МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ
ХАРКІВСЬКИЙ НАЦІОНАЛЬНИЙ ЕКОНОМІЧНИЙ УНІВЕРСИТЕТ
ІМЕНІ СЕМЕНА КУЗНЕЦЯ

ТЕОРІЯ ЙМОВІРНОСТЕЙ ТА МАТЕМАТИЧНА СТАТИСТИКА

методичні рекомендації
до розв'язання завдань для самостійної роботи студентів
першого бакалаврського рівня

Харків, 2018

Тема 1. Емпіричні та логічні основи теорії ймовірностей

У класичній схемі **ймовірність** події визначається як відношення числа випадків m, які сприяють йому, до загальної кількості n рівноможливих, єдино можливих і несумісних результатів випробування, тобто: $P(A)=\dfrac{m}{n}$.

Основні формули комбінаторики, які використовуються в теорії ймовірностей

Перестановки - це комбінації, які складаються з одних і тих самих елементів і відрізняються тільки порядком їх розміщення:

$$P_n = n!,\ \text{где}\ n! = 1\cdot 2\cdot 3\cdot \ldots \cdot n,\ \ 0! = 1.$$

Розміщеннями називають комбінації, які складені з n різних елементів по m элементів, які відрізняються або складом елементів, або їх порядком:

$$A_n^m = n\cdot(n-1)(n-2)\cdot\ldots\cdot(n-m+1).$$

Сполучення - це комбінації, які складені з n різних элементів по m элементів, які відрізняються хоча б одним елементом:

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

Приклад. На конференцію з групи студентів з 20 чоловік (8 дівчат, 12 хлопців) відбирають 5 чоловік. Знайти ймовірність наступних подій:

А – серед відібраних студентів одні хлопці,

В – серед відібраних студентів одні дівчата,

С – серед відібраних 2 дівчини і 3 хлопці.

Розв'язання.

Загальна кількість результатів для всіх трьох подій буде однаковою n= C_{20}^{5}.

Кількість сприяючих результатів:

$$m_A = C_{12}^5,\ m_B = C_8^5,\ m_C = C_8^2 \cdot C_{12}^3.$$

Отже, отримуємо ймовірність появи події А:

$$P(A)=\frac{m_A}{n}=\frac{C_{12}^8}{C_{20}^5}=\frac{12!}{5!\cdot 7!}:\frac{20!}{5!\cdot 15!}=\frac{12!\cdot 5!\cdot 15!}{5!\cdot 7!\cdot 20!}=$$

$$=\frac{7!\cdot 8\cdot 9\cdot 10\cdot 11\cdot 12\cdot 15!}{7!\cdot 15!\cdot 16\cdot 17\cdot 18\cdot 19\cdot 20}=\frac{8\cdot 9\cdot 10\cdot 11\cdot 12}{16\cdot 17\cdot 18\cdot 19\cdot 20}=0{,}051.$$

Знайдемо ймовірність появи події В:

$$P(B)=\frac{m_B}{n}=\frac{C_8^5}{C_{20}^5}=\frac{8!\cdot 5!\cdot 15!}{5!\cdot 3!\cdot 20!}=\frac{3!\cdot 4\cdot 5\cdot 6\cdot 7\cdot 8\cdot 15!}{3!\cdot 15!\cdot 16\cdot 17\cdot 18\cdot 19\cdot 20}=$$

$$\frac{4\cdot 5\cdot 6\cdot 7\cdot 8}{16\cdot 17\cdot 18\cdot 19\cdot 20}=0{,}006.$$

Аналогічно отримуємо:

$$P(C)=\frac{m_C}{n}=\frac{C_8^2\cdot C_{12}^3}{C_{20}^5}=\frac{8!\cdot 12!\cdot 5!\cdot 15!}{2!\cdot 6!\cdot 3!\cdot 9!\cdot 20!}=\frac{7\cdot 8\cdot 10\cdot 11\cdot 12\cdot 3!\cdot 4\cdot 5\cdot 15!}{2!\cdot 6!\cdot 3!\cdot 9!\cdot 15!\cdot 16\cdot 17\cdot 18\cdot 19\cdot 20}=$$

$$=\frac{7\cdot 8\cdot 10\cdot 11\cdot 12\cdot 4\cdot 5}{2\cdot 16\cdot 17\cdot 18\cdot 19\cdot 20}=0{,}0795.$$

Тема 2. Основні теореми теорії ймовірностей

Теореми додавання ймовірностей для несумісних і сумісних подій. Теореми множення ймовірностей

Ймовірність появи суми двох несумісних подій дорівнює сумі ймовірностей цих подій: $P(A+B)=P(A)+P(B)$.

Ймовірність появи суми двох спільних подій дорівнює сумі ймовірностей цих подій мінус ймовірність їх спільної появи:

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

Ймовірність спільної появи двох залежних подій дорівнює добутку ймовірності однієї з цих подій на умовну ймовірність другої за умови, що перша вже відбулась:

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

Ймовірність спільної появи двох незалежних подій дорівнює добутку ймовірностей цих подій: $P(A \cdot B) = P(A) \cdot P(B)$.

Приклад 1. У ящику 15 деталей, серед яких 5 забарвлених. Складальник наугад дістає 3 деталі. Знайти ймовірність того, що з трьох узятих деталей забарвленою виявиться хоча б одна деталь.

Розв'язання. Вимога – хоча б одна з трьох деталей забарвлена – буде здійснено, якщо відбудеться будь-яка з наступних трьох несумісних подій:

B – одна деталь з трьох забарвлена,

C – дві деталі з трьох забарвлено,

D – три деталі забарвлено. Подію, що цікавить нас, A можна представити у вигляді суми подій: A = B + C + D, і по теоремі про ймовірність суми несумісних подій отримуємо:

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

$$P(B) = \frac{C_5^1 \cdot C_{10}^2}{C_{15}^3} = \frac{45}{91} = 0,495;$$

$$P(C) = \frac{C_5^1 \cdot C_{10}^2}{C_{15}^3} = \frac{20}{91} = 0,220;$$

$$P(D) = \frac{C_5^3}{C_{15}^3} = \frac{2}{91} = 0,022;$$

тоді $P(A) = P(B) + P(C) + P(D) = \frac{45}{91} + \frac{20}{91} + \frac{2}{91} = \frac{67}{91} = 0,736.$

Приклад 2. Ймовірність влучення у мішень при одному пострілі першим стрільцем 0,8, а другим – 0,9. Знайти ймовірність того, що:

а) обидва стрільці уразять мішень;

б) тільки перший стрілець уразить мішень;

в) тільки другий стрілець уразить мішень;

г) один стрілець уразить мішень;

д) жоден стрілець не уразить мішень;

е) хоча би один стрілець уразить мішень.

Розв'язання.

Хай подія A – мішень уражена першим стрільцем, подія B – другим. За умовою $P\left(A\right)=0{,}8$, $P\left(B\right)=0{,}9$, $P\left(\bar{A}\right)=0{,}2$, $P\left(\bar{B}\right)=0{,}1.$

а) $P(A\bigcap B)=P(AB)=P(A)\cdot P(B)=0{,}8\cdot 0{,}9=0{,}72$;

б) $P(A\bigcap \bar{B})=P(A)\cdot P(\bar{B})=0{,}8\cdot 0{,}1=0{,}08$;

в) $P(\bar{A}\bigcap B)=P(\bar{A})\cdot P(B)=0{,}2\cdot 0{,}9=0{,}18$;

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

д) $P(\bar{A}\bigcap \bar{B})=P(\bar{A}\cdot \bar{B})=P(A)\cdot P(\bar{B})=0{,}2\cdot 0{,}1=0{,}02$;

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

Формула повної ймовірності. Формула Байєса

Нехай подія A може статися з однією з подій (їх називають ***гіпотезами***) $B_1, B_2, \ldots, B_n$, які утворюють повну групу несумісних подій.

Формула повної ймовірності $P\left(A\right)=\sum\limits_{i=1}^{n}P\left(B_i\right)P_{B_i}\left(A\right).$

Формула Байєса $P_A\left(B_i\right)=\dfrac{P\left(B_i\right)\cdot P_{B_i}\left(A\right)}{P\left(A\right)}.$

Приклад 1. У урну, що містить дві кулі, опущена зелена куля. Знайти ймовірність того, що буде витягнута з урни зелена куля, якщо рівноімовірні первинні уявлення про колір куль.

Розв'язання.

Подія А – витягується зелена куля.

Можливі наступні гіпотези про первинний склад куль:

B_1 – зелених куль не має в урні;

B_2 – була 1 зелена куля;

B_3 – дві кулі зелені.

За умовою завдання гіпотези рівноімовірні і утворюють повну групу подій, отже, ймовірністьь кожної з гіпотез ⅓, тобто $P(B_1) = P(B_2) = P(B_3) = ⅓$. Тоді умовна ймовірінстьь настання події А при появі кожної з гіпотез будуть відповідно:

$$P_{B_1}(A) = ⅓;\ \ P_{B_2}(A) = ⅔;\ \ P_{B_3}(A) = 1.$$

Звідки за формулою повної ймовірності отримуємо:

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

$$P(A) = ⅓ \cdot ⅓ + ⅓ \cdot ⅔ + ⅔ \cdot 1 = ⅔.$$

Приклад 2. Два автомати виробляють однакові деталі, що поступають на загальний конвеєр. Продуктивність першого автомата удвічі більше продуктивності другого. Перший автомат виробляє в середньому 60% деталей відмінної якості, а другий 84%. Наугад узята деталь опинилася відмінної якості. Знайти ймовірність того, що ця деталь зроблена першим автоматом.

Розв'язання..

Розглянемо подію А – деталь відмінної якості.

Можна скласти дві гіпотези:

B_1– деталь вироблена первшим автоматом, $P(B_1) = ⅔$,

B_2– деталь вироблена другим автоматом $P(B_2) = ⅓$.

Умовна ймовірність появи події А за умовою виконання гіпотези B_1 – $P_{B_1}(A) = 0{,}6$.

Умовна ймовірність появи події А за умовою виконання гіпотези B_2 – $P_{B_2}(A) = 0{,}84$.

Тоді $P(A) = ⅔ \cdot 0{,}6 + ⅓ \cdot 0{,}84 = 0{,}68$.

Ймовіріність того, що деталь відмінної якості зроблена першим автоматом, за формулою Байєса:

$$P_A(B_1) = \frac{P(B_1) \cdot P_{B_1}(A)}{P(A)} = \frac{2/3 \cdot 0{,}6}{0{,}68} = \frac{10}{17}.$$

Тема 3. Схема незалежних випробувань

Формула Бернуллі

Нехай відбуваються n однорідних незалежних випробувань, в кожному з яких може відбутися або не відбутися певна подія A (таку серію повторних незалежних випробувань називають ***схемою Бернуллі***). Ймовірність появи події A в кожному випробуванні дорівнює p $(q = 1 - p)$. Тоді ймовірність того, що в результаті n незалежних випробувань подія A відбудеться рівно m раз, обчислюється за ***формулою Бернуллі***:

$$P_n(m) = C_n^m p^m q^{n-m} = \frac{n!}{m!\,(n-m)!} p^m q^{n-m}.$$

Приклад 1. Ймовірність того, що витрата електроенергії протягом однієї доби не перевищить встановленої норми р=0,7. Знайти ймовірність того, що в найближчі 6 діб витрата електроенергії протягом 4 діб не перевищить норми.

Розв'язання.

Ймовірінсть нормальної витрати електроенергії впродовж кожних з 6 діб постійна і рівна р=0,7.

Отже, ймовірінсть перевитрати електроенергії в кожну добу також постійна і рівна q =1 – p = 1 – 0,7 = 0,3.

З умови завдання виходить, що n = 6; k=4.

Тоді:

$$P_6(4) = C_6^4 p^4 q^2 = \frac{6!}{4!\cdot 2!}\cdot(0{,}7)^4\cdot(0{,}3)^2 = 0{,}324.$$

Локальна теорема Муавра – Лапласа.
Інтегральна теорема Муавра – Лапласа

Якщо ймовірність p появи події A в кожному з n незалежних випробувань постійна, а число випробувань досить велике $(npq \geq 20)$, то ймовірність того, що в цих випробуваннях подія A відбудеться m раз, обчислюється за формулою ***Муавра - Лапласа***:

$P_n(m) \approx \dfrac{\varphi(x)}{\sqrt{npq}}$, де $x = \dfrac{m - np}{\sqrt{npq}}$; $\varphi(x) = \dfrac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$ – *диференціальна функція Лапласа*; .

Якщо ймовірність p появи події в кожному випробуванні стала, а число випробувань n досить велике, то ймовірність того, що подія A відбудеться не менш m_1 й не більше m_2 раз ($m_1 < m_2$),наближено дорівнює: $P_n(m_1 \le m \le m_2) \approx \Phi(x_2) - \Phi(x_1)$, де $x_1 = \dfrac{m_1 - np}{\sqrt{npq}}$, $x_2 = \dfrac{m_2 - np}{\sqrt{npq}}$, функція $\Phi(x) = \int\limits_0^x \varphi(t)dt = \dfrac{1}{\sqrt{2\pi}} \int\limits_0^x e^{-\frac{t^2}{2}} dt$ – *інтегральна функція Лапласа*.

Приклад 2. Ймовірність народження хлопчика дорівнює $0{,}515$. Знайти ймовірність того, що з 200 народжених дітей кількість хлопчиків і дівчаток буде однаковою.

Розв'язання.

У даному випадку $n = 200$; $m = 100$; $p = 0{,}515$; $q = 1 - p = 0{,}485$; $\sqrt{npq} = \sqrt{200 \cdot 0{,}515 \cdot 0{,}485} \approx \sqrt{49{,}955} \approx 7{,}068$.

Значення x, що відповідає $m = 100$, дорівнює:

$$x = \frac{100 - 200 \cdot 0{,}515}{7{,}068} \approx -0{,}424.$$

Через те, що $\varphi(-0{,}424) = \varphi(0{,}424) = 0{,}3647$, отримаємо:

$$P_{200}(100) = \frac{\varphi(-0{,}424)}{\sqrt{49{,}955}} \approx \frac{0{,}3647}{7{,}068} \approx 0{,}052.$$

Приклад. Ймовірність того, що деталь виготовлена з порушенням стандартів, $p = 0{,}2$. Знайти ймовірність того, що серед 400 випадково відібраних деталей нестандартних виявиться від 70 до 100.

Розв'язання.

При $m_1 = 70$, $m_2 = 100$, $n = 400$, $p = 0{,}2$, $q = 0{,}8$ визначаємо x_1 і x_2:

$$x_1 = \frac{70 - 400 \cdot 0{,}2}{\sqrt{400 \cdot 0{,}2 \cdot 0{,}8}} = -1{,}25, \quad x_2 = \frac{100 - 400 \cdot 0{,}2}{\sqrt{400 \cdot 0{,}2 \cdot 0{,}8}} = 2{,}5.$$

Обчислимо ймовірність шуканої події:

$$P_{400}(70 \le m \le 100) = \Phi(2{,}5) - \Phi(-1{,}25) = \Phi(2{,}5) + \Phi(1{,}25).$$

За таблицею значень $\Phi(x)$ знаходимо $\Phi(2{,}5) = 0{,}4938$; $\Phi(1{,}25) = 0{,}3944$.

Тоді $P_{400}(70 \le m \le 100) = 0{,}4938 + 0{,}3944 = 0{,}8882$.

Тема 4. Випадкові величини, їх закони розподілу та числові характеристики

Числові характеристики дискретних випадкових величин

Нехай задан ряд розподілу випадкової величини X:

x_i	x_1	x_2	$\dots$	x_i	...	x_n
p_i	p_1	p_2	$\dots$	p_i	...	p_n

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

Дисперсією $D(X)$ дискретної випадкової величини є математичні сподівання квадрата відхилення випадкової величини від свого математичного сподівання, тобто:

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

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

Середнє квадратичне відхилення $\sigma(X)$ (σ_x) випадкової величини X – це квадратний корінь з дисперсії, тобто:

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

Приклад.

Задано ряд розподілу дискретної випадкової величини X:

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

.

Знайти числові характеристики величини X.

Розв'язання.

1. Математичне сподівання:

$$M(X)=x_1p_1+x_2p_2+x_3p_3=1\cdot 0{,}4+2\cdot 0{,}3+3\cdot 0{,}3=1{,}9.$$

2. Дисперсія:

$$D(X)=(x_1-M(X))^2p_1+(x_2-M(X))^2p_2+(x_3-M(X))^2p_3=$$
$$=(1-1{,}9)^2\cdot 0{,}4+(2-1{,}9)^2\cdot 0{,}3+(3-1{,}9)^2\cdot 0{,}3\approx 0{,}69,$$

або

$$D(X)=x_1^2p_1+x_2^2p_2+x_3^2p_3-(M(X))^2=1\cdot 0{,}4+4\cdot 0{,}3+9\cdot 0{,}3=0{,}69$$

.

3. Середнє квадратичне відхилення:

$$\sigma(X)=\sqrt{D(X)}=\sqrt{0{,}69}\approx 0{,}84.$$

Тема 6. Основні закони розподілу неперервних випадкових величин

Числові характеристики неперервної випадкової величини

Диференціальна функція розподілу ймовірностей (щільність розподілу ймовірностей) $f(x)$ є похідна від інтегральної функції розподілу $F(x)$: $f(x)=F'(x)$.

Пошук *інтегральної функції*, якщо задана диференціальна, пов'язаний з розв’язанням оберненої задачі: $F(x)=\int\limits_{-\infty}^{x} f(x)dx$.

Математичне сподівання неперервної випадкової величини X, можливі значення якої належать відрізку $[a,b]$, обчислюється як визначений інтеграл:

$$M(X)=\int\limits_{a}^{b} xf(x)dx.$$

Дисперсія неперервної випадкової величини X обчислюється як математичне сподівання квадрата відхилення випадкової величини від свого математичного сподівання.

Якщо $X\in(-\infty,\infty)$, то:

$$D(X)=\int\limits_{-\infty}^{\infty}(x-M(X))^2 f(x)dx, \quad \text{або} \quad D(X)=\int\limits_{-\infty}^{\infty}x^2 f(x)dx-(M(X))^2.$$

Середнє квадратичне відхилення неперервної випадкової величини X: $\sigma(X)=\sqrt{D(X)}$.

Ймовірність влучення неперервної випадкової величини в заданий інтервал (α,β) дорівнює різниці функцій $F(x)$ на кінцях інтервала:

$$P(\alpha<X<\beta)=F(\beta)-F(\alpha) \text{ або } P(a<X<b)=\int\limits_{a}^{b} f(x)dx.$$

Приклад. Дана функція розподілу неперервної випадкової величини X:

$$F(x)=\begin{cases}0, x\le 0,\\ \dfrac{x^2}{4}, 0<x\le 2,\\ 1, x>2.\end{cases}$$

Потрібно знайти: графік F(x), щільність f(x), графік f(x), математичне сподівання M(X), дисперсію D(X), середнє квадратичне відхилення σ, P(X < –2), P($\frac{1}{2}$ ≤ X < 1) P(X ≥ $\frac{3}{4}$).

Розв'язання.

1.

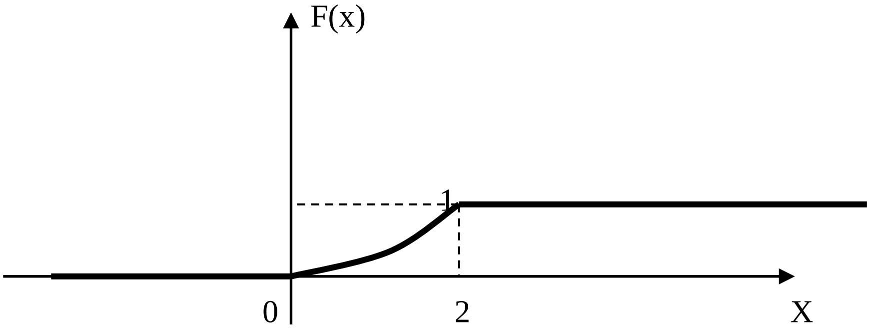

Рис. 1. Графік функції розподілу.

2. f(x)= F′(x), тоді:

$$0' = 0; \left(\frac{x^2}{4}\right)' = \frac{\left(x^2\right)'}{4} = \frac{2x}{4} = \frac{x}{2}; \ 1' = 0.$$

Отже f(x)=$\begin{cases} 0, x \le 0, \\ \frac{x}{2}, 0 < x \le 2, \\ 0, x > 2. \end{cases}$

3.

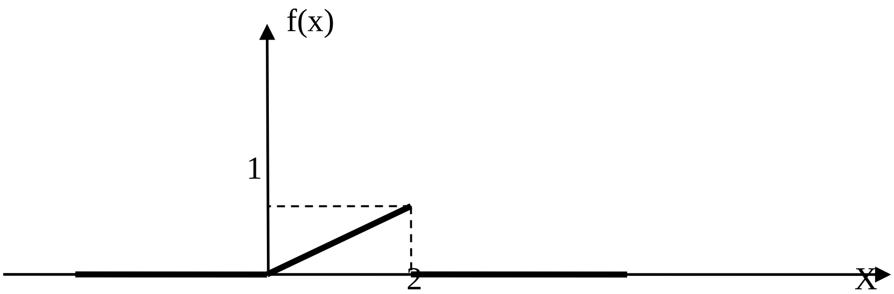

Рис. 2. Графік щільності f(x).

4. M(X)=$\int\limits_0^2 x\cdot f(x)dx = \int\limits_0^2 x\cdot\frac{x}{2}dx = \frac{1}{2}\int\limits_0^2 x^2dx = \frac{1}{2}\cdot\left.\frac{x^3}{3}\right|_0^2 = \frac{8}{6} = \frac{4}{3}$.

5. M(X^2)=$\int\limits_0^2 x^2\cdot f(x)dx = \int\limits_0^2 x^2\cdot\frac{x}{2}dx = \frac{1}{2}\int\limits_0^2 x^3dx = \frac{1}{2}\cdot\left.\frac{x^4}{4}\right|_0^2$ =2.

Тоді: D(X) = M(X^2) – M^2(X) = 2–$\left(\frac{4}{3}\right)^2$ = 2 –1,78 = 0,22.

6. σ(X) =$\sqrt{D(X)} = \sqrt{0{,}22}$ = 0,47.

7. Знайдемо ймовірінсть того, що випадкова величина X прийме значення з інтервалу (– ∞;– 2), тобто P(X< – 2):

P(X< – 2) = F(– 2) = 0,

Ймовірність P($\frac{1}{2}$ ≤ X < 1) знайдемо за формулою

P(a ≤ X < b)= F(b) – F(a): P($\frac{1}{2}$ ≤X<1)= F(1) – F($\frac{1}{2}$)=$\frac{1}{4} - \frac{1}{16} = \frac{3}{16}$.

Оскільки події $\left(X \geq \frac{3}{4}\right)$и $\left(X < \frac{3}{4}\right)$ протилежні, то:

P$\left(X \geq \frac{3}{4}\right)$=1– P$\left(X < \frac{3}{4}\right)$=1– F$\left(\frac{3}{4}\right)$=1– $\frac{9}{64} = \frac{55}{64}$.

Рівномірний закон розподілу ймовірностей і його числові характеристики. Нормальний закон розподілу ймовірностей і його стандартний вигляд

Рівномірним розподілом ймовірностей неперервної випадкової величини називають такий розподіл, при якому диференціальна функція є постійною величиною на інтервалі $[a,b]$, а поза межами цього інтервала дорівнює нулю:

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

$F(x)$ має вигляд:

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

Ймовірність потрапляння випадкової величини X в інтервал $(\alpha,\beta)\in[a,b]$:

$$\boldsymbol{P(\alpha<X<\beta)}=F(\beta)-F(\alpha)=\frac{\beta-a}{b-a}-\frac{\alpha-a}{b-a}=\boldsymbol{\frac{\beta-\alpha}{b-a}}.$$

Числові характеристики випадкової величини, розподіленої за рівномірним законом:

$$\boldsymbol{M(X)}=\int_a^b x\cdot f(x)dx=\int_a^b x\cdot\frac{1}{b-a}dx=\frac{1}{b-a}\left.\frac{x^2}{2}\right|_a^b=\frac{b^2-a^2}{2(b-a)}=\boldsymbol{\frac{a+b}{2}}.$$

$$\boldsymbol{D(X)}=\int_a^b x^2\cdot f(x)dx-(M(X))^2=\frac{1}{b-a}\int_a^b x^2\,dx-\left(\frac{a+b}{2}\right)^2=\boldsymbol{\frac{(b-a)^2}{12}}$$

$$\boldsymbol{\sigma(X)}=\sqrt{D(X)}=\boldsymbol{\frac{b-a}{2\sqrt{3}}}.$$

Випадкова величина X розподілена за нормальним законом, якщо її щільність розподілу ймовірностей при $x\in(-\infty;\infty)$ визначається формулою:

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

де σ, a – параметри розподілу.

Для нормального розподілу: $\boldsymbol{a=M(X)}$, $\boldsymbol{\sigma=\sqrt{D(X)}}$.

Ймовірність потрапляння випадкової величини X, розподіленої за нормальним законом з параметрами a и σ, в інтервал (α,β):

$$\boldsymbol{P}(\boldsymbol{\alpha} < \boldsymbol{X} < \boldsymbol{\beta}) = \boldsymbol{\Phi}\left(\frac{\boldsymbol{\beta}-\boldsymbol{a}}{\boldsymbol{\sigma}}\right) - \boldsymbol{\Phi}\left(\frac{\boldsymbol{\alpha}-\boldsymbol{a}}{\boldsymbol{\sigma}}\right).$$

Приклад 1. Якщо розподіл випадкової величини X – рівномірний і заданий відрізок [2;8], то b – a = 8 – 2 = 6 і

$$f(x)=\begin{cases} 0, x \notin [2;8], \\ \frac{1}{6}, x \in [2;8]. \end{cases}$$

Знайти числові характеристики рівномірного розподілу.

Розв'язання.

M(X)= $\int_a^b xf(x)dx = \frac{b+a}{2}$, тому M(X)= $\frac{b+a}{2} = \frac{8+2}{2} = 5.$

$$D(X) = \int_a^в [x^2 f(x)dx - [M(x)]^2 = \frac{(b-a)^2}{12}.$$

Тоді D(X) = $\frac{(b-a)^2}{12} = \frac{(8-2)^2}{12} = 3.$

Приклад 2. Випадкова величина X розподілена за нормальним законом і має щільність розподілу

$$f(x) = \frac{1}{2\sqrt{2\pi}} e^{-\frac{(x-3)^2}{8}}.$$

Знайти числові характеристики величини X та ймовірність влучення її в інтервал $(1;7)$.

Розв'язання.

З вигляду функції $f(x)$ маємо: $a=3, \sigma=2$. Таким чином, $M(X)=3, \sigma(X)=2, D(X)=4$. Імовірність влучення випадкової величини X в інтервал $(1;7)$ за формулою (7.19) дорівнює:

$$P(1 < X < 7) = \Phi\left(\frac{7-3}{2}\right) - \Phi\left(\frac{1-3}{2}\right) = \Phi(2) - \Phi(-1) = 0{,}477 + 0{,}341 = 0{,}818$$

.

Приклад 3. Довжина деталі, яку виготовляє автомат, є випадковою величиною, що розподілена за нормальним законом з параметрами $a = 10$см, $\sigma^2 = 0{,}0004$. Знайти ймовірність браку, якщо допустимі розміри деталі повинні бути $10 \pm 0{,}05$см.

Розв'язання.

$P_{\text{браку}} + P(9{,}95 < X < 10{,}05) = 1;$

$P_{\text{браку}} = 1 - P(9{,}95 < X < 10{,}05);$ $\alpha = 9{,}95;$ $\beta = 10{,}05;$ $\sigma = 0{,}02$.

$$P(9{,}95 < X < 10{,}05) = \Phi\left(\frac{10{,}05 - 10}{0{,}02}\right) - \Phi\left(\frac{9{,}95 - 10}{0{,}02}\right) =$$

$$= \Phi\left(\frac{0{,}05}{0{,}02}\right) - \Phi\left(\frac{-0{,}05}{0{,}02}\right) = 2\Phi(2{,}5) = 2 \cdot 0{,}494 = 0{,}988.$$

Таким чином, імовірність браку дорівнює: $P_{\text{браку}} = 1 - 0{,}988 = 0{,}012$.

Тема 7. Основні поняття математичної статистики: вибіркові спостереження і вибіркові оцінки

Вибірковий розподіл. Емпірична функція розподілу

Якщо записати вибірку у вигляді зростаючої чи спадаючої послідовностей, то отримаємо дискретний варіаційний ряд. Деякі значення дискретної випадкової величини можуть повторюватися, тоді ***варіаційний ряд*** записують у вигляді таблиці, де вказують ***варіанти*** - можливі значення (x_i) випадкової величини X та їх частоти (m_i). ***Частотою*** називають число появи окремих значень випадкової величини (варіант).

x_i	x_1	x_2	$\ldots$	x_l
m_i	m_1	m_2	$\ldots$	m_l

, $\sum_{i=1}^{l} m_i = n$.

Замість частот можна вказувати відносні частоти.

Відносної частотою w_i називається відношення частоти появи ознаки до загального обсягу вибірки: $w_i = \dfrac{m_i}{n}$, $\sum_{i=1}^{l} w_i = 1$.

Емпіричною функцією розподілу вибірки називається функція $F^*(x)$, яка для будь-якого значення x визначає відносну частоту події, що задовільняє умові $X < x$, тобто випадкова величина прийме значення менше, ніж x: $F^*(x) = \dfrac{m_x}{n}$, де m_x – сума частот варіант для значень аргумента, що меньші, ніж x; n – обсяг вибірки.

Графіком емпіричної функції розподілу є ***кумулята***, або графік накопичених відносних частот. Цей графік - ступінчаста фігура, яка має точки розриву при значеннях абсцис, рівних числовим значенням, які приймає випадкова величина.

Числові характеристики вибірки

Нехай заданий варіаційний ряд:

x_i	x_1	x_2	$\ldots$	x_i	$\ldots$	x_k
m_i	m_1	m_2	$\ldots$	m_i	$\ldots$	m_k

Для характеристики варіаційного ряду використовують такі величини:

1. *Вибіркове середнє* $\overline{x}$:

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

Вибіркове середнє характеризує середнє значення ознаки X.

2. *Вибіркова дисперсія* D: $D = \frac{1}{n}\sum_{i=1}^{k}\left(x_i - \bar{x}\right)^2 m_i$

або

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

Дисперсія в статистиці характеризує міру розсіювання ознаки X.

3. *Вибіркове середньоквадратичне відхилення* σ:

$$\sigma = \sqrt{D}.$$

Вибіркова дисперсія D є зміщеною оцінкою.

Досить легко виправити вибіркову дисперсію:

$$S^2 = \frac{n}{n-1} D_в = \frac{\sum_{i=1}^{l}\left(x_i - \bar{x}\right)^2 m_i}{n-1},$$

де S^2 – виправлена вибіркова дисперсія.

Приклад. Вибіркова сукупність задана таблицею розподілу

x_i	3	3,5	4	4,5	5
m_i	5	4	9	4	8

Знайти виправлену вибіркову дисперсію.

Розв'язання.

За визначенням виправлена вибіркова дисперсія $S_x^2 = \frac{n}{n-1} D$.

Об'єм вибірки: $n = 5 + 4 + 9 + 4 + 8 = 30$.

Вибіркова дисперсія: $D = \frac{1}{n}\sum_{i=1}^{5}\left(x_i - \bar{x}\right)^2 m_i$.

Вибіркова середня : $\bar{x} = \frac{\sum_{i=1}^{5} x_i m_i}{n}$.

Знаходимо вибіркову середню

$$\bar{x} = \frac{3 \cdot 5 + 3{,}5 \cdot 4 + 4 \cdot 9 + 4{,}5 \cdot 4 + 5 \cdot 8}{30} = 4{,}1$$

та вибіркову дисперсію

$$D = \frac{(3-4{,}1)^2 \cdot 5 + (3{,}5-4{,}1)^2 \cdot 4 + (4-4{,}1)^2 \cdot 9 + (4{,}5-4{,}1)^2 \cdot 4 + (5-4{,}1)^2 \cdot 8}{30},$$

$D = 0{,}49$.

Виправлена вибіркова дисперсія: $S_x^2 = \frac{30}{29} \cdot 0{,}49 \approx 0{,}505$.

Інтервальне оцінювання параметрів вибірки

Довірчий інтервал для середнього генеральної сукупності з надійністю $P = 2\Phi(t)$: $\bar{x} - \frac{\sigma_{\text{в}} t}{\sqrt{n}} \le a \le \bar{x} + \frac{\sigma_{\text{в}} t}{\sqrt{n}}$.

При $n > 30$ параметр t знаходят в таблиці функції $\Phi(t)$ (див. додаток В конспекта лекцій), при $n \le 30$ параметр $t = t(\gamma, n)$ – табл. Е.1 з додатка Е (замість $\sigma_{\text{в}}$ слід ставити виправлене середнє квадратичне відхилення $\sqrt{S_x^2}$).

Довірчий інтервал для генерального середньоквадратичного відхилення з надійністю $P = \gamma$:

$$s(1-q) \le \sigma_{\text{ген.}} \le s(1+q), \text{ где } q = \frac{\varepsilon}{s}.$$

Значення величини $q = q(\gamma, n)$, яка залежить від заданої надійності й обсягу вибірки, знаходять в спеціальній таблиці Е.2 (додаток Е конспекта лекцій).

Приклад 1. З генеральної сукупності вилучено вибірку:

Варіанта x_i	5	3	3	1	1	3
Частота n_i	3	1	1	1	2	1

Оцінити з надійністю 0,95 математичне сподівання нормально розподіленої ознаки генеральної сукупності по вибірковій середній за допомогою довірчого інтервалу.

Розв'язання.

$$\overline{x}_B = \frac{\sum n_i \cdot x_i}{n}, \; s = \sqrt{\frac{\sum n_i \cdot (x_i - \overline{x}_B)^2}{n-1}}.$$

Тоді:

$$\overline{x}_B = \frac{3\cdot 5+1\cdot 3+1\cdot 3+1\cdot 1+2\cdot 1+1\cdot 3}{9} = \frac{15+3+3+1+2+3}{9} = \frac{27}{9} = 3.$$

$$s = \sqrt{\frac{3\cdot(5-3)^2+1\cdot(3-3)^2+1\cdot(3-3)^2+1\cdot(1-3)^2+2\cdot(1-3)+1\cdot(3-3)^2}{9-1}} =$$

$$= \sqrt{\frac{3\cdot 2^2+1\cdot 0^2+1\cdot 0^2+1\cdot(-2)^2+2(-2)^2+1\cdot(0)^2}{8}} = \sqrt{\frac{24}{8}} = 1{,}7.$$

Отже $\overline{x_B}$=3, s = 1,7.

Тоді:

$$\overline{x}_B - t_\gamma \cdot s/\sqrt{n} < a < \overline{x}_B + t_\gamma \cdot s/\sqrt{n}.$$

Значення t_γ находимо в таблиці додатка Е конспекта лекцій по заданим n = 9 і γ=0,95: t_γ=2,31.

Отримуємо:

$$3 - 2{,}31\cdot 1{,}7/\sqrt{9} < a < 3 + 2{,}31\cdot 1{,}7/\sqrt{9},$$

1,691 < a < 4.309.

Приклад 2. За даними вибірки об'єму n = 40 з генеральної сукупності знайдено «виправлене» середнє квадратичне відхилення s =1 нормально розподіленої кількісної ознаки. Знайти довірчий інтервал, що покриває генеральне середнє квадратичне відхилення з надійністю 0,99.

Розв'язання.

s·(1 – q) < σ < s·(1 + q) (якщо q < 1)

або

0 < σ < s·(1 + q) (если q > 1).

Значення q находять по таблиці додатка Е конспекта лекції по заданим n=40 та γ=0,99: q=0,35.

q = 0,35 < 1, тоді

1·(1 – 0,35) < σ < 1·(1 + 0,35), отсюда 0,65 < σ < 1,35.

Тема 8. Методы проверки статистических гипотез

Критерій згоди Пірсона

Згідно з цим критерієм спостерігається емпіричний розподіл вибіркової сукупності, яке виражене емпіричними частотами m_i сгруппированного ряду, порівнюється з передбачуваним теоретичним розподілом генеральної сукупності, яке виражене теоретичними частотами $\tilde{m}_i$.

Для перевірки нульової гіпотези H_0 треба обчислити величину $\chi^2 = \sum_{i=1}^{s} \frac{(m_i - \tilde{m}_i)^2}{\tilde{m}_i}$, де s – кількість інтервалів сгрупованого ряду розподілу; m_i – емпіричні частоти; $\tilde{m}_i$ – теоретичні частоти.

Знайдена величина χ^2 порівнюється з критичними значеннями $\chi^2_\alpha(k)$, які знаходяться в спеціальних довідкових таблицях (див. додаток Д конспекта лекцій).

Число ступенів свободи k визначається за формулою: $k = s - r - 1$, де s – кількість укрупнених інтервалів; r – число параметрів теоретичного закону розподілу (для нормального та рівномірного закону $r = 2$, для показникового – $r = 1$).

Величина α визначає рівень значимості. Для критерія Пірсона будемо розглядати два рівня значущості: $\alpha = 0{,}05$ та $\alpha = 0{,}01$.

Якщо $\chi^2 < \chi^2_{0,05}(k)$, то нульова гіпотеза H_0 приймається, тобто передбачуваний закон розподілу відповідає емпіричним даним, при цьому ми помиляємося в п'яти випадках зі ста, приймаючи можливо помилкову гіпотезу (помилка другого роду).

Якщо $\chi^2 > \chi^2_{0,01}(k)$, то нульову гіпотезу слід відкинути, тобто передбачуваний закон розподілу не відповідає емпіричним даним, при цьому ми помиляємося в одному випадку зі ста, відкидаючи можливо правильну гіпотезу (помилка першого роду).

Якщо $\chi^2_{0,05}(k) < \chi^2 < \chi^2_{0,01}(k)$, то це область невизначеності (гіпотезу можна як прийняти так і відкинути) і необхідно використовувати інші критерії.

Обчислення теоретичних частот $\tilde{m}_i$.

Якщо маємо вибірку випадкової величини X, то за її значеннями знаходять $x_{\max}$ и $x_{\min}$. Тоді розмах варіювання $R = x_{\max} - x_{\min}$, округлимо $R \approx R_0$. Знайдемо крок варіювання $h = \frac{R_0}{S}$, де S – число інтервалів, тоді таблиця обчислень $\tilde{m}_i$:

Інтервали	x_i	m_i	$h_i = \Delta x_i$	$t_i = \frac{x_i - \bar{x}}{\sigma}$	$\varphi(t_i)$	$f_i = \frac{\varphi(t_i)}{\sigma}$	$\tilde{p}_i = hf_i$	$\tilde{m}_i = n\tilde{p}_i$
$x_0^* - x_1^*$	x_1	m_1	h_1	t_1	$\varphi(t_1)$	f_1	$\tilde{p}_1$	$\tilde{m}_1$
…	…	…	…	…	…	…	…	…

В таблиці $x_i = \frac{x_{i-1}^* + x_i^*}{2}$ – середина інтервала; m_i – емпірична частота; $t_i = \frac{x_i - \bar{x}}{\sigma}$ – стандартизована величина; $\varphi(t_i)$ – диференціальна функція; $\tilde{p}_i = h_i f_i$ – теоретична ймовірність; $\tilde{m}_i$ – теоретична частота.

Приклад. При рівні значущості α = 0,05 перевірити гіпотезу про нормальний розподіл генеральної сукупності, якщо відомі емпіричні (m_i) і теоретичні $\tilde{m}_i$ частоти:

(m_i)	6	13	38	74	106	85	30	14
$\tilde{m}_i$	3	14	42	82	99	76	37	13

Розв'язання. Для обчислення $\chi^2_{розр}$ заповнимо таблицю:

i	(m_i)	$\tilde{m}_i$	$\dfrac{(m_i-\tilde{m}_i)^2}{\tilde{m}_i}$
1 2	$\left.\begin{matrix}6\\13\end{matrix}\right\}19$	$\left.\begin{matrix}3\\14\end{matrix}\right\}17$	0,24
3	38	42	0,42
4	74	82	0,78
5	106	99	0,49
6	85	76	1,07
7	30	37	1,32
8	14	13	0,08
$\sum$	366	366	$\chi^2 = 4{,}43$

Число ступенів свободи *r* = *7* – 3 = 4. За табл. значень χ^2 (додаток Д конспекта лекцій) за рівнем значущості α = 0,05 знаходимо $\chi^2_{табл}$(0,05,4) = 9,5. Якщо $\chi^2_{розр} < \chi^2_{табл}$,то приймається гіпотеза, що дані спостережень узгоджуються з гіпотезою про нормальний розподіл генеральної сукупності.

Тема 9. Елементи теорії кореляції

Нехай маємо емпіричні дані:

x	x_1	x_2	$\cdots$	x_i	$\cdots$	x_n
$y_{эмп.}$	y_1	y_2	$\cdots$	y_i	$\cdots$	y_n

Теоретичне рівняння регресії Y на X має вигляд:

$$\hat{y}=\rho_{y/x}x+b \text{, где } \boldsymbol{\rho}_{y/x}=\frac{\boldsymbol{\mu}_{xy}}{\boldsymbol{\sigma}_x^2}=\frac{\overline{xy}-\overline{x}\cdot\overline{y}}{\overline{x^2}-\left(\overline{x}\right)^2},\ b=\overline{y}-\rho_{y/x}\overline{x}.$$

Оскільки кореляційний момент (коваріація) μ_{xy} – характеристика розмірна, то вводять додаткову безрозмірну характеристику - ***коефіцієнт кореляції*** - показник тісноти лінійного зв'язку: $r_{xy}=\frac{\mu_{xy}}{\sigma_x\sigma_y}$.

Приклад. Дана таблиця, дані в якій не згруповані

x_i	y_i
1,5	2,2
1,4	3,5
1,2	3,7
1,1	3,8
0,9	4,9
0,8	5,7

Необхідно записати теоретичне рівняння регресії, оцінити тісноту зв'язку.

Розв'язання.

Рівняння залежності шукаємо у вигляді: $\overline{y}_x=\rho_{y/x}x+b$. Для знаходження параметрів $\rho_{y/x}$ та b заповнюємо таблицю.

x_i	y_i	$x_i\cdot y_i$	x_i^2	y_i^2
1,5	2,2	3,3	2,25	4,84
1,4	3,5	4,5	1,96	12,25
1,2	3,7	4,44	1,44	13,69
1,1	3,8	4,18	1,21	14,44
0,9	4,9	4,41	0,81	24,01
0,8	5,7	4,56	0,64	32,49
$\sum=6{,}9$	$\sum=23{,}8$	$\sum=25{,}79$	$\sum=8{,}31$	$\sum=101{,}72$

Далі знаходимо:

$$\bar{x} = \frac{\sum x_i}{n} = \frac{6{,}9}{6} = 1{,}150; \qquad \bar{y} = \frac{\sum y_i}{n} = \frac{23{,}8}{6} = 3{,}967;$$

$$\overline{x^2} = \frac{1}{n}\sum x_i^2 = \frac{8{,}31}{6} = 1{,}385; \qquad \overline{y^2} = \frac{1}{n}\sum y_i^2 = \frac{101{,}72}{6} = 16{,}953;$$

$$s_x^2 = \overline{x^2} - (\bar{x})^2 = 1{,}385 - (1{,}15)^2 = 0{,}0625; \quad s_x = 0{,}25;$$

$$s_y^2 = \overline{y^2} - (\bar{y})^2 = 16{,}953 - (3{,}967)^2 = 1{,}216; \quad s_y = 1{,}103.$$

$$\overline{xy} = \frac{\sum x_i \cdot y_i}{n} = \frac{25{,}79}{6} = 4{,}298, \ \mu_{xy} = \overline{xy} - \bar{x} \cdot \bar{y} = -0{,}264.$$

$$\rho_{y/x} = \frac{\mu_{xy}}{s_x^2} = \frac{-0{,}264}{0{,}0625} = -4{,}224$$

$$b = \bar{y} - \rho_{y/x} \cdot \bar{x} = 3{,}967 + 4{,}224 \cdot 1{,}15 = 8{,}825.$$

Рівняння лінії регресії має вигляд : $\overline{y_x} = -4{,}224 \cdot x + 8{,}825$.

Набуте значення коефіцієнта регресії $\rho_{y/x} = -4{,}224$ означає, що із збільшенням чинника X на одну одиницю свого вимірювання чинник Y в середньому зменшується на 4,224 своїх одиниць вимірювання.

Оцінимо тісноту зв'язку за допомогою коефіцієнта кореляції:

$$r = \frac{\overline{xy} - \bar{x} \cdot \bar{y}}{s_x \cdot s_y} = \frac{-0{,}264}{0{,}25 \cdot 1{,}103} = -0{,}957.$$

Оскільки, значення коефіцієнта кореляції близьке до 1, то можна стверджувати, що між чинниками існує тісна лінійна кореляційна залежність.