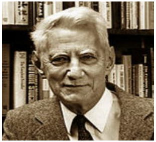
Мера количества информации по Шеннону



Клод Шеннон(1916–2001)

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

$$i(S) = -\sum_{i=1}^{N} p_i \cdot \log_2 p_i,$$

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

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: $i_{s1} = -0,5 \cdot \log_2 0,5 - 0,5 \cdot \log_2 0,5 = 1$ бит.

Пример 2. При подбрасывании монеты со смещённым центром тяжести количество непредсказуемости становится меньше:
$i_{s2} = -0,75 \cdot \log_2 0,75 - 0,25 \cdot \log_2 0,25 \approx 0,8$ 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?

$$\text{Вероятность вытащить}\left\{\begin{array}{c}\text{джокера}\\ \text{туза}\\ \text{короля}\\ \text{даму}\\ \text{валета}\end{array}\right\}\text{ равна }\begin{array}{c}3/9 = 1/3\\ 3/9 = 1/3\\ 1/9\\ 1/9\\ 1/9\end{array}$$

Количество информации, выраженное в тритах, равно:

$$i(S) = -(\frac{1}{3}\cdot\log_3\frac{1}{3} + \frac{1}{3}\cdot\log_3\frac{1}{3} + \frac{1}{9}\cdot\log_3\frac{1}{9} + \frac{1}{9}\cdot\log_3\frac{\bar{1}}{9} + \frac{1}{9}\cdot\log_3\frac{1}{9}) =$$

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

Нестрогий вывод формулы Шеннона

Задача. Монета имеет смещённый 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 ($N \to \infty$), из которых «решка» выпала M раз, «орёл» — K раз (очевидно, что $N = M + K$).
- Количество информации в N подбрасываниях: $i_N = M \cdot i(\text{«решка»}) + K \cdot i(\text{«орёл»})$.
- Тогда среднее количество информации в одном подбрасывании: $i_1 = i_N/N = (M/N) \cdot i(\text{«решка»}) + (K/N) \cdot i(\text{«орёл»}) = p(\text{«решка»}) \cdot i(\text{«решка»}) + p(\text{«орёл»}) \cdot i(\text{«орёл»})$.
- Подставив формулу Шеннона для i, окончательно получим: $i_1 = -p(\text{«решка»}) \cdot \log_x p(\text{«решка»}) - p(\text{«орёл»}) \cdot \log_x p(\text{«орёл»}) \approx 0,8$ бит.

Приставки для единиц измерения количества информации/данных: проблема

Рис.: Linux Ubuntu 14

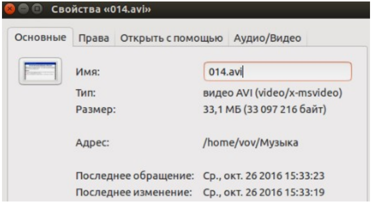
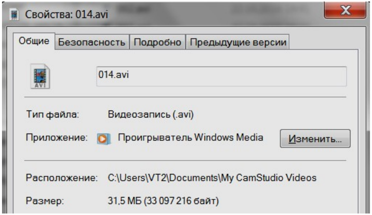


Рис.: Microsoft Windows 7



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

Приставки для единиц измерения количества информации/данных: решение

1. **IEEE 1541-2002** – Институт инженеров по электротехнике и радиоэлектронике.
2. **ISO/IEC 80000-13:2008** – Международная организация по стандартизации.
3. **ГОСТ IEC 60027-2-2015** – Международная электротехническая комиссия.

Приставки единиц СИ	Новые двоичные префиксы	$\Delta, \%$
килобайт (kB) $= 10^3$ байт	кибибайт (KiB, КиБ) $= 2^{10}$ байт	2
мегабайт (MB) $= 10^6$ байт	мебибайт (MiB, МиБ) $= 2^{20}$ байт	5
гигабайт (GB) $= 10^9$ байт	гибибайт (GiB, ГиБ) $= 2^{30}$ байт	7
терабайт (TB) $= 10^{12}$ байт	тебибайт (TiB, ТиБ) $= 2^{40}$ байт	10

Краткое обозначение битов и байтов: b = bit = бит, B = Б = байт
1024 B = 1024 Б = 8192 b = 8192 бит = 8 Кибит = 1 КиБ = 1 KiB