Міністерство освіти і науки України

ДЕРЖАВНИЙ БІОТЕХНОЛОГІЧНИЙ УНІВЕРСИТЕТ

Факультет мехатроніки та інжинірингу

Кафедра «Сільськогосподарські машини та інженерія тваринництва»

Вибір закону розподілу випадкової величини

МЕТОДИЧНІ ВКАЗІВКИ

до виконання практичної роботи з навчальної дисципліни
«Моделювання технологічних процесів та систем»

для здобувачів другого (магістерського) рівня вищої освіти
спеціальності 208 «Агроінженерія»

Затверджено рішенням
науково-методичної комісії
факультету мехатроніки та
інжинірингу ДБТУ
Протокол № 2 від 27.12.2023 р.

Харків – 2023

УДК 519.21(072)
В 41

Схвалено
на засіданні кафедри сільськогосподарських машин та інженерії тваринництва
Протокол № 5 від 19.12.2023 р.

В 41 Вибір закону розподілу випадкової величини: методичні вказівки до виконання практичної роботи з навчальної дисципліни «Моделювання технологічних процесів та систем» для здобувачів другого (магістерського) рівня вищої освіти спеціальності 208 «Агроінженерія» / Державний біотехнологічний університет; уклад. Р.В. Кириченко – Харків: [б. в.], 2023. – 23 с.

До методичних вказівок за темою «Вибір закону розподілу випадкової величини» включено вибір закону розподілу випадкової величини, оцінка параметрів закону розподілу, оцінка відповідності вибраного закону розподілу випадкової величини експериментальним даним, порядок виконання роботи, зміст звіту та контрольні запитання для самоперевірки.

Видання призначене здобувачам другого (магістерського) рівня вищої освіти спеціальності 208 «Агроінженерія».

Рецензенти:
Р.В. Антощенков, доктор технічних наук, професор, завідувач кафедри мехатроніки, безпеки життєдіяльності та управління якістю Державного біотехнологічного університету.
М.Л. Шуляк, доктор технічних наук, професор, завідувач кафедри агроінжинірингу Сумського національного аграрного університету.

УДК 519.21(072)
Відповідальний за випуск: М.М. Крекот

ЗМІСТ

Мета роботи……………………………………………………………4
1. Вибір закону розподілу випадкової величини……………..………4
2. Оцінка параметрів закону розподілу……………………..………10
3. Оцінка відповідності вибраного закону розподілу випадкових величин експериментальним даним…………………………...….12
3.1 Критерій Пірсона……………………………...…………..13
3.2 Критерій Колмогорова…………………………………………19
4. Порядок виконання роботи……………………...……………..21
5. Зміст звіту………..…………………….…………………..21
6. Контрольні запитання для самоперевірки………….…………….22

ВИБІР ЗАКОНУ РОЗПОДІЛУ ВИПАДКОВОЇ ВЕЛИЧИНИ

Мета роботи

Вибрати закон розподілу випадкової величини для отриманої статистичної вибірки, перевірити правильність вибраної гіпотези про відповідність експериментального закону розподілу випадкової величини теоретичному.

1. Вибір закону розподілу випадкової величини

Більшість величин, з якими доводиться мати справу в інженерній практиці, відомі з деякою похибкою, що обумовлюється неточностями вимірювання і розрахунку. В цьому значенні ці величини є випадковими величинами. Зменшити вказану похибку або хоча б оцінити її з потрібною довірчою імовірністю дозволяють методи математичної статистики.

Будь-яка випадкова величина X повністю характеризується своєю функцією розподілу:

$$F(x) = P(X < x), \tag{1}$$

яка являє собою імовірність того, що в окремому експерименті (досліді) X прийме значення, менше x**.** Тому знання $F(x)$ дуже важливо і дозволяє вирішити багато задач, пов'язаних з оцінкою похибки параметрів технічних засобів, якістю технологічних процесів, прогнозуванням відсотка браку тощо.

Наприклад, імовірність того, що певний параметр знаходитиметься в межах інтервалу (a, b), можна знайти із співвідношення:

$$P(a < X < b) = F(b) - F(a). \tag{2}$$

На практиці функція розподілу іноді може бути відома наперед або знайдена виходячи з фізичної природи досліджуваної величини. Наприклад, відомо, що похибка дискретизації (коли неперервна крива замінюється послідовністю випадкових значень, що відповідають точкам перетину січних площин, проведених з певним кроком відносно осі абсцис) при вимірюванні має рівномірний розподіл. Проте частіше робити висновок про вигляд функція розподілу випадкової величини доводиться маючи в розпорядженні тільки емпіричні дані, наприклад, результати вимірювань, що виражені чисельно. Розглянемо, як це можна зробити.

Наприклад припустимо, що в результаті n дослідів були, отримані значення досліджуваної величини x:

$$x_1, x_2, ..., x_n,$$

Ці величини, розташовані в порядку їх отримання, називаються *простою вибіркою*, або *статистичним рядом*. Кількість елементів вибірки називається *об'ємом вибірки*. Вибірка повинна мати достатній для її дослідження об'єм, або, як то кажуть, повинна бути представницькою. Так, для знаходження закону розподілу випадкової величини необхідно мати не менше декілька десятків значень.

Якщо об'єм вибірки записати в порядку зростання, то отримана послідовність називається *варіаційним рядом*. Така форма представлення вибірки значно зручніше для подальшої обробки.

Якщо об'єм вибірки великий, її обробка стає трудомісткою. В цьому випадку дані заздалегідь групують і при розрахунках використовують груповану вибірку.

Для цього з вибірки вибирають мінімальне x_{min} і максимальне x_{max} значення і визначають *розмах варіювання* R, який розбивають на k інтервалів:

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

Одностайного розбиття розмаху варіювання на інтервали не існує. Інтервали можуть бути як рівними за величиною, так і різними. Але при цьому важливо пам'ятати, що групування вибірки, звичайно, робить менш точною (грубою) початкову інформацію, оскільки після цього про елемент відомо тільки до якого з інтервалів він належить. Чим довше інтервали, тим менша достовірність інформації. Тому, по можливості, потрібно прагнути того, щоб інтервалів було більше. З іншого боку, в кожному інтервалі має міститися не менше шести елементів, і якщо при розбитті в деяких інтервалах опинилося менше елементів, потрібно об'єднати їх з сусідніми. Якщо значення вибірки розподілені нерівномірно, доцільно вибирати інтервали різної довжини – більш довгі на ділянці рідкісного розташування значень.

Як первинне розбиття може бути рекомендований, наприклад, наступний підхід:

$$k = 1 + 3{,}3 \cdot \lg(n), \qquad (3)$$

де n – об'єм вибірки.

Іноді розбиття проводять з припущення, щоб в інтервал потрапило в середньому не менше 10 значень, тоді:

$$k = 0{,}1n - 1. \qquad (4)$$

При цьому середні значення отриманих інтервалів називають *представниками інтервалів*:

$$\overline{x_i'} = \frac{(x_{i+1} + x_i)}{2}. \qquad (5)$$

Для кожного i-го інтервалу $(x_i\,,\, x_{i+1}\,)$ підраховують *частоту* – число n_i, значень випадкової величини, що потрапили в цей інтервал, а також *відносну частоту* (*частотність*):

$$p_i{*} = \frac{n_i}{n}, \qquad (6)$$

і *щільність відносної частоти*:

$$f_i^* = \frac{n_i}{n \cdot \Delta x_i}, \tag{7}$$

де $\Delta x_i = \dfrac{R}{k}$ – довжина інтервалу.

Отримані дані, звичайно, зводять в таблицю. Наприклад, для певної вибірки експериментальних даних одержуємо таблицю 1.

Таблиця 1 - Данні для побудови статистичних даних

Номер інтервалу k_i	Довжина інтервалу Δx_i	Межі інтервалу x_i'	Представник інтервалу $\overline{x}_i'$	Частота n_i	Відносна частота p_i^*	Щільність відносної частоти f_i^*
1	0,5	5,0-5,5	5,25	2	0,02	0,04
2	0,5	5,5-6,0	5,75	7	0,07	0,14
3	0,5	6,0-6,5	6,25	8	0,08	0,16
4	0,5	6,5-7,0	6,75	13	0,13	0,26
5	0,5	7,0-7,5	7,25	15	0,15	0,30
6	0,5	7,5-8,0	7,75	16	0,16	0,32
7	0,5	8,0-8,5	8,25	17	0,17	0,34
8	0,5	8,5-9,0	8,75	22	0,22	0,44
Сума				100	1,00	

Правильність складання таблиці можна проконтролювати, перевіривши виконання рівності:

$$\sum_{i=1}^{k} n_i = n \quad \text{та} \quad \sum_{i=1}^{k} p_i^* = 1. \tag{8}$$

Геометричною ілюстрацією групованої вибірки служить *гістограма*, яка будується з прямокутників, основами яких служать інтервали, а висоти дорівнюють відповідній щільності відносних частот *f**.

Крім того, іноді будують *полігон*. Для цього по горизонтальній осі, як і в гістограмі, відкладають інтервали, а з середини інтервалів – значення щільності відносних частот f_1*. Отримані точки з'єднують ламаною лінією. Гістограма і полігон представлені на рисунках 1 і 2.

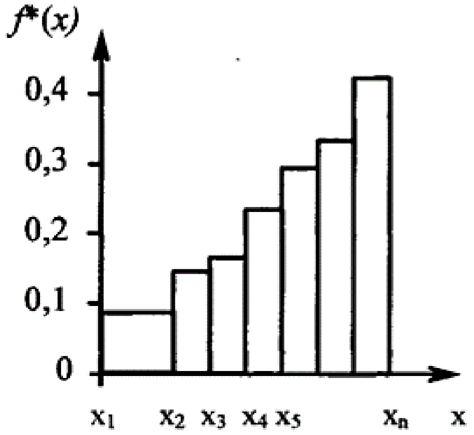

Рис. 1 – Гістограма

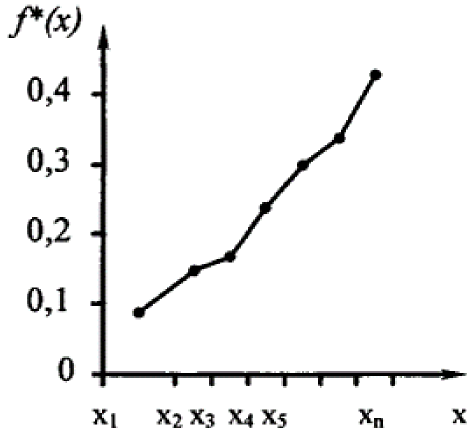

Рис. 2 – Полігон

Відзначимо деякі властивості гістограми. Загальна площа гістограми дорівнює одиниці, тобто:

$$\sum_{i=1}^{k} f_i^*(x) = 1. \qquad (9)$$

Якщо об'єм вибірки спрямувати до нескінченності, а величину інтервалу до нуля, гістограма буде наближена до кривої щільності розподілу імовірності (ЩРІ).

Таким чином, гістограма служить деяким наближенням графіка ЩРІ досліджуваної величини. Тому по вигляду гістограми, так само як і полігону, можна зробити припущення про тип розподілу випадкової величини. Розподіл, який вибирається при висуванні гіпотези, надалі називатимемо *теоретичним розподілом*.

Статистичною функцією розподілу (СФР) випадкової величини X називається функція $F^*(x)$, що дорівнює відносній частоті події ($X<x$):

$$F^*(x) = \frac{n_x}{n}, \tag{10}$$

де n_x – кількість значень вибірки, меншої за x;

n – об'єм вибірки.

Функція розподілу (ФР) легко обчислюється за варіаційним рядом. Вона має стрибки, що кратні $1/n$ в точках значень елементів вибірки.

Як вже відмічалося, при великому об'ємі вибірки зручніше використовувати груповані дані, для яких СФР обчислюється тільки на межах інтервалів і має стрибків, що дорівнюють n_i/n.

Прикладом СФР є групована і негрупована статистичні функції (рис. 3 і 4).

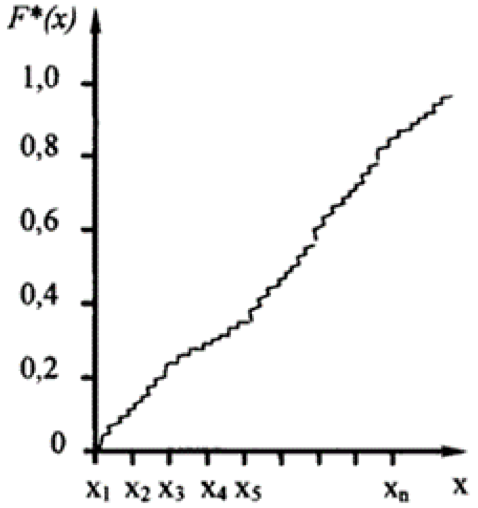

Рис. 3 – Негрупована статистична функція

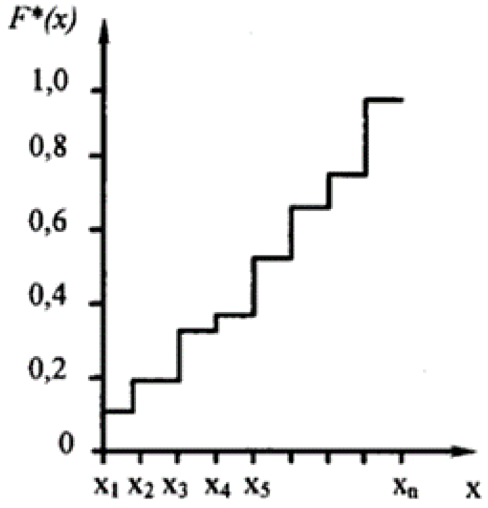

Рис. 4 – Групована статистична функція

Вочевидь, що СФР, як і ФР, не можуть бути спадаючими величинами, а її значення при верхній межі області визначення (для негрупованої СФР – це x_{max}, а для групованої – крайній правий інтервал) дорівнює 1.

Згідно закону великих чисел, при збільшенні числа дослідів, частота p_i^* сходиться до p, тобто при $n \to \infty$ СФР наближається до істинної ФР, і по ній також може бути вибраний теоретичний розподіл.

2. Оцінка параметрів закону розподілу

Після вибору типу теоретичного розподілу потрібно визначити параметри, що входять в нього. Так,

– *для експоненціального розподілу*:

$$f(x) = ae^{-a\,x}, \text{ де } x > 0 \tag{11}$$

потрібно знайти один параметр *a*,

– *для нормального розподілу*:

$$f(x) = \frac{1}{\sqrt{2\pi\sigma}} e^{-\frac{(x-a)^2}{2\sigma^2}}, \text{ де } -\infty < x < \infty \tag{12}$$

необхідно знайти два параметра – *a* та *σ*.

Іноді параметри можуть бути відомі заздалегідь або визначені з фізичної сутності досліджуваного процесу. Проте частіше параметри доводиться оцінювати, використовуючи значення самої вибірки.

Відомі декілька підходів до розв’язання цієї задачі. Найпоширенішим з них є *метод моментів*.

Сутність методу моментів полягає в тому, що параметри теоретичного розподілу вибираються так, щоб деякі його моменти співпадали із статистичними моментами вибірки.

Нагадаємо, що першим початковим моментом теоретичного розподілу є *математичне очікування*:

$$M(X) = m'_x = \int_{\infty}^{-\infty} x \cdot f(x)dx, \tag{13}$$

другим центрованим моментом є *дисперсія*:

$$D(X) = \sigma_x^2 = \int_{\infty}^{-\infty} (x - m'_x)^2 \cdot f(x)dx. \tag{14}$$

Оцінити за вибіркою наведені вище статистичні моменти можна, використовуючи наступні співвідношення:

– *оцінка математичного очікування*:

$$\overline{m'_x} = \frac{1}{n} \cdot \sum_{i=1}^{n} x_i \text{ – для простої вибірки;} \tag{15}$$

$$\overline{m'_x} = \frac{1}{n} \cdot \sum_{i=1}^{n} (x_i \cdot m_i) \text{ – для групованої вибірки;} \tag{16}$$

– *оцінка дисперсії:*

$$\overline{\sigma}_x^2 = \frac{1}{n-1} \cdot \sum_{i=1}^{n} (x_i - \overline{m'_x})^2 \text{ – для простої вибірки;} \tag{17}$$

$$\overline{\sigma}_x^2 = \frac{1}{n-1} \cdot \sum_{i=1}^{n} (x_i - \overline{m'_x})^2 \cdot m'_i \text{ – для групованої вибірки;} \tag{18}$$

Аналогічно можуть бути визначені моменти і більш високих порядків.

Таким чином, якщо теоретичний розподіл має два параметри, вони можуть бути визначені з системи рівнянь:

$$\begin{cases} m'_x = \overline{m'_x}; \\ \sigma_x = \overline{\sigma}_x. \end{cases} \tag{19}$$

Якщо параметр один – може бути використане будь-яке з наведених рівнянь. Наприклад,

– *для експоненціального розподілу*:

$$m'_x = \frac{1}{a}, \tag{20}$$

тоді

$$a = \frac{1}{\overline{m'_x}}; \tag{21}$$

– для нормального розподілу:

$$m'_x = m' \text{ та } \sigma_x^2 = \sigma^2, \tag{22}$$

тоді

$$m' = \overline{m'_x} \text{ та } \sigma = \sqrt{\sigma_x^2}\,. \tag{23}$$

3. Оцінка відповідності вибраного закону розподілу вибіркової величини експериментальним даним

Після знаходження параметрів теоретичного закону розподілу необхідно оцінити відповідність висунутої гіпотези про тип закону вихідним (експериментальним) даним.

Для вирішення поставленої задачі користуються критеріями згоди. В якості міри узгодження теоретичного і емпіричного (експериментального) законів розподілу обчислюється деяка функція, область можливих значень якої ділиться на дві області: S_0 – *область прийняття рішення* та S_1 – *критична область.* Якщо обчислене значення належить області S_0, гіпотеза приймається, в протилежному випадку – відкидається.

При перевірці гіпотез можливі наступні помилки:

а) гіпотеза вірна, але відкидається (помилка першого роду);

б) гіпотеза не вірна, але приймається (помилка другого роду).

Імовірність цих помилок повинна бути мінімальна, проте прагнення зменшити імовірність однієї з помилок неминуче призводить до збільшення іншої. Насправді, для зменшення до нуля імовірності помилки першого роду доведеться приймати гіпотезу при будь-якій вибірці, а щоб досягти того ж результату для імовірності помилки другого роду – не приймати ніколи. На практиці використовується розумний компроміс. Для цього імовірність однієї з помилок задається (як правило, помилки першого роду), а імовірність іншої – прагнуть мінімізувати. Тоді «кращим» критерієм буде той, який забезпечує цей мінімум.

При перевірці гіпотези про закон розподілу задається імовірність помилки першого роду (яка називається *рівнем значущості*), тобто імовірності відкинути правильну гіпотезу.

Якщо гіпотеза прийнята, це ще не значить, що вона вірна, ми тільки зробили висновок про те, що вона не суперечить експериментальним даним.

Якщо гіпотеза знехтувана (відкинута), необхідно висунути нову гіпотезу, вибравши інший тип закону розподілу. В теорії імовірності відомий широкий клас розподілів, з яких майже завжди можна вибрати що-небудь відповідне.

Крім того, можна спробувати і власні розподіли: проте для виконання даного завдання достатньо з’ясувати, чи не належить досліджувана вибірка одному з розподілів.

3.1 Критерій Пірсона

Найбільш поширеним критерієм при виборі типу закону розподілу є критерій χ^2 (хі-квадрат), розроблений К. Пірсоном.

Якщо гіпотеза про тип закону розподілу вірна, то теоретичну імовірність p_i потрапляння випадкової величини в i-й інтервал можна знайти як:

$$p_i = F(x_{i+1}) - F(x_i), \text{ де } i=1,2,\ldots,k. \qquad (24)$$

Статистична ж імовірність (відносна частота), знайдена за вибіркою:

$$p_i{*} = \frac{n_i}{n}, \text{ де } i = 1, 2, \ldots, k. \quad (25)$$

В якості міри розбіжності між p_i та p_i*, Пірсон запропонував використовувати величину:

$$\chi^2 = n\sum_{i=1}^{k}\frac{(p_i{*} - p_i)^2}{p_i}. \quad (26)$$

При цьому передбачається, що $\sum p_i = 1$. Тому при розрахунку p_i необхідно враховувати область визначення вибраного теоретичного розподілу. Так, наприклад, для нормального розподілу початком першого інтервалу є – ∞, а кінцем останнього – +∞.

Величина χ^2 – зростає при збільшенні відмінностей між p_i і p_i*, а також із збільшенням *n*, що цілком логічно, оскільки при великому об'ємі вибірки навіть невелика розбіжність між p_i і p_i* повинна насторожувати.

Наведене вище співвідношення можна записати в більш зручному для розрахунків вигляді:

$$\chi^2 = n\sum_{i=1}^{k}\frac{(n_i - n \cdot p_i)^2}{n \cdot p_i}. \quad (27)$$

Пірсон показав, що розподіл χ^2 не залежить від виду теоретичного розподілу ВВ. Проте, складові додатки в наведеному вище виразі залежні.

Цей факт враховується числом ступенів вільності, яке відображає число незалежних параметрів. Число ступенів вільності z визначається за формулою:

$$z = k - s - 1, \quad (28)$$

де s – число зв’язків, накладених на вибірку і теоретичний закон розподілу.

Як правило, s – це прирівнювання одного або двох моментів розподілу.

Якщо прирівняні середні значення і дисперсії, то $s = 2$; якщо тільки середні значення або дисперсії – $s = 1$. Віднімання одиниці в наведеній вище формулі відображає завжди зв’язок, що існує між n_i та $(n \cdot p_i)$:

$$n_1 + n_2 + ... + n_k = n \cdot p_1 + n \cdot p_2 + ... + n \cdot p_k \,, \qquad (29)$$

або

$$\sum_{i=1}^{k} p_i * = \int_0^{\infty} f(x)dx = 1 \,. \qquad (30)$$

Таблиця χ^2-розподілу в залежності від числа ступенів вільності наведена в Додатку Б.

Таким чином, застосування критерію χ^2 зводиться до наступного. З групованої вибірки знаходяться частоти n_i – (вони повинні бути не менше 6). За вибраним теоретичним розподілом обчислюються теоретичні імовірності p_i за допомогою формули (24). Далі визначається число ступенів вільності z. З таблиці за числом ступенів вільності і заданому рівню значущості p знаходиться $\chi^2_{кр}$ За співвідношенням (27) обчислюється розрахунковий критерій χ^2_p. Якщо $\chi^2_p \le \chi^2_{кр}$, то гіпотеза про тип вибраного теоретичного розподілу приймається, якщо $\chi^2_p > \chi^2_{кр}$ – відкидається.

Обчислення зручно оформити у вигляді таблиці 2 при $k = 8$ і $z = 5$.

$$\widetilde{x} = \frac{\sum_{i=1}^{k} (\overline{x}_i' \cdot n_i)}{N} = \frac{\sum_{i=1}^{k} (\overline{x}_i' \cdot n_i)}{\sum_{i=1}^{k} n_i} \,. \qquad (31)$$

$$\widetilde{x} = \frac{754{,}00}{100} = 7{,}54\,.$$

$$D = \frac{\sum_{i=1}^{k} (\overline{x}'_i - \widetilde{x})^2 \cdot n_i}{N} = \frac{\sum_{i=1}^{k} (\overline{x}'_i - \widetilde{x})^2 \cdot n_i}{\sum_{i=1}^{k} n_i}\,. \tag{32}$$

$$D = \frac{97{,}09}{100} = 0{,}9709\,.$$

$$\sigma = \sqrt{\frac{\sum_{i=1}^{k} (\overline{x}'_i - \widetilde{x})^2 \cdot n_i}{N}} = \sqrt{\frac{\sum_{i=1}^{k} (\overline{x}'_i - \widetilde{x})^2 \cdot n_i}{\sum_{i=1}^{k} n_i}}\,. \tag{33}$$

$$\sigma = \sqrt{\frac{97{,}09}{100}} = 0{,}9853\,.$$

$$N \cdot \frac{j}{\sigma} = \sum_{i=1}^{k} n_i \cdot \frac{j}{\sigma}\,, \tag{34}$$

де j – ширина інтервалу.

$$j = x'_{i,\max} - x'_{i,\min}\,. \tag{35}$$

Виконуючи розрахунок для першого рядочку таблиці 2, тобто $i = 1$, будемо мати:

$$N \cdot \frac{j}{\sigma} = 100 \cdot \frac{5{,}5 - 5{,}0}{0{,}9853} = 50{,}75\,.$$

$$n' = N \cdot \frac{j}{\sigma} \cdot f_i(t) \tag{36}$$

$$n' = 50{,}75 \cdot 0{,}0270 = 1{,}37 \approx 1\,.$$

Оскільки, частота теоретичної n' кривої нормального закону розподілу та експериментальних даних n не співпадають, то їх дисперсії не можуть дорівнювати одна одній, відповідно, вважаємо що існує зв'язок між їх абсолютними значеннями, тоді $s = 1$, а $z=k-s-1=8-1-1=6$.

Відповідно, фактичне значення критерію Пірсона дорівнює $\chi^2_{ф} = 19{,}73$, а табличні значення становлять:

– для довірчої імовірності $\alpha = 95\%$ (рівень значущості $p = 0{,}05$) і ступені вільності $z = 6$ – $\chi^2_{табл}(z = 6; p = 0{,}05) = 12{,}6$, тобто $\chi^2_{ф} = 19{,}73 > \chi^2_{табл} = 12{,}6$, що свідчить про необхідність спростування гіпотези щодо правильності вибраного типу теоретичного розподілу;

– для довірчої імовірності $\alpha = 99{,}9\%$ (рівень значущості $p = 0{,}001$) і ступені вільності $z = 6$ – $\chi^2_{табл}(z = 6; p = 0{,}001) = 22{,}5$, тобто $\chi^2_{ф} = 19{,}73 < \chi^2_{табл} = 22{,}5$, що свідчить про прийняття гіпотези щодо типу вибраного теоретичного розподілу. Таким чином, можна вважати доведеним, що з імовірністю 99,9% розподіл значень частоти n відповідає нормальному закону розподілу.

Якщо розглянути згладжування фактичного розподілу абсолютної маси насіння, тобто маси 1000 насінин (значень частоти n) в залежності від їх довжини (представниками інтервалу $\overline{x_i'}$) за допомогою теоретичної кривої нормального розподілу, то на графіку можна простежити достатню близькість кривих розподілу фактичних частот до теоретичних, що свідчить про правильність вибраного типу розподілу. Такого самого висновку можна дійти і за допомогою інших критеріїв узгодження.

Таблиця 2 - Обчислення критерію Пірсона

Номер інтервалу k	Межі інтервалу	Представник інтервалу $\overline{x}'_i$	Частота n_i	$\overline{x}'_i \cdot n_i$	$\lvert\overline{x}'_i - \tilde{\overline{x}}\rvert$	$(\lvert\overline{x}'_i - \tilde{\overline{x}}\rvert)^2$	$(\lvert\overline{x}'_i - \tilde{\overline{x}}\rvert)^2 \cdot n_i$	$t = \frac{\lvert\overline{x}'_i - \tilde{\overline{x}}\rvert}{\sigma}$	$p_i = f(t)$	Частота теоретичної кривої нормального закону розподілу n_i'	$n_i - n_i'$	$(n_i - n_i')^2$	$\frac{(n_i - n_i')^2}{n_i'}$
1	5,0-5,5	5,25	2	10,50	2,29	5,2441	10,4882	2,3242	0,0270	1	1	1	1
2	5,5-6,0	5,75	7	40,25	1,79	3,2041	22,4287	1,8167	0,0761	4	3	9	2,25
3	6,0-6,5	6,25	8	50,00	1,29	1,6641	13,3128	1,3092	0,1691	9	-1	1	0,11111
4	6,5-7,0	6,75	13	87,75	0,79	0,6241	8,1133	0,8018	0,2897	15	-2	4	0,26667
5	7,0-7,5	7,25	15	108,75	0,29	0,0841	1,2615	0,2943	0,3825	19	-4	16	0,84211
6	7,5-8,0	7,75	16	124,00	0,21	0,0441	0,7056	0,2131	0,3902	20	-4	16	0,8
7	8,0-8,5	8,25	17	140,25	0,71	0,5041	8,5697	0,7206	0,3079	16	1	1	0,0625
8	8,5-9,0	8,75	22	192,50	1,21	1,4641	32,2102	1,2281	0,1872	10	12	144	14,4
Сума			$N =$ =100	$\sum_{i=1}^{k}(\overline{x}'_i \cdot n_i) =$ = 754,00	–	12,8328	97,0900	–	–	94	6	36	19,73238

3.2 Критерій Колмогорова

Припустимо, що $F(x)$ – теоретична ФР, а $F^*(x)$ – статистична функція розподілу. В якості міри розбіжності між ними О.М. Колмогоров запропонував використовувати максимальне значення модуля їх різниці:

$$D = \max\left|F(x) - F^*(x)\right| \tag{37}$$

Приклад геометричної інтерпретації наведений на рисунку 5.

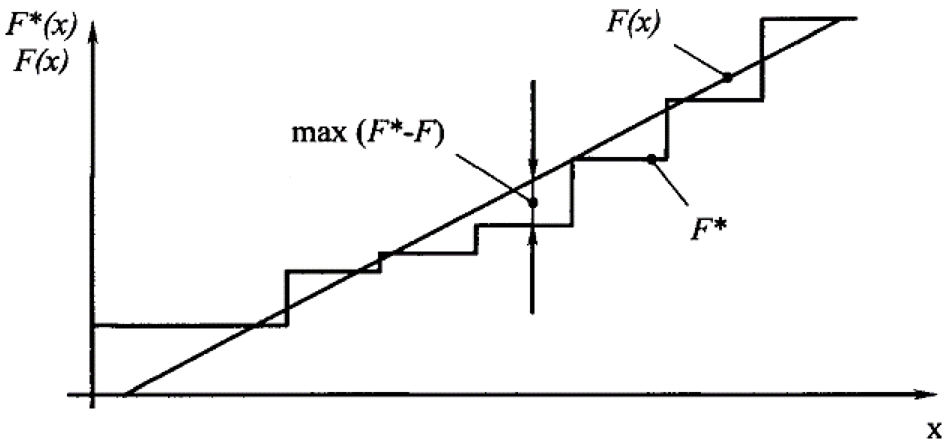

Рис. 5 – Геометрична інтерпретація критерію Колмогорова

Якщо гіпотеза про вид закону розподілу справедлива, то розподіл випадкової величини D, при $n \to \infty$, спрямований до деякого граничного розподілу, не залежного від розподілу $F(x)$, а імовірність нерівності:

$$D\sqrt{n} > \lambda \tag{38}$$

спрямовується до межі:

$$p(\lambda) = 1 - \sum_{i=-\infty}^{\infty} (-1)^i \cdot e^{-2 \cdot i \cdot \lambda^2}. \tag{39}$$

Існують таблиці розподілу Колмогорова.

Методика застосування критерію Колмогорова дуже проста. Будуються $F(x)$ і $F^*(x)$. Визначається максимум модуля різниці D між ними. За заданим рівнем значущості $p(\lambda)$ знаходиться критичне значення $\lambda_{кр}$, яке порівнюється із значенням:

$$\lambda = D\sqrt{n}\ , \tag{40}$$

де D – абсолютна максимальна різниця кумулятивних часток або частот емпіричного і теоретичного розподілів;

n – число спостережень (кількість одиниць сукупності).

Якщо розподіл задано в частотах, то:

$$\lambda = \frac{D}{\sqrt{n}}. \tag{41}$$

Гіпотеза про відповідність експериментальних даних вибраному теоретичному розподілу приймається, якщо $\lambda \le \lambda_{кр}$, і відкидається в протилежному випадку.

Слід, проте, відзначити, що розподіл Колмогорова знайдений в припущенні, що параметри теоретичного розподілу вибрані наперед, без використання вибірки, тобто без накладання додаткових зв’язків. Якщо ж такі зв’язки є, тобто якщо вибірка при визначенні параметрів закону використовувалася, то застосування критерію дає завищене значення імовірності $p(\lambda)$. Тому підвищується ризик прийняти як правдоподібну гіпотезу, насправді таку, що погано узгоджується з вибіркою. Тому, якщо параметри теоретичного розподілу невідомі наперед і не можуть бути визначені з фізичної природи досліджуваного процесу, використання критерію Колмогорова не доцільно.

4. Порядок виконання роботи

1. Відповідно до індивідуального завдання необхідно провести експеримент в результаті якого отримати вибірку.

2. Побудувати варіаційний ряд.

3. Побудувати груповану вибірку з числом інтервалів 5-10.

4. Побудувати гістограму і полігон, згруповану і негруповану функції.

5. За групованою вибіркою знайти оцінки математичного очікування і середнєквадратичного відхилення.

6. Вибрати в якості теоретичного один із законів розподілу.

7. Знайти параметри теоретичного закону розподілу. При цьому, якщо область визначення закону обмежена, то в якості оцінки кінцевої межі використовувати мінімальне (максимальне) значення вибірки. Решту параметрів знаходити по методу моментів. Побудувати на одному графіку теоретичний розподіл спільно з гістограмою.

8. За заданим рівнем значущості і числом ступенів вільності знайти критичне значення $\chi^2_{кр}$.

9. Перевірити гіпотезу про теоретичний розподіл по критерію $\chi^2_{кр}$. Якщо гіпотеза виявилася знехтуваною, перевірити інший розподіл.

5. Зміст звіту

1. Навести таблиці вибірки і варіаційного ряду.

2. Побудувати графіки полігону, гістограми, групованої і негрупованої статистичних функцій розподілу.

3. Навести таблицю перевірки відповідності вибраної функції теоретичного розподілу даним експерименту за критерієм χ^2.

4. Накреслити графіки теоретичних функцій $F(x)$ і $f(x)$ спільно, відповідно, з гістограмою і негрупованою статистичною функцією розподілу.

5. Зробити висновки щодо виконаної роботи.

6. Контрольні запитання для самоперевірки

1. Що називається випадковою величиною?
2. Що називається функцією розподілу і щільністю розподілу імовірності випадкової величини?
3. Як визначають оцінку математичного очікування і дисперсії випадкової величини?
4. Що називається простою вибіркою і варіаційним рядом?
5. Що називається групованою вибіркою?
6. Що являє собою статистична функція розподілу?
7. Як побудувати гістограму?
8. Як побудувати полігон?
9. Що називається гістограмою?
10. Як визначити параметри теоретичного розподілу за методом моментів?
11. Що називається похибками першого і другого роду?
12. Що таке рівень значущості?
13. Для чого застосовується критерій χ^2?
14. Як визначити число ступенів свободи для критерію χ^2?
15. Для чого застосовується критерій Колмогорова?

Навчальне видання

ВИБІР ЗАКОНУ РОЗПОДІЛУ ВИПАДКОВОЇ ВЕЛИЧИНИ

Методичні вказівки
до виконання практичної роботи з навчальної дисципліни
«Моделювання технологічних процесів та систем»

Укладач:

КИРИЧЕНКО Роман Васильович

Формат 60x84/16. Гарнітура Times New Roman.
Папір для цифрового друку. Друк ризографічний.
Ум. друк. арк. 0,96.
Тираж 100 пр.
Державний біотехнологічний університет.
61002, м. Харків, вул. Алчевських, 44.