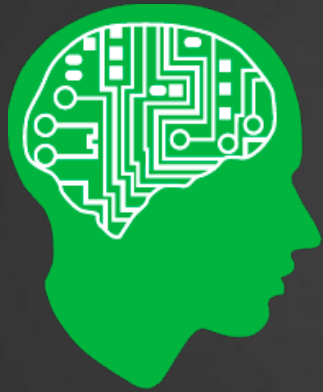




SKILLFACTORY



SKILLFACTORY

Вероятность

Случайный эксперимент (испытание) – явление с несколькими возможными исходами, когда заранее неизвестно, какой исход наступит

Случайная величина X – количественная характеристика результата случайного эксперимента

Вероятность $P(A)$ – численно выраженная оценка шанса, что событие A наступит. $0 \leq P(A) \leq 1$

$$P(X = 5) = 0.63, P(0 \leq X \leq 4) = 0.21$$

Случайная величина

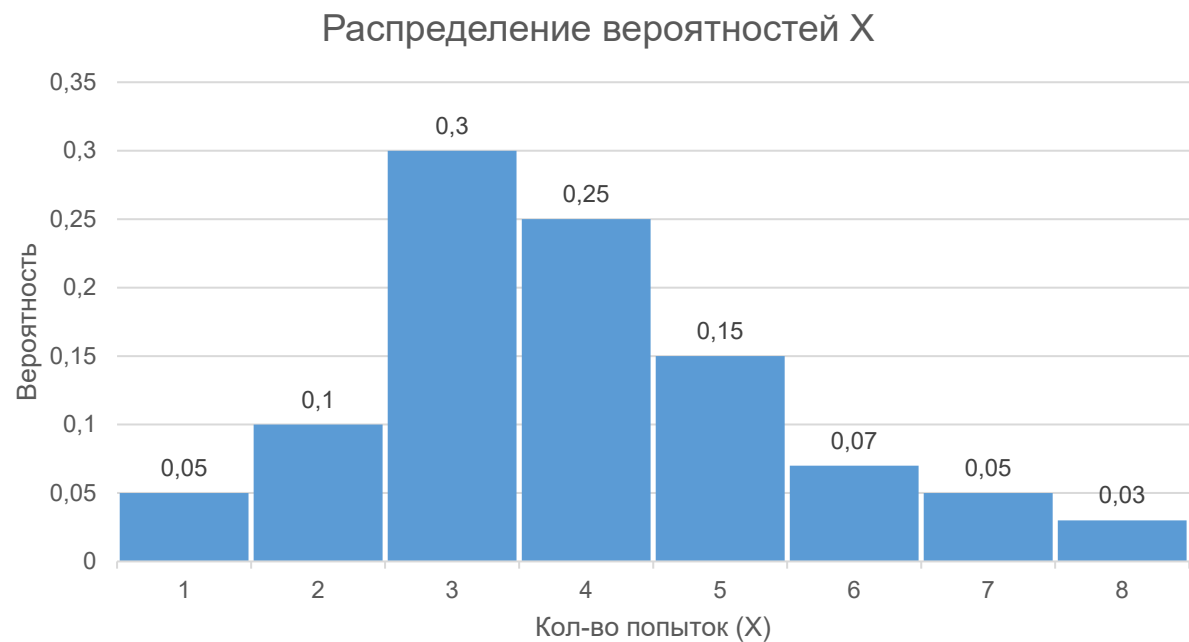
Дискретная случайная величина имеет конечное число возможных значений

Непрерывная случайная величина имеет бесконечное число возможных значений

Распределение вероятностей дискретной случайной величины — набор ее возможных значений и их соответствующих вероятностей

Распределение вероятностей

x	1	2	3	4	5	6	7	8
$P(X = x)$	0.05	0.1	0.3	0.25	0.15	0.07	0.05	0.03



$$P(5 \leq X \leq 7) = 0.15 + 0.07 + 0.05 = 0.27$$

Математическое ожидание

Математическое ожидание $E(X)$ или $M(X)$:
какого значения X ожидать в среднем? $E(X)$ -
сумма значений X , взвешенных по их
вероятностям:

$$E(X) = \sum_{i=1}^N x_i \cdot P(X = x_i)$$

x	1	2	3	4	5	6	7	8
$P(X = x)$	0.05	0.1	0.3	0.25	0.15	0.07	0.05	0.03

Это константа!

$$E(X) = 1 \cdot 0.05 + 2 \cdot 0.1 + \dots + 8 \cdot 0.03 = 3.91$$

Это истинное **среднее** μ_X величины X

$$E(X) = \mu_X$$

Свойства мат. ожидания

-

$$E(X) = \sum_{i=1}^N x_i \cdot P(X = x_i)$$

1. Мат.ожидание константы равно этой константе:

$$E(c) = c$$

2. Константу можно вынести за знак мат.ожидания:

$$E(cX) = c \cdot E(X)$$

3. Мат.ожидание суммы (разности) случайных величин равно сумме (разности) мат.ожиданий эти величин:

$$E(X + Y) = E(X) + E(Y)$$

$$E(X - Y) = E(X) - E(Y)$$

Дисперсия и стандартное отклонение

Дисперсия $\text{Var}(X)$ или $D(X)$: Как далеко от среднего μ в среднем попадает X ?

$$\text{Var}(X) = E(X - \mu)^2 = \sum_{i=1}^N (x_i - \mu)^2 \cdot p_i$$
$$\text{Var}(X) = E(X^2) - E^2(X) = E(X^2) - \mu^2$$

Стандартное отклонение σ ('сигма')

$$\sigma = \sqrt{\text{Var}(X)}$$

Дисперсия и стандартное отклонение

-

x	1	2	3	4	5	6	7	8
$P(X = x)$	0.05	0.1	0.3	0.25	0.15	0.07	0.05	0.03

$$\begin{aligned} \text{Var}(X) &= \sum_{i=1}^N (x_i - 3.91)^2 \cdot p_i = \\ &= (1 - 3.91)^2 \cdot 0.05 + (2 - 3.91)^2 \cdot 0.1 + \dots + (8 - 3.91)^2 \cdot 0.03 \\ &= 2.5019 \end{aligned}$$

$$\sigma = \sqrt{\text{Var}(X)} = \sqrt{2.5019} \approx 1.582$$

- σ - константа!
- $\sigma \geq 0$

Свойства дисперсии

$$Var(X) = E(X - \mu)^2 = \sum_{i=1}^N (x_i - \mu)^2 \cdot p_i$$

1. Дисперсия константы равна нулю:

$$Var(c) = 0$$

2. Константу можно вынести за знак дисперсии в квадрате:

$$Var(cX) = c^2 \cdot Var(X)$$

3. Дисперсия суммы (разности) случайных величин равна сумме дисперсий этих величин **только если** они независимы:

$$Var(X \pm Y) = Var(X) + Var(Y) \pm Cov(X, Y)$$

Непрерывная случайная величина

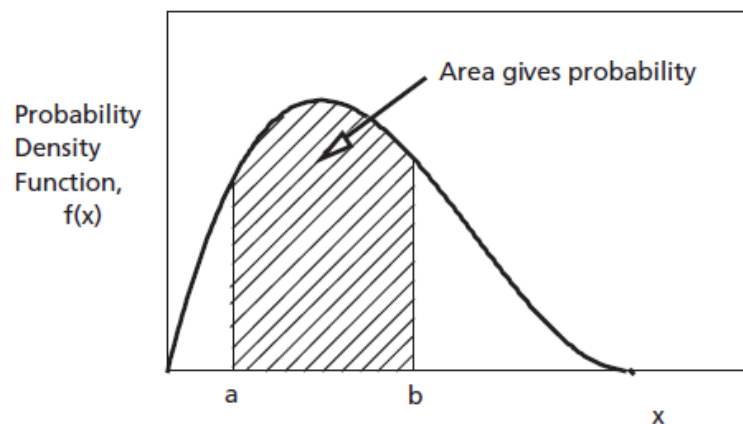
- Бесконечное количество возможных значений x_i
- $P(X = x_i) = 0$

Как задать распределение вероятностей X

???

Обобщая...

Функция плотности распределения $f(x)$



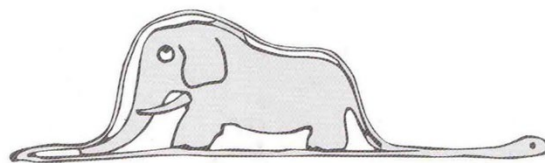
1. Это функция от x : $f(x)$
2. $f(x) \geq 0$
3. Вероятность=площадь: $\int_a^b f(x)dx = P(a \leq X \leq b)$
4. Вся площадь под кривой равна 1:

$$\int_{-\infty}^{+\infty} f(x)dx = 1$$

$f(x)$ может иметь любую форму

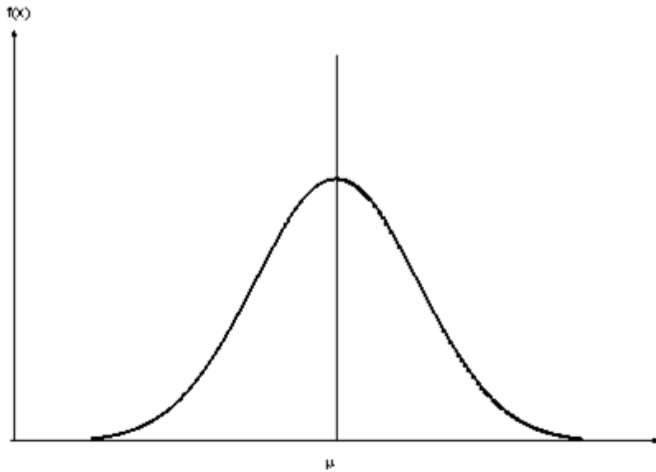


СЕРГЕЙ



Le Petit Prince

Нормальная случайная величина

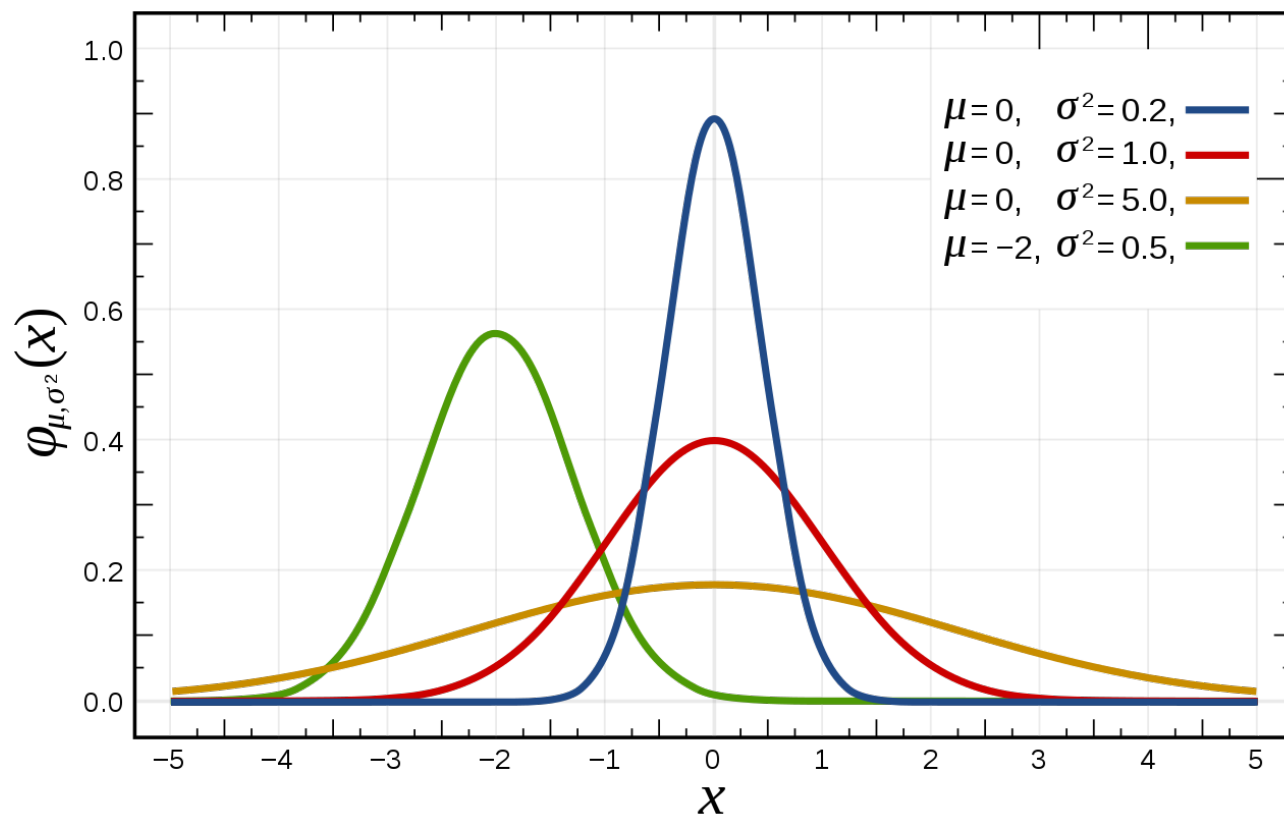


1. Симметричная, форма колокола
2. Имеет центр в среднем значении = медиане
3. Может принимать значения от $-\infty$ до $+\infty$
4. Описывается функцией:

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

Параметры нормальной величины

$$X \sim N(\mu, \sigma)$$



Например, рост $X \sim N(182, 4)$

Стандартная нормальная величина

- Частный случай нормальной величины:

$$E(Z) = 0, Var(Z) = 1$$

$$Z \sim N(0,1)$$

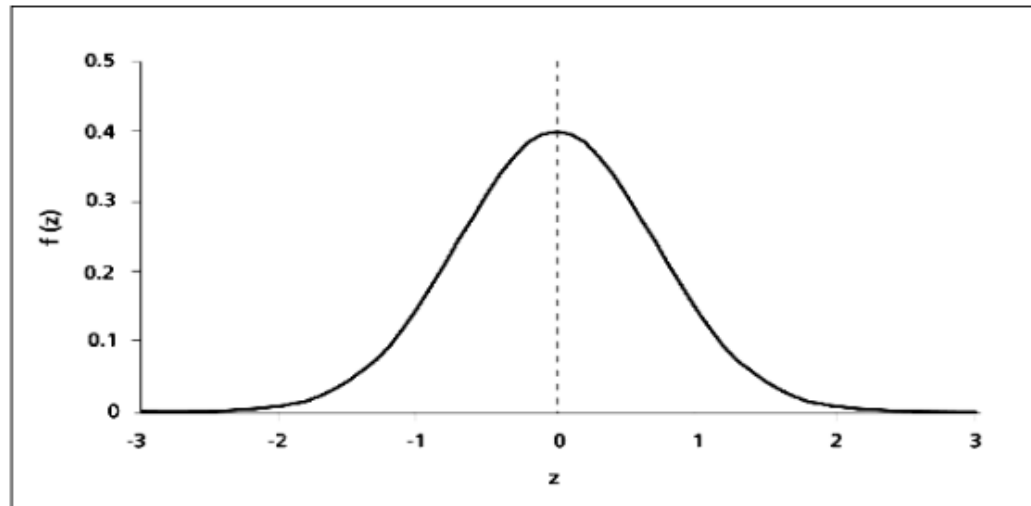


Таблица вероятностей Z

Z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07
0.0	0.5000	0.5040	0.5080	0.5120	0.5160	0.5199	0.5239	0.5279
0.1	0.5398	0.5438	0.5478	0.5517	0.5557	0.5596	0.5636	0.5675
0.2	0.5793	0.5832	0.5871	0.5910	0.5948	0.5987	0.6026	0.6064
0.3	0.6179	0.6217	0.6255	0.6293	0.6331	0.6368	0.6406	0.6443
0.4	0.6554	0.6591	0.6628	0.6664	0.6700	0.6736	0.6772	0.6808
0.5	0.6915	0.6950	0.6985	0.7019	0.7054	0.7088	0.7123	0.7157
0.6	0.7257	0.7291	0.7324	0.7357	0.7389	0.7422	0.7454	0.7486
0.7	0.7580	0.7611	0.7642	0.7673	0.7704	0.7734	0.7764	0.7794
0.8	0.7881	0.7910	0.7939	0.7967	0.7995	0.8023	0.8051	0.8079
0.9	0.8159	0.8186	0.8213	0.8239	0.8264	0.8289	0.8315	0.8340

Приведение к Z

-

$$Z = \frac{X - \mu}{\sigma}$$

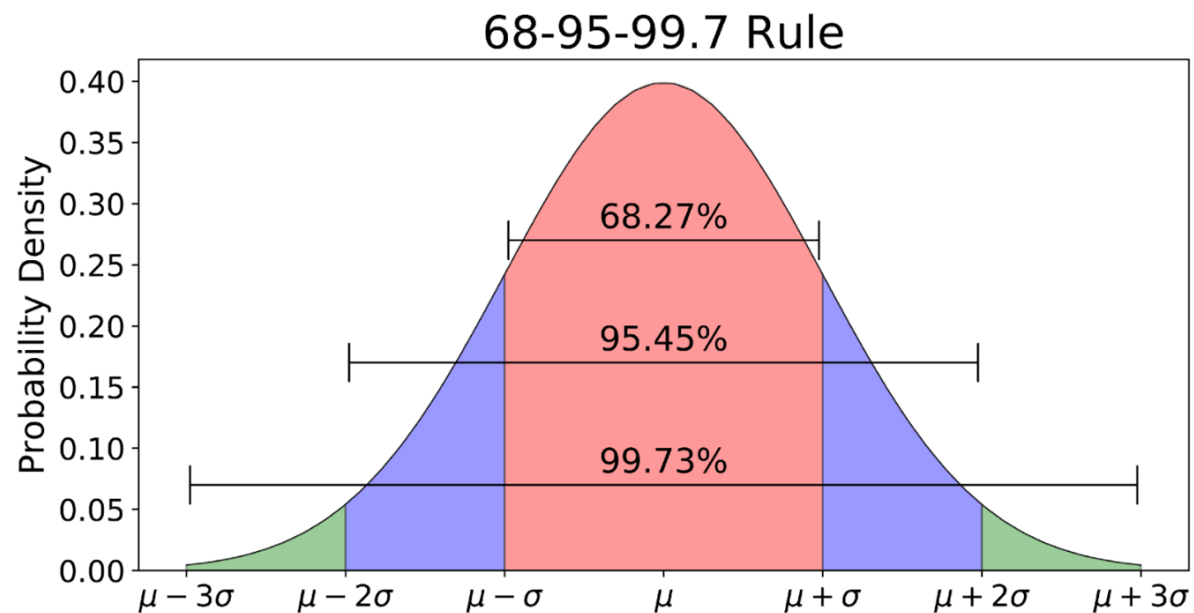
На сколько стандартных отклонений (σ) X отклонился от среднего μ ?

Пример. $X \sim N(182, 4)$, $x_i = 190$, $z_i = ?$

$$z_i = \frac{190 - 182}{4} = 2$$

$$P(X > 190) = P(Z > 2) = 0.0228$$

Правило «трёх сигм»



$$P(\mu - \sigma < X < \mu + \sigma) = P(-1 < Z < 1) \approx 68\%$$

$$P(\mu - 2\sigma < X < \mu + 2\sigma) = P(-2 < Z < 2) \approx 95\%$$

$$P(\mu - 3\sigma < X < \mu + 3\sigma) = P(-3 < Z < 3) \approx 99.7\%$$

Центральная предельная теорема

Сумма независимых одинаково распределенных величин имеет приблизительно нормальное распределение, если n (число слагаемых в сумме) достаточно велико.

Центральная предельная теорема



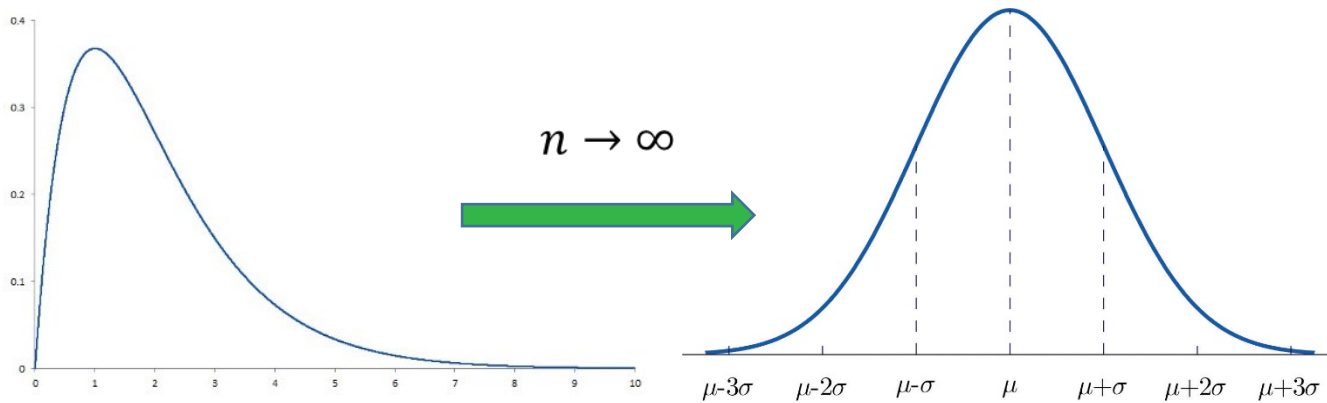
Центральная предельная теорема

Сумма независимых одинаково распределенных величин имеет приблизительно нормальное распределение, если n (число слагаемых в сумме) достаточно велико.

1. Сумма
2. независимых
3. одинаково распределенных величин
 $\approx N$
4. если n достаточно велико.

Среднее распределено нормально

1. Выборочное среднее – **сумма**: $\bar{X} = \frac{\sum X_i}{n}$
2. **независимых** величин X_i
3. **Одинаково** распределенных величин X_i



$X \sim \text{'как угодно'}$

$\bar{X} \approx N$

Пропорция распределена нормально

Выборочная пропорция: $\hat{p} = \frac{Y}{n}$,

$$Y = X_1 + X_2 + \dots + X_n,$$

1. Пропорция – **сумма** $\hat{p} = \frac{Y}{n}$
2. **независимых** величин X_i : X_1, X_2, \dots, X_n
$$X_i = \begin{cases} 1, & \text{если глаза голубые} \\ 0, & \text{если не голубые} \end{cases}$$
3. **одинаково** распределенных величин X_i

Если $n \rightarrow \infty$, $\hat{p} \approx N$

ЦПТ по сути

Если выборка случайная и большая
- все нормально!

Сколько орехов – это куча?

-

Минимальные требования для нормального приближения по ЦПТ:

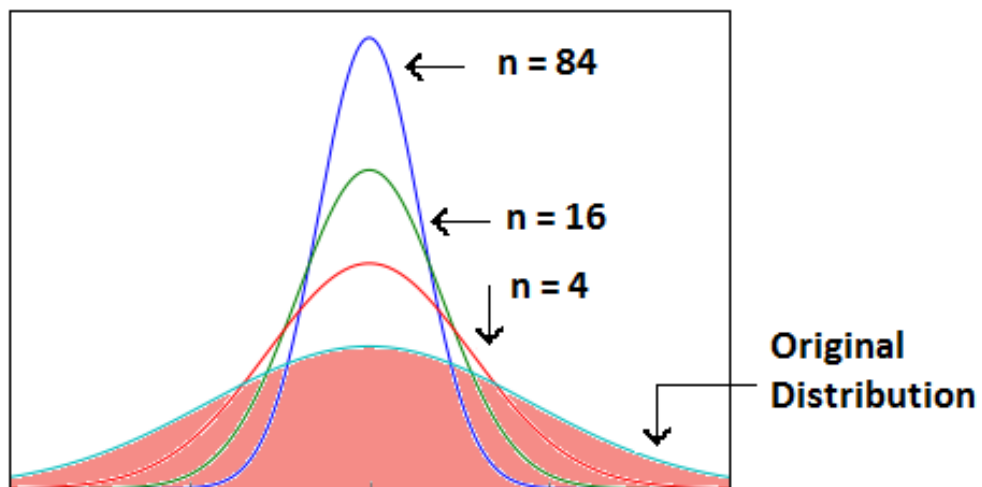
- Для среднего \bar{X} : $n \geq 30$
- Для пропорции \hat{p} : $n\hat{p} \geq 5$ и $n(1 - \hat{p}) \geq 5$

Рост выборки – рост точности. \bar{X}

$$\bar{X} \approx N\left(\mu, \frac{\sigma}{\sqrt{n}}\right)$$

$$\sigma(X) = \sigma, \quad \sigma(\bar{X}) = \frac{\sigma}{\sqrt{n}}$$

$$\sigma > \frac{\sigma}{\sqrt{n}}$$



$$n \rightarrow \infty, \quad \frac{\sigma}{\sqrt{n}} \rightarrow 0$$

Рост выборки – рост точности. \hat{p}

$$E(\hat{p}) = p, \quad \sigma(\hat{p}) = \sqrt{\frac{p(1-p)}{n}}$$

$$\hat{p} \approx N\left(p, \sqrt{\frac{p(1-p)}{n}}\right)$$

$$n \rightarrow \infty, \quad \sqrt{\frac{p(1-p)}{n}} \rightarrow 0$$