

Информация = — неопределенность

Рассмотрим вероятностное пространство  $\Omega = \{\omega_1 \dots \omega_n\}$   $p_1 \dots p_n$   $\sum_{i=1}^n p_i = 1$

**Определение.**  $H$  — мера неопределенности случайного источника (**энтропия**), если это отображение удовлетворяет следующему:

$$H(p_1 \dots p_n) : \mathbb{R}^n \rightarrow \mathbb{R}_+$$

$$H\left(\frac{1}{n}, \frac{1}{n} \dots \frac{1}{n}\right) < H\left(\frac{1}{n+1} \dots \frac{1}{n+1}\right)$$

$H$  — непр. от всех аргументов

Рассмотрим два эксперимента, где у первого исходы  $p_i$ , у второго  $q_{ij}$ .

$$H(p_1 q_{11}, p_2 q_{12} \dots p_1 q_{1m_1}, p_2 q_{21} \dots p_n q_{nm_n}) = H(p_1 \dots p_n) + \sum_{i=1}^n p_i H(q_{i1} \dots q_{im_i})$$

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

**Лемма 1.**  $h(n) = c \log_2 n$

$$p_i = \frac{1}{n} \quad m_i = m \quad a_{ij} = \frac{1}{m}$$

$$h(mn) = h(n) + h(m)$$

$$h(2) = c - \text{бит.}$$

$$2^i \leq n^k < 2^{i+1}$$

$$i \leq k \log_2 n < i + 1$$

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

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

$$ci \leq h(n^k) < c(i+1)$$

$$ci \leq k \cdot h(n) < c(i+1)$$

$$\frac{i}{k} \leq \frac{h(n)}{c} < \frac{i+1}{k}$$

Т.к.  $\frac{h(n)}{c}$  и  $\log_2 n$  зажимаются  $\frac{i}{k}$  и  $\frac{i+1}{k} \Rightarrow \frac{h(n)}{c} = \log_2 n$

$$]p_i = \frac{a_i}{b} \quad \triangleleft q_{ij} = \frac{1}{a_i}, m_i = a_i$$

$$h(b) = H(p_1 \dots p_n) + \sum_{i=1}^n p_i h(a_i)$$

$$\begin{aligned}
 H(p_1 \dots p_n) &= c \log_2 b - c \sum_{i=1}^n p_i \log_2 a_i = -c \left( \sum_{i=1}^n p_i (\log_2 a_i - \log_2 b) \right) = \\
 &= -c \sum_{i=1}^n p_i \log_2 p_i
 \end{aligned}$$

Рассмотрим арифметическое кодирование:

$$H(p_1 \dots p_n) = - \sum_{i=1}^n p_i \log_2 p_i$$

$$L \geq \frac{1}{2^q} \Rightarrow -\log_2 L \leq q$$

$$\prod_{i=1}^n \left( \frac{f_i}{m} \right)^{f_i}$$

$$-\log_2 \prod_{i=1}^n \left( \frac{f_i}{m} \right)^{f_i} = - \sum_{i=1}^n f_i \log_2 p_i = m \left( - \sum_{i=1}^n p_i \log_2 p_i \right) = m H(p_1 \dots p_n)$$

## 1 Симуляция одного распределения другим

Рассмотрим распределение  $1 \dots n$  с вероятностями  $p_1 \dots p_n$ ,  $n \geq 2, p_i > 0$

Задача: сгенерировать распределение с вероятностями  $q_1 \dots q_m$

Поделит отрезок  $[0, 1]$  в пропорциях  $p_i$ . Если отрезок  $p_i$  не лежит полностью в одном отрезке, то делают зум по всем отрезкам  $q_j$ , которым соответствует какое-то число из  $p_i$ .