# Математические основы искусственного интеллекта
# Нормальное распределение

Солодушкин Святослав Игоревич

Кафедра вычислительной математики и компьютерных наук,
УрФУ имени первого Президента России Б.Н. Ельцина

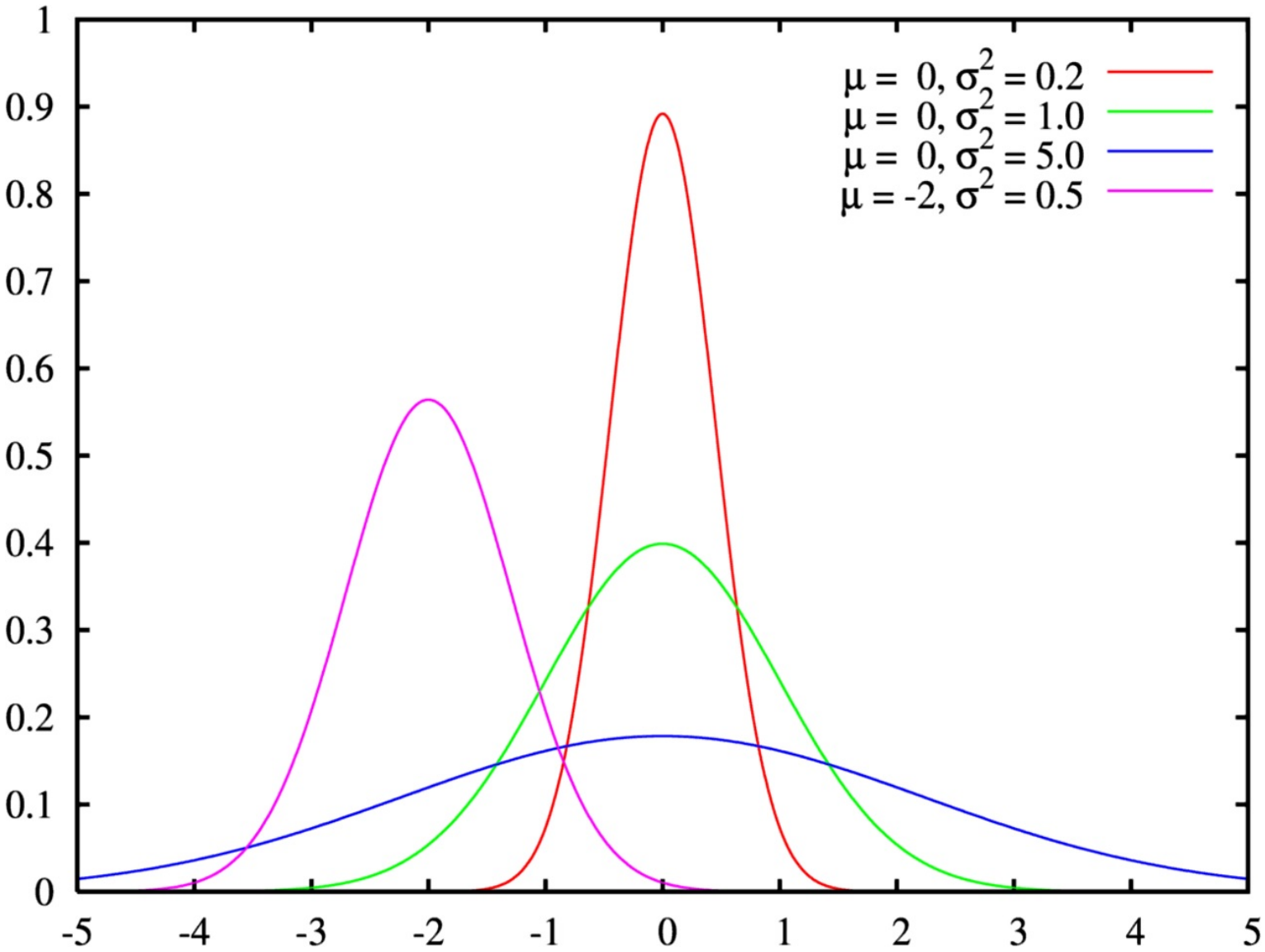
Ноябрь 2021

## Определение

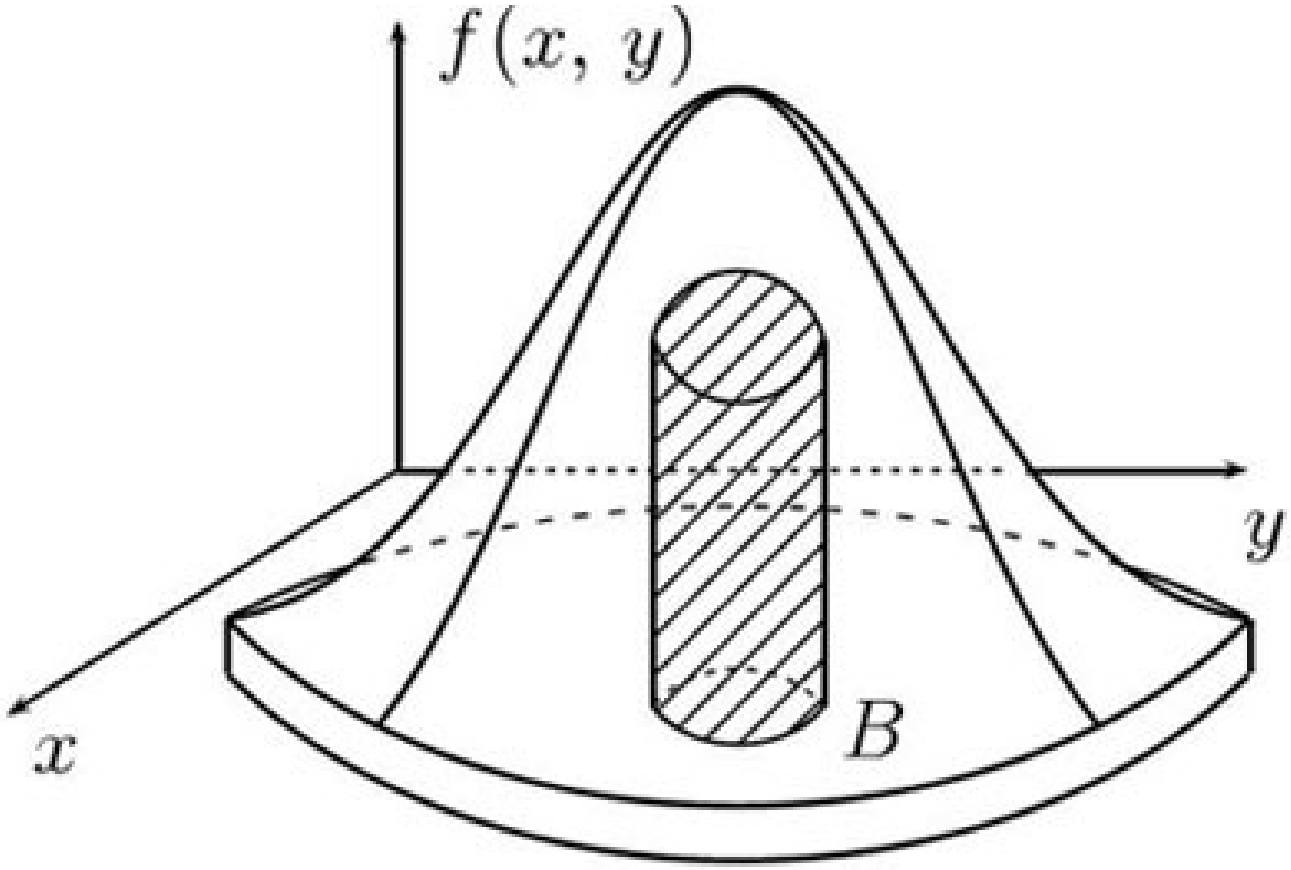
Говорят, что $\xi$ имеет нормальное (гауссовское) распределение с параметрами $m$ и $\sigma^2$, где $m \in \mathbb{R}, \sigma > 0$, и пишут: $\xi \sim \mathrm{N}_{m,\,\sigma^2}$, если $\xi$ имеет следующую плотность распределения:

$$f_\xi(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-m)^2}{2\sigma^2}}, \quad x \in \mathbb{R}.$$

# Плотность распределения



# Плотность распределения



# Интеграл Пуассона и функция Лапласа

Интеграл Пуассона

$$I = \int_{-\infty}^{\infty} e^{-x^2/2}\, dx = \sqrt{2\pi}.$$

Функция Лапласа

$$\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{0}^{x} e^{-t^2/2}\, dt.$$

В частности,

$$\Phi(5) = \frac{1}{\sqrt{2\pi}} \int_{0}^{5} e^{-t^2/2}\, dt \approx 0.5.$$

# Функция Лапласа. Таблица

Значение функции $\Phi_0(x) = \frac{1}{\sqrt{2\pi}} \int_0^x e^{-\frac{t^2}{2}}\, dt$

| $x$ | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 |
|---|---|---|---|---|---|---|---|---|---|---|
| | Сотые доли $x$ | | | | | | | | | |
| 0,0 | 0,0000 | 0040 | 0080 | 0112 | 0160 | 0199 | 0239 | 0279 | 0319 | 0359 |
| 0,1 | 0398 | 0438 | 0478 | 0517 | 0557 | 0596 | 0636 | 0675 | 0714 | 0754 |
| 0,2 | 0793 | 0832 | 0871 | 0910 | 0948 | 0987 | 1026 | 1064 | 1103 | 1141 |
| 0,3 | 1179 | 1217 | 1255 | 1293 | 1331 | 1368 | 1406 | 1443 | 1480 | 1517 |
| 0,4 | 1554 | 1591 | 1628 | 1664 | 1700 | 1736 | 1772 | 1808 | 1844 | 1879 |
| 0,5 | 1915 | 1950 | 1985 | 2019 | 2054 | 2088 | 2123 | 2157 | 2190 | 2224 |
| 0,6 | 2258 | 2291 | 2324 | 2357 | 2389 | 2422 | 2454 | 2486 | 2518 | 2549 |
| 0,7 | 2580 | 2612 | 2642 | 2673 | 2704 | 2734 | 2764 | 2794 | 2823 | 2852 |
| 0,8 | 2881 | 2910 | 2939 | 2967 | 2996 | 3023 | 3051 | 3079 | 3106 | 3133 |
| 0,9 | 3159 | 3186 | 3212 | 3238 | 3264 | 3289 | 3315 | 3340 | 3365 | 3389 |
| 1,0 | 3413 | 3438 | 3461 | 3485 | 3508 | 3531 | 3553 | 3577 | 3599 | 3621 |
| 1,1 | 3643 | 3665 | 3686 | 3708 | 3729 | 3749 | 3770 | 3790 | 3810 | 3830 |
| 1,2 | 3849 | 3869 | 3888 | 3907 | 3925 | 3944 | 3962 | 3980 | 3997 | 4015 |
| 1,3 | 4032 | 4049 | 4066 | 4082 | 4099 | 4115 | 4131 | 4147 | 4162 | 4177 |
| 1,4 | 4192 | 4207 | 4222 | 4236 | 4251 | 4265 | 4279 | 4292 | 4306 | 4319 |
| 1,5 | 4332 | 4345 | 4357 | 4370 | 4382 | 4394 | 4406 | 4418 | 4430 | 4441 |
| 1,6 | 4452 | 4463 | 4474 | 4485 | 4495 | 4505 | 4515 | 4525 | 4535 | 4545 |
| 1,7 | 4554 | 4564 | 4573 | 4582 | 4591 | 4599 | 4608 | 4616 | 4625 | 4633 |
| 1,8 | 4641 | 4649 | 4656 | 4664 | 4671 | 4678 | 4686 | 4693 | 4700 | 4706 |
| 1,9 | 4713 | 4719 | 4726 | 4732 | 4738 | 4744 | 4750 | 4756 | 4762 | 4767 |
| | Десятые доли $x$ | | | | | | | | | |
| 2, | 4773 | 4821 | 4861 | 4893 | 4918 | 4938 | 4953 | 4965 | 4974 | 4981 |
| 3, | 4987 | 4990 | 4993 | 4995 | 4997 | 4998 | 4998 | 4999 | 4999 | 5000 |

# Вероятность попадания в заданный интервал

$$\mathrm{P}(a < \xi < b) = \frac{1}{\sigma\sqrt{2\pi}} \int_a^b e^{-(x-m)^2/2\sigma^2} dx = \frac{1}{\sqrt{2\pi}} \int_{(a-m)/\sigma}^{(b-m)/\sigma} e^{-z^2/2} dz =$$

$$= \frac{1}{\sqrt{2\pi}} \int_0^{(b-m)/\sigma} e^{-z^2/2} dz - \frac{1}{\sqrt{2\pi}} \int_{(a-m)/\sigma}^0 e^{-z^2/2} dz =$$

$$= \Phi\left(\frac{b-m}{\sigma}\right) - \Phi\left(\frac{a-m}{\sigma}\right).$$

Замена переменных $z = \frac{x-m}{\sigma}$, $\sigma\, dz = dx$.

# Вероятность попадания в заданный интервал

Случайная величина $\xi$ имеет нормальное распределение с параметрами $m = 30$, $\sigma = 10$. Найти вероятность того, что $\xi$ примет значение из промежутка $(10, 50)$.

$$\mathrm{P}(a < \xi < b) = \frac{1}{\sigma\sqrt{2\pi}} \int_a^b e^{-(x-m)^2/2\sigma^2} dx =$$

$$= \Phi\left(\frac{b-m}{\sigma}\right) - \Phi\left(\frac{a-m}{\sigma}\right).$$

$$\mathrm{P}(10 < \xi < 50) = \Phi\left(\frac{50-30}{10}\right) - \Phi\left(\frac{10-30}{10}\right) =$$

$$= \Phi(2) - \Phi(-2) = 2\,\Phi(2) = 20.4773 = 0.9546.$$

# Веротятность заданого отклонения

Необходимо вычислить вероятность того, что случайная величина $\xi$ отклонится от своего мет. ожидания не более чем на $\delta$.
$-\delta < \xi - m < \delta$

$$\mathrm{P}(-\delta+m < \xi < \delta+m) = \Phi\left(\frac{\delta+m-m}{\sigma}\right) - \Phi\left(\frac{-\delta+m-m}{\sigma}\right) =$$

$$= \Phi\left(\frac{\delta}{\sigma}\right) - \Phi\left(\frac{-\delta}{\sigma}\right) = 2\,\Phi\left(\frac{\delta}{\sigma}\right).$$

Правило трех сигм
Если $\xi \sim \mathrm{N}_{m,\,\sigma^2}$, т. е. распределена нормально с параметрами $m$ и $\sigma$ то $\mathrm{P}(|\xi - a| \geq 3\sigma) = 0.0027$, что совсем мало.

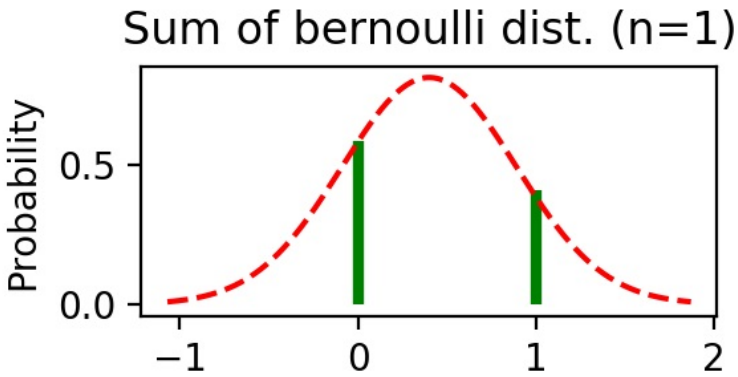
# Центральная предельная теорема

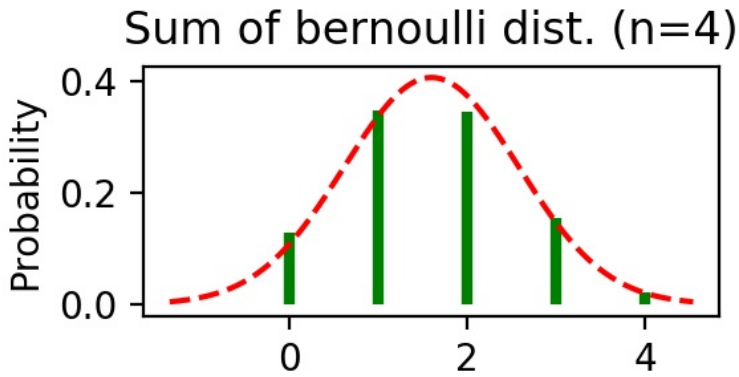
Пусть $X_1, \ldots, X_n, \ldots$ — бесконечная последовательность независимых одинаково распределенных случайных величин, имеющих конечное математическое ожидание $\mu$ и дисперсию $\sigma^2$. Пусть также $S_n = \sum_{i=1}^{n} X_i$. Тогда

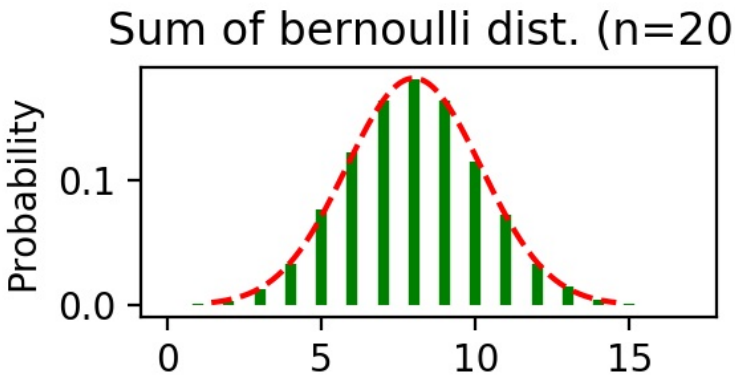
$$\frac{S_n - \mu n}{\sigma\sqrt{n}} \to N(0,1)$$

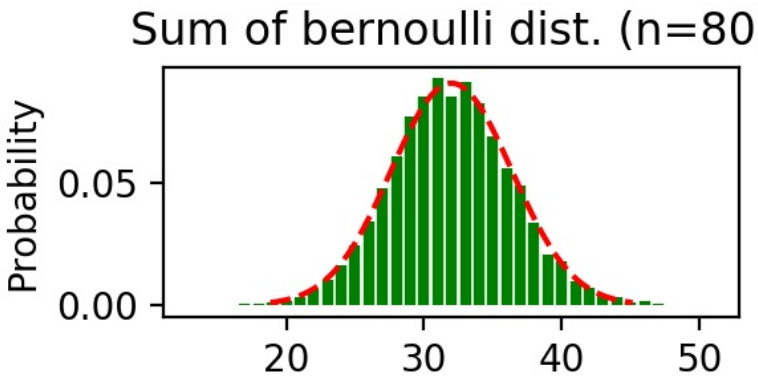
по распределению при $n \to \infty$, где $\mathrm{N}_{0,1}$ — нормальное распределение с нулевым математическим ожиданием и стандартным отклонением, равным единице.

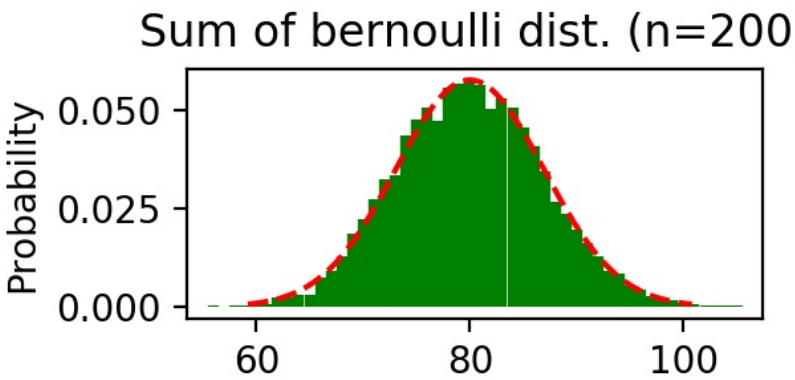
# Иллюстрация к ЦПТ

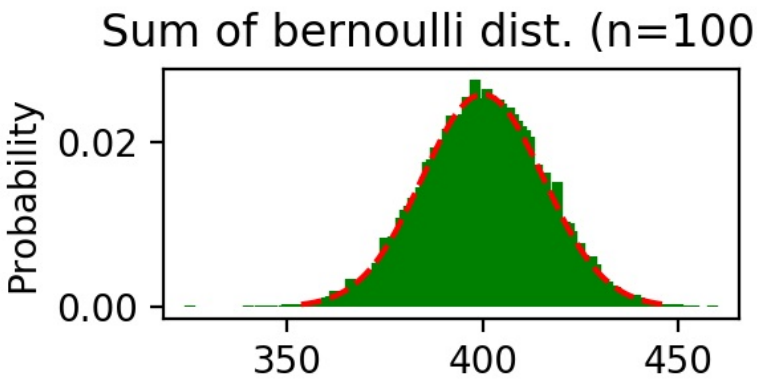


# Воспоминания о распределении Бернулли

Случайна величина $\xi$ имеет распределение Бернулли, если в результате эксперимента она может принять одно их двух значений: 1 с вероятностью $p$ и 0 с вероятностью $q = 1 - p$.

| $\xi$ | 0 | 1 |
|---|---|---|
| P | q | p |

Математическое ожидание $\mathrm{E}\xi = 0(1-p) + 1p = p$.
Дисперсия $\mathrm{D} = \mathrm{E}(\xi - \mathrm{E}\xi)^2 = (0-p)^2(1-p) + (1-p)^2 p = pq$.
Среднее квадратическое отклонение $\sigma = \sqrt{\mathrm{D}} = \sqrt{pq}$.

# $n$ случайных величин с распределением Бернулли

Пусть $\xi_1, \xi_2, \ldots, \xi_n$ — независимые одинаково распределенные случайные величины с распределение Бернулли.

$$\mathsf{E}(\xi_1 + \ldots + \xi_n) = np.$$

$$\mathsf{D}(\xi_1 + \ldots + \xi_n) = npq, \qquad \sigma(\xi_1 + \ldots + \xi_n) = \sqrt{npq}.$$

$$\mathsf{E}\left(\frac{\xi_1 + \ldots + \xi_n}{n}\right) = \frac{\mathsf{E}\left(\xi_1 + \ldots + \xi_n\right)}{n} = \frac{np}{n} = p.$$

$$\mathsf{D}\left(\frac{\xi_1 + \ldots + \xi_n}{n}\right) = \frac{\mathsf{D}\left(\xi_1 + \ldots + \xi_n\right)}{n^2} = \frac{npq}{n^2} = \frac{pq}{n}.$$

$$\sigma\left(\frac{\xi_1 + \ldots + \xi_n}{n}\right) = \frac{\sigma}{\sqrt{n}}.$$

# $n$ случайных величин с распределением Бернулли

Вопрос. Во сколько раз надо увеличить объем выборки, чтобы точность измерений возрасла в 10 раз?

Чтобы увеличить точность измерений в 10 раз, т. е. уменьшить среднее квадратическое отклонение в 10 раз объем выборки нужно увеличить в $10^2 = 100$ раз.

# Предельная теорема Муавра – Лапласа

Пусть событие $A$ может произойти в любом из $n$ независимых испытаний с одной и той же вероятностью $p$. Найдем вероятность того, что в $n$ испытаниях событие $A$ наступит от $k_1$ до $k_2$ раз.

Пусть $\nu_n(A)$ — число наступлений событие $A$ в $n$ испытаниях.

$$\frac{\nu_n(A) - np}{\sqrt{np(1-p)}} \Rightarrow \mathrm{N}_{0,1} \quad \text{при} \quad n \to \infty,$$

т. е. для любых вещественных $k_1 < k_2$ имеет место сходимость

$$\mathrm{P}(k_1 \le \nu_n(A) \le k_2) = \mathrm{P}\left(x_1 \le \frac{\nu_n(A) - np}{\sqrt{np(1-p)}} \le x_2\right) \to \Phi(x_2) - \Phi(x_1),$$

где

$$x_1 = \frac{k_1 - np}{\sqrt{np(1-p)}}, \quad x_2 = \frac{k_2 - np}{\sqrt{np(1-p)}}.$$

# Предельная теорема Муавра – Лапласа. Пример

Проводят тестирование системы контроля на входе в здание. Вероятность того, что камера не распознает пропуск на входе равна $p = 0.2$. Найти вероятность того, что из 400 случайно проходящих сотрудников камера не распознает от 70 до 100 сотрудников.

$$\mathrm{P}(70, 100) = \Phi(x_2) - \Phi(x_1),$$

$$x_1 = \frac{k_1 - np}{\sqrt{npq}} = -1.25$$

$$x_2 = \frac{k_2 - np}{\sqrt{npq}} = 2.5$$

$$\Phi(1.25) = 0.3944, \quad \Phi(2.5) = 0.4938$$

# Ковариация

Ковариацией случайных величин $\xi$ и $\eta$ называется математическое ожидание от произведения их отклонений

$$\mu_{\xi\eta} = \mathrm{E}\Big([\xi - \mathrm{E}\xi]\,[\eta - \mathrm{E}\eta]\Big).$$

Для дискретных случайных величин ковариацию находят по формуле

$$\mu_{\xi\eta} = \sum_{i=0}^{n}\sum_{j=0}^{m}\Big([x_i - \mathrm{E}\xi]\,[y_j - \mathrm{E}\eta]\Big)\,p(x_i, y_j).$$

Для непрерывных случайных величин ковариацию находят по формуле

$$\mu_{\xi\eta} = \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}[x - \mathrm{E}\xi]\,[y - \mathrm{E}\eta]\,dx\,dy.$$

# Ковариация

Ковариацией случайных величин $\xi$ и $\eta$ называется математическое ожидание от произведения их отклонений

$$\mu_{\xi\eta} = \mathsf{E}\Big([\xi - \mathsf{E}\xi]\,[\eta - \mathsf{E}\eta]\Big).$$

Несложно убедиться, что

$$\mu_{\xi\eta} = \mathsf{E}\Big(\xi\eta\Big) - \mathsf{E}\xi\,\mathsf{E}\eta.$$

**Теорема**

Ковариация независимых случайных величин $\xi$ и $\eta$ равна нулю.

Замечания. Обратное утверждение неверно.

# Коэффициент корреляции

Коэффициент корреляции — это мера линейной связи между случайнфми величинами, показывает тесноту связи.

$$r_{\xi\eta} = \frac{\mu_{\xi\eta}}{\sigma_\xi \, \sigma_\eta}.$$

Значения коэффициента корреляции находятся в диапазоне от -1.0 до 1.0.
Коэффициент корреляции величина безразмерная.

# Коэффициент корреляции

Миша и Катя воспитывают троих детей. Кто-то должен сидеть с детьми, а кто-то работать.
Считая заработок Миши случайной величиной $\xi$, а заработок Кати — случайной величиной $\eta$, найти коэффициент корреляции их дневного заработка. Распределение двумерной случайной величины $(\xi, \eta)$ задано таблицей

| $\xi \mid \eta$ | 1 | 2 | 4 |
|---|---|---|---|
| 1 | 0.06 | 0.20 | 0.36 |
| 5 | 0.18 | 0.10 | 0.10 |