

Лекція 27.
Розподіл Пуасона

Найбільш загальним випадком різного роду імовірнісних розподілів є біноміальний розподіл. Скористаємося його універсальністю для визначення найбільше що часто зустрічаються на практиці приватних видів розподілів.

Біноміальний розподіл

Нехай є якась подія A . Імовірність появи події A дорівнює p , імовірність неяви події A дорівнює $1 - p$, іноді неї позначають як q . Нехай n — число випробувань, m — частота появи події A у цих n випробуваннях.

Відомо, що сумарна ймовірність всіх можливих комбінацій результатів дорівнює одиниці, тобто:

$$1 = p^n + n \cdot p^{n-1} \cdot (1-p) + C_n^{n-2} \cdot p^{n-2} \cdot (1-p)^2 + \dots + C_n^m \cdot p^m \cdot (1-p)^{n-m} + \dots + (1-p)^n \dots$$

p^n — ймовірність того, що в n випробуваннях подія A відбудеться n раз;

$n \cdot p^{n-1} \cdot (1-p)$ — ймовірність того, що в n випробуваннях подія A відбудеться $(n-1)$ раз і не відбудеться 1 раз;

$C_n^{n-2} \cdot p^{n-2} \cdot (1-p)^2$ — ймовірність того, що в n випробуваннях подія A відбудеться $(n-2)$ рази й не відбудеться 2 рази;

$P_m = C_n^m \cdot p^m \cdot (1-p)^{n-m}$ — імовірність того, що в n випробуваннях подія A відбудеться m раз і не відбудеться $(n-m)$ раз;

$(1-p)^n$ — імовірність того, що в n випробуваннях подія A не відбудеться жодного разу;

— число сполучень із n по m .

Математичне очікування M біноміального розподілу дорівнює:

$$M = n \cdot p,$$

де n — число випробувань, p — імовірність появи події A .

Середньоквадратичне відхилення σ :

$$\sigma = \text{sqrt}(n \cdot p \cdot (1 - p)).$$

Приклад 1. Обчислити ймовірність того, що подія, що має ймовірність $p = 0.5$, в $n = 10$ випробуваннях відбудеться $m = 1$ раз. Маємо: $C_{10}^1 = 10$, і далі:
 $P_1 = 10 \cdot 0.5^1 \cdot (1 - 0.5)^{10-1} = 10 \cdot 0.5^{10} = 0.0098$. Як бачимо, ймовірність настання цієї події досить мала. Порозумівається це, по-перше, тим, що абсолютно не ясно, чи відбудеться чи подія ні, оскільки ймовірність дорівнює 0.5 і шанси тут «50 на 50»; а по-друге, потрібно обчислити те, що подія відбудеться саме один раз (не більше й не менше) з десяти.

Приклад 2. Обчислити ймовірність того, що подія, що має ймовірність $p = 0.5$, в $n = 10$ випробуваннях відбудеться $m = 2$ рази. Маємо: $C_{10}^2 = 45$, і далі:
 $P_2 = 45 \cdot 0.5^2 \cdot (1 - 0.5)^{10-2} = 45 \cdot 0.5^{10} = 0.044$. Ймовірність настання цієї події стала більше!

Приклад 3. Збільшимо ймовірність настання самої події. Зробимо його більше ймовірним. Обчислити ймовірність того, що подія, що має ймовірність $p = 0.8$, в $n = 10$ випробуваннях відбудеться $m = 1$ раз. Маємо: $C_{10}^1 = 10$, і далі:

$$P_1 = 10 \cdot 0.81 \cdot (1 - 0.8)^{10-1} = 10 \cdot 0.81 \cdot 0.29 = 0.000004.$$

Ймовірність стала менше, ніж у першому прикладі! Відповідь, на перший погляд, здається дивним, але оскільки подія має досить більшу ймовірність, навряд чи воно відбудеться тільки один раз. Більш імовірно, що воно відбудеться більше, ніж один, кількість разів.

Дійсно, підраховуючи $P_0, P_1, P_2, P_3, \dots, P_{10}$ (імовірність того, що подія в $n = 10$ випробуваннях відбудеться 0, 1, 2, 3, ..., 10 раз), ми побачимо:

$$C_{10}^0 = 1, C_{10}^1 = 10, C_{10}^2 = 45, C_{10}^3 = 120, C_{10}^4 = 210, C_{10}^5 = 252,$$

$$C_{10}^6 = 210, C_{10}^7 = 120, C_{10}^8 = 45, C_{10}^9 = 10, C_{10}^{10} = 1;$$

$$P_0 = 1 \cdot 0.80 \cdot (1 - 0.8)^{10-0} = 1 \cdot 1 \cdot 0.2^{10} = 0.0000\dots;$$

$$P_1 = 10 \cdot 0.8^1 \cdot (1 - 0.8)^{10-1} = 10 \cdot 0.81 \cdot 0.29 = 0.0000\dots;$$

$$P_2 = 45 \cdot 0.8^2 \cdot (1 - 0.8)^{10-2} = 45 \cdot 0.82 \cdot 0.28 = 0.0000\dots;$$

$$P_3 = 120 \cdot 0.8^3 \cdot (1 - 0.8)^{10-3} = 120 \cdot 0.83 \cdot 0.27 = 0.0008\dots;$$

$$P_4 = 210 \cdot 0.8^4 \cdot (1 - 0.8)^{10-4} = 210 \cdot 0.84 \cdot 0.26 = 0.0055\dots;$$

$$P_5 = 252 \cdot 0.8^5 \cdot (1 - 0.8)^{10-5} = 252 \cdot 0.85 \cdot 0.25 = 0.0264\dots;$$

$$P_6 = 210 \cdot 0.8^6 \cdot (1 - 0.8)^{10-6} = 210 \cdot 0.86 \cdot 0.24 = 0.0881\dots;$$

$$P_7 = 120 \cdot 0.8^7 \cdot (1 - 0.8)^{10-7} = 120 \cdot 0.87 \cdot 0.23 = 0.2013\dots;$$

$$P_8 = 45 \cdot 0.88^8 \cdot (1 - 0.8)^{10-8} = 45 \cdot 0.88 \cdot 0.22 = 0.3020\dots \text{ (сама більша ймовірність!)};$$

$$P_9 = 10 \cdot 0.89 \cdot (1 - 0.8)^{10-9} = 10 \cdot 0.89 \cdot 0.21 = 0.2684\dots;$$

$$P_{10} = 1 \cdot 0.8^{10} \cdot (1 - 0.8)^{10-10} = 1 \cdot 0.8^{10} \cdot 0.20 = 0.1074\dots$$

Зрозуміло, $P_0 + P_1 + P_2 + P_3 + P_4 + P_5 + P_6 + P_7 + P_8 + P_9 + P_{10} = 1$.

Нормальний розподіл

Якщо зобразити величини $P_0, P_1, P_2, P_3, \dots, P_{10}$, які ми підраховали в прикладі 3, на графіку, то виявиться, що їхній розподіл має вигляд, близький до нормального закону розподілу (див. мал. 27.1) (див. лекцію 25).

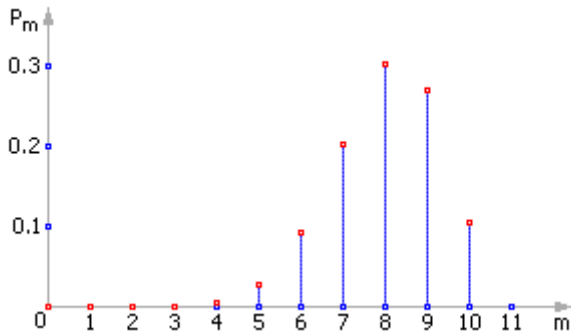


Рис. 27.1. Вид біноміального розподілу ймовірностей для різних m при $p = 0.8$, $n = 10$

Біноміальний закон переходить у нормальний, якщо ймовірності появи й не появи події A приблизно однакові, тобто, умовно можна записати: $p \approx (1 - p)$. Для приклада візьмемо $n = 10$ й $p = 0.5$ (тобто $p = 1 - p = 0.5$).

За змістом до такої задачі ми прийдемо, якщо, наприклад, захочемо теоретично порахувати, скільки буде хлопчиків і скільки дівчинок з 10 народжених у роддомі в один день дітей. Точніше, уважати будемо не хлопчиків і дівчинок, а ймовірність, що народяться тільки хлопчики, що народиться 1 хлопчик й 9 дівчинок, що народиться 2 хлопчика й 8 дівчинок і так далі. Приймемо для простоти, що ймовірність народження хлопчика й дівчинки однакова й дорівнює 0.5 (але насправді, якщо чесно, це не так, див. курс «Системи та методи штучного інтелекту»).

Ясно, що розподіл буде симетричне, тому що ймовірність народження 3 хлопчиків й 7 дівчинок дорівнює ймовірності народження 7 хлопчиків й 3 дівчинок. Найбільша ймовірність народження буде в 5 хлопчиків й 5 дівчинок. Ця ймовірність дорівнює 0.25, до речі, не така вуж вона й більша по абсолютній величині. Далі, ймовірність того, що народиться відразу 10 або 9 хлопчиків набагато менше, ніж ймовірність того, що народиться 5 ± 1 хлопчик з 10 дітей. Саме біноміальний розподіл нам допоможе зробити цей розрахунок. Отже.

$$C_{10}^0 = 1, C_{10}^1 = 10, C_{10}^2 = 45, C_{10}^3 = 120, C_{10}^4 = 210, C_{10}^5 = 252,$$

$$C_{10}^6 = 210, C_{10}^7 = 120, C_{10}^8 = 45, C_{10}^9 = 10, C_{10}^{10} = 1;$$

$$P_0 = 1 \cdot 0.50 \cdot (1 - 0.5)^{10-0} = 1 \cdot 1 \cdot 0.5^{10} = 0.000977\dots;$$

$$P_1 = 10 \cdot 0.5^1 \cdot (1 - 0.5)^{10-1} = 10 \cdot 0.5^{10} = 0.009766\dots;$$

$$P_2 = 45 \cdot 0.5^2 \cdot (1 - 0.5)^{10-2} = 45 \cdot 0.5^{10} = 0.043945\dots;$$

$$P_3 = 120 \cdot 0.5^3 \cdot (1 - 0.5)^{10-3} = 120 \cdot 0.5^{10} = 0.117188\dots;$$

$$P_4 = 210 \cdot 0.5^4 \cdot (1 - 0.5)^{10-4} = 210 \cdot 0.5^{10} = 0.205078\dots;$$

$$P_5 = 252 \cdot 0.5^5 \cdot (1 - 0.5)^{10-5} = 252 \cdot 0.5^{10} = 0.246094\dots;$$

$$P_6 = 210 \cdot 0.5^6 \cdot (1 - 0.5)^{10-6} = 210 \cdot 0.5^{10} = 0.205078\dots;$$

$$P_7 = 120 \cdot 0.5^7 \cdot (1 - 0.5)^{10-7} = 120 \cdot 0.5^{10} = 0.117188\dots;$$

$$P_8 = 45 \cdot 0.5^8 \cdot (1 - 0.5)^{10-8} = 45 \cdot 0.5^{10} = 0.043945\dots;$$

$$P_9 = 10 \cdot 0 \dots 5^9 \cdot (1 - 0.5)^{10-9} = 10 \cdot 0.5^{10} = 0.009766\dots;$$

$$P_{10} = 1 \cdot 0 \dots 5^{10} \cdot (1 - 0.5)^{10-10} = 1 \cdot 0.5^{10} = 0.000977\dots$$

Зрозуміло, $P_0 + P_1 + P_2 + P_3 + P_4 + P_5 + P_6 + P_7 + P_8 + P_9 + P_{10} = 1$.

Відіб'ємо на графіку величини $P_0, P_1, P_2, P_3, \dots, P_{10}$ (см. мал. 27.2).

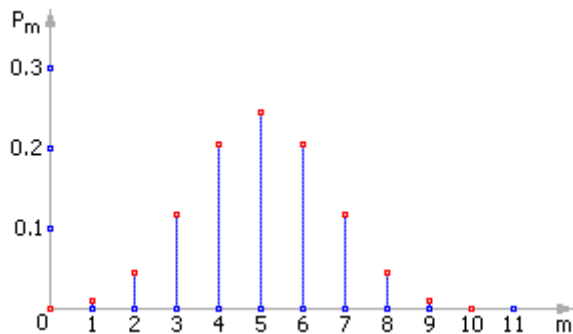


Рис. 27.2. **Графік** біноміального розподілу при параметрах $p = 0.5$ й $n = 10$, що наближають його до нормального закону

Отже, при умовах $m \approx n/2$ й $p \approx 1 - p$ або $p \approx 0.5$ замість біноміального розподілу можна використати нормальне. При більших значеннях n графік зрушується вправо й стає усе більше пологим, тому що математичне очікування й дисперсія зростають зі збільшенням n : $M = n \cdot p$, $D = n \cdot p \cdot (1 - p)$.

До речі, біноміальний закон прагне до нормального й при збільшенні n , що цілком природно, відповідно до центральної граничної теореми.

Тепер розглянемо, як зміниться біноміальний закон у випадку, коли $p \neq q$, тобто $p \rightarrow 0$. У цьому випадку застосувати гіпотезу про нормальність розподілу не можна, і біноміальний розподіл переходить у розподіл Пуасона.

Розподіл Пуасона

Розподіл Пуасона — це окремий випадок біноміального розподілу (при $n \gg 0$ і при $p \rightarrow 0$ (рідкі події)).

З математики відома формула, що дозволяє приблизно підрахувати значення будь-якого члена біноміального розподілу:

$$P_m = C_n^m \cdot p^m \cdot (1-p)^{n-m} \cong \frac{a^m \cdot e^{-a}}{m!}$$

де $a = n \cdot p$ — параметр Пуасона (математичне очікування), а дисперсія дорівнює математичному очікуванню.

Приведемо математичні викладення, що пояснюють цей перехід. Біноміальний закон розподілу

$$P_m = C_n^m \cdot p^m \cdot (1-p)^{n-m}$$

може бути написаний, якщо покласти $p = a/n$, у вигляді

$$P_m = C_n^m \cdot \left(\frac{a}{n}\right)^m \cdot \left(1 - \frac{a}{n}\right)^{n-m}$$

або

$$P_m = \left(1 - \frac{1}{n}\right) \cdot \left(1 - \frac{2}{n}\right) \cdot \dots \cdot \left(1 - \frac{m-1}{n}\right) \cdot \left(1 - \frac{a}{n}\right)^{-m} \cdot \left(1 - \frac{a}{n}\right)^n \cdot \frac{a^m}{m!}$$

Тому що p дуже мало, те варто брати до уваги тільки числа m , малі в порівнянні з n . Добуток

$$\left(1 - \frac{1}{n}\right) \cdot \left(1 - \frac{2}{n}\right) \cdot \dots \cdot \left(1 - \frac{m-1}{n}\right)$$

досить близько до одиниці. Це ж ставиться до величини

$$\left(1 - \frac{a}{n}\right)^{-m}$$

Величина

$$\left(1 - \frac{a}{n}\right)^n$$

дуже близька до e^{-a} . Звідси одержуємо формулу:

$$P_m = C_n^m \cdot p^m \cdot (1-p)^{n-m} \cong \frac{a^m \cdot e^{-a}}{m!}$$

Приклад. У ящику перебуває $n = 100$ деталей, як якісних, так і бракованих. Імовірність дістати бракований виріб становить $p = 0.01$. Допустимо, що ми виймаємо виріб, визначаємо, браковане воно чи ні, і кладемо його обернено. Надходячи таким чином, вийшло, що з 100 виробів, які ми перебрали, два виявилися бракованими. Яка ймовірність цього?

По біноміальному розподілі одержуємо:

$$\begin{aligned} P_2 &= C_n^m \cdot p^m \cdot (1 - p)^{n - m} = C_{100}^2 \cdot 0.01^2 \cdot (1 - 0.01)^{100 - 2} = \\ &= \frac{100!}{2!} \cdot \frac{1}{98!} \cdot 0.01^2 \cdot 0.99^{98} \cong 0.185 \end{aligned}$$

По розподілі Пуасона одержуємо:

$$P_2 = \frac{a^m \cdot e^{-a}}{m!} = \frac{(n \cdot p)^m \cdot e^{-n \cdot p}}{m!} = \frac{(100 \cdot 0.01)^2 \cdot e^{-100 \cdot 0.01}}{2!} \cong 0.184$$

Як видно, величини вийшли близькими, тому у випадку рідких подій цілком припустимо застосовувати закон Пуасона, тим більше що він вимагає менших обчислювальних витрат.

Покажемо графічно вид закону Пуасона. Візьмемо для прикладу параметри $p = 0.05$, $n = 10$. Тоді:

$$C_{10}^0 = 1, C_{10}^1 = 10, C_{10}^2 = 45, C_{10}^3 = 120, C_{10}^4 = 210, C_{10}^5 = 252,$$

$$C_{10}^6 = 210, C_{10}^7 = 120, C_{10}^8 = 45, C_{10}^9 = 10, C_{10}^{10} = 1;$$

$$P_0 = 1 \cdot 0.05^0 \cdot (1 - 0.05)^{10-0} = 1 \cdot 1 \cdot 0.95^{10} = 0.5987\dots;$$

$$P_1 = 10 \cdot 0.05^1 \cdot (1 - 0.05)^{10-1} = 10 \cdot 0.05^1 \cdot 0.95^9 = 0.3151\dots;$$

$$P_2 = 45 \cdot 0.05^2 \cdot (1 - 0.05)^{10-2} = 45 \cdot 0.05^2 \cdot 0.95^8 = 0.0746\dots;$$

$$P_3 = 120 \cdot 0.05^3 \cdot (1 - 0.05)^{10-3} = 120 \cdot 0.05^3 \cdot 0.95^7 = 0.0105\dots;$$

$$P_4 = 210 \cdot 0.05^4 \cdot (1 - 0.05)^{10-4} = 210 \cdot 0.05^4 \cdot 0.95^6 = 0.00096\dots;$$

$$P_5 = 252 \cdot 0.05^5 \cdot (1 - 0.05)^{10-5} = 252 \cdot 0.05^5 \cdot 0.95^5 = 0.00006\dots;$$

$$P_6 = 210 \cdot 0.05^6 \cdot (1 - 0.05)^{10-6} = 210 \cdot 0.05^6 \cdot 0.95^4 = 0.0000\dots;$$

$$P_7 = 120 \cdot 0.05^7 \cdot (1 - 0.05)^{10-7} = 120 \cdot 0.05^7 \cdot 0.95^3 = 0.0000\dots;$$

$$P_8 = 45 \cdot 0.05^8 \cdot (1 - 0.05)^{10-8} = 45 \cdot 0.05^8 \cdot 0.95^2 = 0.0000\dots;$$

$$P_9 = 10 \cdot 0.05^9 \cdot (1 - 0.05)^{10-9} = 10 \cdot 0.05^9 \cdot 0.95^1 = 0.0000\dots;$$

$$P_{10} = 1 \cdot 0.05^{10} \cdot (1 - 0.05)^{10-10} = 1 \cdot 0.05^{10} \cdot 0.95^0 = 0.0000\dots$$

Зрозуміло, $P_0 + P_1 + P_2 + P_3 + P_4 + P_5 + P_6 + P_7 + P_8 + P_9 + P_{10} = 1$.

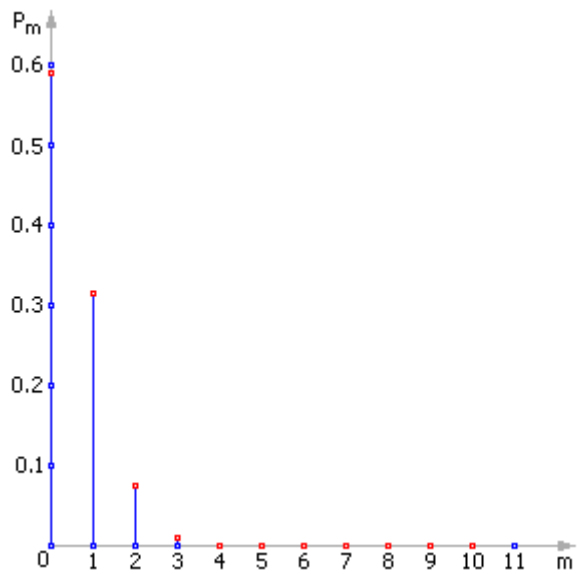


Рис. 27.3. **Графік** розподілу Пуасона при $p = 0.05$ й $n = 10$

При $n \rightarrow \infty$ розподіл Пуасона переходить у нормальний закон, відповідно до центральної граничної теореми.