4. Кодирование источников сообщений

4.1. Источники сообщений

$S = \{A, P\}$ — дискретный стационарный источник без памяти

Вообще вероятн-ть порождения сообщ-я зависит от времени и уже порождённого, но мы такие рассм-ть не будем

Опр Источник Маркова — источник с памятью

4.2. Кодирование источников

Опр $S = \{A, P\}$ $\quad A = \{x_1, \dots, x_N\}$ $\quad P = \{p_1, \dots, p_N\}$ $\quad \sum^N p_i = 1$ $\quad 0 < p_i \leq 1$

A - алфавит $\quad x_i$ - буква $\quad B$ - алфавит кодовых символов $\quad B = \{0, 1\}$

B^* - мн-во слов в алфавите B

Опр Алфавитное кодирование источника — $\varphi: A \to B^*$

$\varphi(x_i)$ - кодовое слово для буквы x_i

$|\varphi(x_i)|$ - длина кодового слова

Опр Стоимость кодирования: $\boxed{C(\varphi, S) = \sum_{i=1}^{N} p_i |\varphi(x_i)|}$

(скорость кодир-я)

Опр Кодирование равномерное, если $|\varphi(x_i)| = |\varphi(x_j)| \quad \forall i \neq j$

Опр Кодирование разделимо (однозначно декодируемо), если любая посл-ть код. слов однозначно раздел. на код. слова.

Опр Кодирование префиксное, если никакое код. слово не явл. началом (префиксом) никакого другого код. слова.

Зам. Префиксные коды предст. в виде деревьев

Префиксн. коды всегда разделимые.

Ex S: $A = \{a, b, c, d, e\}$ $\quad P = \{0{,}3,\ 0{,}2,\ 0{,}2,\ 0{,}15,\ 0{,}15\}$ $\quad B = \{0, 1\}$

φ_1: a - 000, b - 001, c - 010, d - 011, e - 100 — равномерное код-е

$C(\varphi_1, S) = 3$

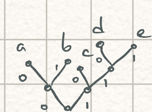
φ_2: a - 00, b - 01, c - 10, d - 11, e - 001 — неразделимое

контр. пр.: ee = acb

$C(\varphi_2, S) = 2{,}15$

φ_3: a - 00, b - 01, c - 10, d - 110, e - 111 — префиксное

$C(\varphi_3, S) = 2{,}3$ ← оптимальная (меньше для разд. кода нельзя)



φ_4: a - 0, b - 01, c - 011, d - 0111, e - 01111 — не префиксн., но разделимый

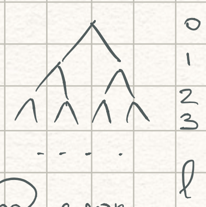
$C(\varphi_4, S) = 2{,}65$

Задача: строить для источника разделимое кодир-е с наим. стоимостью

4.3. Неравенство Крафта для префиксных кодов (будет на коллоке)

Теорема.
Пусть $l_1, \dots, l_N$ — положит. целые числа.
Префиксный код для источника $S=(A,P)$ с длинами $l_1, \dots, l_N$ существует $\Leftrightarrow$
$\Leftrightarrow \sum_{i=1}^{N} 2^{-l_i} \le 1$

Д-во: Рассмотрим полное бинарное дерево высоты $l := \max_{i=\overline{1,N}} l_i$
на слое l — 2^l верш.



$\Rightarrow$ Рассм. преф. код на полн. бин. дереве
Если нашли код. слово, то в его поддереве нет код. слов.
Т.е. $\underbrace{01\dots1}_{l_i}|\overbrace{*\dots*}^{l-l_i}$ (всего l) Надо, чтобы теги поддеревьев не пересекались:

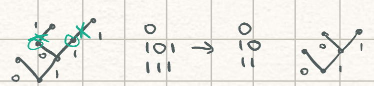
$$2^l \ge \sum_{i=1}^{N} 2^{l-l_i} \Rightarrow \sum_{i=1}^{N} 2^{-l_i} \le 1$$

$\Leftarrow$ Построим преф. код с длинами $l_1, \dots, l_N$.
Пусть s_i — число эл-тов l_j, равных i, $i=1,2,\dots,l$
$\sum_{i=1}^{N} 2^{-l_i} \le 1 \Rightarrow \sum_{i=1}^{l} s_i \cdot 2^{-i} \le 1$ $\langle$ Ex: $l_1=1, l_2=1, l_3=3, l_4=5, l_5=5$; $s_1=2, s_2=0, s_3=1, s_4=0, s_5=2$ $\rangle$

Строим код:
1. Возьмём любые s_1 двоичных слов длины 1 в кач-ве кодовых
(это возможно, т.к. $s_1 \le 2$)
Вычеркнем в полн. бин. дереве запрещ. (из-за префиксности) вершины
2. Возьмём любые s_2 двоичн. слов длины 2 из оставшихся
(это возможно, т.к. $s_1 \cdot 2^{-1} + s_2 \cdot 2^{-2} \le 1 \rightarrow s_2 \le 2^2 - 2s_1$)
3. и т.д. Вот и построили!

Опр Префиксный код **примитивный**, если его невозможно сократить

Ex



примитивность $\ne$ оптимальность

Опр Код **оптимальный**, если стоимость минимальна

4.4. Неравенство Крафта для разделимых кодов.

Теорема:
Пусть $l_1, \dots, l_N$ — положит. целые числа.
Разделимый код для источника $S=(A,P)$ с длинами $l_1, \dots, l_N$ существует $\Leftrightarrow$
$\Leftrightarrow \sum_{i=1}^{N} 2^{-l_i} \le 1$

Д-во: $\Leftarrow$ По т.1 сущ. префиксн. код, а он разделимый.
$\Rightarrow$ Пусть M — нктр полож. целое число. Тогда рассмотрим сумму:

$$\left(\sum_{i=1}^{N} 2^{-l_i}\right)^M = \left(\sum_{i_1=1}^{N} 2^{-l_{i_1}}\right) \dots \left(\sum_{i_M=1}^{N} 2^{-l_{i_M}}\right) = \sum_{i_1=1}^{N} \dots \sum_{i_M=1}^{N} 2^{-(l_{i_1}+\dots+l_{i_M})} \,(=)$$

$l_{i_1} + \dots + l_{i_M}$ — суммарн. длина соотв. посл-ти M код. слов
Пусть A_j — число послед-тей M код. слов сумм. длины j

Ex: a - 0 $l_1 - 1$
b - 01 $l_2 - 2$
c - 011 $l_3 - 3$
d - 0111 $l_4 - 4$
e - 01111 $l_5 - 5$

$M = 3$
$l_{i_1} + l_{i_2} + l_{i_3}$
$l_2 + l_3 + l_5 = \underline{10} \leftarrow j$

надо перебирать посл-ти A_j найдём

$= \sum\limits_{j=1}^{M \cdot l} A_j \cdot 2^{-j}$, где $l = \max\limits_{i=\overline{1,N}} l_i$

Заметим, что $A_j \leq 2^j$ (т.к. двоичн. посл-тей длины j $= 2^j$). Тогда:

$$\left(\sum_{i=1}^{N} 2^{-l_i}\right)^M = \sum_{j=1}^{Ml} A_j \cdot 2^{-j} \leq \sum_{j=1}^{Ml} 1 = Ml$$

$$\Rightarrow \sum_{i=1}^{N} 2^{-l_i} \leq \sqrt[M]{Ml} = 2^{\frac{1}{M}\log_2 Ml} \xrightarrow[M \to \infty]{} 1$$, ура!