ПРАКТИЧНЕ ЗАНЯТТЯ 3

Тема 6. Випадкові величини

Означення. Розглянемо деяку функцію ξ, яка діє на множині елементарних подій Ω та приймає дійсні значення:

$$\xi : \Omega \to \mathbb{R},$$

тобто кожній випадковій події ця функція ставить у відповідність деяке дійсне число. Таку функцію називають *випадковою величиною*.

Дискретні випадкові величини

Означення. *Випадкова величина* ξ називається *дискретною*, якщо всі її можливі значення можна занумерувати натуральними числами.

Можна використовувати ще й наступне означення.

Означення. *Дискретною* називається *випадкова величина*, можливі значення якої є ізольовані числа.

Кожному з чисел $\xi_i, i = \overline{1,n}$ можна зіставити ймовірність $p_i, i = \overline{1,n}$, яка дорівнює ймовірності появи випадкової події, відповідної цьому числу:

$$p_i = P(\xi = \xi_i) = P(\omega_i).$$

Таблиця розподілу дискретної випадкової величини

Значення ξ_i вип. вел. ξ	ξ_1	ξ_2	$\ldots$	ξ_{n-1}	ξ_n
Ймовірність $p_i = P_\xi(\xi_i)$	p_1	p_2	$\ldots$	p_{n-1}	p_n

Зауваження. Властивість $\sum\limits_{i=1}^{n} p_i = 1$ часто називається *умовою нормування.*

Закон розподілу ймовірності дискретної випадкової величини можна також задати *графічно. Графічне представлення* закону називається *багатокутником розподілу ймовірностей.* Для графічного представлення необхідно в декартовій системі координат відмітити точки (ξ_i, p_i) та з'єднати їх відрізками, в результаті чого отримаємо ламану, яку називають багатокутником розподілу.

Означення. *Функцією розподілу ймовірностей* дискретної випадкової величини називається

$$F_\xi(x) = P(\xi < x)$$

(тобто ймовірність того, що випадкова величина набуде значення від $-\infty$ до x не включно).

Приклад 1. Стрілок робить 3 постріли у мішень. Ймовірність влучити у мішень при кожному пострілі дорівнює 0,4. За кожне попадання стрілку зараховується 5 очок. Треба побудувати таблицю розподілу числа отриманих очок та багатокутник розподілу. Визначити функцію розподілу ймовірностей дискретної випадкової величини та побудувати її графік.

Розв'язання

Позначимо через випадкову величину ξ – число зарахованих очок, відповідно вона може приймати наступні значення:

$$\xi_1 = 0, \xi_2 = 5, \xi_3 = 10, \xi_4 = 15.$$

Обчислимо відповідні кожному з $\xi_i, i = \overline{1,4}$ ймовірності $p_i, i = \overline{1,4}$ *за формулою Бернуллі* $p_n(m) = C_n^m \cdot p^m \cdot q^{n-m}$, де $n = 3$, $m = \overline{0,3}$, $p = 0,4$, $q = 0,6$:

$$p_1 = P(\xi = 0) = C_3^0 \cdot 0,4^0 \cdot 0,6^3 = 0,216;$$

$$p_2 = P(\xi = 5) = C_3^1 \cdot 0{,}4^1 \cdot 0{,}6^2 = 0{,}432;$$

$$p_3 = P(\xi = 10) = C_3^2 \cdot 0{,}4^2 \cdot 0{,}6^1 = 0{,}288;$$

$$p_4 = P(\xi = 15) = C_3^3 \cdot 0{,}4^3 \cdot 0{,}6^0 = 0{,}064.$$

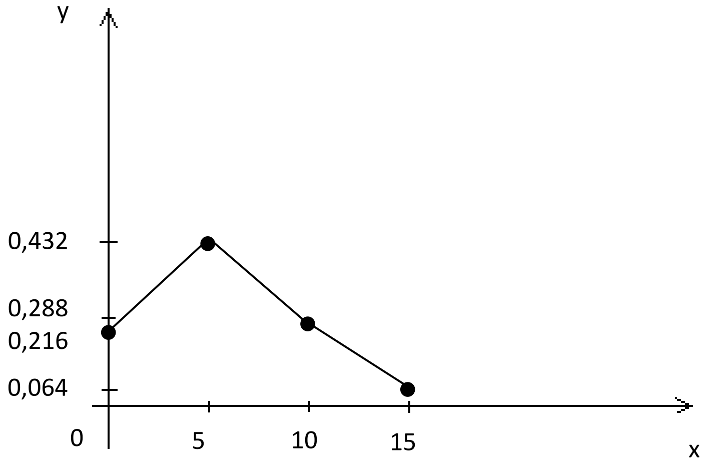
Побудуємо таблицю розподілу дискретної випадкової величини ξ – числа зарахованих очок.

ξ_i	0	5	10	15
p_i	0,216	0,432	0,288	0,064

Перевіримо виконання властивості $\sum_{i=1}^{4} p_i = 1$ (*умови нормування*):

$$\sum_{i=1}^{4} p_i = 0{,}216 + 0{,}432 + 0{,}288 + 0{,}064 = 1.$$

Представимо закон розподілу ймовірності дискретної випадкової величини *графічно* у вигляді *багатокутнику розподілу ймовірностей.* Для графічного представлення необхідно в декартовій системі координат відмітити точки (ξ_i, p_i) та з'єднати їх відрізками:

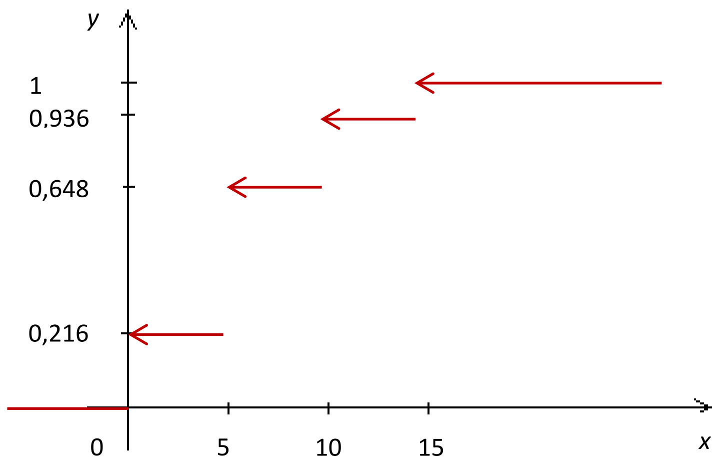


Визначимо *функцію розподілу ймовірностей* $F_{\xi}(x) = P(\xi < x)$:

$$F_{\xi}(x) = P(\xi < x) = \begin{cases} 0, \ \textit{при} \ x \le 0; \\ 0{,}216, \ \textit{при} \ 0 < x \le 5; \\ 0{,}216 + 0{,}432, \ \textit{при} \ 5 < x \le 10; \\ 0{,}216 + 0{,}432 + 0{,}288, \ \textit{при} \ 10 < x \le 15; \\ 0{,}216 + 0{,}432 + 0{,}288 + 0{,}064, \ \textit{при} \ x > 15, \end{cases} =$$

$$= \begin{cases} 0, \ \textit{при} \ x \le 0; \\ 0{,}216, \ \textit{при} \ 0 < x \le 5; \\ 0{,}648, \ \textit{при} \ 5 < x \le 10; \\ 0{,}936, \ \textit{при} \ 10 < x \le 15; \\ 1, \ \textit{при} \ x > 15. \end{cases}$$

Після чого побудуємо її графік:



Задачі для практичного заняття та самостійної роботи

6.1. Стрілок має 4 патрони та робить постріли до першого влучення у мішень. Ймовірність влучити у мішень при кожному пострілі дорівнює 0,6.

Треба побудувати таблицю розподілу боєзапасу, що залишається невитраченим, та багатокутник розподілу. Визначити функцію розподілу ймовірностей дискретної випадкової величини та побудувати її графік.

6.2. Серед 10 лотерейних білетів є 4 виграшних. На удачу купують 2 білети. Треба побудувати таблицю розподілу числа виграшних білетів серед куплених та багатокутник розподілу. Визначити функцію розподілу ймовірностей дискретної випадкової величини та побудувати її графік.

6.3. В партії з 25 шкіряних курток 5 мають прихований дефект. На удачу купують 3 куртки. Треба побудувати таблицю розподілу числа дефектних курток серед куплених та багатокутник розподілу. Визначити функцію розподілу ймовірностей дискретної випадкової величини та побудувати її графік.

6.4. Пристрій складається з 3 незалежно працюючих елементів. Ймовірність відмови кожного елементу дорівнює 0,15. Треба побудувати таблицю розподілу числа елементів, що відмовили, та багатокутник розподілу. Визначити функцію розподілу ймовірностей дискретної випадкової величини та побудувати її графік.

6.5. Баскетболіст робить 3 штрафних кидка. Ймовірність влучення при кожному кидку дорівнює 0,7. Треба побудувати таблицю розподілу числа влучень м'яча та багатокутник розподілу. Визначити функцію розподілу ймовірностей дискретної випадкової величини та побудувати її графік.

Основні числові характеристики дискретної випадкової величини

Означення. *Математичним сподіванням* або *середнім* $M\xi$ (або $M(\xi)$) дискретної випадкової величини ξ називається

$$M\xi = \xi_1 \cdot p_1 + \xi_2 \cdot p_2 + \ldots + \xi_n \cdot p_n = \sum_{i=1}^{n} \xi_i \cdot p_i.$$

Властивості математичного сподівання

1. $M(C) = C$, де $C = const \in \mathbb{R}$.
2. $M(C \cdot \xi) = C \cdot M(\xi)$, де $C = const \in \mathbb{R}$.
3. Якщо ξ та η незалежні випадкові величини, то $M(\xi \cdot \eta) = M(\xi) \cdot M(\eta)$.
4. Якщо ξ та η незалежні випадкові величини, то $M(\xi + \eta) = M(\xi) + M(\eta)$.

Означення. *Центральним моментом порядку* $n \in \mathbb{N}$ дискретної випадкової величини ξ називається математичне сподівання $M((\xi - M\xi)^n)$, якщо воно існує.

Означення. *Дисперсією* дискретної випадкової величини ξ називається центральний момент порядку 2:

$$D\xi = M((\xi - M\xi)^2).$$

Ця формула не зручна для використання на практиці, найчастіше використовують формулу, яка є її наслідком:

$$D\xi = M(\xi^2) - (M\xi)^2.$$

Дисперсія випадкової величини ξ – це випадкова величина, яка характеризує величину відхилення випадкової величини ξ від її середнього значення $M\xi$.

Властивості дисперсії

1. $D(C) = 0$, де $C = const \in \mathbb{R}$.
2. $D(C \cdot \xi) = C^2 \cdot M(\xi)$, де $C = const \in \mathbb{R}$.
3. Якщо ξ та η незалежні випадкові величини, то $D(\xi + \eta) = D(\xi) + D(\eta)$.

Означення. Для отримання характеристики розсіяння випадкової величини ξ від її середнього значення $M\xi$ використовують ще *середнє квадратичне відхилення*, яке визначається за формулою:

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

Означення. *Модою* $(Mo(\xi))$ називають те значення ξ_i випадкової величини ξ, ймовірність p_i появи якого є максимальною.

Означення. *Медіаною* $(Me(\xi))$ називають те значення ξ_i, для якого ймовірність появи випадкової величини, меншої або більшої за це значення, є однаковою:

$$P(\xi < Me(\xi)) = P(\xi > Me(\xi)).$$

Приклад 2. Обчислити математичне сподівання, дисперсію та середнє квадратичне відхилення дискретної випадкової величини з Прикладу 1.

Розв'язання

Для обчислення числових характеристик розглянемо таблицю розподілу дискретної випадкової величини ξ.

ξ_i	0	5	10	15
p_i	0,216	0,423	0,288	0,064

За формулою обчислення математичного сподівання $M\xi$

$$M\xi = \xi_1 \cdot p_1 + \xi_2 \cdot p_2 + \ldots + \xi_n \cdot p_n = \sum_{i=1}^{n} \xi_i \cdot p_i.$$

У нашому випадку:

$$M\xi = \xi_1 \cdot p_1 + \xi_2 \cdot p_2 + \xi_3 \cdot p_3 + \xi_4 \cdot p_4 =$$

$$= 0 \cdot 0{,}216 + 5 \cdot 0{,}423 + 10 \cdot 0{,}288 + 15 \cdot 0{,}064 = 5{,}955.$$

Для обчислення дисперсії скористуємось формулою $D\xi = M(\xi^2) - (M\xi)^2$.

Обчислимо спочатку $M(\xi^2)$, для зручності запишемо таблицю розподілу дискретної випадкової величини ξ^2:

ξ_i^2	0	25	100	225
p_i	0,216	0,423	0,288	0,064

Звідки:

$$M(\xi^2) == 0 \cdot 0{,}216 + 25 \cdot 0{,}423 + 100 \cdot 0{,}288 + 225 \cdot 0{,}064 = 54.$$

Також для обчислення дисперсії знайдемо

$$(M\xi)^2 = (5{,}955)^2 = 35{,}462025.$$

Підставимо отримані значення у формулу для знаходження дисперсії:

$$D\xi = M(\xi^2) - (M\xi)^2 = 54 - 35{,}462025 \approx 18{,}54.$$

І нарешті знайдемо середнє квадратичне відхилення дискретної випадкової величини за формулою:

$$\sigma(\xi) = \sqrt{D\xi} = \sqrt{18{,}54} \approx 4{,}31.$$

6.6. Обчислити математичне сподівання, дисперсію та середнє квадратичне відхилення дискретної випадкової величини для задач 6.1 – 6.5.

Тема 7. Первинне опрацювання статистичних даних

Генеральна та вибіркова сукупності

Нехай потрібно вивчити сукупність однорідних об’єктів відносно деякої якісної або кількісної ознаки, яка характеризує ці об’єкти. Наприклад, якщо досліджується партія товарів, то якісною ознакою може слугувати стандартність товару, я кількісною – наприклад, розмір товару.

Іноді проводять повне дослідження, тобто аналізується кожний об'єкт сукупності відносно ознаки, якою цікавляться. Проте на практиці, таке повне дослідження проводиться відносно рідко, і найчастіше із сукупності обирається та досліджується лише якась контрольна частина об'єктів у обмеженій кількості. Введемо відповідні означення.

Означення. Сукупність всіх можливих об'єктів даного виду, над якими проводяться спостереження, або сукупність всіх можливих спостережень, які проводяться в незмінних умовах над деякою випадковою величиною, називається *генеральною сукупністю.*

Генеральна сукупність може містити скінченну або нескінченну кількість елементів.

Означення. Відібрані з генеральної сукупності об'єкти (або результати спостережень) називаються *вибірковою сукупністю* або просто *вибіркою.*

Означення. Число N елементів *генеральної сукупності* та число n елементів *вибіркової сукупності* будемо називати відповідно *об'ємами генеральної* та *вибіркової сукупностей*.

Повторна та безповторна вибірки. Репрезентативна вибірка

При складанні вибірки можна діяти двома способами: після того, як об'єкт був відібраний та над ним було проведене спостереження, він може бути повернутий або не повернутий у генеральну сукупність. Відповідно до цього вибірки розділяють на *повторні* та *безповторні.*

Означення. *Повторною* називають *вибірку*, при якій відібраний об'єкт (перед відібранням наступного) повертають до генеральної сукупності.

Означення. *Безповторною* називають *вибірку*, при якій відібраний об'єкт не повертають до генеральної сукупності.

На практиці зазвичай користуються безповторним випадковим відбором.

Для того щоб за даними вибірки можна було достатньо впевнено судити про ознаку генеральної сукупності, яка нас цікавить, необхідно, щоб об'єкти вибірки правильно її представляли. Цю вимогу формулюють так: вибірка повинна бути *репрезентативною* (або *представницькою*). Вважається, що якщо кожний об'єкт вибірки відібраний із генеральної сукупності випадково, тобто всі об'єкти мають однакову ймовірність потрапити у вибірку, і кількість об'єктів, відібраних для спостереження, є досить великою, то вибірка буде репрезентативною. Різниця між показниками вибіркової та генеральної сукупностей становить *помилку репрезентативності*. Ці помилки виникають тому, що вибіркова сукупність неточно відображає генеральну сукупність.

Означення. Вибірка називається *репрезентативною*, якщо вона достатньо добре відтворює генеральну сукупність.

Це означення не дозоляє робити конкретні виводи, бо не вказана загальна міра відповідності між репрезентативною вибіркою та генеральною сукупністю, тому питання про репрезентативність необхідно вирішувати у конкретних задачах, спираючись на конкретні критерії відповідності.

Дискретний варіаційний ряд розподілу

Нехай із генеральної сукупності проведена вибірка об'єму *n*, і досліджувана випадкова величина в цій вибірці прийняла різні *k* значень $(x_1, x_2, \ldots, x_k)$ – ці елементи називають *варіантами*:

$$(\underbrace{\underbrace{x_1,\ldots,x_1}_{n_1\ \text{разів}},\underbrace{x_2,\ldots,x_2}_{n_2\ \text{разів}},\ldots,\underbrace{x_k,\ldots,x_k}_{n_k\ \text{разів}}}_{n\ \text{разів}})$$

Нехай у вибірці $(x_1, x_2, \ldots, x_k)$ варіанта x_1 спостерігалась n_1 разів, x_2 спостерігалась n_2 разів, і так далі відповідно до x_k, яка спостерігалась n_k разів, тобто з того, що об'єм вибірки дорівнює *n*, випливає:

$$n_1 + n_2 + \ldots + n_k = \sum_{i=1}^{k} n_i = n.$$

Означення. Значення $n_1, n_2, \ldots, n_k$, тобто чисельність окремої групи згрупованого ряду вибірки, називають *частотами* або *вагами варіант*. Відношення n_i до загального об'єму вибірки n, називається *відносною частотою варіанти* x_i та позначається:

$$p_i^* = \frac{n_i}{n}.$$

З означення відносної частоти випливає:

$$p_1^* + p_2^* + \ldots + p_k^* = \sum_{i=1}^{k} p_i^* = 1.$$

Означення. Якщо значення випадкової величини, яке відповідає окремій групі згрупованого ряду даних, називається *варіантою*, то змінення цього значення – *варіюванням*.

Означення. Розташування вибіркових спостережень значень випадкової величини в порядку неспадання називається *ранжуванням*.

Означення. Різниця між максимальним та мінімальним значеннями варіант, тобто інтервал варіювання, називається *розмахом вибірки* та позначається:

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

Означення. *Дискретним варіаційним рядом розподілу* (або *дискретним статистичним розподілом,* або *розподілом частот*) називається ранжована сукупність варіант x_i з відповідними частотами або відносними частотами.

У *табличній формі* він має такий вигляд:

x_i	x_1	x_2	x_3	$\ldots$	x_k
n_i	n_1	n_2	n_3	$\ldots$	n_k
p_i^*	p_1^*	p_2^*	p_3^*	$\ldots$	p_k^*

Приклад. Проводяться спостереження над значеннями грошових виграшів у миттєвій лотереї. У результаті отримані наступні значення (у тис. грн.):

0, 1, 0, 0, 5, 0, 10, 0, 1, 0, 0, 1, 5, 1, 0, 0, 0, 1, 0, 1, 0, 0, 0, 0, 5, 0, 5, 0, 0, 1, 1, 1, 5, 10, 0, 1, 1, 0, 5, 0, 0, 0, 0, 1, 0, 1, 0, 5, 0, 0, 0, 0, 1, 0.

Складемо варіаційний ряд розподілу у табличній формі, для цього визначимо які різні значення прийняла досліджувана випадкова величина в цій вибірці, тобто визначимо варіанти, обрахуємо загальний об’єм вибірки *n,* відповідні частоти та відносні частоти варіант:

x_i	0	1	5	10
n_i	31	14	7	2
p_i^*	$\frac{31}{54}$	$\frac{14}{54}$	$\frac{7}{54}$	$\frac{2}{54}$

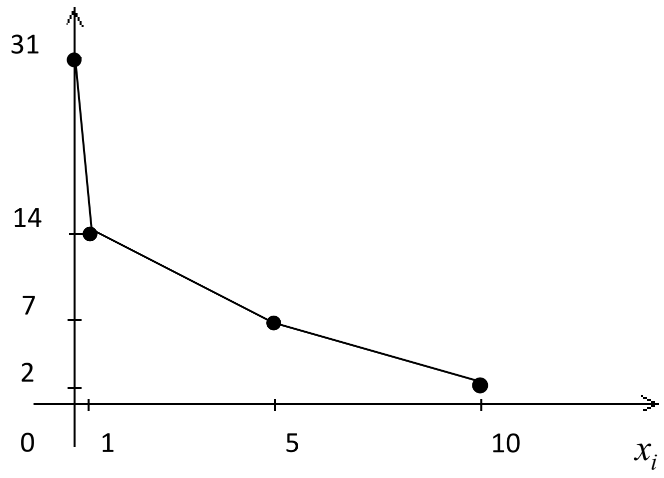
$$n = 31 + 14 + 7 + 2 = 54,$$

$$\sum_{i=1}^{k} p_i^* = 1.$$

Для графічного представлення дискретного варіаційного ряду розподілу будують *полігон (або багатокутник) частот* та *полігон (або багатокутник) відносних частот*.

Означення. Ламана лінія, відрізки якої послідовно з’єднують точки з координатами (x_i, n_i), $i = \overline{1,k}$, називається *полігоном частот*, а ламана лінія, відрізки якої послідовно з’єднують точки з координатами (x_i, p_i^*), $i = \overline{1,k}$, називається *полігоном відносних частот.*

Приклад. Побудуємо полігон частот для попереднього прикладу. Для цього відмітимо на декартовій площині точки з координатами (x_i, n_i), $i = \overline{1,k}$, які у нашому випадку будуть дорівнювати: (0,31), (1,14), (5,7), (10,2), та з’єднаємо їх послідовно відрізками прямих.



Зауважимо, що полігон, побудований за дискретним варіаційним рядом, є вибірковим аналогом багатокутника розподілу дискретної випадкової величини.

Емпірична функція розподілу

Означення. *Вибірковою (емпіричною) функцією розподілу* називається функція $F^*(x)$, яка задає для кожного значення *x* відносну частоту події $X < x$. Отже, за означенням $F^*(x) = \frac{n_x}{n}$, де n_x – число вибіркових значень величини *X*, менших за *x*, а *n* – об’єм вибірки.

Вибіркову функцію можна задати у табличному та у графічному виді. Розберемо це на прикладі.

Приклад. Побудуємо вибіркову функцію розподілу для розглянутого вище прикладу.

Об’єм вибірки $n = 54$, найменша варіанта дорівнює 0, отже $n_x = 0$ при $x \le 0$, а $F^*(x) = \frac{n_x}{n} = \frac{0}{54}$, при $x \le 0$.

При $0 < x \le 1$ нерівність $X < x$ виконується для варіанти $x_1 = 0$, і відповідно $n_x = n_1 = 31$, а $F^*(x) = \frac{n_x}{n} = \frac{31}{54} = p_1^*$.

При $1 < x \le 5$ нерівність $X < x$ виконується для варіант $x_1 = 0$ та $x_2 = 1$, і відповідно $n_x = n_1 + n_2 = 31 + 14 = 45$, а $F^*(x) = \frac{n_x}{n} = \frac{45}{54} = p_1^* + p_2^*$.

При $5 < x \le 10$ нерівність $X < x$ виконується для варіант $x_1 = 0$, $x_2 = 1$ та $x_3 = 5$, і відповідно $n_x = n_1 + n_2 + n_3 = 31 + 14 + 7 = 52$, а $F^*(x) = \frac{n_x}{n} = \frac{52}{54} = p_1^* + p_2^* + p_3^*$.

При $10 < x$ нерівність $X < x$ виконується для варіант $x_1 = 0$, $x_2 = 1$, $x_3 = 5$ та $x_4 = 10$, і відповідно $n_x = n_1 + n_2 + n_3 + n_4 = 31 + 14 + 7 + 2 = 54$, а $F^*(x) = \frac{n_x}{n} = \frac{54}{54} = 1 = p_1^* + p_2^* + p_3^* + p_4^*$.

Результати обчислень занесемо у таблицю (для зручності).

x	$F^*(x)$
$x \le 0$	0
$0 < x \le 1$	$p_1^* = \frac{31}{54}$
$1 < x \le 5$	$p_1^* + p_2^* = \frac{45}{54}$
$5 < x \le 10$	$p_1^* + p_2^* + p_3^* = \frac{52}{54}$
$10 < x$	$p_1^* + p_2^* + p_3^* + p_4^* = 1$

Можна замість таблиці записувати вибіркову функцію розподілу так само, як записували функцію розподілу дискретної випадкової величини.

Побудуємо графік вибіркової функції розподілу.