§ 符号のクラス

符号の分類とその特徴づけを勉強しよう．
情報源アルファベットを実際に伝達する場合，情報源記号を符号に変換することが必要なる．
これを符号化とよぶ．具体的には情報源アルファベット

$$S = \begin{Bmatrix} s_1, & s_2, & \cdots & s_n \\ P(s_1), & P(s_2), & \cdots & P(s_n) \end{Bmatrix}$$

に対する符号を構成することが符号化である．

例1

$S = \begin{Bmatrix} a, & b, & c, & d \\ 0.4, & 0.3, & 0.2, & 0.1 \end{Bmatrix}$ としよう．例えばa→0, b→10, c→110, d→1110とすることが符号化である．

平均符号長

符号語 c_k を構成する記号の個数 l_k を**符号語長**とよぶ．**平均符号長** L は

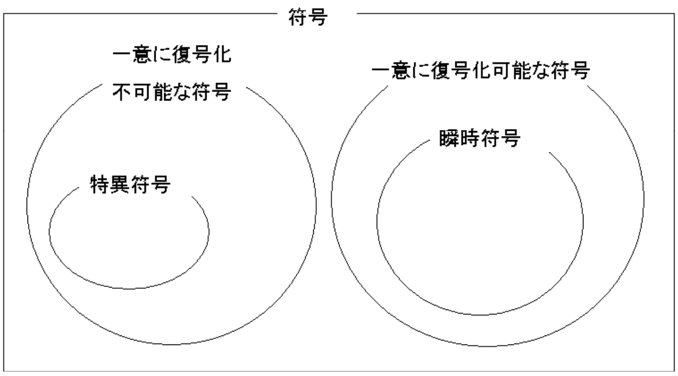
$L = \sum_{k=1}^{n} l_k P(s_k)$ で表される．

例えば例１の平均符号語長は $l_a = 1, l_b = 2, l_c = 3, l_d = 4$ だからL=１×0.４+2×0.３+3×0.２+4×0.１=2である．

平均符号長を小さくするには大きい出現確率をもつアルファベットに短い符号，小さい出現確率をもつアルファベットに長い符号を割り当てればよい．

問１　例1でa→1110，b→110，c→10，d→0とすると平均符号長はどうなるか？

符号のクラス



特異符号とは１つの符号語を異なる記号に割り当てた符号である．瞬時符号とはその符号が送られた瞬間に復号できる符号のことである．

<u>問2</u>　次の符号は上の図のどこにあてはまるか？
(1) a→0, b→10, c→10, d→11
(2) a→0, b→01, c→10, d→11
(3) a→0, b→10, c→110, d→1110
(4) a→0, b→01, c→011, d→0111

（答）
(1) bとcに同じ符号を割り当てているから特異符号である.
(2) 特異符号ではないが, 0110がbcなのかadaなのか区別がつかないから<u>一意に復号化不可能な</u>符号である.
(3) ある符号を得たときその符号に続きがあって他の符号になるということはないから瞬時符号である.
(4) 例えばaを得ているとしよう. このとき後に1が来ればc,dの可能性がある. 運良く0が来ればaと判断できる. 直後に0を得た時点でどの符号を得ているかわかるということは全ての場合にあてはまる. このことからこの符号は<u>一意に復号可能だが瞬時符号でない</u>ことがわかる.

<u>問３</u>　（３）と（４）の違いは符号木を描いてみるとよくわかる．符号木を描いてみよ．

問3の答

図から次のことがわかる．

<u>定理1</u>　次の（１），（２）は同値である．
（１）瞬時符号である．
（２）全ての符号語が符号木の枝の末端(葉)に割り振られている．

<u>定理2</u>　等長符号(符号の長さが全て等しい符号)は特異符号でなければ, 瞬時符号である.

(証明)　等長符号は符号木の枝の末端(葉)に割り振られている.

問4 情報源S={a, b, c, d, e}を次の表のように符号化した．符号C1～C6について以下の問に答えよ．

	C_1	C_2	C_3	C_4	C_5	C_6
a	0	1	0	0	100	1
b	10	110	10	01	101	01
c	110	001	110	011	110	000
d	1110	011	1110	0111	111	0010
e	1011	101	11110	01111	000	0011

（１）C_1，C_2 は一意に復号化不可能な符号らしい．理由を述べよ．
（２）瞬時符号を列挙せよ
（３）情報源Sの発生確率を$S=\begin{Bmatrix} a, & b, & c, & d, & e \\ \frac{1}{2}, & \frac{1}{4}, & \frac{1}{8}, & \frac{1}{16}, & \frac{1}{16} \end{Bmatrix}$とする．
瞬時符号でこの情報源に最も適した符号はどれか?（ヒント　平均符号長を計算せよ）

問4の答

§クラフトの不等式

定理　符号語長 $\{l_1, l_2, \cdots, l_n\}$ をもつr元瞬時符号が存在するための必要十分条件は次の不等式をみたすことである．

$$\sum_{k=1}^{n} r^{-l_k} \leq 1 \qquad \text{（クラフトの不等式）}$$

注意　クラフトの不等式は瞬時符号であるための必要十分条件ではなく瞬時符号が**存在するための**必要十分条件である．つまり符号語長 がクラフトの不等式をみたしていれば，**賢く符号を構成すれば**瞬時符号にできるということである．

注意の例
問2の
（３）a→0，b→10，c→110，d→1110
（４）a→0，b→01，c→011，d→0111
で（３）は瞬時符号で（４）はそうではなかった．しかし，共に となりクラフトの不等式をみたす．よって

$$\sum_{k=1}^{4} 2^{-l_i} = \frac{1}{2} + \frac{1}{2^2} + \frac{1}{2^3} + \frac{1}{2^4} = \frac{15}{16} < 1$$

「瞬時符号である」　　　　　「クラフトの不等式をみたす」

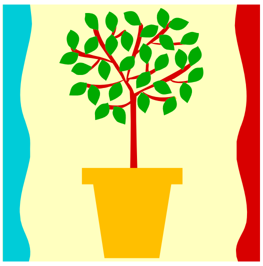
という関係式が成り立つことがわかる．

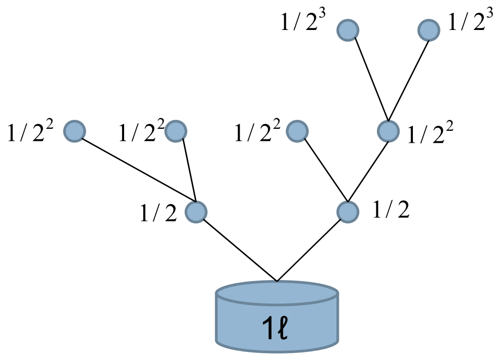
定理の証明

２元符号の場合を考えよう．木の根に量１の水をあたえることにしよう．この水は最初の枝で1/2と1/2に分流し，次の分岐でもまた２分され$1/2^2$と$1/2^2$に分流される．同様にしてm次の分岐では$1/2^m$と$1/2^m$に分流されることがわかる．符号語に対応するノード(葉)の下にバケツをおき，バケツに流れてくる水量をすべてのバケツについて足し合わせたものがクラフトの不等式の左辺となる．

明らかに瞬時符号ではバケツの水の総和は１以下となることがわかる．

逆にクラフトの不等式がみたされるならば，l_k に対して $l_k - 1$ 次の葉を選んでやれば瞬時符号が構成できることがわかる．





問5
情報源S={a, b, c, d}に対する次の符号長を持つ瞬時符号が構成できるかどうか判定せよ．
構成できるものについてはその例を挙げよ．
（１）（１，２，２，２）の２元符号
（２）（１，２，２，２）の３元符号

問5の答