**Probability Theory**

**Chapter 2**

**Discrete probability distributions.**

Lecturer: Nataliia Kruglova.

# Лекція 7. Дискретні ймовірнісні розподіли.

### Розподіл Бернуллі

Випробуваннями Бернуллі називають стохастичний експеримент, що може закінчитися тільки одним із двох результатів: "успіхом"$(A)$ чи "невдачею"$(\bar{A})$.
Наприклад, може бути успіхом випадіння герба при підкиданні монети, випадіння шістки при підкиданні грального кубика. Простором елементарних подій буде: $\Omega = \{A, \bar{A}\}$.
Введемо ймовірнісну міру:

$$P(A) = p, \qquad P(\bar{A}) = 1 - p, \qquad 0 < p < 1.$$

Випадкова величина буде визначатися таким чином: $X(\omega) = I_A(\omega) = \begin{cases} 1, & \omega \in A, \\ 0, & \omega \in \bar{A}. \end{cases}$

Зведемо розподіл у таблицю:

| | | |
|---|---|---|
| X | 0 | 1 |
| P | 1-p | p |

Запишемо функцію розподілу:

$$F(x) = \begin{cases} 0, & x \le 0, \\ 1 - p, & x \in (0, 1], \\ 1, & x > 1. \end{cases}$$

Знайдемо моменти випадкової величини $X$:

$$m_1 = E[X] = 0 \cdot (1-p) + 1 \cdot p = p,$$

$$m_2 = E[X^2] = 0 \cdot (1-p) + 1 \cdot p = p,$$

$$\mu_2 = D(X) = 0 \cdot (1-p) + 1 \cdot p - p^2 = p(1-p).$$

Випадкову величину із розподілом Бернуллі ще називають "індикаторною" - її значення можна використовувати для визначення, чи відбулась деяка подія A, чи ні. Подією A може бути, наприклад, чи одержить людина матеріальну допомогу на роботі протягом певного періоду.

**Біноміальний закон розподілу.**

Проводиться $n$ незалежних випробувань за схемою Бернуллі і $p = P(A)$ — імовірність появи події $A$ в кожному випробуванні. Потрібно визначити закон розподілу дискретної випадкової величини $X$ — числа появ події $A$ в $n$ випробуваннях.

Випадкова величина $X$ може приймати наступні значення: $x_0 = 0$, $x_1 = 1$, $x_2 = 2$, …, $x_n = n$. Імовірності цих значень $x_k$ випадкової величини $X$ порахуємо за біноміальною формулою:

$$p_k = P_n(k) = C_n^k p^k q^{n-k}, \qquad q = 1-p$$

і отримаємо закон розподілу випадкової величини $X$, що називається біноміальним.

| $X = x_k$ | $0$ | $1$ | $2$ | $\ldots$ | $n$ |
|---|---|---|---|---|---|
| $P = p_k$ | $q^n$ | $C_n^1 pq^{n-1}$ | $C_n^2 p^2 q^{n-2}$ | $\ldots$ | $p^n$ |

Ми вже перевіряли раніше, що

$$\sum_{k=0}^{n} p_k = 1.$$

**Приклад.** Прилад містить 4 плати і ймовірність виявлення деяких технічних неполадок в кожній з них дорівнює 0,5. Запишіть закон розподілу випадкової величини $X$ — кількості плат, в яких виявлено технічні неполадки.

**Розв'язання.** Можливими значеннями дискретної випадкової величини $X$ будуть числа $x_0 = 0$, $x_1 = 1$, $x_2 = 2$, $x_3 = 3$, $x_4 = 4$. За біноміальною формулою знайдемо відповідні ймовірності цих значень, враховуючи, що $p = q = \frac{1}{2}$.

$$p_0 = C_4^0 p^0 q^4 = \left(\frac{1}{2}\right)^4 = \frac{1}{16}, \quad p_1 = C_4^1 p^1 q^3 = 4\left(\frac{1}{2}\right)\left(\frac{1}{2}\right)^3 = \frac{1}{4},$$

$$p_2 = C_4^2 p^2 q^2 = 6\left(\frac{1}{2}\right)^2\left(\frac{1}{2}\right)^2 = \frac{3}{8},$$

$$p_3 = C_4^3 p^3 q^1 = 4\left(\frac{1}{2}\right)\left(\frac{1}{2}\right)^3 = \frac{1}{4}, \quad p_4 = C_4^4 p^4 = \frac{1}{16}.$$

Перевіримо:

$$p_0 + p_1 + p_2 + p_3 + p_4 = \frac{1}{16} + \frac{1}{4} + \frac{3}{8} + \frac{1}{4} + \frac{1}{16} = 1.$$

Зведемо закон розподілу даної випадкової величини у таблицю:

| $X = x_k$ | 0 | 1 | 2 | 3 | 4 |
|---|---|---|---|---|---|
| $P = p_k$ | 1/16 | 1/4 | 3/8 | 1/4 | 1/16 |

З таблиці бачимо, що найбільш імовірна кількість плат, в яких виявлено технічні неполадки, $k_0 = 2$.

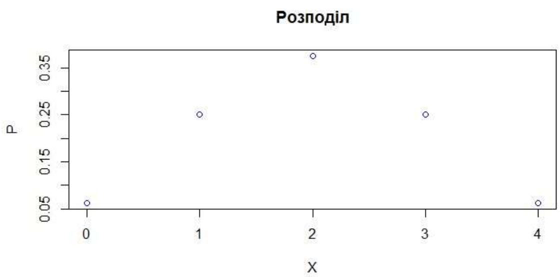


Рис. 1: Розподіл біноміальної випадкової величини

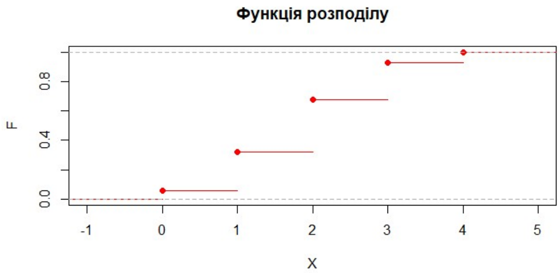


Рис. 2: Функція розподілу біноміальної випадкової величини

### Дисперсія і математичне сподівання випадкової величини, розподіленої за біноміальним законом.

Подібним чином розподілена кількість успіхів $\mu$ в $n$ випробуваннях за схемою Бернуллі, якщо ймовірність успіху в кожному випробуванні дорівнює $p$ і $q = 1 - p$. Введемо позначення $\mu_k$ — кількість успіхів у $k$-тому випробуванні. Тоді $\mu = \mu_1 + \ldots + \mu_n$, $E[\mu_k] = p$, $\mu_k^2 = \mu_k$, оскільки кількість успіхів в одному випробуванні може приймати тільки значення 1 або 0. Тому,

$$E[\mu_k^2] = E[\mu_k] = p, \quad D(\mu_k) = E[\mu_k^2] - (E[\mu_k])^2 = p - p^2 = p(1-p) = pq.$$

Випадкові величини $\mu_1, \ldots, \mu_n$ будуть попарно незалежними (оскільки випробування за схемою Бернуллі є незалежними), отже,

$$E[\mu_1 + \ldots + \mu_n] = nE[\mu_k] = np,$$

$$D(\mu_1 + \ldots + \mu_n) = D(\mu_1) + \ldots + D(\mu_n) = npq.$$

### Геометричний розподіл

Розглянемо схему випробувань Бернуллі з імовірністю успіху $p, \quad 0 < p < 1$, яка відбувається до здійснення першого успіху. Наприклад, така задача виникає, якщо людина підкидає монету до першого випадіння цифри.

Якщо успіху ставити у відповідність 1, а невдачі – 0, то простором елементарних подій буде: $\Omega = \{(0), (0,1), (0,0,1), (0,0,0,1), \ldots\}$. Задамо випадкову величину $X$ як номер випробування, в якому відбувся успіх. Тоді

$$P(X = k) = p(1-p)^{k-1}, k = 1, 2, 3, \dots$$

Побудуємо графіки розподілу випадкової величини для різних значень параметру.

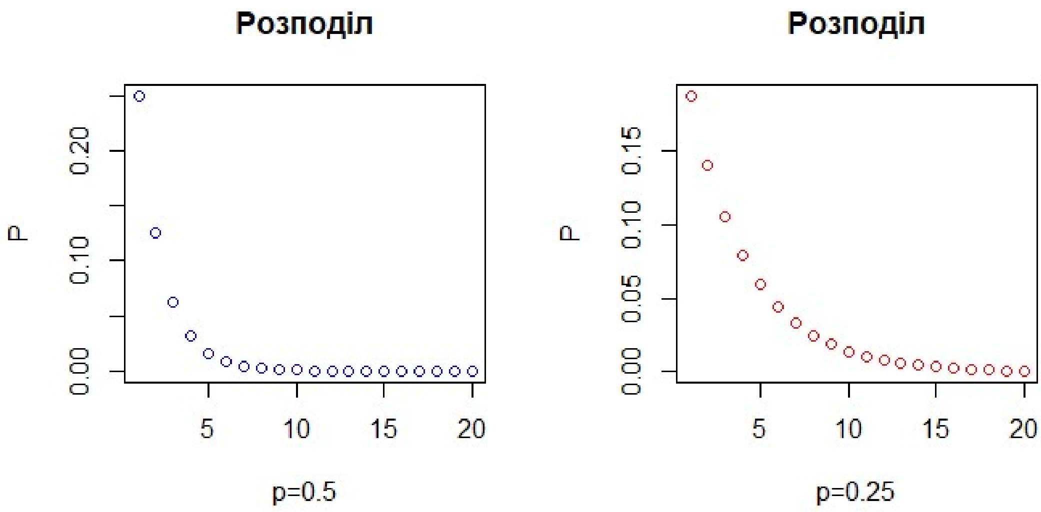


Рис. 3: Геометричний розподіл з різними параметрами

Обчислимо числові характеристики такої випадкової величини.

$$E[X] = \sum_{n=1}^{\infty} np(1-p)^{n-1} = p\sum_{n=1}^{\infty} n(1-p)^{n-1}.$$

$$\sum_{n=1}^{\infty}((1-p)^n)' = -\sum_{n=1}^{\infty} n(1-p)^{n-1}.$$

Оскільки $\sum_{n=1}^{\infty}(1-p)^n = \frac{1-p}{p}$, а $\left(\frac{1-p}{p}\right)' = -\frac{1}{p^2}$, то

$$E[X] = \frac{p}{p^2} = \frac{1}{p}.$$

$$E[X^2] = \sum_{n=1}^{\infty} n^2 p(1-p)^{n-1} = p\sum_{n=1}^{\infty} n(n-1)(1-p)^{n-1} + p\sum_{n=1}^{\infty} n(1-p)^{n-1} =$$

$$= p(1-p)\sum_{n=1}^{\infty} n(n-1)(1-p)^{n-2} + \frac{1}{p}.$$

$$\sum_{n=1}^{\infty}((1-p)^n)'' = \sum_{n=1}^{\infty} n(n-1)(1-p)^{n-2} = \frac{2}{p^3}.$$

Тоді $E[X^2] = \frac{2p(1-p)}{p^3} + \frac{1}{p} = \frac{2-p}{p^2}$. А дисперсія:

$$D(X) = \frac{2-p}{p^2} - \frac{1}{p^2} = \frac{1-p}{p^2}.$$

**Розподіл Пуассона.**

Розподіл імовірностей дискретної випадкової величини $X$, що приймає значення

$$x_k : 0, 1, 2, \ldots, , n, \ldots$$

з імовірностями

$$p = P\{X = x_k\} = \frac{\lambda^k e^{-\lambda}}{k!}, \quad k = 0, 1, \ldots \tag{24}$$

називається законом розподілу Пуассона з параметром $\lambda$. Розподіл Пуассона можна задати у формі таблиці:

| $X = x_k$ | $0$ | $1$ | $2$ | $\ldots$ | $n$ | $\ldots$ |
|---|---|---|---|---|---|---|
| $P = p_k$ | $e^{-\lambda}$ | $\lambda e^{-\lambda}$ | $\frac{\lambda^2 e^{-\lambda}}{2!}$ | $\ldots$ | $\frac{\lambda^n e^{-\lambda}}{n!}$ | $\ldots$ |

Просумувавши всі ймовірності розподілу Пуассона і використавши рівність

$$e^{\lambda} = \sum_{k=0}^{\infty} \frac{\lambda^k}{k!},$$

отримаємо підтвердження факту, що це дійсно буде розподіл ймовірностей:

$$\sum_{k=0}^{\infty} p_k = e^{-\lambda} \sum_{k=0}^{\infty} \frac{\lambda^k}{k!} = e^{-\lambda} e^{\lambda} = 1.$$

Під час вивчення схеми Бернуллі було зазначено, що для великих $n$ при обчисленні ймовірностей $P_n(k)$ краще використовувати асимптотичні формули Пуассона, які полегшують ці обрахунки. Доречі, з асимптотичної формули Пуассона слідує, що за допомогою розподілу Пуассона можна апроксимувати біноміальний закон розподілу, якщо кількість експериментів $n$ необмежено зростає $(n \to \infty)$ й одночасно ймовірність успіху в одному експерименті необмежено зменшується $(p \to 0)$ так, що їх добуток $np$ наближається до числа $\lambda : \lim_{n\to\infty} np = \lambda$.

**Приклад.** Електронна пошта банку підтримує зв'язки із сотнею абонентів. Імовірність того, що за годину на електронну пошту надійде повідомлення від абонента, дорівнює 0,02. Задайте закон розподілу випадкової величини $X$ — кількості повідомлень від абонентів за годину.

**Розв'язання.** У даному випадку проводиться $n = 100$ експериментів за схемою Бернуллі, і випадкова величина $X$ може приймати значення $x_0 = 0$, $x_1 = 1$, $x_2 = 2$, $x_3 = 3$, $x_4 = 4$, ..., $x_{100} = 100$. Імовірність події $A$ — надходження повідомлення від одного абонента буде дуже малою, а число $n = 100$ — велике і $\lambda = 100 \cdot 0,02 = 2$, тому відповідні ймовірності обчислюємо за формулою (24):

$$p_0 = P_{100}(0) \approx e^{-2} \approx 0,1353; \quad p_1 = P_{100}(1) \approx 2e^{-2} \approx 0,2707; \quad p_2 = P_{100}(2) \approx 2e^{-2} \approx 0,2707;$$

$$p_3 = P_{100}(3) \approx \frac{4}{3}e^{-2} \approx 0,1804; \qquad p_4 = P_{100}(4) \approx \frac{2}{3}e^{-2} \approx 0,0902;$$

....................................

$$p_{100} = P_{100}(100) = 1 - \sum_{k=0}^{99} P_{100}(k).$$

Закон розподілу заданої в задачі випадкової величини $X$ записуємо у формі таблиці:

| $X = x_k$ | 0 | 1 | 2 | 3 | 4 | ... |
|---|---|---|---|---|---|---|
| $P = p_k$ | 0,1353 | 0,2707 | 0,2707 | 0,1804 | 0,0902 | ... |

З таблиці бачимо, що найбільш імовірна кількість повідомлень від абонентів за годину - один або два.

### Математичне сподівання

Для випадкової величини Пуассона $X$:

$$p_k = P\{X = x_k\} = \frac{\lambda^k e^{-\lambda}}{k!}, \quad k = 0, 1, 2, \dots,$$

$$E[X] = \sum_{k=0}^{\infty} k p_k = \sum_{k=0}^{\infty} \frac{k\lambda^k e^{-\lambda}}{k!} = \lambda e^{-\lambda} \sum_{k=1}^{\infty} \frac{\lambda^{k-1}}{(k-1)!} = \lambda e^{-\lambda} \sum_{m=0}^{\infty} \frac{\lambda^m}{m!} = \lambda e^{-\lambda} e^{\lambda} = \lambda.$$

Таким чином, математичне сподівання випадкової величини Пуассона $X$ дорівнює параметру цього розподілу $\lambda$.

**Дисперсія випадкової величини, розподіленої за законом Пуассона.**

$$E[X^2] = \sum_{k=0}^{\infty} k^2 p_k = \sum_{k=0}^{\infty} \frac{k^2 \lambda^k e^{-\lambda}}{k!} = e^{-\lambda} \sum_{k=1}^{\infty} \frac{k\lambda^k}{(k-1)!} = e^{-\lambda} \sum_{k=1}^{\infty} \frac{(k-1+1)\lambda^k}{(k-1)!} =$$

$$= e^{-\lambda} \sum_{k=1}^{\infty} \frac{\lambda^k}{(k-2)!} + e^{-\lambda} \sum_{k=1}^{\infty} \frac{\lambda^k}{(k-1)!} = \lambda^2 e^{-\lambda} \sum_{k=2}^{\infty} \frac{\lambda^{k-2}}{(k-2)!} + \lambda e^{-\lambda} \sum_{k=1}^{\infty} \frac{\lambda^{k-1}}{(k-1)!} =$$

$$= \lambda^2 e^{-\lambda} e^{\lambda} + \lambda e^{-\lambda} e^{\lambda} = \lambda^2 + \lambda.$$

Таким чином, $D(X) = E[X^2] - (E[X])^2 = \lambda^2 + \lambda - \lambda^2 = \lambda$, тобто, аналогічно до математичного сподівання, дисперсія випадкової величини , розподіленої за законом Пуассона, дорівнює параметру цього розподілу $\lambda$.

**Гіпергеометричний розподіл**

Пояснимо цей розподіл на конкретному прикладі.

Нехай задано деяку множину елементів, число яких дорівнює $N$; з них $K$ елементів мають, наприклад, ознаку А (колір, стандартність, наповнення), а решта $N-K$ елементів – ознаку В. З цієї множини навмання беруть $n$ елементів. Випадкова величина $X$ - число елементів з ознакою А, що трапляється серед $n$ навмання взятих елементів. Тоді $X$ набуває значень $k = 0, 1, 2, ..., min\{n, K\}$ , а ймовірність їх появи визначається за гіпергеометричним законом:

$$P(X = k) = \frac{C_K^k \cdot C_{N-K}^{n-k}}{C_N^n}, \quad k \le K, n \le N.$$

Позначимо через $M = \min\{n, K\}$. Тоді числові характеритики цієї випадкової величини обчислюються за наступними формулами:

$$E[X] = \sum_{k=0}^{M} \frac{kC_K^k \cdot C_{N-K}^{n-k}}{C_N^n} = \frac{nK}{N}.$$

$$D(X) = \sum_{k=0}^{M} \frac{k^2 C_K^k \cdot C_{N-K}^{n-k}}{C_N^n} - \left(\frac{nK}{N}\right)^2 = \frac{n \cdot \frac{K}{N}\left(1 - \frac{K}{N}\right)(N-n)}{(N-1)}.$$

**Завдання на додатковий бал.** Довести ці формули, використовуючи означення математичного сподівання і дисперсії дискретної випадкової величини.

**Приклад.** В ящику міститься 10 однотипних деталей, із них 7 стандартних, а решта – браковані. Навмання із ящика вибирають 3 деталі. Побудувати закон розподілу випадкової величини X — числа появи стандартних деталей серед 3 навмання вибраних і обчислити математичне сподівання $E[X]$, дисперсію $D(X)$.

**Розв'язання.** $P(X = 0) = \frac{C_3^3}{C_{10}^3} = \frac{1}{120}$, $P(X = 1) = \frac{C_3^2 \cdot C_7^1}{C_{10}^3} = \frac{7}{40}$, $P(X = 2) = \frac{C_3^1 \cdot C_7^2}{C_{10}^3} = \frac{21}{40}$, $P(X = 3) = \frac{C_7^3}{C_{10}^3} = \frac{7}{24}$.

| $X = x_k$ | 0 | 1 | 2 | 3 |
|---|---|---|---|---|
| $P = p_k$ | $\frac{1}{120}$ | $\frac{7}{40}$ | $\frac{21}{40}$ | $\frac{7}{24}$ |

Обчислимо числові характеристики:

$$E[X] = \frac{7}{40} + \frac{42}{40} + \frac{21}{24} = \frac{84}{40} = 2.1,$$

$$E[X^2] = \frac{7}{40} + \frac{84}{40} + \frac{63}{24} = \frac{196}{40} = 4.9,$$

$$D(X) = E[X^2] - (E[X])^2 = \frac{49}{10} - \frac{441}{100} = \frac{49}{100}.$$

Порівняємо з теоретичними значеннями.

$N = 10$, $n = 3$, $K = 7$. Тоді

$$E[X] = \frac{nK}{N} = \frac{3 \cdot 7}{10} = 2.1.$$

$$D(X) = \frac{n \cdot \frac{K}{N}\left(1 - \frac{K}{N}\right)(N - n)}{(N - 1)} = \frac{3 \cdot \frac{7}{10}\frac{3}{10} \cdot 7}{9} = \frac{49}{100}.$$

## Література

1. Теорія ймовірностей : конспект лекцій, В.В. Булдигін, Ю.П. Буценко, О.О. Диховичний/ Київ, 1999, 97с.

2. Теорія ймовірностей та математична статистика: навч. посіб., Н. Д. Федоренко, О. І. Баліна, І. С. Безклубенко/К.: КНУБА, 2007, 104 с.

3. Теорія ймовірностей, математична статистика та імовірнісні процеси : навч. посіб., Ю. М. Слюсарчук, Й. Я. Хром'як, Л. Л. Джавала, В. М. Цимбал/ Львів : Вид-во Львів. політехніки, 2015, 364 с.

4. Курс теорії ймовірностей: навч. посіб., Б. В. Гнєденко/ К.: ВПЦ Київський університет, 2010, 464 с.

5. Теория вероятностей и математическая статистика. Изд.9. Учебное пособие для вузов, В. Е. Гмурман/М: Высшая школа, 2003, 479с.