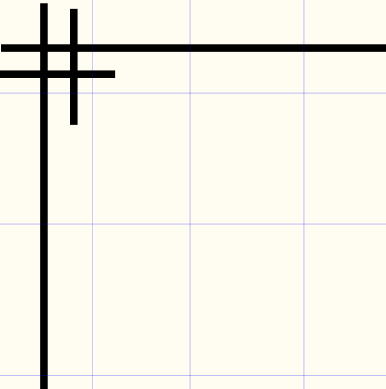




Информация и вероятность. Формула Шеннона

Информация и вероятность

Доля символов в русских текстах

из 1000
символов
около 175
пробелов

	0,175
О	0,090
Е	0,072
А	0,063
И	0,062
Т	0,053
Н	0,052
С	0,045
р	0,040
В	0,038
Л	0,035
К	0,028
М	0,026
Д	0,025
П	0,023
У	0,021

вероятность p
появления символа

Я	0,018
Ы	0,017
З	0,016
Ь	0,015
Б	0,014
Г	0,013
Ч	0,012
Й	0,010
Х	0,009
Ж	0,007
Ю	0,006
Ш	0,005
Ц	0,004
Щ	0,003
Э	0,002
Ф	0,001

Вероятность

Вероятность события – число от 0 до 1, показывающее, как часто случается это событие в большой серии одинаковых опытов.

$$0 \leq p \leq 1$$

$$x^2 < 0$$

$$p = 0$$

событие **никогда** не происходит
(нет неопределенности)

$$p = 0,5$$

событие происходит в половине случаев (**есть неопределенность**)



$$p = 1$$

событие происходит **всегда**
(нет неопределенности)

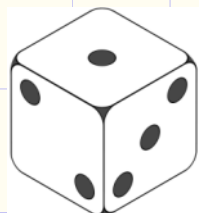
$$x^2 \geq 0$$

Вероятность

N – количество испытаний

m – сколько раз произошло событие

$$p = \frac{m}{N}$$



ровно 2: $p = \frac{1}{6}$

чётное: $p = \frac{3}{6} = \frac{1}{2}$

меньше 3: $p = \frac{2}{6} = \frac{1}{3}$



2 и 2: $p = \frac{1}{36}$

2 чётных: $p = \frac{3 \cdot 3}{36} = \frac{1}{4}$

оба меньше 3: $p = \frac{4}{36} = \frac{1}{9}$

Вероятность и информация

Чем более неожиданно событие, тем больше получено информации.

$$p = 1$$

...AAAAAAAAAAAAAAAAAAAA

получили букву «А»: $I = 0$

$$p \rightarrow 0$$

...**В**AAAAAAAAAAAAAAAAAAAA

получили букву «В»: $I \rightarrow \infty$

В 10 опытах будет получено в 10 раз больше информации, чем в одном (**аддитивность**).

Вероятность и информация

$$f(p) = -K \cdot \log_2 p \quad \text{при } K=1 \Rightarrow \text{информация в битах}$$

Если событие имеет вероятность p , то количество информации в битах, полученное в сообщении об этом событии, равно

$$I = -\log_2 p = \log_2 \frac{1}{p}$$

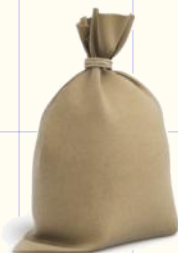
$$p=1 \Rightarrow I = \log_2 1 = 0$$

$$p \rightarrow 0 \Rightarrow I \rightarrow \log_2 \infty = \infty$$

Вероятность и информация

Аддитивность:

по 8 шариков разного цвета



$$p = \frac{1}{8}$$



всего $8 \cdot 8 = 64$ варианта

$$p = \frac{1}{64}$$

$$I_1 = I_2 = \log_2 \frac{1}{p} = \log_2 8 = 3 \text{ бита}$$

$$I = I_1 + I_2 = 6 \text{ битов}$$

$$I = \log_2 64 = 6 \text{ битов}$$



Аддитивность выполняется!

Связь с формулой Хартли

N равновероятных событий $\Rightarrow p = \frac{1}{N}$

$$I = \log_2 \frac{1}{p} = \log_2 N$$

совпадает с
формулой
Хартли

Если вероятности разные:



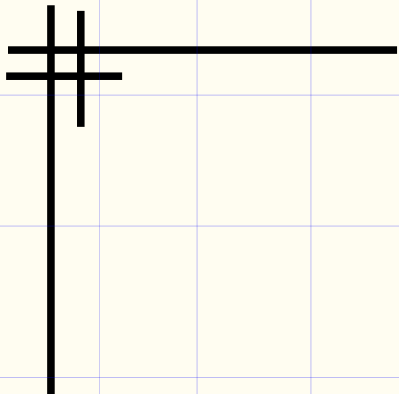
«**Васе достался зелёный шарик**».

$$p = \frac{6}{8} = \frac{3}{4}$$

$$I = \log_2 \frac{4}{3} \approx 0,415 \neq 0,5$$

?

А если красный шарик?



Задачи 1, 2, 6 и 7, стр 15 учебника

Формула Шеннона

Количество полученной информации равно уменьшению неопределенности.

$$I = \Delta H = H_{\text{нач}} - H_{\text{кон}}$$

? Как вычислить H ?

Неопределённость знаний об источнике данных (N событий, вероятности p_i):

$$H = \sum_{i=1}^N p_i \cdot \log_2 \frac{1}{p_i} = p_1 \cdot \log_2 \frac{1}{p_1} + \dots + p_N \cdot \log_2 \frac{1}{p_N}$$

информационная
ЭНТРОПИЯ



Клод Шеннон

Формула Шеннона

«Идёт ли сейчас снег?» (1 – да, 2 – нет)

зимой: $p_1 = \frac{1}{2}$? Как вычислить p_2 ?

$$p_2 = 1 - p_1 = \frac{1}{2}$$

! Сумма вероятностей всех событий, составляющих полную систему, равна 1!

$$H = \frac{1}{2} \cdot \log_2 2 + \frac{1}{2} \cdot \log_2 2 = \log_2 2 = 1 \text{ бит}$$

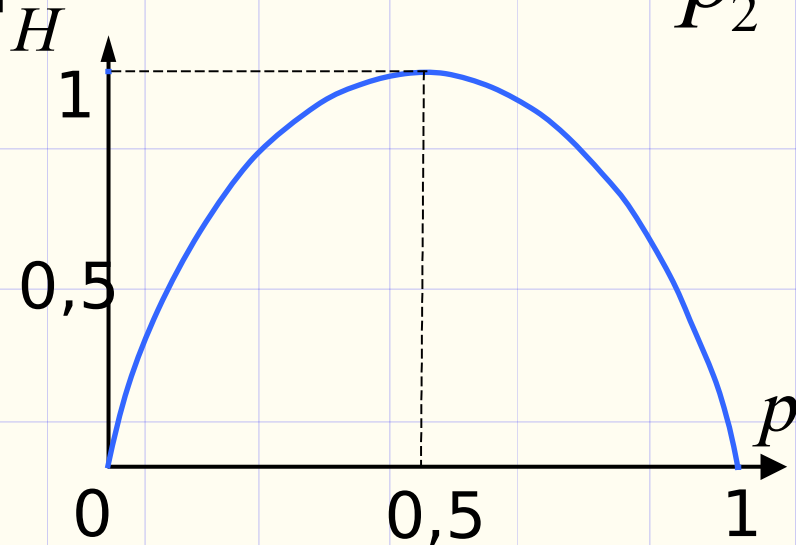
летом: $p_1 = 0,0001$, $p_2 = 0,9999$

$$H = 0,0001 \cdot \log_2 \frac{1}{0,0001} + 0,9999 \cdot \log_2 \frac{1}{0,9999} \approx 0,00156 \text{ бит}$$

Когда неопределённость наибольшая?

Система двух событий:

$$p_2 = 1 - p_1$$

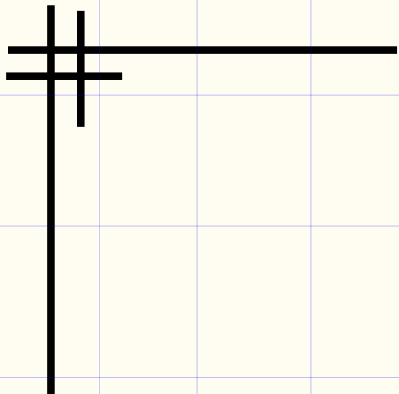


Неопределенность **максимальна**, когда все события **равновероятны**.

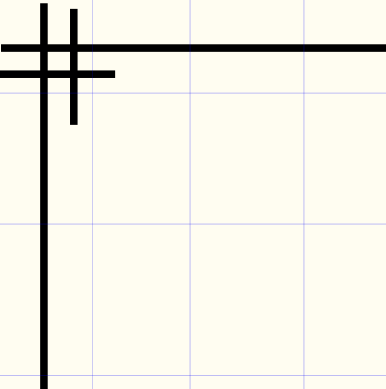
совпадает с формулой Хартли!

$$p_1 = p_2 = \dots = p_N = \frac{1}{N}$$

$$H = \sum_{i=1}^N \frac{1}{N} \log_2 N = \log_2 N$$



Задачи 1, 2 и 3, стр 20 учебника



Задание на дом:

§ 1