Анализ свойств меры Хартли

Экспериментатор одновременно подбрасывает монету (М) и кидает игральную кость (К).
Какое количество информации содержится в эксперименте (Э)?

Аддитивность:
$i(\text{Э}) = i(\text{М}) + i(\text{К}) \Rightarrow i(12) = i(2) + i(6) : \log_x 12 = \log_x 2 + \log_x 6$

Неотрицательность:
Функция $\log_x N$ неотрицательно при любом $x > 1$ и $N \geq 1$.

Монотонность:
С увеличением $p(\text{М})$ или $p(\text{К})$ функция $i(\text{Э})$ монотонно возрастает.

Принцип предопределённости:
При наличии всегда только одного исхода (монета и кость с магнитом) количество информации равно нулю: $\log_x 1 + \log_x 1 = 0$.

Мера количества информации по Шеннону

Мера Хартли подходит лишь для систем м равновероятными состояниями.
Если состояния системы S не равновероятны, используют меру Шеннона:

$$i(S) = \sum p_i * \log_2 p_i,$$

где N – число состояний системы,
p_i – вероятность того, что система S находится в состоянии i (сумма всех p_i равна 1).

Клод Шеннон
(1916–2001)

Формула Хартли является частным случаем формулы Шеннона!

Пример 1. Количество информации в акте подбрасывания обычной монеты по формуле Хартли равно $\log_2 2 = 1$ бит. По формуле Шеннона получим то же: $i_{s1} = -0,5 * \log_2 0,5 - 0,5 * \log_2 0,5 = 1$ бит.
Пример 2. При подбрасывании монеты со смещенным центром тяжести количество непредсказуемости становится меньше: $i_{s2} = -0,75 * \log_2 0,75 - 0,25 * \log_2 0,25 \approx 0,8$ бит.

Пример использования меры Шеннона

Шулер наугад вытаскивает одну карту из стопки, содержащей 9 известных ему карт: 3 джокера, 3 туза, 1 король, 1 дама и 1 валет. Какое количество информации для шулера содержится в этом событии s?

$$\text{Вероятность вытащить} = \begin{cases} \text{джокера} \\ \text{туза} \\ \text{короля} \\ \text{даму} \\ \text{валета} \end{cases} \quad \text{равна} \begin{cases} 3/9 = 1/3 \\ 3/9 = 1/3 \\ 1/9 \\ 1/9 \\ 1/9 \end{cases}$$

Количество информации, выраженное в тритах, равно:

$i(s) = -(\frac{1}{3}*\log_3 \frac{1}{3} + \frac{1}{3} * \log_3 \frac{1}{3} + \frac{1}{9} * \log_3 \frac{1}{9} + \frac{1}{9}*\log_3 \frac{1}{9} + \frac{1}{9}*\log_3 \frac{1}{9} =$

$$= \frac{1}{3} + \frac{1}{3} + \frac{2}{9} + \frac{2}{9} + \frac{2}{9} = 1\frac{1}{3} \approx \log_3 5 \text{ vs } \log_3 14$$

Нестрогий вывод формулы Шеннона

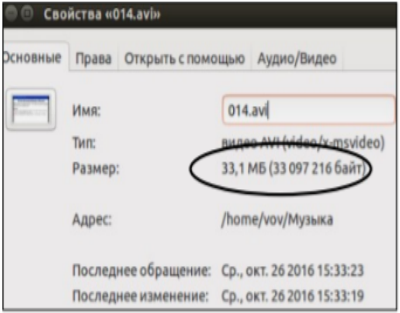
Задача. Монета имеет смещенный центр тяжести. Вероятность выпадения «орла» – 0,25, вероятность выпадения «решки» – 0,75. Какое количество информации содержится в одном подбрасывании?

Решение

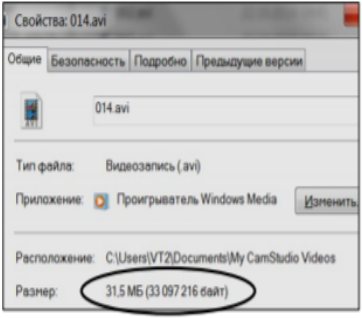
- Пусть монета была подброшена N раз ($N \Longrightarrow \infty$) из которых «решка» выпала M раз, «орел» $-K$ раз (очевидно, что $N = M + K$).
- Количество информации и в N подбрасываниях: $i_N = M$*$i($ «решка»$)$ $+ K$*$i($«орел»$)$.
- Тогда среднее количество информации в одном подбрасывании: $i_1 = i_N/ =$ (M/N)*i(«решка»)+(K/N)*i(«орел») = p(«решка»)* *i(«решка»)+p(«орел»)*i(«орел»).
- Подставив формулу Шеннона для i, окончательно получим: $i_1 =$ -p(«решка»)*$\log_x p$(«решка») - p(«орел»)*$\log_x p$(«орел») $\approx 0{,}8$ бит.

Приставки для единиц измерения количества информации/данных

Linux Ubuntu 14



Microsoft Windows 7



33 097 216 байт – это 33,1 МБ или 31,5 МБ?