

Лекція 3. Послідовності незалежних випробувань Бернуллі. Граничні теореми.

Розглянемо стохастичний експеримент, який полягає в багатократному повторенні деякого досліду. Вважатимемо, що множина можливих наслідків, кожного досліду скінченна, а самі досліди відбуваються незалежно і при однакових умовах. Такий стохастичний експеримент називають **послідовністю незалежних однорідних випробувань**. Для спрощення міркувань розглянемо спочатку ситуацію, коли кожний дослід має лише два можливих варіанти завершення, які умовно позначимо, як «успіх» та «невдача», що стаються відповідно з ймовірностями p та q . В такій постановці задачі цікавляться не результатами кожного окремого досліду, а кількістю «успіхів» в серії із n випробувань, приміром кількість влучань в серії пострілів. Саме таку задачу успішно дослідив Я. Бернуллі. На його честь її називають схемою Бернуллі.

Теорема 1. В серії із n незалежних і однорідних випробувань ймовірність $P_n(k)$ події, що кількість «успіхів» рівна k ($0 \leq k \leq n$), визначається формулою:

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

Набір ймовірностей $P_n(k)$ для $0 \leq k \leq n$ називається **біноміальним** (інколи кажуть **біномним**) **розподілом**.

Доведення. Простір елементарних подій послідовності незалежних випробувань можна описати, як набір ланцюжків довжини n , складених із літер «У» та «Н», поставлених у відповідності з результатом кожного стохастичного експерименту. Зверніть увагу, що елементарні події даного стохастичного експерименту не є рівноможливими.

Наприклад, при $n=2$ $\Omega = \{HH, UV, HU, VH\}$ і, взагалі кажучи, $P(HH) = q^2 \neq P(UV) = p^2$. Знайдемо ймовірність події $A_n(i_1, i_2, \dots, i_k)$, що успіхи мали місце у дослідах з номерами i_1, i_2, \dots, i_k . Із-за незалежності дослідів маємо $P(A_n(i_1, i_2, \dots, i_k)) = p^k q^{n-k}$, і дана ймовірність не залежить від номерів випробувань, у яких сталися «успіхи», а визначається лише їх кількістю. Отже, ймовірність $P_n(k)$ є сумою ймовірностей $P(A_n(i_1, i_2, \dots, i_k))$ по всіх можливих виборах номерів i_1, i_2, \dots, i_k . Кількість таких виборів, очевидно, рівна C_n^k , що й доводить справедливості формули (1).

Зауваження 1. Назва даного розподілу пов'язана з тим фактом, що кожна ймовірність $P_n(k)$ є елементом біному:

$$(p+q)^n = \sum_{k=0}^n C_n^k p^k q^{n-k} = \sum_{k=0}^n P_n(k) = 1.$$

Зауваження 2. Ця теорема може бути узагальнена на випадок довільної скінченної кількості варіантів завершення кожного випробування, а саме, має місце наступне твердження (доведення пропонується провести самостійно).

Теорема 2. Нехай в серії n незалежних однорідних випробувань кожне з них може завершитись випадковими наслідками v_1, v_2, \dots, v_m з ймовірностями рівними відповідно p_1, p_2, \dots, p_m (при цьому $p_1 + p_2 + \dots + p_m = 1$). Тоді ймовірність $P_n(k_1, k_2, \dots, k_m)$ події, що наслідки v_1, v_2, \dots, v_m сталися відповідно k_1, k_2, \dots, k_m разів, є доданком розкладу полінома $(p_1 + p_2 + \dots + p_m)^n$ і визначається формулою

$$P_n(k_1, k_2, \dots, k_m) = \frac{n!}{k_1! k_2! \dots k_m!} p_1^{k_1} p_2^{k_2} \dots p_m^{k_m} \quad (2)$$

Набір ймовірностей $P_n(k_1, k_2, \dots, k_m)$ для всіх допустимих значень k_1, k_2, \dots, k_m називається **поліноміальним розподілом**. Для них справедливо, що $\sum_{k_1, \dots, k_m} P_n(k_1, k_2, \dots, k_m) = (p_1 + \dots + p_m)^n = 1$.

Приклад. На деякому виробництві ймовірність виготовлення бракованого виробу складає $p=0,01$. Визначити ймовірність, що в партії із 100 виробів бракованих буде рівно 2, не більше ніж 2. Яку партію виробів треба взяти, щоб ймовірність знаходження в ній принаймні однієї бракованої була більша за 0,5?

Ймовірність, що в партії із 100 виробів бракованих буде рівно 2, визначимо за формулою (1): $P_{100}(2) = C_{100}^2 p^2 q^{100-2} = \frac{100 \cdot 99}{2} (0,01)^2 (0,99)^{98} \approx 0,185$. Ймовірність, що в даній партії кількість бракованих не перевищує двох, знайдемо як суму $P_{100}(2) + P_{100}(1) + P_{100}(0) \approx 0,185 + 0,370 + 0,366 = 0,921$

Для відповіді на останнє запитання необхідно розв'язати відносно n нерівність $\sum_{k=1}^n P_n(k) = 1 - P_n(0) > 0,5$, або

$P_n(0) < 0,5$, що означає виконання нерівності $(0,99)^n < 0,5$. Логарифмуючи, одержимо: $n > -\frac{\lg 2}{\lg 0,99}$, звідки $n \geq 69$.

*Дослідимо поведінку послідовності значень ймовірностей $P_n(k)$ при зміні k від 0 до n . З цією метою розглянемо відношення $\frac{P_n(k)}{P_n(k-1)} = \frac{C_n^k p^k q^{n-k}}{C_n^{k-1} p^{k-1} q^{n-k+1}} = \frac{(n-k+1)p}{kq} = \frac{(n+1)p + k(1-p) - k}{kq} = 1 + \frac{(n+1)p - k}{kq}$. Звідси видно, що при $(n+1)p < k$ виконується $P_n(k) < P_n(k-1)$, а при $(n+1)p > k$ — $P_n(k) > P_n(k-1)$. Таким чином, значення $P_n(k)$ спочатку зростають, а потім спадають, при цьому $\max_{0 \leq k \leq n} P_n(k) = P_n(k_0)$ досягається при $k_0 = \lceil (n+1)p \rceil$ — **найбільш ймовірна**

кількість «успіхів» — і називається **максимальною ймовірністю**. Якщо ж число $(n+1)p$ — ціле, то максимальних значень ймовірності два: $P_n(k_0) = P_n(k_0 - 1)$.

Приклад. Яку найбільш ймовірну кількість м'ячів кине в кільце баскетболіст в серії із 50 кидків, якщо він в ймовірністю 0,8 при кожному підкиданні м'яча, і яка ймовірність такої кількості закинутих м'ячів?

Маємо: $n = 50, p = 0,8[(n+1)p] = [51 \cdot 0,8] = 40$. Відповідна ймовірність дорівнює $P_{50}(40) = C_{50}^{40} 0,8^{40} 0,2^{10} \approx 0,14$.

Граничні теореми в серії незалежних випробувань Бернуллі.

Три великих кількостях незалежних випробувань безпосередні обчислення за формулою (1) стають обтяжливими, і виникає питання про використання наближених формул. Перші граничні теореми для незалежних випробувань Бернуллі довів А. Муавр у 1733 р. в припущенні, що $p = q = 0,5$.

Введемо наступні позначення: $\Phi(x) = \int_{-\infty}^x \varphi(t) dt$ – функція нормального розподілу, $\varphi(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$ – щільність

нормального розподілу. Властивості та зміст цих важливих функцій будуть визначені дещо пізніше.

Теорема 3 (Гранична теорема Муавра-Лапласа). Припустимо, що в схемі Бернуллі p та q відповідно означають ймовірності «успіху» та «невдачі» відповідно, нехай $n \rightarrow \infty$, причому k та $n-k$ також досить великі. Тоді

$$P_n(k) \approx \frac{1}{\sqrt{npq}} \varphi\left(\frac{k-np}{\sqrt{npq}}\right) \quad \text{– локальна теорема Мавра-Лапласа} \quad (4)$$

$$P(n_1 \leq k \leq n_2) = \sum_{k=n_1}^{n_2} P_n(k) \approx \Phi\left(\frac{n_2-np}{\sqrt{npq}}\right) - \Phi\left(\frac{n_1-np}{\sqrt{npq}}\right) \quad \text{– інтегральна теорема Мавра-Лапласа} \quad (5)$$

Доведення. Для доведення скористаємося формулою Стірлінга: $n! \approx \sqrt{2\pi n} n^n e^{-n}$ справедливою при великих n .

$$\begin{aligned} P_n(k) &= \frac{n!}{k!(n-k)!} p^k q^{n-k} \approx \frac{\sqrt{2\pi n} n^n e^{-n} p^k q^{n-k}}{\sqrt{2\pi k} k^k e^{-k} \sqrt{2\pi(n-k)} (n-k)^{n-k} e^{-n+k}} = \frac{\sqrt{nn} p^k q^{n-k}}{\sqrt{2\pi k(n-k)} k^k (n-k)^{n-k}} = \\ &= \frac{1}{\sqrt{2\pi}} \sqrt{\frac{n}{k(n-k)}} \left(\frac{k}{np}\right)^{-k} \left(\frac{n-k}{nq}\right)^{-(n-k)} \end{aligned}$$

Позначимо через $x_k = x = \frac{k-np}{\sqrt{npq}}$. Якщо значення x знаходиться у скінченному проміжку, то k та $n-k$ зростають

разом з n . Тоді $k = np + x\sqrt{npq}$, а $n-k = n - np - x\sqrt{npq} = nq - x\sqrt{npq}$. Враховуючи це, розглянемо логарифм виразу

$$R_n = \left(\frac{k}{np}\right)^{-k} \left(\frac{n-k}{nq}\right)^{-(n-k)} : \ln R_n = -k \ln\left(\frac{k}{np}\right) - (n-k) \ln\left(\frac{n-k}{nq}\right) = -(np + x\sqrt{npq}) \ln\left(1 + x\sqrt{\frac{q}{np}}\right) - (nq - x\sqrt{npq}) \ln\left(1 - x\sqrt{\frac{p}{nq}}\right)$$

Далі скористаємось розкладом в ряд Тейлора функції $\ln(1+t) \approx t - \frac{t^2}{2} + O(t^3)$ для $t \rightarrow 0$:

$$\begin{aligned} \ln R_n &\approx -(np + x\sqrt{npq}) \left(x\sqrt{\frac{q}{np}} - \frac{x^2 q}{2np} + O\left(\frac{1}{n^{3/2}}\right)\right) - (nq - x\sqrt{npq}) \left(-x\sqrt{\frac{p}{nq}} - \frac{x^2 p}{2nq} + O\left(\frac{1}{n^{3/2}}\right)\right) = \\ &= -x\sqrt{npq} - x^2 q + \frac{x^2 q}{2} + O\left(\frac{1}{\sqrt{n}}\right) + x\sqrt{npq} - x^2 p + \frac{x^2 p}{2} + O\left(\frac{1}{\sqrt{n}}\right) = -\frac{x^2}{2} + O\left(\frac{1}{\sqrt{n}}\right) \end{aligned}$$

Таким чином, $R_n \approx e^{-\frac{x^2}{2}}, n \rightarrow \infty$. Крім того, $\sqrt{\frac{n}{k(n-k)}} = \left(\left(\frac{k}{n}\right)\left(1 - \frac{k}{n}\right)n\right)^{-1/2} \approx (np(1-p))^{-1/2} = \frac{1}{\sqrt{npq}}, n \rightarrow \infty$. Остаточно

одержимо, що при $n \rightarrow \infty$ $P_n(k) \approx \frac{1}{\sqrt{2\pi}} \frac{1}{\sqrt{npq}} e^{-\frac{x^2}{2}} = \frac{1}{\sqrt{npq}} \varphi(x) = \frac{1}{\sqrt{npq}} \varphi\left(\frac{k-np}{\sqrt{npq}}\right)$. Тим самим твердження локальної

граничної теореми доведено. (Правда, тут не досліджується оцінка точності збіжності).

Покажемо, звідки випливає справедливість інтегральної граничної теореми Мавра-Лапласа. Розглянемо

$$\text{ймовірність: } P(n_1 \leq k \leq n_2) = \sum_{k=n_1}^{n_2} P_n(k) \approx \sum_{k=n_1}^{n_2} \frac{1}{\sqrt{npq}} \varphi\left(\frac{k-np}{\sqrt{npq}}\right) = \sum_{k=n_1}^{n_2} \frac{1}{\sqrt{npq}} \varphi(x_k)$$

Оскільки $x_{k+1} - x_k = \frac{1}{\sqrt{npq}}$, то при $n \rightarrow \infty$ ймовірність $P(n_1 \leq k \leq n_2)$ є інтегральною сумою, що прямуватиме до

інтегралу $\int_a^b \varphi(x) dx$, де $a = \frac{n_1 - np}{\sqrt{npq}}, b = \frac{n_2 - np}{\sqrt{npq}}$. Враховуючи прийняті позначення, одержимо, що

$$P(n_1 \leq k \leq n_2) \approx \Phi\left(\frac{n_2 - np}{\sqrt{npq}}\right) - \Phi\left(\frac{n_1 - np}{\sqrt{npq}}\right).$$

Слід зазначити, що значення функцій $\Phi(x)$ та $\phi(x)$, що дають граничні наближення, табульовані. При використанні таблицями слід враховувати властивості цих функцій (перевірте їх самостійно), а саме: $\phi(-x) = \phi(x)$, $\int_{-\infty}^{\infty} \phi(x) dx = 1$,

$\Phi(x) = \frac{1}{2} + \operatorname{sgn}(x) \Phi_0(|x|)$, де $\Phi_0(x) = \int_0^x \phi(t) dt$ – саме її значення можна знайти в таблицях.

Теорема 4 (Гранична теорема Пуассона). Нехай кількість незалежних випробувань $n \rightarrow \infty$, причому $np_n \rightarrow \lambda$. Тоді $P_n(k) = C_n^k p_n^k q_n^{n-k} \rightarrow e^{-\lambda} \frac{\lambda^k}{k!}$ ($k = 0, 1, 2, \dots, n, \dots$)

Доведення. Дійсно,

$$P_n(k) = \frac{n!}{k!(n-k)!} p_n^k q_n^{n-k} = \frac{n(n-1)\dots(n-k+1)}{k! n^k (1-p_n)^k} (np_n)^k (1-p_n)^n = \frac{(1-\frac{1}{n})(1-\frac{2}{n})\dots(1-\frac{k-1}{n})}{k! (1-p_n)^k} (np_n)^k (1-p_n)^n$$

При $n \rightarrow \infty$, очевидно, $\frac{(1-\frac{1}{n})(1-\frac{2}{n})\dots(1-\frac{k-1}{n})}{(1-p_n)^k} \rightarrow 1$, $(np_n)^k \rightarrow \lambda^k$, $(1-p_n)^n = \left(1-p_n\right)^{-\frac{1}{p_n}} \rightarrow e^{-\lambda}$, отже,

$$P_n(k) \rightarrow e^{-\lambda} \frac{\lambda^k}{k!} = \pi_k(\lambda), \quad k = 0, 1, 2, \dots \quad \text{при } n \rightarrow \infty. \quad (3)$$

Граничні ймовірності $\pi_k(\lambda)$ мають очевидну властивість $\sum_{k=0}^{\infty} \pi_k(\lambda) = 1$ і називаються **розподілом Пуассона**. Значення

цих ймовірностей можна знайти в спеціальних таблицях. Теорема Пуассона справедлива при $p_n \cong \frac{\lambda}{n}$, $n \rightarrow \infty$, отже, при p_n близьких до нуля, тому теорему Пуассона часто називають законом **рідких подій**. Досить точні наближення досягаються лише при $p_n \in (0, 10)$. Для ймовірностей успіху поза межами рекомендованого інтервалу варто користатись наближенням, яке дає гранична теорема Мавра-Лапласа.

Розглянемо приклади застосування граничних теорем в схемі Бернуллі.

Приклад. Обчислити ймовірність того, що при 200 підкиданнях симетричної монети кількість випадін герба відрізнятиметься від 100 не більше, ніж на 5.

Нехай ймовірність випадіння герба $p = 0,5$, кількість повторень експерименту $n = 200$, отже, значення $np = 100$ – досить велике, тому використовуватимемо граничну теорему Мавра-Лапласа.

Позначимо k – кількість випадін герба при 200 підкиданнях монети. Знайдемо $P(|k - 100| \leq 5) = P(95 \leq k \leq 105) \approx$

$$\approx \Phi\left(\frac{105 - 100}{\sqrt{200 \cdot 0,5 \cdot 0,5}}\right) - \Phi\left(\frac{95 - 100}{\sqrt{200 \cdot 0,5 \cdot 0,5}}\right) = \Phi\left(\frac{1}{\sqrt{2}}\right) - \Phi\left(-\frac{1}{\sqrt{2}}\right) = 2\Phi_0\left(\frac{\sqrt{2}}{2}\right) = 2\Phi_0(0,7778) \approx 0,56331$$

Приклад. Ймовірність виготовлення бракованої деталі на виробництві рівна $p = 0,005$. Яка ймовірність, що серед 10000 випадково обраних деталей кількість бракованих дорівнює 40? Не перевищує 40?

Оскільки в даному випадку $np = 50$, хоча й ймовірність p зовсім мала, знову використовуємо наближення Мавра-Лапласа. Тут $\sqrt{npq} = \sqrt{10000 \cdot 0,005 \cdot 0,995} \approx 7,05$; $\frac{k - np}{\sqrt{npq}} = \frac{40 - 50}{7,05} \approx -1,42$, тому $P_{10000}(40) \approx \frac{1}{7,05} \phi(1,42) \approx \frac{0,1456}{7,05} \approx 0,0206$

Для відповіді на друге питання шукаємо ймовірність

$$P(0 \leq k \leq 40) = \sum_{k=0}^{40} P_{10000}(k) \approx \Phi\left(\frac{40 - 50}{7,05}\right) - \Phi\left(\frac{0 - 50}{7,05}\right) \approx 0,5 - \Phi_0(1,42) \approx 0,5 - 0,4222 = 0,0778$$

Приклад. Ймовірність влучити в ціль рівна $p = 0,001$. Для поразки цієї цілі потрібно не менше двох влучань. Знайти ймовірність враження цілі при 5000 пострілів.

Тут $np = 5$, тому в якості наближення використовуємо розподіл Пуассона. Знайдемо

$$P(k \geq 2) = \sum_{k=2}^{5000} P_{5000}(k) = 1 - \sum_{k=0}^1 P_{5000}(k) = 1 - P_{5000}(0) - P_{5000}(1) \approx 1 - \pi_0(5) - \pi_1(5) = 1 - e^{-5} - 5e^{-5} \approx 0,9596$$

Приклад. Нехай в серії з n випробувань Бернуллі успіх стається з ймовірністю p . Оцінити ймовірність, що відхилення частоти успіху від його ймовірності не перевищує заданої величини ε .

Отже, маємо оцінити $P\left(\left|\frac{k}{n} - p\right| \leq \varepsilon\right) = P\left(\left|\frac{k - np}{\sqrt{npq}}\right| \leq \varepsilon \sqrt{\frac{n}{pq}}\right) \approx \Phi\left(\varepsilon \sqrt{\frac{n}{pq}}\right) - \Phi\left(-\varepsilon \sqrt{\frac{n}{pq}}\right) = 2\Phi_0\left(\varepsilon \sqrt{\frac{n}{pq}}\right)$. Якщо $n \rightarrow \infty$, то

$P\left(\left|\frac{k}{n} - p\right| \leq \varepsilon\right) \rightarrow 1$. Цей результат відомий як **теорема Бернуллі**, яка показує, що частота успіху в схемі Бернуллі із зростанням кількості випробувань відрізняється від ймовірності успіху менше від будь-якого малого числа ε з ймовірністю, як завгодно близькою до 1.