Теория информации.

Пр.:

1) $\exists$ 8 карандашей 3 бита

Если 1000, то 10 бит

$\exists$ 9 бит $\Rightarrow 2^9$ сообщений < 1000

Неопределенность

[?] $\to N_1$ неоп — посмотрели $\to$ [!] $\to N_2$ $\Rightarrow$ $N_2 < N_1$

$\Delta N = N_2 - N_1$

Информация $= -\Delta N$

Случайный источник

$\Omega = \{\omega_1, \ldots, \omega_n\}$; $p_1, \ldots, p_n$

С.И. — $\xi_1, \xi_2 \ldots$

$P(\xi_i = a) = p_a$, $a \in (1 \ldots n)$

Энтропия:

$H(p_1, \ldots, p_n) : RS \to \mathbb{R}^+$ (случ. источник)

$$\left.\begin{aligned} &1)\ f(1) = e \\ &2)\ f(a+b) = f(a)f(b) \\ &3)\ f \in C(\mathbb{R}) \end{aligned}\right\} f(x) = e^x \leftarrow \text{единственное решение}$$

$$H(p_1,\dots,p_n) = -\sum_{i=1}^{n} p_i \cdot \log p_i$$

$\exists\ h(n) = H\left(\frac{1}{n},\dots,\frac{1}{n}\right)$

Свойства:

1) монотонность

$$h(n+1) > h(n)$$

2) Аддитивность

$$\Omega = \{(1,1),(1,2),\dots,(1,m_1),$$
$$(2,1),(2,2),\dots,(2,m_2),$$
$$\vdots \quad \vdots \quad \vdots$$
$$(k,1),(k,2),\dots,(k,m_k)\}$$

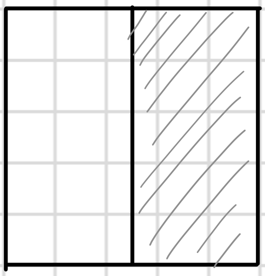
выпало $i\ j$

$$p((i\ j)) = q_{ij}$$

$$\sum q_{ij} = 1$$

$$p_i = \sum_{j=1}^{m_i} q_{ij}$$

$$p_1,\dots,p_k \rightarrow \sum p_i = 1$$



$$p((1,1) \mid (1,x)) = \frac{q_{11}}{p_1}$$

$$\Rightarrow p((i,j) \mid (i,x)) = \frac{q_{ij}}{p_i}$$

$$H(q_{11}, q_{12},\dots,q_{km_k}) = H(p_1,p_2,\dots,p_k) + \sum_{i=1}^{k} p_i \cdot H\left(\frac{q_{i1}}{p_i}, \frac{q_{i2}}{p_i},\dots,\frac{q_{im_i}}{p_i}\right)$$

3) Непрерывность

$H(p_1,\dots,p_n)$ — непрерывна как ф-ия из R^n

Докажем что все правда

1) $h(m\cdot n) = h(m) + h(n)$ // $h(n) = \log n$

▷

$$\overbrace{\begin{matrix} \frac{1}{mn} & \frac{1}{mn} & \dots & \frac{1}{mn} \\ \vdots & \vdots & & \vdots \\ \frac{1}{mn} & \frac{1}{mn} & \dots & \frac{1}{mn} \end{matrix}}^{n} \left.\right\} m \qquad \frac{1}{m}$$

$$h(m \cdot n) = h(m) + \underbrace{\sum \frac{1}{m} \cdot H\left(\frac{1}{n}, \frac{1}{n}, \dots, \frac{1}{n}\right)}_{h(n)} \circleddash$$

$$q_{ij} = \frac{1}{m \cdot n}; \quad p_i = \frac{1}{m}; \quad \frac{q_{ij}}{p_i} = \frac{1}{n}$$

$$\circleddash\ h(m) + h(n) \qquad \blacktriangleleft$$

Следствие: $h(n^r) = r \cdot h(n), \; r \in \mathbb{N}$

$$\left.\begin{matrix} x \geq 0 \\ \forall n: \; x \leq \frac{1}{n} \end{matrix}\right\} \Rightarrow x = 0$$

$$h(n) = \log_2 n \cdot h(2)$$

$\triangleright$ Пусть $2^k \leq n^r < 2^{k+1}, \; r \in \mathbb{N}$

$$\frac{k}{r} \leq \log_2 n < \frac{k+1}{r}$$

$$h(2^k) \leq h(n^k) \leq h(2^{k+1})$$

$$\exists\, h(2) = \alpha$$

$$\frac{h(2) \cdot k}{r} \leq h(n) \leq \frac{(k+1)h(2)}{r} \rightarrow \frac{\alpha k}{r} \leq h(n) \leq \frac{\alpha(k+1)}{r}$$

$$\frac{\alpha k}{r} \leq \alpha \log_2 n \leq \frac{\alpha(k+1)}{r}$$

$$\forall r: \; r \in \mathbb{N}, \; |\alpha \log_2 n - h(n)| \leq \frac{\alpha}{r}$$

$$\Rightarrow \alpha \log_2 n = h(n) \qquad \blacktriangleleft$$

$$p_i \in \mathbb{Q} \qquad p_i = \frac{q_i}{b}, \quad \sum q_i = b$$

$$H(p_1, \dots, p_n) = ?$$

$\frac{1}{b}, \frac{1}{b}, \dots, \frac{1}{b} \leftarrow q_1$ штук

$\frac{1}{b}, \frac{1}{b}, \dots, \frac{1}{b} \leftarrow q_2$ штук

$\vdots$

$\leftarrow q_n$ штук

$$h(b) = H(p_1, \dots, p_n) + \sum_i p_i \cdot h(q_i) \qquad \alpha = h(2)$$

$$\sum p_i \cdot \log_2 b \cdot \alpha = H(p_1, \dots, p_n) + \sum p_i \cdot \log_2 q_i \cdot \alpha$$

$$\log_2 q_i - \log_2 b = \log_2 \frac{q_i}{b} = \log p_i$$

$\rightarrow$ $H(p_1, \dots, p_n) = -\alpha \sum p_i \cdot \log p_i$ ✓

$\alpha = h(2)$?

$h(2) = 1$

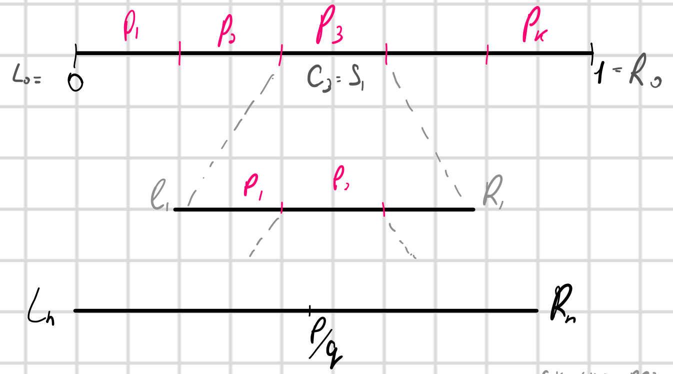
Опр.: бит – энтропия честной монеты

$[h(1000)] = [9\dots] = 10$

Арифметическое кодирование

$\exists\ S \in \Sigma^*$ (алфавит) $\quad |S| = n \quad |\Sigma| = *$

f_i – кол-во c_i в S; $p_i = \frac{f_i}{n}$



$-\log_2(R_n - L_n)$ $\leftarrow$ $\log q$ надо бит

$1 \cdot p_{s_1} \cdot p_{s_2} \cdot \ldots \cdot p_{s_n} = \prod_{i=1}^{k} p_i^{f_i}$ (сколько раз встретился p_i)

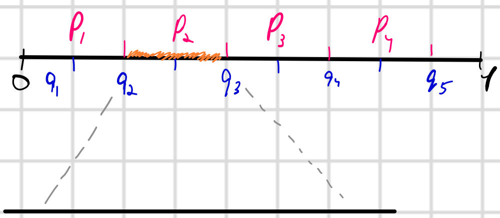
$-\log\left(\prod_{i=1}^{k} p_i^{f_i}\right) = -n \sum p_i \cdot \log_2 p_i$

$|a| \leq n \cdot H(p_1, \dots, p_k)$

Эмуляция случайных источников.

$\exists\ p_1, \dots, p_n$

$q_1, \dots, q_m$



мат. ожидание кол-ва шагов: $\approx \dfrac{H(q_1, \ldots, q_k)}{H(p_1, \ldots, p_n)}$