4. Независимые испытания. Схема Бернулли

Допустим, что проводится серия из n независимых экспериментов (испытаний), в каждом из которых с вероятностью p, $0 < p < 1$, может наступить и с вероятностью $q = 1 - p$ может не наступить некоторое событие A. Появление события A обычно называют ”<успехом”>, а не появление — ”<неудачей”>. Основной вопрос в схеме Бернулли состоит в вычислении вероятности события $B_n(k)$, которое заключается в том, что в проведенной серии испытаний произошло ровно k успехов, $k = 0, 1, \ldots, n$.

Для решения поставленной задачи построим вероятностное пространство $(\Omega, \mathcal{A}, \mathbb{P})$, соответствующее схеме Бернулли. Каждый результат серии экспериментов можно представить n-мерной точкой $\omega = (\delta_1, \ldots, \delta_n)$, где δ_i принимает значение 1 в случае успеха в i-том испытании и значение 0 в случае неуспеха. В силу независимости испытаний вероятность элементарного события $\omega = (\delta_1, \ldots, \delta_n)$ будет определяться равенством

$$\mathbb{P}(\{\omega\}) = p^{\sum_{i=1}^n \delta_i} q^{n - \sum_{i=1}^n \delta_i}. \tag{4.1}$$

Если $\omega \in B_n(k)$, то из (4.1) следует, что $\mathbb{P}(\{\omega\}) = p^k q^{n-k}$. Замечая теперь, что $B_n(k)$ состоит из C_n^k элементарных событий, приходим к формуле

$$\mathbb{P}(B_n(k)) = C_n^k p^k q^{n-k}. \tag{4.2}$$

События $B_n(k)$, $k = 0, 1, \ldots, n$, образуют разбиение. Это подтверждается равенством

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

Формула (4.2) выражает распределение вероятностей числа успехов в n независимых испытаниях. Это распределение называют *биномиальным.*

При больших значениях n реализация формулы (4.2) сопряжена с трудоемкими вычислениями. Поэтому её пытаются заменить приближенными формулами. Следующий результат относится к случаю, когда p мало, а n велико. В связи с малостью p этот результат иногда называют законом редких событий.

Теорема 4.1 *[Пуассона.] Если $n \to \infty$ и $p \to 0$ так, что $np \to \lambda$, $0 < \lambda < \infty$, то при всех $k = 0, 1, 2, \ldots$ выполняется соотношение*

$$\mathbb{P}(B_n(k)) = C_n^k p^k q^{n-k} \to \frac{\lambda^k}{k!} e^{-\lambda}.$$

Доказательство *Из преобразований*

$$\begin{aligned}\frac{n!}{k!(n-k)!}p^k q^{n-k} &= \frac{1}{k!}\frac{(n-1)\cdot\ldots\cdot(n-k+1)}{n^{k-1}}n^k p^k (1-p)^{n-k}\\ &= \frac{(np)^k}{k!}\left(1-\frac{1}{n}\right)\cdot\ldots\cdot\left(1-\frac{k-1}{n}\right)(1-p)^{-k}(1-p)^{\frac{1}{p}np}\end{aligned}$$

и предельного соотношения

$$\lim_{p\to 0}(1-p)^{1/p} = e^{-1}$$

следует утверждение теоремы. □

Заметим, что в условиях теоремы Пуассона 4.1 имеет место оценка

$$\left|\mathbb{P}(B_n(k)) - \frac{\lambda^k}{k!}e^{-\lambda}\right| \le \frac{\lambda^2}{n}, \qquad \lambda = np,$$

т. е. результат применим, когда np^2 мало.

Доказанная теорема относится к так называемым предельным теоремам в схеме Бернулли. Распределение вероятностей

$$p_k = e^{-\lambda}\frac{\lambda^k}{k!}, \qquad k = 0,1,2,\ldots,$$

к которому стремится в условиях теоремы биномиальное распределение, называют *пуассоновским распределением.* В случае, когда вероятность успеха фиксирована, а число экспериментов стремится к бесконечности, используют предельные теоремы Муавра"—Лапласа.

Теорема 4.2 *[Локальная теорема Муавра-Лапласа.] Пусть в схеме Бернулли $0 < p < 1$ фиксировано, $\sigma = \sqrt{npq}$, $x = x(k) = (k-np)/\sigma$, $k = 0,1,2,\ldots$. Тогда для любого $M > 0$ равномерно по всем k таким, что $|x(k)| \le M$, при $n \to \infty$ имеет место асимптотическое равенство*

$$\mathbb{P}(B_n(k)) = \frac{1}{\sigma\sqrt{2\pi}}e^{-x^2/2}(1+o(1)).$$

Доказательство *В основе доказательства теоремы лежит формула Стирлинга*

$$n! = \sqrt{2\pi n}\cdot n^n \cdot e^{-n}\cdot e^{\frac{\theta}{12n}}, \qquad 0 < \theta < 1.$$

Из этой формулы следует, что

$$\ln n! = \ln\sqrt{2\pi n} + n\ln n - n + O\left(\frac{1}{n}\right).$$

Пусть k таково, что $|x(k)| \le M$. Тогда из равенства

$$k \,=\, np + x\sigma \,=\, np\left(1 + \frac{xq}{\sigma}\right)$$

следует, что $k \to \infty$ при $n \to \infty$ ($\sigma \to \infty$). Аналогично, из равенства

$$n - k \,=\, nq - x\sigma \,=\, nq\left(1 - \frac{xp}{\sigma}\right)$$

видно, что $(n-k) \to \infty$ при $n \to \infty$. Следовательно, к $(n-k)!$ и $k!$ также можно применить формулу Стирлинга. Кроме того,

$$\frac{1}{n} \,=\, O\left(\frac{1}{\sigma^2}\right), \quad \frac{1}{k} \,=\, O\left(\frac{1}{\sigma^2}\right), \quad \frac{1}{n-k} \,=\, O\left(\frac{1}{\sigma^2}\right).$$

Таким образом,

$$\begin{aligned}
\ln \mathbb{P}(B_n(k)) \,&=\, \ln n! - \ln k! - \ln(n-k)! + k\ln p + (n-k)\ln q \\
&=\, \frac{1}{2}\ln\frac{n}{2\pi k(n-k)} + n\ln n - k\ln k - (n-k)\ln(n-k) \\
&+ k\ln p + (n-k)\ln q + O\left(\frac{1}{\sigma^2}\right) \\
&=\, \frac{1}{2}\left(\ln\frac{1}{2\pi} + \ln\frac{n}{k(n-k)}\right) - k\ln\frac{k}{np} - (n-k)\ln\frac{n-k}{nq} + O\left(\frac{1}{\sigma^2}\right).
\end{aligned}$$

Далее,

$$\ln\frac{n}{k(n-k)} \,=\, \ln\frac{1}{npq} - \ln\left(1 + \frac{xq}{\sigma}\right) - \ln\left(1 - \frac{xp}{\sigma}\right) \,=\, 2\ln\frac{1}{\sigma} + O\left(\frac{1}{\sigma}\right),$$

а также

$$k\ln\frac{k}{np} + (n-k)\ln\frac{n-k}{nq} \,=\, (np+x\sigma)\ln\left(1 + \frac{xq}{\sigma}\right) + (nq-x\sigma)\ln\left(1 - \frac{xp}{\sigma}\right) \,=$$

$$=\, (np+x\sigma)\left(\frac{xq}{\sigma} - \frac{x^2q^2}{2\sigma^2} + o\left(\frac{1}{\sigma^2}\right)\right) + (nq-x\sigma)\left(-\frac{xp}{\sigma} - \frac{x^2p^2}{2\sigma^2} + o\left(\frac{1}{\sigma^2}\right)\right) \,=$$

$$=\, x\sigma - \frac{x^2q}{2} + x^2q - x\sigma - \frac{x^2p}{2} + x^2p + O\left(\frac{1}{\sigma}\right) \,=\, \frac{x^2}{2} + O\left(\frac{1}{\sigma}\right).$$

В результате получаем

$$\ln \mathbb{P}(B_n(k)) \,=\, \ln\frac{1}{\sigma\sqrt{2\pi}} - \frac{x^2}{2} + O\left(\frac{1}{\sigma}\right),$$

что и доказывает теорему. □

Теорема 4.3 *[Интегральная теорема Муавра-Лапласа.] Пусть в схеме Бернулли $0 < p < 1$ фиксировано и S_n — число успехов в серии из n независимых экспериментов. Тогда для $-\infty < a < b < +\infty$ при $n \to \infty$ имеет место предельное соотношение*

$$\lim_{n\to\infty} \mathbb{P}\left(a \le \frac{S_n - np}{\sqrt{npq}} \le b\right) = \frac{1}{\sqrt{2\pi}} \int\limits_a^b e^{-\frac{x^2}{2}} dx.$$

Доказательство этой теоремы следует из центральной предельной теоремы, которая будет рассмотрена позже.

Предельные теоремы Пуассона и Муавра"—Лапласа используют для приближённых вычислений вероятностей $\mathbb{P}(S_n = k)$ и $\mathbb{P}(k_1 \le S_n \le k_2)$ в схеме Бернулли, когда n велико. В теореме Пуассона появилось распределение Пуассона: $p_k = e^{-\lambda}\lambda^k/k!$, $k = 0, 1, 2, \ldots$. В теореме Муавра"—Лапласа возникает нормальное распределение, которое описывается плотностью $\varphi(x) = e^{-x^2/2}/\sqrt{2\pi}$. Для распределения Пуассона и интеграла Лапласа

$$\Phi_0(x) = \frac{1}{\sqrt{2\pi}} \int\limits_0^x e^{-\frac{u^2}{2}} du$$

имеются таблицы.

Допустим, что нам нужно вычислить вероятность $\mathbb{P}(k_1 \le S_n \le k_2)$ в схеме Бернулли с n независимыми испытаниями и вероятностью p успеха в отдельном испытании. Если n велико, то можно использовать интегральную теорему Муавра"—Лапласа следующим образом:

$$\mathbb{P}(k_1 \le S_n \le k_2) = \mathbb{P}\left(\frac{k_1 - np}{\sqrt{npq}} \le \frac{S_n - np}{\sqrt{npq}} \le \frac{k_2 - np}{\sqrt{npq}}\right) \approx \Phi(x_2) - \Phi(x_1),$$

где

$$x_1 = \frac{k_1 - np}{\sqrt{npq}}, \quad x_2 = \frac{k_2 - np}{\sqrt{npq}}, \quad \Phi(x) = \frac{1}{\sqrt{2\pi}} \int\limits_{-\infty}^x e^{-\frac{u^2}{2}} du.$$

Если считать, что $\Phi_0(x) = -\Phi_0(|x|)$ при отрицательных x, то $\Phi(x_2) - \Phi(x_1) = \Phi_0(x_2) - \Phi_0(x_1)$ для любых значений x_1, x_2. Это следует из чётности функции $\varphi(x)$.

Полиномиальная схема является обобщением схемы Бернулли. Здесь результатом каждого испытания может быть один из r взаимоисключающих исходов $A_1, \ldots, A_r$ с вероятностями появления $p_1, \ldots, p_r$, соответственно, $p_1 + \ldots + p_r = 1$. Элементарное событие, соответствующее серии независимых n испытаний, можно представить в виде $\omega = (\delta_1, \ldots, \delta_n)$, где каждое δ_i может принимать одно из значений $1, \ldots, r$.

Вероятность элементарного события в силу независимости испытаний будет определяться равенством

$$\mathbb{P}(\{\omega\}) = p_{\delta_1} \cdot \ldots \cdot p_{\delta_n}.$$

Подобно событию $B_n(k)$, которое рассматривалось в схеме Бернулли, в полиномиальной схеме вводится $B_n(k_1, \ldots, k_r)$ — событие, состоящее в том, что в серии из n экспериментов произошло k_1 исходов с номером 1, ..., k_r исходов с номером r, $k_1 + \ldots + k_r = n$. Из общей формулы видно, что если $\omega \in B_n(k_1, \ldots, k_r)$, то

$$\mathbb{P}(\{\omega\}) = p_1^{k_1} \cdot \ldots \cdot p_r^{k_r}.$$

Используя основное правило комбинаторики, легко подсчитывается количество элементарных событий в $B_n(k_1, \ldots, k_r)$:

$$C_n^{k_1} \cdot C_{n-k_1}^{k_2} \cdot \ldots \cdot C_{k_r}^{k_r} = \frac{n!}{k_1! \cdot \ldots \cdot k_r!}.$$

Таким образом, мы приходим к основному результату полиномиальной схемы — *полиномиальному распределению*

$$\mathbb{P}(B_n(k_1, \ldots, k_r)) = \frac{n!}{k_1! \cdot \ldots \cdot k_r!} p_1^{k_1} \cdot \ldots \cdot p_r^{k_r}, \qquad k_1 + \ldots + k_r = n.$$

В случае $r = 2$ мы снова получаем схему Бернулли.