3.2 コンパクト符号の構成法

次の補助定理 2 は，コンパクト符号を構成するための十分条件を与える．

【補助定理 2】（コンパクトな瞬時符号 C を構成するための十分条件）
3 個以上の情報源記号をもつ情報源 S の情報源記号を $A=\{a_1,\cdots,a_n\}$, 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 $1\le j\le n-2$ に対して $p_j \ge p_{n-1} \ge p_n$ とする．

このとき， a_n と a_{n-1} を一つの記号 b_{n-1} に統合して，情報源記号の集まり $A'=\{a_1,\cdots,a_{n-2},b_{n-1}\}$，各情報源記号の出現確率を $\{p_1,\cdots,p_{n-2},p_{n-1}+p_n\}$ とする情報源 S' のコンパクトな瞬時符号 C' が得られたとする．

すると，C' において，情報源記号 b_{n-1} に割り付けられた符号語 c_b のかわりに，$a_{n-1}\leftarrow c_b 0, a_n \leftarrow c_b 1$ という割り付けを加えた符号 C もまたコンパクトな瞬時符号である．

証明をする前に，例を用いてこの補助定理がどう使えるか示そう．

【補助定理 2】の使い方：

情報源 S: 情報源記号発生確率　〈A: 0.6, B : 0.2, C: 0.1, D: 0.07 , E: 0.03〉

情報源 S_1: 情報源記号発生確率　〈A: 0.6, B : 0.2, C: 0.1, D: 0.1〉

情報源 S_2: 情報源記号発生確率　〈A: 0.6, B : 0.2, C: 0.2〉

情報源 S_3: 情報源記号発生確率　〈A: 0.6, B : 0.4〉

〈A←0, B ←1〉は情報源 S_3 のコンパクト符号

〈A←0, B ←10, C ←11 〉は，情報源 S_2 のコンパクト符号

〈A←0, B ←10, C ←110, D ←111〉は，情報源 S_1 のコンパクト符号

〈A←0, B ←10, C ←110, D ←1110, E ←1111〉は，情報源 S のコンパクト符号

補助定理 2 の証明の骨子は次のようになる．

情報源 S_n の情報源記号発生確率を〈a_1: p_1, …, a_{n-1} : p_{n-1}, a_n : p_n 〉としよう．

T_{n-1} がコンパクトであるにも関わらず， T_n がコンパクトでないと仮定する．すると，$A=\{a_1,\cdots,a_n\}$ に対するコンパクト符号 C_n' が存在し，その平均符号長 L_n' は T_n の平均符号長 L_n より小さいことになる．補助定理 1 に関わる議論から， C_n' と平均符号長の等しいコンパクト符号 C_n'' が存在し，C_n'' において，最も深い節点には出現確率 p_n と p_{n-1} をもつ符号語が対応づけられている． C_n'' において， a_n と a_{n-1} に対する符号語割り付けの代わりに，出現確率 p_n+p_{n-1} の情報源記号 b_{n-1} に対して符号語 c を割り付ける符号 C_{n-1}' を構成する．すると， C_{n-1}' は，その作り方から C_{n-1} と同じ情報源に対する瞬時符号であり， C_{n-1}' の平均符号長 $L_{n-1}'=L_n'-p_n-p_{n-1}$ が T_{n-1} の平均符号長 $L_{n-1}=L_n-p_n-p_{n-1}$ より短いことになり， T_{n-1} がコンパクトであるという前提に矛盾する．■

核心部分を図示すると，次のようになる．