

# Теория информации

Энтропия.

# Информация

- (Философия) **Знания** об окружающем мире, помогающие преобразовывать этот мир под свои нужды, чаще всего эти знания представляют собой различные тексты
- (Информатика) **Данные**, последовательность битов в памяти компьютера
- (Теория систем) «Связи», **передача вещества и энергии** от одной части системы к другой
- (Юриспруденция) **Сведения** (сообщения, данные) независимо от формы их представления
- (Кибернетика, Н. Винер) Обозначение **содержания**, полученное нами из внешнего мира в процессе приспособления к нему нас и наших чувств
- (Физика) **Состояние** группы физических систем

# Передача информации, кодирование

- **Сигнал** — это любая функция, аргументом которой является время, а значением — величина, которую можно представить числом или СИМВОЛОМ.



ойство, предназначенное для  
ремени

- **Кодирующее устройство** — преобразователь сигнала в форму, допускающую передачу по линии связи
- **Декодирующее устройство** — преобразователь сигнала в первоначальную форму, т.е. в тот вид, который был получен из источника сигнала.
- **Обратимое кодирование** - преобразование информации из одной формы в другую, допускающее восстановление исходной формы информации (сигнала) без искажений

# Код

- **Код** — правило (алгоритм) сопоставления каждому конкретному сообщению строго определённой комбинации символов (знаков) (или сигналов).



●●●●●●●●●●  
●●●●●●●●●●  
●●●●●●●●●●  
●●●●●●●●●●  
●●●●●●●●●●  
●●●●●●●●●●



# Количество информации

			Коды		
0	1	1	1		
1	2		0, 1	1	1
2	4		00, 01, 10, 11	2	2
3	8		000, 001, ..., 110, 111	3	3
4	16		0000, 0001, ..., 1110, 1111	4	4
...	...	...	...	...	...

- $I \geq 0$
- $I(x) = I(p(x))$
- $I(p^m) = mI(p)$ ,  $I(p_4) = I(p_2^2)$ ,  $I(p_1^4) = 2I(p_2) = 4I(p_1)$ .
- $I(x_1, \dots, x_n) = I(x_1) + \dots + I(x_n)$  для независимых сообщений  $x_1, \dots, x_n$ .

# Собственная информация

- $I(x_i) = -\log(p(x_i))$ 
  - Свойства:
    - Неотрицательность  $I \geq 0$
    - Монотонность  $I(x_1) \geq I(x_2)$ , если  $p(x_1) \leq p(x_2)$
    - Аддитивность  $I(x_1, \dots, x_n) = \sum_{i=1}^n I(x_i)$
- Бит (англ. binary digit, англ. bit — кусочек, частица) — единица измерения количества информации. Существуют также нат ( $\ln$ ), хартли ( $\lg$ ), трит ( $\log_3$  ).

# Энтропия

- (Физика) мера незнания, мера хаоса, мера неорганизованности, мера скрытой информации.
- Среднее значение количества информации в сообщении.
- Формула Шеннона

$$H(x) = - \sum_{i=1}^N p_i \log(p_i) = \sum_{i=1}^N p_i \log\left(\frac{1}{p_i}\right)$$

- Формула Хартли  $H(x) = \log N$ , где  $N$  – количество альтернатив
- Энтропия ограничивает максимально возможное сжатие без потерь (равна минимальному количеству бит, которое необходимо для записи этого сообщения без искажений).

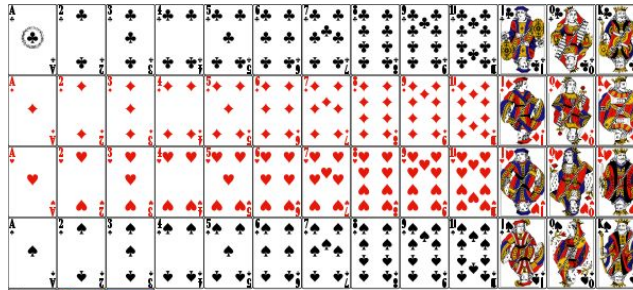
# Количество информации, получаемой в процессе сообщения

- Энтропию  $H$  можно заменять на  $I$ , если количество информации, получаемое при полном снятии неопределенности некоторой ситуации, количественно равно начальной энтропии этой ситуации. Но неопределенность может быть снята только частично, поэтому количество информации  $I$ , получаемой из некоторого сообщения, вычисляется как уменьшение энтропии, произошедшее в результате получения данного сообщения.
- $I = H_{\text{до}} - H_{\text{после}}$
- Для равновероятного случая (формула Хартли) получаем  $I = \log \frac{N_{\text{до}}}{N_{\text{после}}}$



# Примеры

1. Подсчитать количество информации в событии подбрасывания игрального кубика (для 4-, 6-, 8-, 12-, 20-гранника)
2. Подсчитать количество информации в сообщении о доставании из игральной колоды карты определенного а) цвета б) масти с) одну из «старших» д) одна из карт е) конкретная карта



1. Шарик находится в одной из трех урн А, В или С. Определить сколько бит информации содержит сообщение о том, что он находится в урне В
2. Вероятность первого события составляет 0,5, а второго и третьего 0,25. Чему для такого распределения равна информационная энтропия.

- 1) В результате определения того, что интересующий нас человек живет на втором этаже, было получено 3 бита информации. Сколько этажей в его доме?
- 2) В доме 10 этажей, какое количество информации мы получили, узнав, что интересующий нас человек живет на втором этаже?
- 3) Пусть в некотором учреждении состав работников распределяется так: а)  $\frac{3}{4}$  - мужчины,  $\frac{1}{4}$  - женщины б)  $\frac{1}{2}$ ,  $\frac{1}{2}$ . Какова неопределенность относительно того, кого вы встретите первым, зайдя в учреждение? Какое количество информации несет в себе сообщение, что мы первым встретили мужчину/женщину?
- 4) Имеются два источника информации, алфавиты и распределения вероятностей которых заданы матрицами:

$$\begin{matrix} \|X\| \\ \|P\| \end{matrix} = \begin{matrix} x_1 & x_2 \\ p_1 & p_2 \end{matrix} \quad \begin{matrix} \|Y\| \\ \|Q\| \end{matrix} = \begin{matrix} y_1 & y_2 & y_3 \\ q_1 & q_2 & q_3 \end{matrix}$$

Определить, какой источник дает большее количество информации, если а)  $p_1 = p_2, q_1 = q_2 = q_3$  б)  $p_1 = q_1, p_2 = q_2 + q_3$

1) Вот список сотрудников некоторой организации:

<i>Год рождения</i>	<i>Фамилия</i>	<i>Имя</i>
1970	Иванова	Марина
1970	Иванова	Наталья
1970	Петрова	Татьяна
1970	Звягина	Ирина

Определите количество информации, недостающее для того, чтобы выполнить следующие просьбы:

а) Пожалуйста, позвоните к телефону Иванову.

б) Меня интересует одна ваша сотрудница она 1970 года рождения.

2) Какое из сообщений несет больше информации:

- В результате подбрасывания монеты (орел, решка) выпала решка.

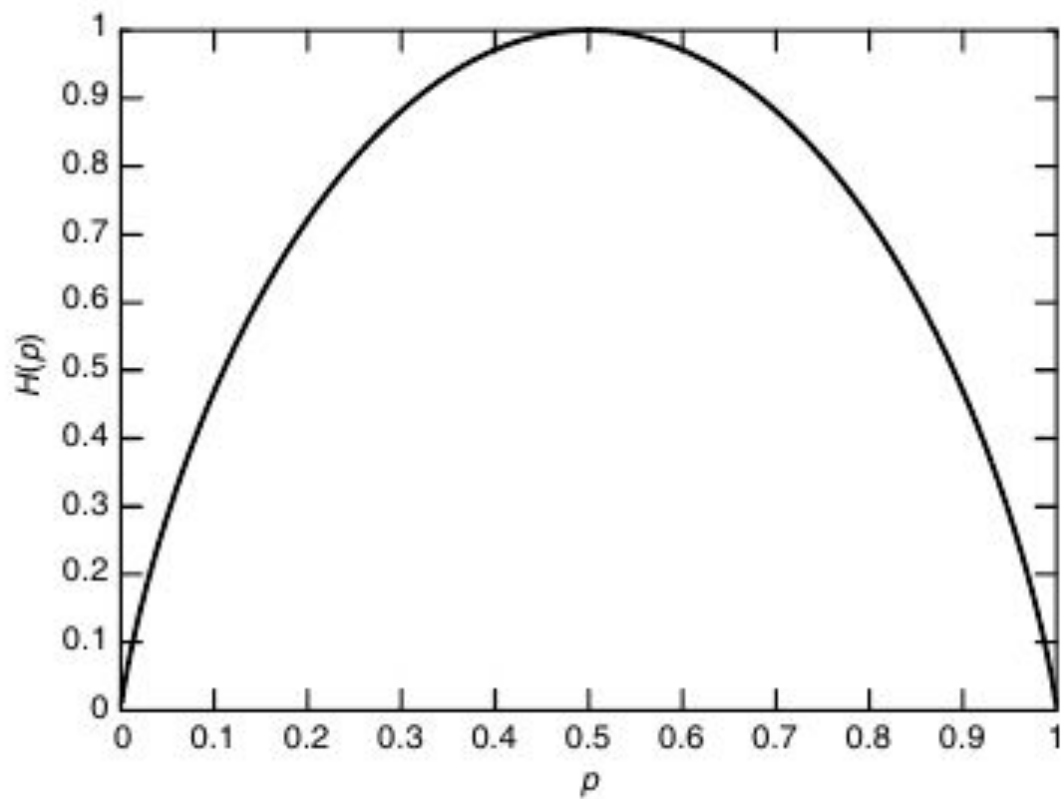
- На светофоре (красный, желтый, зеленый) сейчас горит зеленый свет.

- В результате подбрасывания игральной кости (1...6) выпало 3 очка.

3) Какое из соотношений несет в себе больше информации  $x = 5$  или  $x > 3$ ?

- Вероятности победы каждой из восьми лошадей на скачках  $(\frac{1}{2}, \frac{1}{4}, \frac{1}{8}, \frac{1}{16}, \frac{1}{64}, \frac{1}{64}, \frac{1}{64}, \frac{1}{64})$ . Вычислите энтропию.

$$\therefore X = \begin{cases} p \\ q = 1 - p \end{cases}$$



# Другие меры измерения количества информации

- 1) Геометрическая
- 2) Комбинаторная

Число размещений с повторениями

$$\bar{A}_n^k = n^k$$

Без повторений

$$A_n^k = n(n-1) \dots (n-k+1) = \frac{n!}{(n-k)!} = \binom{n}{k} k!$$

Число сочетаний

$$C_n^k = \binom{n}{k} = \frac{n!}{k!(n-k)!}$$

- 3) Аддитивная мера (формула Хартли)

- **Целесообразность информации**

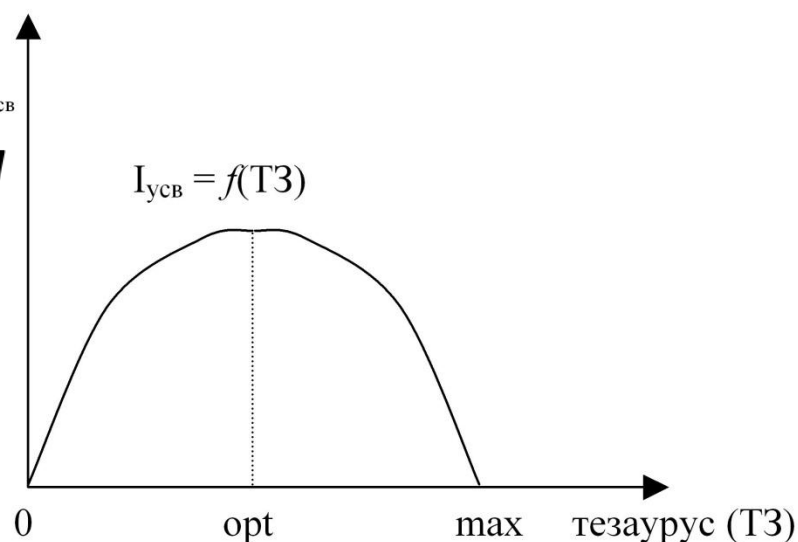
$$I = \log \frac{p_1}{p_2}, \text{ где } p_1, p_2 \text{ – вероятности}$$

достижения цели после и до получения сообщения

- **Полезность информации**

- **Истинность**

усваиваемая  
потребителем  
информация  $I_{\text{усв}}$



# Взаимная информация

- Для д.с.в.  $X$  и  $Y$ , заданных законом распределения  $P(X = X_i) = p_i$ ,  $P(Y = Y_j) = q_j$  и совместным распределением  $P(X = X_i, Y = Y_j) = p_{ij}$ , количество информации, содержащейся в  $X$  относительно  $Y$ , равно

$$I(X, Y) = \sum_{i,j} \log \frac{p_{ij}}{p_i q_j}$$



# Условная энтропия

- $H(X, Y) = - \sum_{ij} p_{ij} \log p_{ij}$

Свойства меры информации и энтропии:

- 1)  $I(X, Y) \geq 0$ ,  $I(X, Y) = 0 \Leftrightarrow X$  и  $Y$  независимы;
- 2)  $I(X, Y) = I(Y, X)$ ;
- 3)  $HX = 0 \Leftrightarrow X$  — константа;
- 4)  $I(X, Y) = HX + HY - H(X, Y)$ , где  $H(X, Y) = - \sum_{i,j} p_{ij} \log_2 p_{ij}$ ;
- 5)  $I(X, Y) \leq I(X, X)$ . Если  $I(X, Y) = I(X, X)$ , то  $X$  — функция от  $Y$ .