Блочное кодирование с ошибками.

$A = \{a_1, a_2, \dots, a_k\}$

Буквы появл. независимо a_i появл. с вер. p_i

$\sum_{i=1}^{k} p_i = 1$.

$E_n : A^n \to \{0,1\}^{L_n}$ $\quad L_n = \lceil h \cdot n \rceil$

$D_n : \{0,1\}^{L_n} \to A^n$

$\varepsilon_n = \Pr_X [D(E(X)) \neq X]$ вер. то ошибки.

<u>Теорема</u> (Шеннон)

1) если $h > H(p_1, p_2, \dots, p_k)$ Тогда

$\exists E_n$ и D_n : $\varepsilon_n \xrightarrow[n\to\infty]{} 0$

2) если $h < H(p_1, p_2, \dots, p_k)$, тогда

$\forall E$ и D $\quad \varepsilon_n \xrightarrow[n\to\infty]{} 1$

<u>Д-во</u> (1)

<u>Опр.</u> $z \in A^n$ назыв. δ-типичным, если

$\forall i$ частота q_i символа e_i в z

$p_i - \delta \le q_i \le p_i + \delta$

<u>План</u> 1) $\Pr[\text{слово } \delta\text{-типичное}] = 1 - \underbrace{\lambda_n}_{\text{мало}}$ $\quad \lambda_n = o(1)$

$\delta = \delta(n)$

2) Число δ-типичных слов $< 2^{L_n}$

a_1 зафикс. $\Pr[q_1 \notin [p_1 - \delta; p_1 + \delta]]$

$X \leftarrow A^n$ $\quad q_1$ - частота a_1

$Y_i = \begin{cases} 1, & \text{если } i\text{-я буква} = a_1 \\ 0, & \text{иначе} \end{cases}$ $\quad q_1 = \frac{\sum Y_i}{n}$

$E[Y_i] = \Pr[Y_i = 1] = p_1$

Y_i - нез.,

$\Pr\left[\left|\frac{\sum_{i=1}^{n} Y_i}{n} - p_i\right| \ge \delta\right] \le \underbrace{2e^{-2\delta^2 n}}_{= 2e^{-2\sqrt{n}}}$ оценка Чернова

$\delta = 1/n^{1/4}$

$\Pr[\exists \text{ у одной буквы частота отличается на } \delta] \le \underbrace{K \cdot 2e^{-2\sqrt{n}}}_{\to 0}$

2) Оценка числа δ-типичных слов

$\begin{array}{ccccc} a_1 & a_2 & - & - & a_k \\ n_1 & n_2 & & & n_k \end{array}$ | Коэф $\frac{n!}{n_1!\, n_2! \dots n_k!}$

$$n! = \sqrt{2\pi n}\left(\frac{n}{e}\right)^n (1+o(1))$$

$$\log\left(\frac{n!}{n_1!\,n_2!\cdots n_k!}\right) = \log\left(\frac{\left(\frac{n}{e}\right)^n}{\left(\frac{n_1}{e}\right)^{n_1}\left(\frac{n_2}{e}\right)^{n_2}\cdots\left(\frac{n_k}{e}\right)^{n_k}}\right) +$$

$$+ O(\log n) = \log\left(\frac{1}{\left(\frac{n_1}{n}\right)^{n_1}\left(\frac{n_2}{n}\right)^{n_2}\cdots\left(\frac{n_k}{n}\right)^{n_k}}\right) +$$

$$+ O(\log n) = \log\left(\frac{1}{q_1^{n_1}\,q_2^{n_2}\cdots q_k^{n_k}}\right) + O(\log n) =$$

$$\boxed{q_i = \frac{n_i}{n}} = n\left(\frac{n_1}{n}\log\frac{1}{q_1} + \frac{n_2}{n}\log\frac{1}{q_2} + \dots + \frac{n_k}{n}\log\frac{1}{q_k}\right) + O(\log n)$$

$$= n\,H(q_1, q_2, \dots, q_k) + O(\log n) \quad \leq$$

Пусть слово δ-типичное $\quad |q_i - p_i| \leq \delta$

$$|H(q_1, q_2, \dots, q_k) - H(p_1, p_2, \dots, p_k)| = O(\delta)$$

$$q_i \log\frac{1}{q_i} - p_i \log\frac{1}{q_i}$$

$$\leq \underbrace{n \cdot H(p_1, p_2, \dots, p_k) + O(\log n) + O(\delta n)}_{n\left(H(p_1, \dots, p_k) + \underbrace{O(\delta) + O\left(\frac{\log n}{n}\right)}\right)} < L_n$$

#δ-типичных слов $\leq$

число типов $\cdot\, 2^{n\left(H(p_1, p_2, \dots, p_k) + O(\delta) + O\left(\frac{\log n}{n}\right)\right)}$

$\|$

poly(n)

$n_1 \quad a_1$

$n_2 \quad a_2$

$n_k \quad a_k$

(2) z $\qquad q_1 \ldots q_k$

α δ-типичное слово

$$Pr[z] = p_1^{q_1 n} \cdot p_2^{q_2 n} \cdots p_k^{q_k \cdot n} =$$

$$= 2^{q_1 n \log p_1 + q_2 n \log p_2 + \ldots + q_k n \log p_k}$$

$$= 2^{n\left(-\sum q_i \log \frac{1}{p_i}\right)} \leq$$

$$\leq 2^{-H(p_1, \ldots, p_k) n + O(\delta n)}$$

Всего можно корректно декодировать $\leq 2^{Ln}$ δ-типичных слов

$$2^{Ln} \cdot 2^{-H(p_1, \ldots, p_k) n + O(\delta n)} \longrightarrow 0$$

Комбинаторное q b

$$\boxed{H(X,Y) \leq H(X) + H(Y)}$$

$a_1 \ldots a_{k_1}$ $\qquad (a_i, b_j)$

$b_1 \ldots b_{k_2}$ $\qquad q_{i,j}$

$$H(X) + H(Y)$$

$H(X)$

$$H(X,Y) = H(X) + \underline{H(Y|X)}$$

$$0 \le H(Y|X) \le H(Y)$$

Взаимная информация

X и Y сл. велич.

кол-во инф. в X об Y

$$I(X:Y) = H(Y) - H(Y|X)$$

Св-ва

1) $I(X:Y) = H(X) + H(Y) - H(X,Y)$

Дво $I(X:Y) = H(Y) - H(Y|X) =$

$= H(Y) - (H(X,Y) - H(X)) =$

$= H(X) + H(Y) - H(X,Y)$

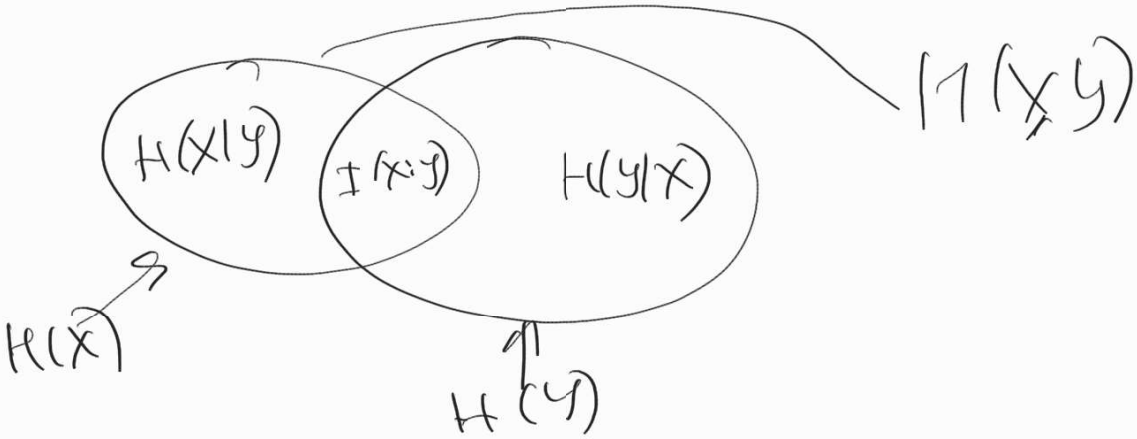
2) $I(X:Y) \ge 0 \quad = 0 \iff X$ и Y независ.

3) $I(X:Y) = I(Y:X)$

4) $I(X:Y) \le H(Y)$

$I(X:Y) \le H(X)$

5) $I(X:X) = H(X)$



Условная взаимная информация

$$I(X:Y|Z) = H(Y|Z) - H(Y|XZ) =$$
$$= H(X|Z) + H(Y|Z) - H(X,Y|Z)$$

Утв. $I(X:Y|Z) = H(X,Z) + H(Y,Z) -$
$$- H(X,Y,Z) - H(Z)$$

$$H(X|Z) = H(X,Z) - H(Z)$$
$$H(Y|Z) = H(Y,Z) - H(Z)$$
$$H(X,Y|Z) = H(X,Y,Z) - H(Z)$$

Утв. $I(X:Y|Z) \geq 0$

$I(X:Y) \geq 0$

Примеры

① X, Y, Z — случ. биты,
$X \oplus Y \oplus Z = 0$

$$I(X:Y) = H(X) - H(X|Y) =$$
$$= 1 - 1 = 0$$

$$I(X:Y,Z) = \underset{1}{H(X)} - \underset{0}{H(X|YZ)} = 1$$

$$I(X:Y|Z) = \underset{1}{H(X|Z)} - \underset{0}{H(X|Y,Z)} = 1$$

(2) $X = \underbrace{x_1 x_2}$ $\quad Y = \underbrace{y_1 y_2}$

$$x_1 \oplus x_2 = y_1 \oplus y_2$$

$$Z = x_1 \oplus x_2 = y_1 \oplus y_2$$

$$I(X:Y) = H(X) - H(X|Y) = 2 - 1 = 1$$

$$I(X:Y|Z) = \underset{1}{H(X|Z)} - \underset{1}{H(X|ZY)} = 0$$

Лемма (правило цепочки)

$$I(XY:Z) = \underbrace{I(X:Z)} + I(Y:Z|X)$$

Д-во

$$I(XY:Z) = H(XY) - H(XY|Z) =$$

$$= \underbrace{H(X)} + H(Y|X) - \underbrace{H(X|Z)} - H(Y|XZ)$$

$H(X)$, $H(Y)$ $\quad H(X,Y)$

Энтропийные пространства

Лемма $\forall h \geq 0 \ \exists$ случ. вел. X:

$H(X) = h$.

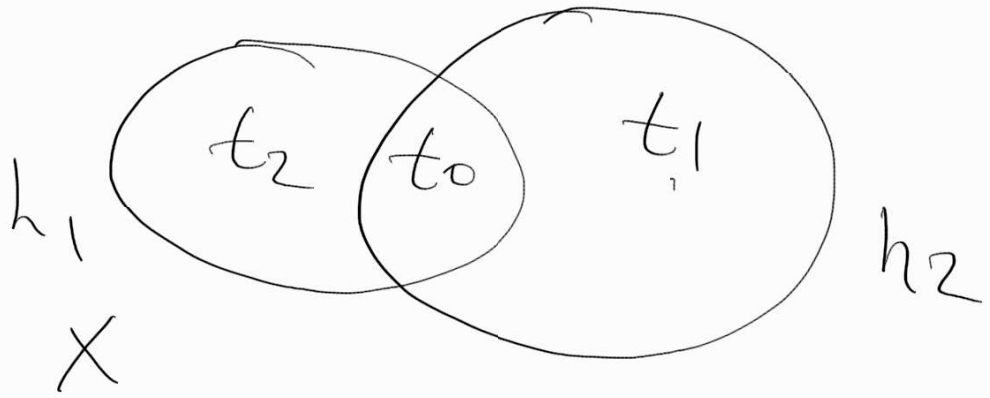
$0 \leq h \leq \log n$

Д-во

$$\lambda \underbrace{(1, 0, \ldots, 0)}_{0} + (1-\lambda)\underbrace{\left(\frac{1}{n}, \frac{1}{n}, \ldots, \frac{1}{n}\right)}_{\log n}$$

$$\underset{h_1}{H(X)} \qquad \underset{h_2}{H(Y)} \qquad \underset{h_{1,2}}{H(X,Y)}$$

$$(*) \begin{cases} h_{1,2} \leq h_1 + h_2 \\ h_1 \leq h_{1,2} \\ h_2 \leq h_{1,2} \end{cases} \qquad \begin{aligned} t_0 &= I(X:Y) \geq 0 \\ t_1 &= H(Y|X) \geq 0 \\ t_2 &= H(X|Y) \geq 0 \end{aligned}$$



Лемма Если

вып усл (*), то

$\exists X, Y, Z$: $H(X) = h_1$

$H(Y) = h_2$, $H(X,Y) = h_{1,2}$

Д-во

ξ_0 ξ_1 ξ_2

независ. случ. вел.

$H(\xi_0) = t_0$, $H(\xi_1) = t_1$

$H(\xi_2) = t_2$

$Y = (\xi_0, \xi_1)$ $X = (\xi_0, \xi_2)$

$H(Y) = t_0 + t_1 = h_2$

$H(X) = t_0 + t_2 = h_1$

$$H(x,y) = t_0 + t_1 + t_2$$
$$= h_{1,2}$$