§ハフマン符号化

<u>定義</u>　（コンパクト符号）ある与えられた情報源 S に対して（N 次拡大情報源でない）に対し一意復号可能な符号に符号化するとき平均符号長を最小にする符号を**コンパクト符号**という．
このコンパクト符号の構成法がハフマン(Huffman)によって与えられている．

2 元ハフマン符号の構成法

（１）　各情報源記号に対する対応する葉を作る．おのおのの葉には，情報源記号の発生確率を記しておく．
（２）　確率の最も小さい 2 枚の葉に対して，1 つの節点を作りその節点と 2 枚の葉を枝で結ぶ．この 2 本の枝の一方には０，他方には１を割り当てる．さらにこの節点に 2 枚の葉の確率の和を記し，この節点を新たな葉と考える．
（３）　葉が 1 枚しか残っていなければ, 符号の構成が完了している. そうでなければ,（２）へ戻る.

<u>問 1</u>　情報源 S を $S=\begin{Bmatrix} a, & b, & c, & d \\ 0.6, & 0.25, & 0.1, & 0.05 \end{Bmatrix}$ とする．a,b,c,d のハフマン符号を求めよ．

問２　情報源Ｓを$S=\begin{Bmatrix} a, & b, & c, & d \\ 0.35, & 0.3, & 0.2, & 0.15 \end{Bmatrix}$とする．a,b,c,d のハフマン符号を求めよ．

多分皆さんの答えは２通りに分かれただろう．しかし，どちらの符号化も共に平均符号長は２である．このようにハフマン符号は与えられた情報源に対して一意的には決まらない．しかし，そのような場合でも平均符号長は一致する（そうでないとコンパクト符号にならないから）．
さて次にＮ次拡大情報源をハフマン符号化することを考えてみよう．

問３　無記憶情報源　$S=\begin{Bmatrix} a, & b \\ \frac{1}{5}, & \frac{4}{5} \end{Bmatrix}$を考える．この情報源に対する２次拡大情報源$S^2$を求めハフマン符号化し，１情報源記号あたりの平均符号長を求めよ．同様に３次拡大情報源S^3を求めハフマン符号化し，１情報源記号あたりの平均符号長を求めよ．さらにエントロピーH(S)を求め，１次，２次，３次の順に１情報源記号あたりの平均符号長がエントロピーH(S)に近づくことを確かめよ．$\log_2 5 = 2.322$とする．

一定個数の情報源記号ごとにまとめて符号化する方法を**ブロック符号化**といい，それによって構成される符号を**ブロック符号**という．N 次拡大情報源に対する符号化は N 個の符号をまとめて行うブロック符号化である．特に N 次拡大情報源に対するハフマン符号化は**ハフマンブロック符号化**という．

ところで，ハフマン符号はコンパクト符号でかつ瞬時符号であるから当然平均符号長は前回の 1 次エントロピーに対する不等式

$$H_r(S) \leq L < H_r(S) + 1$$

をみたしている．

N 次拡大情報源に対するハフマン符号化（ハフマンブロック符号化）を行えば

$$\frac{H_r(S^N)}{N} \leq L < \frac{H_r(S^N)}{N} + \frac{1}{N}$$

をみたしていることになる．よって次が成り立つ．

<u>重要</u>

ハフマンブロック符号化を行うと 1 情報源記号あたりの平均符号長は約 1/N の誤差でエントロピーH(S)に近づいて行く．これは，シャノンーファノ符号化によるエントロピーへの近づき方よりも速い．