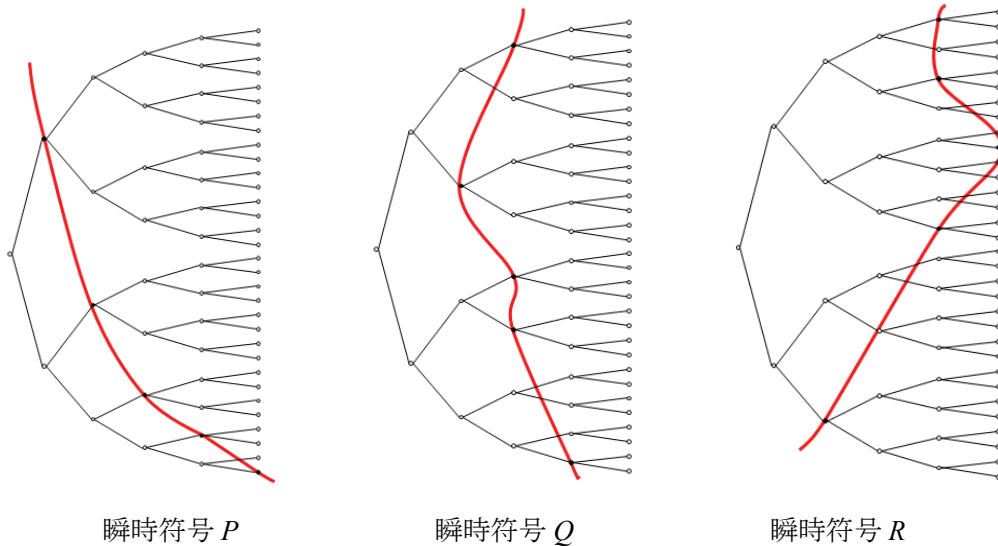


3.1 コンパクト符号

所与の情報源 S から発生する情報源記号に一つずつ符号語を割り当てる符号化によってできる一意復号可能な符号のうち、平均符号長が最小となる符号をコンパクト符号という。マクミラン不等式とクラフト不等式が同形だから、任意の一意復号可能なコンパクト符号に対して、それと同じ符号語長セットを持つ瞬時符号が存在する。情報源 S に対するコンパクト符号が複数存在することがあるので注意したい。

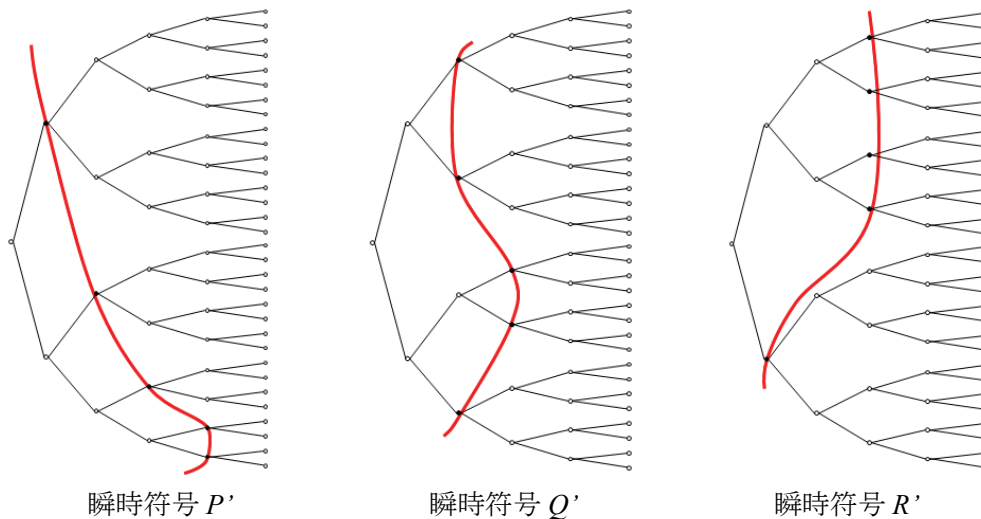
原理的には、与えられた情報源 S に対するコンパクト符号を見つけるためには、すべての瞬時符号を枚挙し、そのなかで平均符号長が最少になるものを選べばよい。後述するハフマン符号化などのようにコンパクト符号をつくる作業を効率化する方法が知られているが、その前に、どのような符号がコンパクト符号になるか、2章であげた3つの瞬時符号を例にとり、原理に基づいて考えてみよう。



上の例はいずれも冗長である。たとえば瞬時符号 P については、現状では符号長バッグは $\{1,2,3,4,5\}$ である。符号の木を吟味すれば、同じ5個の情報源記号に対して、符号長バッグ $\{1,2,3,4,4\}$ となる瞬時符号 P' を構成できることがわかる。瞬時符号 P' は第5番目の構成要素に対応する符号語だけについて P の符号語より符号長が短く、それ以外の構成要素に対応する符号語については、 P とまったく同じ符号語であるから、各符号語に対する情報源記号の出現確率がゼロでないいかなる値であっても、瞬時符号 P' の平均符号長は P の平均符号長よりも短くなる。同様に、瞬時符号 Q についても、現状の符号長バッグ $\{3,2,3,3,4\}$ よりも平均符号長の短い符号長バッグ $\{2,2,3,3,2\}$ をもつ符号を構成できる。

問い 瞬時符号 R についても、現状の符号長バッグよりも平均符号長の短い符号長バッグをもつ符号を構成できることを示せ。

このように、与えられた瞬時符号の符号長バッグを構成する要素のなかに、それをより小さな値に置き換えた符号長バッグに対応する瞬時符号が構成可能なものが一つでもあれば、その瞬時符号は冗長であり、コンパクト符号ではありえない。このような検討に基づいて、瞬時符号 P , Q , R のそれぞれを改良して作った次の瞬時符号 P' , Q' , R' はコンパクト符号の候補になる。



以下に示すように、情報源の性質によってその平均符号長は異なるので、これらのうちのどれがコンパクト符号になるかは情報源の性質に依存する。

(例 1) 情報源記号発生確率が $\langle A: 0.2, B: 0.2, C: 0.2, D: 0.2, E: 0.2 \rangle$ の場合は、

$$\text{瞬時符号 } P' \text{ の平均符号長} = 0.2 \times 1 + 0.2 \times 2 + 0.2 \times 3 + 0.2 \times 4 + 0.2 \times 4 = 2.8$$

$$\text{瞬時符号 } Q' \text{ の平均符号長} = 0.2 \times 2 + 0.2 \times 2 + 0.2 \times 3 + 0.2 \times 3 + 0.2 \times 2 = 2.4$$

$$\text{瞬時符号 } R' \text{ の平均符号長} = 0.2 \times 3 + 0.2 \times 3 + 0.2 \times 3 + 0.2 \times 3 + 0.2 \times 1 = 2.6$$

この場合は瞬時符号 Q' の平均符号長が 3 つの中で一番短い。

(例 2) 情報源記号発生確率が $\langle A: 0.6, B: 0.2, C: 0.1, D: 0.07, E: 0.03 \rangle$ の場合は、

$$\text{瞬時符号 } P' \text{ の平均符号長} = 0.6 \times 1 + 0.2 \times 2 + 0.1 \times 3 + 0.07 \times 4 + 0.03 \times 4 = 1.7$$

$$\text{瞬時符号 } Q' \text{ の平均符号長} = 0.6 \times 2 + 0.2 \times 2 + 0.1 \times 3 + 0.07 \times 3 + 0.03 \times 2 = 2.17$$

$$\text{瞬時符号 } R' \text{ の平均符号長} = 0.6 \times 3 + 0.2 \times 3 + 0.1 \times 3 + 0.07 \times 3 + 0.03 \times 1 = 2.94$$

この場合は瞬時符号 P' の平均符号長が 3 つの中で一番短い。

一般に、2元符号瞬時符号 C がコンパクト符号であるならば、次の補助定理 1 の命題が成立する。

【補助定理 1】 (瞬時符号 C がコンパクトであるための必要条件)

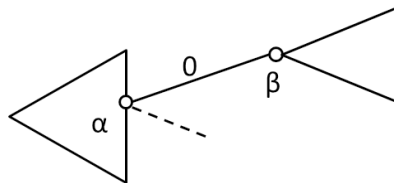
2 個以上の情報源記号をもつ情報源 S が与えられたとする。 S に対する瞬時符号 C がコンパクトであるならば、 C に対する符号の木 T において、

- (1) 葉以外の節点は必ず 2 個の子節点を持つ。
- (2) 情報源記号 α, β の出現確率をそれぞれ P_α, P_β とする。 $P_\alpha < P_\beta$ ならば情報源記号 α, β に割り付けられる符号語をそれぞれ c_α, c_β とすると、 $|c_\alpha| \geq |c_\beta|$ である。

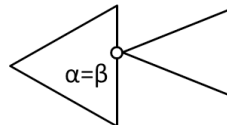
証明骨子は次の通りである。

(1) について：

「葉以外の節点が必ず 2 個の子節点を持つ」 でないとすると、 T の節点のなかに、 1 個しか子節点を持たないものが存在するというを意味する。 すると、 次のように、 より平均符号長が短い瞬時符号が構成されることになり、 次のように符号の木 T がコンパクト符号であることに反する。



ここで、 ノード α は分岐 0 に対応する子節点 β しか持っていないとしたら、 下のようになるほうが平均符号長が短くなることは明らか。

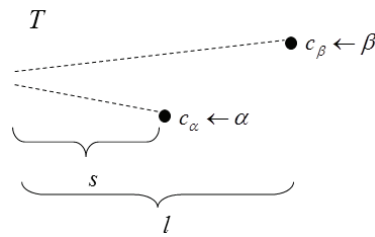


(2) について：

仮にコンパクトな瞬時符号の木 T において、 補助定理の主張が成立しないと仮定してみよう。 すると次のように T がコンパクトであることと矛盾が生じてしまう。

コンパクト符号に対する符号 C の木 T において $P_\alpha < P_\beta$ なる出現確率を持つ情報源記号 α, β に割り付けられる符号語 c_α, c_β の深さを s, l とすると、 $s \geq l$ となる。なぜならば、 $s < l$ と仮定すると、 T において符号語 c_α, c_β への情報源記号の割り付けを入れ替えた符号の木 T' に対応する符号 C' の平均符号長 L' が T に対応する符号の平均符号長 L より小さくなり、 C がコンパクトであるという前提に反するからである。

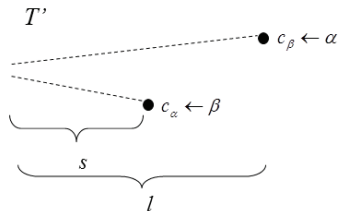
つまり，次図：



に示すように， T において情報源記号 α ， β に対して，符号語 c_α ， c_β をそれぞれ割り当てたときの平均符号長 を L とすると，

$$L = \dots + sp_\alpha + lp_\beta$$

となるが，次図：



に示すように， T における情報源記号 α ， β に対する符号語 c_α ， c_β の割り当てを入れ替えたときの平均符号長 を L' とすると，

$$L' = \dots + sp_\beta + lp_\alpha$$

であり， $s < l$ ， $p_\alpha < p_\beta$ ならば，

$$L' - L = (sp_\beta + lp_\alpha) - (sp_\alpha + lp_\beta) = (l - s)(p_\alpha - p_\beta) < 0$$

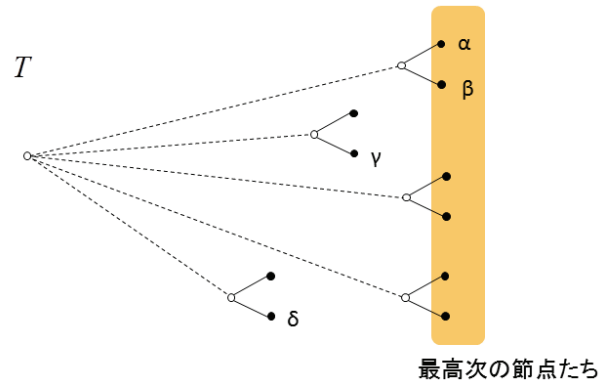
により $L' < L$ が導かれ， C がコンパクトであるという前提に対する矛盾が生じる．補助定理の主張が成立しないと仮定すると矛盾が生じるので，背理法により補助定理が証明された

■

以上の通り，コンパクト符号は補助定理 1 を満足しなければならないことがわかった．そのもとの，各情報源記号にどのように符号語が割り当てられるか，イメージしてみよう．すなわち，3 個以上の情報源記号をもつ情報源 S の情報源記号を $A = \{a_1, \dots, a_n\}$ ，各情報源記号 a_i の出現確率を p_i とする．また，すべての $1 \leq j \leq n-2$ に対して $p_j \geq p_{n-1} \geq p_n$ とする． S のコンパクトな瞬時符号 T において， a_{n-1} と a_n はどの節点に対応づけられているか，考えてみよう．

補助定理 1 で得られた知見を使うと，典型的には， a_{n-1} と a_n は最高次の節点（例えば，下図 α と β ）に割り付けられる．しかし， $1 \leq \{i, j\} \leq n-2$ に対して $p_i = p_j = p_{n-1} = p_n$ と

なっているときは、下図節点 α や β には a_i や a_j が割り付けられ、 a_{n-1} と a_n は最高次でない節点（例えば、数 γ や δ ）に割り付けられることもある。



以上をまとめると、いずれの場合でも、平均符号長を変えることのないように、節点への符号語の割り付けを変えることにより、下図の T' に対応するコンパクト符号を構成できる。

