tag{3.38}$$

б) для лівосторонньої критичної області:

$$P(K < k_{кр}) = \alpha \quad (k_{кр} < 0); \tag{3.39}$$

в) для двосторонньої симетричної критичної області:

$$P\left(K < k_{кр}^{(1)}\right) + P\left(K > k_{кр}^{(2)}\right) = \alpha. \tag{3.40}$$

Однак знайти дві критичні точки $k_{кр}^{(1)}$ і $k_{кр}^{(2)}$ з одного рівняння можна нескінченним числом способів. Тому виявляється можливим ввести додаткову умову. Найчастіше двосторонню критичну область будують як симетричну, розділяючи α порівну між «хвостами» розподілу, тобто визначаючи $k_{кр}^{(1)} = -k_{кр}$ і $k_{кр}^{(2)} = k_{кр}$ із рівняння:

$$P(K > k_{кр}) = \alpha/2, \quad k_{кр} > 0. \tag{3.40'}$$

Отже, перевірка статистичної гіпотези проводиться за такою схемою:

1. Формулюють нульову гіпотезу H_0, альтернативну гіпотезу H_1 і задають рівень значущості α для перевірки гіпотези H_0.

2. Визначають критерій *K* для перевірки гіпотези H_0, який є випадковою величиною з відомим розподілом її ймовірностей.

3. Визначають критичні області відносно даних критерію *K* та рівня значущості α. Для визначення критичної області достатньо знайти критичні точки $k_{кр}$ за відповідними рівностями (3.38)-(3.40).

4. Знаходять емпіричне (спостережуване) значення критерію $K_{емп}$ на основі конкретної вибірки.

5. Приймають рішення: якщо емпіричне значення критерію $K_{емп}$ потрапляє в критичну область, то нульову гіпотезу H_0 відхиляють; якщо ж значення $K_{емп}$ потрапляє в область допустимих значень, то нульову гіпотезу H_0 приймають.

Зрозуміло, що для певної гіпотези можна побудувати багато різних критеріїв її перевірки, і за кожним таким критерієм можемо одержувати різні результати щодо прийняття нульової гіпотези H_0 на основі тієї самої вибірки. Тому для означення кращого критерію вводиться характеристика, яка називається *потужністю* критерію.

Означення. ***Потужністю критерію*** *називають імовірність належності критерію критичній області за умови, що істинна конкуруюча гіпотеза.*

Тобто потужність критерію (наскільки вибраний критерій добрий) визначається як імовірність не допустити помилку другого роду при вибраному критерії.

2.10. Критерій узгодження Пірсона

Припустімо, що вивчається генеральна випадкова величина *X*, стосовно якої висунуто гіпотезу H_0 про те, що її закон розподілу описується гіпотетичною функцією розподілу $F(x)$ або густиною розподілу $f(x)$. Зрозуміло, що у випадку неперервної ознаки величини *X* висунута гіпотеза буде рівнозначна гіпотезі: H_0: {густина розподілу ймовірностей випадкової величини $X \in f(x)$}; а у випадку дискретної ознаки *X* гіпотезу H_0 можна сформулювати так: H_0: {закон розподілу випадкової величини *X* визначається функцією ймовірності $P(X = x_i) = p_i,\ i \in N\}$.

У даних гіпотезах F, f, P можуть бути повністю відомими функціями або визначеними лише з точністю до невідомих параметрів.

Критерії, які призначені для перевірки сформульованих гіпотез, називаються *критеріями узгодження.* Критерії узгодження дозволяють відповісти на питання про те, чи розбіжність між емпіричними і теоретичними розподілами є настільки незначною, що вона може бути приписана впливу випадковості, чи ні.

У даному пункті ми розглянемо один з таких критеріїв, який називається *критерієм Пірсона,* або *критерієм* χ^2 *(«хі-квадрат»).*

Нехай дані вибірки $x_1, \dots, x_n$, які отримані внаслідок *n* незалежних спостережень над випадковою величиною *X*, згруповані і подані у формі дискретного варіаційного ряду *(табл. 3.1* або *3.1')*, якщо ознака *X* — дискретна, або інтервального варіаційного ряду (*табл. 3.2* або *3.2'*), якщо ознака — неперервна. На підставі побудованого статистичного розподілу вибірки і за наперед заданим рівнем значущості а треба перевірити гіпотезу:

- H_0: {закон розподілу *X* описується функцією розподілу $F(x)$ або густиною розподілу $f(x)$}, якщо альтернативна гіпотеза:
- H_1: {закон розподілу *X* не описується функцією розподілу $F(x)$ чи густиною розподілу $f(x)$}.

Згідно з критерієм Пірсона для перевірки гіпотези H_0 вводиться випадкова величина (статистика) *К:*

$$K = \chi^2 = \sum_{i=1}^{m} \frac{(n_i - n_i')^2}{n_i'} = n \sum_{i=1}^{m} \frac{(w_i - p_i)^2}{p_i}, \qquad (3.41)$$

де m — число груп у статистичному розподілі вибірки;

$n_i(w_i)$ — спостережувана частота (відносна частота) ознаки X в i-тій групі;

$n'_i = n \cdot p_i$ — теоретична частота;

p_i — імовірність того, що значення X належить до i-Ї групи, і вона розрахована за допомогою гіпотетичної функції розподілу $F(x)$ або густини розподілу $f(x)$.

У більш повних курсах математичної статистики доводиться, що статистика K при $n \to \infty$ прямує за розподілом до випадкової величини, розподіленої за законом χ^2 числом $k = m - s - 1$ ступенів вільності, де m — число інтервалів статистичного розподілу вибірки; s — число параметрів, що входять до гіпотетичного розподілу F або f і які оцінюються на підставі спостережуваних даних. Так, наприклад, коли перевіряється узгодження досліджуваного розподілу з розподілом Пуассона, єдиний параметр якого невідомий і оцінюється за вибірковими даними, то $k = m - 2$ (коли параметр λ відомий і не оцінюється за вибірковими даними, то $k = m - 1$**);** якщо перевіряється узгодження з нормальним розподілом, для якого за вибірковими даними оцінюються два параметри α *і* σ, то $k = m - 3$ і т. д.

Якщо маємо повне узгодження теоретичного і статистичного розподілів, то $K = 0$, у протилежному випадку $K > 0$. Обчисливши за формулою (3.41) $K_{емп} = \chi^2_{емп}$ і визначивши за заданим рівнем значущості α та числом ступенів вільності k із *додатка 5* критичне значення $\chi^2_{кр} = k_{кр}$, маємо такі висновки:

а) гіпотезу H_0 відхиляємо, якщо $K_{емп} \geq k_{кр}$;

б) приймаємо гіпотезу H_0, якщо $K_{емп} < k_{кр}$. **Застосування критерію χ^2 вимагає дотримання таких умов:** 1) експериментальні дані мають бути незалежними, тобто вибірка повинна бути випадковою; 2) обсяг вибірки має бути достатньо великим (практично не меншим ніж 50 одиниць), а частота кожної групи — не меншою за 5. Якщо остання умова не виконується, то проводиться попереднє об'єднання нечисленних груп.

Опишемо детальніше схему перевірки за критерієм Пірсона статистичної гіпотези про гіпотетичний закон розподілу ймовірностей випадкової величини ***X*** для випадку, коли ця величина є неперервною.

Схема перевірки гіпотези про закон розподілу ймовірностей неперервної випадкової величини за критерієм Пірсона:

1. *Вихідні статистичні дані (результати вибірки)* $x_1, \dots, x_n$ *групують і записують як інтервальний варіаційний ряд:*

Таблиця 3.3

$(z_{i-1}, z_i]$	$(z_0, z_1]$	$(z_1, z_2]$	…	$(z_{m-1}, z_m]$
n_i	n_i	n_i	…	n_m
w_i	w_1	w_2	…	w_m

У першому рядку відзначені часткові інтервали згрупованого варіаційного ряду $(z_{i-1}, z_i]$, у другому — їх частоти n_i, у третьому — їх відносні частоти (статистичні ймовірності) w_i, тобто числа

$$w_i = \frac{n_i}{n}, \qquad i = \overline{1, m},$$

де *n* — обсяг вибірки, n_i — число спостережень, в яких наставала подія $X = x_i \epsilon (z_{i-1}, z_i]$, т — число груп (частинних інтервалів), на які розбито вибірку. При цьому $\sum_{i=1}^{m} n_i = n, \sum_{i=1}^{m} w_i = 1$.

2. Оскільки перевіряється гіпотеза про те, що розподіл генеральної ознаки Х описується певною (конкретною) функцією розподілу $F(x)$, або, що те ж саме, густиною розподілу $f(x)$, то для кожного інтервалу $(z_{i-1}, z_i]$ можна визначити теоретичні ймовірності p_i попадання значень випадкової величини Х у цей інтервал, а отже, і теоретичні частоти $n'_i = n \cdot p_i$.

Для обчислення ймовірностей p_i використовують формули:

$$p_i = P(z_{i-1} \le Z < z_i) = F(z_i) - F(z_{i-1}) \quad (3.42)$$

або

$$p_i = \int_{z_{i-1}}^{z_i} f(x)dx, \quad i = 1,2, \dots, m. \quad (3.42')$$

Зазначимо, що для обчислення ймовірностей p_1 і p_m у формулах (3.42), (3.42') покладають, відповідно, $z_0 = -\infty$, $z_m = +\infty$.

Тоді $\sum_{i=1}^{m} p_i = 1$

Отримані результати обчислень зручно записати у формі таблиці (табл. 3.4).

Таблиця 3.4

$(z_{i-1}, z_i]$	$(z_0, z_1]$	$(z_1, z_2]$	…	$(z_{m-1}, z_m]$
n_i	n_1	n_2	…	n_m
w_i	w_1	w_2	…	w_m
p_i	p_1	p_2	…	p_m
$n'_i = n \cdot p_i$	n'_1	n'_2		n'_m

Зауваження. Гіпотетичні функції розподілу $F(x)$ або густина розподілу $f(x)$, як правило, характеризуються деякими чисельними параметрами, точні значення яких можуть бути невідомими. Тоді для обчислення теоретичних імовірностей p_i ці невідомі параметри замінюються їх точковими оцінками, визначеними за допомогою даних вибірки.

3. *На підставі даних із табл. 3.4 обчислюють спостережуване значення критерію Пірсона:*

$$K_{емп} = \sum_{i=1}^{m} \frac{(n_i - n_i')^2}{n_i'}, \quad n_i' = np_i; \tag{3.43}$$

$$K_{емп} = n \sum_{i=1}^{m} \frac{(w_i - p_i)^2}{p_i}. \tag{3.43’}$$

4. *За даним рівнем значущості α і числом ступенів вільності $k = m - s - 1$, де s — число параметрів гіпотетичного розподілу, обчислених за даними вибірки, із таблиці критичних значень розподілу χ^2 знаходять критичну точку $k_{kp} = \chi_{kp}^2(\alpha, k)$.*

Зауваження. χ_{kp}^2 є розв'язком рівняння $P(\chi^2 > \chi_{kp}^2) = \int_{\chi_{kp}^2}^{\infty} R(x,m)\,dx = \alpha.$

5. *Зіставляємо значення $K_{емп}$ і $k_{кр}$: якщо $K_{емп} \geq k_{кр}$, то гіпотезу H_0 про вигляд густини розподілу відхиляють; якщо ж $K_{емп} < k_{кр}$, то гіпотезу H_0приймають.*

Задачі для самостійного розв’язування

1. Зібрано статистичні дані про число відвідувачів кінотеатра протягом вечірніх сеансів та записано результати у вигляді інтервальної таблиці частот:

$[x_{i-1};x_i)$	[4;9)	[9;14)	[14;19)	[19;24)	[24;29)	[29;34)	[34;38)
n_i	6	7	16	42	15	6	5

Користуючись критерієм Пірсона, при рівні значущості $\alpha = 0,05$ перевірити, чи узгоджується гіпотеза про нормальний розподіл випадкової величини X – числа відвідувачів кінотеатра протягом вечірніх сеансів – з емпіричним розподілом вибірки?

2. Зібрано дані про кількість x_i підприємств харчової промисловості у 200 районах України:

x_i	5	7	9	11	13	15	17	19	21
n_i	15	26	25	30	26	21	24	20	13

Користуючись критерієм Пірсона, при рівні значущості $\alpha = 0,05$ перевірити гіпотезу про нормальний розподіл ознаки X – кількості підприємств харчової промисловості.

3. На основі інтервального статистичного розподілу вибірки:

$[x_{i-1};x_i)$	[0;5)	[5;10)	[10;15)	[15;20)	[20;25)	[25;30)	[30;35)
n_i	144	52	41	25	14	12	10

з допомогою критерія Пірсона здійснити перевірку гіпотези H_0 про показниковий розподіл кількісної ознаки генеральної сукупності, якщо рівень значущості $\alpha = 0,01$.

4. У результаті випробування 417 ламп було отримано емпіричний розподіл тривалості їх роботи:

$[x_{i-1};x_i)$	n_i
[0;300)	166
[300;600)	93
[600;900)	63
[900;1200)	39
[1200;1500)	31
[1500;1800)	17
[1800;2100)	8

При рівні значущості $\alpha = 0{,}01$ за критерієм Пірсона перевірити гіпотезу про показниковий закон розподілу ознаки X – часу роботи лампочки.

5. За даними про контрольні виміри товщини (в міліметрах) 1000 вкладишів шатунних підшипників:

x_i	0	1	2	3	4	5	6
n_i	15	26	25	30	26	21	24

Перевірити гіпотезу H_0 про те, що випадкова величина X – товщина вкладишів підшипників – розподілена за законом Пуассона.

6. У результаті перевірки 400 ящиків з скляними виробами було встановлено, що кількість пошкоджених виробів X має статистичний розподіл:

x_i	0	1	2	3	4	5	6	7
n_i	159	139	67	21	9	3	1	1

при рівні значущості $\alpha = 0{,}05$ перевірити гіпотезу про те, що випадкова величина X – кількість пошкоджених виробів – розподілена за законом Пуассона.

7. Відділ технічного контролю перевірив 170 партій виробів по 10 виробів у кожній партії. У результаті отримано такий статистичний розподіл ознаки X:

x_i	0	1	2	3	4	5	6
n_i	29	49	41	26	18	5	10

(x_i – кількість нестандартних виробів в одній партії, n_i – кількість партій, які містять x_i нестандартних виробів). Користуючись критерієм Пірсона, за рівня значимості 0,01, перевірити гіпотезу про те, що випадкова величина X розподілена за біноміальним законом розподілу.

8. На підприємстві випадково відібрано 100 вибірок по 4 деталі. Реєструвалась кількість деталей з браком. У результаті отримано такий статистичний розподіл ознаки X:

x_i	0	1	2	3	4
n_i	31	36	24	6	3

(x_i – кількість бракованих деталей в одній вибірці). Користуючись критерієм Пірсона, при рівні значимості 0,05 перевірити гіпотезу про те, що випадкова величина X розподілена за біноміальним законом розподілу.

9. Вимірювали швидкість руху мотоцикліста x_i на певній ділянці шляху. Результати вимірювання наведено у вигляді статистичного розподілу:

x_i	55	60	65	70	75	80	85	90
n_i	2	5	6	7	9	8	6	7

Вважаючи, що X – швидкість руху автомобіля є випадковою величиною, перевірити за критерієм Колмогорова за рівня значимості 0,01 гіпотезу про нормальний закон розподілу X із математичним сподіванням $\bar{x} = 80$ і середнім квадратичним відхиленням $\sigma = 10$.

10. Вимірювання довжини ділянок шляху між зупинками тролейбуса в місті представлено у вигляді статистичного розподілу:

x_i	740,5	742,5	744,5	746,5	748,5	750,5	752,5
n_i	7	9	13	11	6	3	1

Вважаючи, що ознака X – довжина ділянки шляху – є випадковою величиною, яка має нормальний закон розподілу, при рівні значущості 0,001 перевірити гіпотезу H_0 : a=750 при конкуруючій гіпотезі: а) H_1 : $a \neq 750$; б) H_1 : a>750; в) H_1 : a<750.

11. Результати вимірювання зросту чоловіків на підприємстві записано у вигляді інтервального закону розподілу ознаки X – зросту чоловіків:

$[x_{i-1};x_i)$	n_i
[158;162)	3
[162;166)	5
[166;170)	6
[170;174)	7
[174;178)	5
[178;182)	4

Вважаючи, що випадкова величина X має нормальний закон розподілу, за рівня значущості 0,001 перевірити гіпотезу H_0 : a=170 за можливих альтернативних гіпотез: а) H_1 : $a \neq 170$; б) H_1 : a>170; в) H_1 : a<170.

12. За вибіркою обсягу n=50 знайдено середній розмір $\bar{x} = 20{,}1$ мк діаметра оптичного волокна, виготовленого автоматом №1, а за вибіркою m=50 знайдено середній розмір $\bar{x} = 19{,}8$ мк діаметра оптичного волокна, виготовленого автоматом №2. Генеральні дисперсії відомі: D(X)=1,750 мк^2, D(X)=1,375 мк^2. При рівні значущості 0,01 перевірити нульову гіпотезу H_0 : M(X)=M(Y) при конкуруючій гіпотезі а) H_1 : $M(X) \neq M(Y)$; б) H_1 : M(X)>M(Y); в) H_1 : M(X)<M(Y). Припускається, що випадкові величини X, Y розподілені нормально і вибірки незалежні.

13. Для дослідження розтягування певного типу гуми після хімічної обробки було відібрано 5 її мотків, кожний з яких було розділено навпіл і одна його половина була піддана хімічній обробці, а друга ні. Потім за допомогою приладу, що вимірює розтягування матеріалу, мотки гуми були виміряні і результати вимірювання наведені у вигляді двох статистичних розподілів ознак X та Y, які мають нормальний закон розподілу з відомими значеннями генеральних дисперсій D(X)=0,24, D(X)=0,21:

x_i	14,6	15,0	15,4	15,8	16,2	16,6
n_i	1	1	3	6	2	1

y_i	15,5	16,2	16,9	17,6
m_i	2	3	3	1

При рівні значущості $\alpha = 0{,}05$ перевірити гіпотезу H_0: M(X)=M(Y) при конкуруючій гіпотезі а) H_1: $M(X) \neq M(Y)$; б)H_1: M(X)>M(Y);
в)H_1: M(X)<M(Y).

14. На підприємстві розроблено два методи виготовлення скла. Щоб перевірити, чи однаково матеріалоємнісні ці методи, зібрано статистичні дані про витрати сировини на одиницю готової продукції в процесі роботи двома методами:

X (І-й метод)	2,2	2,9	2,7	3,1	2,5	2,8
Y (ІІ-й метод)	2,5	3,1	3,6	3,9	3,4	-

Припускаючи, що витрати сировини розподілені в обидвох методах за нормальним законом, перевірити гіпотезу H_0: M(X)=M(Y) при альтернативній гіпотезі а)H_1: M(X)<M(Y) і рівні значущості $\alpha = 0{,}01$.

15. Протягом доби двома приладами вимірювали напругу в електромережі. Результати вимірювання подано у вигляді статистичних розподілів:

x_i	228	232	237	240
n_i	2	3	5	2

y_i	211	215	222	226	231	236
m_i	1	1	5	3	2	2

Припускаючи, що випадкові величини X та Y (виміряні на двох приладах, напруга у вольтах) є незалежними і мають нормальний закон розподілу ймовірностей, за рівня значущості 0,001, перевірити правильність нульової гіпотези H_0: M(X)=M(Y) при конкуруючій гіпотезі а) H_1: $M(X) \neq M(Y)$; б) H_1: $M(X)>M(Y)$; в)H_1: $M(X)<M(Y)$.

16. Результати вимірювання споживання цукру за добу одним жителем у двох різних областях країни задаються двома статистичними розподілами:

x_i	14,3	16,3	19,3	20,3	22,3	23,3
n_i	1	3	5	4	2	1

y_i	13,5	17,5	22,5	28,5
m_i	2	6	4	1

Ознаки X та Y (добове споживання цукру в грамах) є незалежними випадковими величинами, які мають нормальний закон розподілу ймовірностей. Перевірити вірність гіпотези H_0: $D(X)=D(Y)$ при конкуруючій гіпотезі а) H_1 : $D(X) \neq D(Y)$ при рівні значущості 0,02; б) H_1: $D(X)>D(Y)$ при рівні значущості 0,05.

17. Вимірювалась жива маса бичків, які відгодовувались на двох фермах. Результати вимірювання задаються двома статистичними розподілами:

x_i	195	205	211	215
n_i	2	6	4	1

y_i	194	202	192	205	208	203
m_i	1	2	6	2	1	1

Ознаки X та Y (жива маса бичків в кг) є незалежними випадковими величинами, які мають нормальний закон розподілу ймовірностей. При рівні значущості 0,001 перевірити гіпотези H_0: $D(X)=D(Y)$ при альтернативних гіпотезах а) H_1 : $D(X) \neq D(Y)$ при рівні значущості 0,1; б) H_1: $D(X)>D(Y)$ при рівні значущості 0,01.

Навчально-методичне видання

Леся Андріївна Вотякова

Оксана Степанівна Турржанська

Теорія ймовірностей і математична статистика

Підписано до друку 12.09.2020 р.
Формат 60×80/16. Папір офсетний. Друк різографічний.
Гарнітура Times New Roman.
Наклад 100 прим.
ТОВ Фірма «Планер»