

Лекция 4

Серия испытаний Бернулли

Схемой Бернулли - называется серия одинаковых независимых экспериментов, каждый из которых имеет 2 исхода: произошло интересующее нас событие или нет

$p = p(A)$ - вероятность успеха при одном испытании

$q = 1 - p$ - вероятность неудачи

v_n - число успехов в серии из n испытаний

$p(v_n = k) = p_n(k)$

Из этого получаем формулу Бернулли:

Th. Вероятность того, что при n испытаниях произойдет ровно k успехов, равна

$$p_n(k) = C_n^k p^k q^{n-k}$$

□

Рассмотрим один из элементарных исходов, благоприятных данному событию:

$A_n = \underbrace{УУУ \dots УН \dots ННН}_k - k \text{ успехов, } n - k \text{ неудачи}$

$$p(Y) = p, p(H) = q$$

Так как испытания независимы, то $p(A_n) = p^k q^{n-k}$

Остальные элементарные исходы имеют ту же вероятность, перебираем все расстановки исходов, получаем C_n^k , в итоге, получаем формулу Бернулли

□

Ex. Вероятность попадания стрелка при одном выстреле - 0.8. Какова вероятность того, что из пяти выстрелов точными будут три

$$n = 5 \quad p = 0.8 \quad q = 1 - p = 0.2 \quad k = 3$$

$$p_5(3) = C_5^3 p^3 q^2 = 0.2048$$

Наиболее вероятное число успехов

Выясним, при каком значении k вероятность предшествующего числа успехов $k - 1$ будет не более, чем вероятность k успехов

$$p_n(k - 1) \leq p_n(k)$$

$$C_n^{k-1} p^{k-1} q^{n-k+1} \leq C_n^k p^k q^{n-k}$$

$$\frac{n!}{(k-1)!(n-k+1)!} q \leq \frac{n!}{(k)!(n-k)!} p$$

$$\frac{q}{(k-1)!(n-k+1)!} \leq \frac{p}{(k)!(n-k)!}$$

$$\frac{q}{n-k+1} \leq \frac{p}{k}$$

$$k(1-p) \leq p(n-k+1)$$

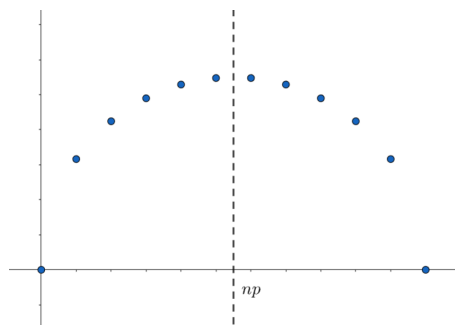
$$k \leq np + p$$

Отсюда $np + p - 1 \leq k \leq np + p$

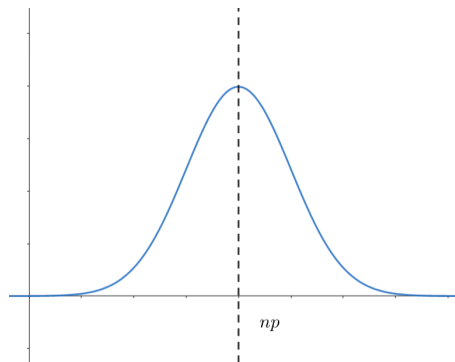
Рассмотрим 3 ситуации:

- 1) np - целое, тогда $np + p$ - нецелое, и $k = np$ - наиболее вероятное
- 2) $np + p$ - нецелое, тогда $k = \lfloor np + p \rfloor$
- 3) $np + p$ - целое, тогда $np + p - 1$ - целое, и 2 наиболее вероятных числа успеха

Геометрическая интерпретация:



При увеличении числа n точки превращаются в кривую Гаусса



При увеличении числа испытаний n формула Бернулли вырождается в следующие асимптотические формы (применяем, если требуется найти вероятность точного числа успеха)

- 1) локальная формула Муавра-Лапласа

$$p_n(k) \xrightarrow{n \rightarrow \infty} \frac{1}{\sqrt{npq}} \varphi(x), \text{ где } \varphi(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} - \text{функция гаусса}$$

$$x = \frac{k - np}{\sqrt{npq}}$$

Свойства $\varphi(x)$:

1. $\varphi(x) = \varphi(-x)$ - функция четная
2. при $x > 5$ $\varphi(x) \approx 0$

2) Интегральная формула Муавра-Лапласа (если требуется найти вероятность того, что число успехов в данном диапазоне)

$$p_n(k_1 \leq k \leq k_2) \xrightarrow{n \rightarrow \infty} \Phi(x_2) - \Phi(x_1), \text{ где } \Phi(x) = \frac{1}{\sqrt{2\pi}} \int_0^x e^{-\frac{z^2}{2}} dz - \text{функция Лапласа}$$

$$x_1 = \frac{k_1 - np}{\sqrt{npq}} - \text{отклонение от левой границы, } x_2 = \frac{k_2 - np}{\sqrt{npq}} - \text{отклонение от правой}$$

Свойства $\Phi(x)$

1. $\Phi(-x) = -\Phi(x)$ - функция нечетная

2. при $x > 5$ $\Phi(x) \approx 0.5$

Nota. Эти формулы обычно можно применять при $n \geq 100$ и $0.1 \leq p \leq 0.9$

Nota. В некоторых источниках под функцией Лапласа подразумевают другую функцию: $F_0(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^x e^{-\frac{t^2}{2}} dt$ - стандартное отклонение. Эта функция отличается от $F_0 = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^0 e^{-\frac{t^2}{2}} dt +$

$$\Phi(x) = \frac{1}{2} + \Phi(x)$$

Так как $\int_{-\infty}^{\infty} e^{-x^2} dx = \sqrt{\pi}$ - интеграл Пуассона

Ex. Вероятность попадания стрелка в цель 0.8, стрелок сделал 400 выстрелов. Найти вероятность того, что:

а) произошло ровно 330 попаданий

б) произошло от 312 до 336 попаданий

$$\text{а) } x = \frac{k - np}{\sqrt{npq}} = \frac{330 - 400 \cdot 0.8}{\sqrt{400 \cdot 0.8 \cdot 0.2}} = \frac{330 - 320}{8} = 1.25$$

$$p_{400}(330) \approx \frac{1}{\sqrt{npq}} \varphi(1.25) = \frac{1}{8} \varphi(1.25) \approx \frac{1}{8} \cdot 0.1826 \approx 0.0228$$

$$\text{б) } x_1 = \frac{312 - 320}{8} = -1, x_2 = \frac{336 - 320}{8} = 2$$

$$p_{400}(312 \leq k \leq 336) \approx \Phi(2) - \Phi(-1) = \Phi(2) + \Phi(1) \approx 0.4772 + 0.3413 = 0.8185$$

Статистическое понятие вероятности

Пусть проводим n реальных экспериментов, n_A - число появления события A , $\frac{n_A}{n}$ - относительная частота события A .

Эксперименты с монетой показали, что при больших n , $\frac{n_A}{n} \approx p(A)$ - явление стабилизации

Вероятность отклонения относительной частоты от вероятности события

n - число испытаний, $p = p(A)$, $\frac{n_A}{n}$ - экспериментальная частота

$$p\left(\left|\frac{n_A}{n} - p\right| \leq \varepsilon\right) = p\left(-\varepsilon \leq \frac{n_A}{n} - p \leq \varepsilon\right) = p(-n\varepsilon \leq n_A - np \leq n\varepsilon) = p(np - n\varepsilon \leq n_A \leq n\varepsilon + np) \xrightarrow{n \rightarrow \infty} [\text{по интегральной формуле Лапласа}] \xrightarrow{n \rightarrow \infty} \Phi\left(\frac{n\varepsilon}{\sqrt{npq}}\right) - \Phi\left(-\frac{n\varepsilon}{\sqrt{npq}}\right)$$

$$= \Phi\left(\frac{\sqrt{n}\varepsilon}{\sqrt{pq}}\right) - \Phi\left(-\frac{\sqrt{n}\varepsilon}{\sqrt{pq}}\right)$$

$$= 2\Phi\left(\frac{\sqrt{n}\varepsilon}{\sqrt{pq}}\right)$$

Итак, получили, что нужная нам вероятность $p\left(\left|\frac{n_A}{n} - p\right| \leq \varepsilon\right) \approx 2\Phi\left(\frac{\sqrt{n}\varepsilon}{\sqrt{pq}}\right)$

Закон больших чисел Бернулли

Итак, $p\left(\left|\frac{n_A}{n} - p\right| \leq \varepsilon\right) \xrightarrow{n \rightarrow \infty} 2\Phi\left(\frac{\varepsilon}{\sqrt{pq}}\sqrt{n}\right)$

при $n \rightarrow \infty$, $\sqrt{n} \rightarrow \infty$, $\frac{\varepsilon}{\sqrt{pq}}\sqrt{n} \rightarrow \infty$, $\Phi\left(\frac{\varepsilon}{\sqrt{pq}}\sqrt{n}\right) \rightarrow 0.5$, $p\left(\left|\frac{n_A}{n} - p\right| \leq \varepsilon\right) \rightarrow 2 \cdot 0.5 = 1$ - закон больших чисел показывает, что вероятность попадания относительной частоты в ε -трубу вероятность события приближается к 1

$\lim_{n \rightarrow \infty} p\left(\left|\frac{n_A}{n} - p\right| \leq \varepsilon\right) = 1$ или $\frac{n_A}{n} \xrightarrow{n \rightarrow \infty} p$ - сходимости по вероятности

Ех. Для оценки доли p курящих людей берется выборка объема n , и делается оценка доли курящих людей по формуле $p^* = \frac{n_A}{n}$. Каким должен быть объем n , чтобы с вероятностью $\gamma = 0.95$ данная оценка отличалась от истинного значения не более, чем на $\varepsilon = 0.01$

По формуле вероятности отклонения частоты от вероятности $p(|p^* - p| \leq \varepsilon) = p\left(\left|\frac{n_A}{n} - p\right| \leq \varepsilon\right) \approx$

$$2\Phi\left(\frac{\varepsilon}{\sqrt{pq}}\sqrt{n}\right) = 0.95$$

$$\Phi\left(\frac{\varepsilon}{\sqrt{pq}}\sqrt{n}\right) = 0.475$$

$$\frac{\varepsilon}{\sqrt{pq}}\sqrt{n} = 1.96$$

$$\frac{1}{\sqrt{pq}}\sqrt{n} = 196$$

$$\frac{n}{pq} = 38416$$

$$n \geq 38416pq$$

В самой худшей ситуации $pq \leq 0.5^2 = \frac{1}{4}$

$$n \geq \frac{38416}{4} = 9604$$