

情報理論レポート課題 5(情報源符号化法)解答例

次の情報源 S に関して、問いに答えよ。

$$S = \left\{ \begin{array}{cccccc} a & , & b & , & c & , & d & , & e & , & f \\ \frac{6}{24} & , & \frac{6}{24} & , & \frac{5}{24} & , & \frac{3}{24} & , & \frac{2}{24} & , & \frac{2}{24} \end{array} \right\}$$

1. シャノン・ファノ符号化 (算術符号化)

(1) 情報源 S に対して、シャノン・ファノ符号 ϕ_S を求めよ。

(ステップ 1-2) 各情報源記号 $\alpha \in S$ に対して、符号 ϕ_S による符号語の長さを $l_s(\alpha)$ と表す。

生成確率を降順に並べ、生成確率より符号語長を求める。

$$l_s(a) = \left\lceil -\log \frac{6}{24} \right\rceil = \lceil \log 4 \rceil = 2$$

$$l_s(b) = \left\lceil -\log \frac{6}{24} \right\rceil = 2$$

$$l_s(c) = \left\lceil -\log \frac{5}{24} \right\rceil = \lceil 2.26 \dots \rceil = 3$$

$$l_s(d) = \left\lceil -\log \frac{3}{24} \right\rceil = \lceil \log 8 \rceil = 3$$

$$l_s(e) = \left\lceil -\log \frac{2}{24} \right\rceil = \lceil 3.58 \dots \rceil = 4$$

$$l_s(f) = \left\lceil -\log \frac{2}{24} \right\rceil = 4$$

よって、符号 ϕ_S の符号語長の集合は $L_S = \{l_s(a), l_s(b), l_s(c), l_s(d), l_s(e), l_s(f)\} = \{2, 2, 3, 3, 4, 4\}$ と求められる。

(ステップ 3) 各情報源記号 $\alpha \in S$ に対して、累積確率を $P^+(\alpha)$ を求める。

$$P^+(a) = 0.0$$

$$P^+(b) = \frac{6}{24} = 0.25$$

$$P^+(c) = \frac{6}{24} + \frac{6}{24} = \frac{12}{24} = 0.5$$

$$P^+(d) = \frac{6}{24} + \frac{6}{24} + \frac{5}{24} = \frac{17}{24} \simeq 0.7083$$

$$P^+(e) = \frac{6}{24} + \frac{6}{24} + \frac{5}{24} + \frac{3}{24} = \frac{20}{24} \simeq 0.8333$$

$$P^+(f) = \frac{6}{24} + \frac{6}{24} + \frac{5}{24} + \frac{3}{24} + \frac{2}{24} = \frac{22}{24} \simeq 0.9167$$

(ステップ 4-5) 累積確率を基に 2 進小数化し、小数点以下符号語長 $L_S = \{2, 2, 3, 3, 4, 4\}$ 求めることで、符号 ϕ_S を得る。

$$\begin{aligned}
P^+(a) &= (0.0)_{10} = (0.00)_2 & \therefore \phi_S(a) &= 00 \\
P^+(b) &= (0.25)_{10} = (0.01)_2 & \therefore \phi_S(b) &= 01 \\
P^+(c) &= (0.5)_{10} = (0.100)_2 & \therefore \phi_S(c) &= 100 \\
P^+(d) &\simeq (0.7083)_{10} = (0.101\cdots)_2 & \therefore \phi_S(d) &= 101 \\
P^+(e) &\simeq (0.8333)_{10} = (0.1101\cdots)_2 & \therefore \phi_S(e) &= 1101 \\
P^+(f) &\simeq (0.9167)_{10} = (0.1110\cdots)_2 & \therefore \phi_S(e) &= 1110
\end{aligned}$$

よって、次のように符号が求められる。

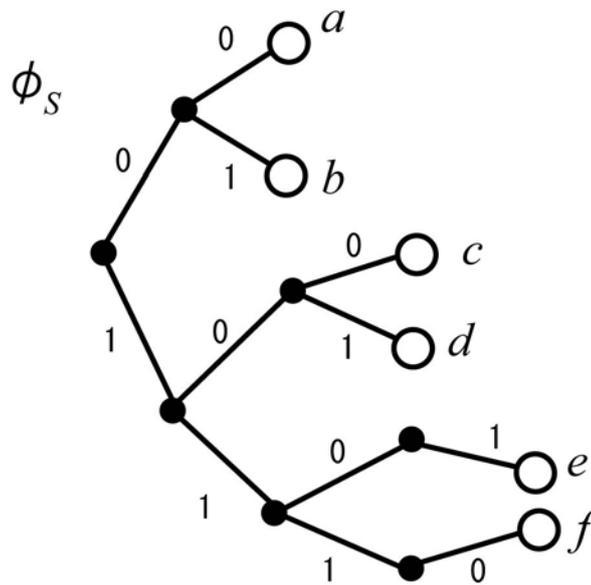
$$\phi_S = \{a \mapsto 00, b \mapsto 01, c \mapsto 100, d \mapsto 101, e \mapsto 1101, f \mapsto 1110\}$$

(2) 符号 ϕ_S に対して符号の木を示し、瞬時複号可能であるかを判別せよ。

符号の木が次のように得られる。

- 節点：符号記号の切れ目
- 符号記号の切れ目
- 符号語

辺：符号記号（上に伸びる辺が0、下が1）



全ての符号語が葉に割り当てられているので、瞬時複号可能な符号（瞬時符号）である。

(3) 符号 ϕ_S に対して、クラフトの不等式が成立するか判別せよ。

$\sum_{\alpha \in S} 2^{-l_S(\alpha)}$ を計算し、1 との大小比較を行う。

$$\begin{aligned}\sum_{\alpha \in S} 2^{-l_S(\alpha)} &= \frac{1}{2^2} + \frac{1}{2^2} + \frac{1}{2^3} + \frac{1}{2^3} + \frac{1}{2^4} + \frac{1}{2^4} \\ &= \frac{4 + 4 + 2 + 2 + 1 + 1}{16} \\ &= \frac{14}{16} \leq 1\end{aligned}$$

よって、クラフトの不等式は成立する。

(4) 符号 ϕ_S の平均符号長と効率を求めよ。

まず、エントロピーを求める。(レポート課題4も参照すること。)

$$\begin{aligned}H(S) &= -\sum_{\alpha \in S} P(\alpha) \log P(\alpha) \\ &= -\frac{6}{24} \log \frac{6}{24} - \frac{6}{24} \log \frac{6}{24} - \frac{5}{24} \log \frac{5}{24} - \frac{3}{24} \log \frac{3}{24} - \frac{2}{24} \log \frac{2}{24} - \frac{2}{24} \log \frac{2}{24} \\ &= \log 24 - \frac{1}{24} (12 \log 6 + 5 \log 5 + 3 \log 3 + 4 \log 2) \\ &= \log 24 - \frac{15}{24} \log 3 - \frac{5}{24} \log 5 - \frac{16}{24} \\ &\simeq 4.58 - 0.990 - 0.484 - 0.667 \\ &\simeq 2.44 \quad [\text{bit / 記号}]\end{aligned}$$

次に、平均符号長 \bar{L}_S を求める。

$$\begin{aligned}\bar{L}_S &= \sum_{\alpha \in S} P(\alpha) l_S(\alpha) \\ &= \frac{6}{24} \times 2 + \frac{6}{24} \times 2 + \frac{5}{24} \times 3 + \frac{3}{24} \times 3 + \frac{2}{24} \times 4 + \frac{2}{24} \times 4 \\ &= \frac{12 + 12 + 15 + 9 + 8 + 8}{24} \\ &= \frac{64}{24} \\ &\simeq 2.67 \quad [\text{bit / 記号}]\end{aligned}$$

よって、効率 $e(\phi_S)$ は次のように求められる。

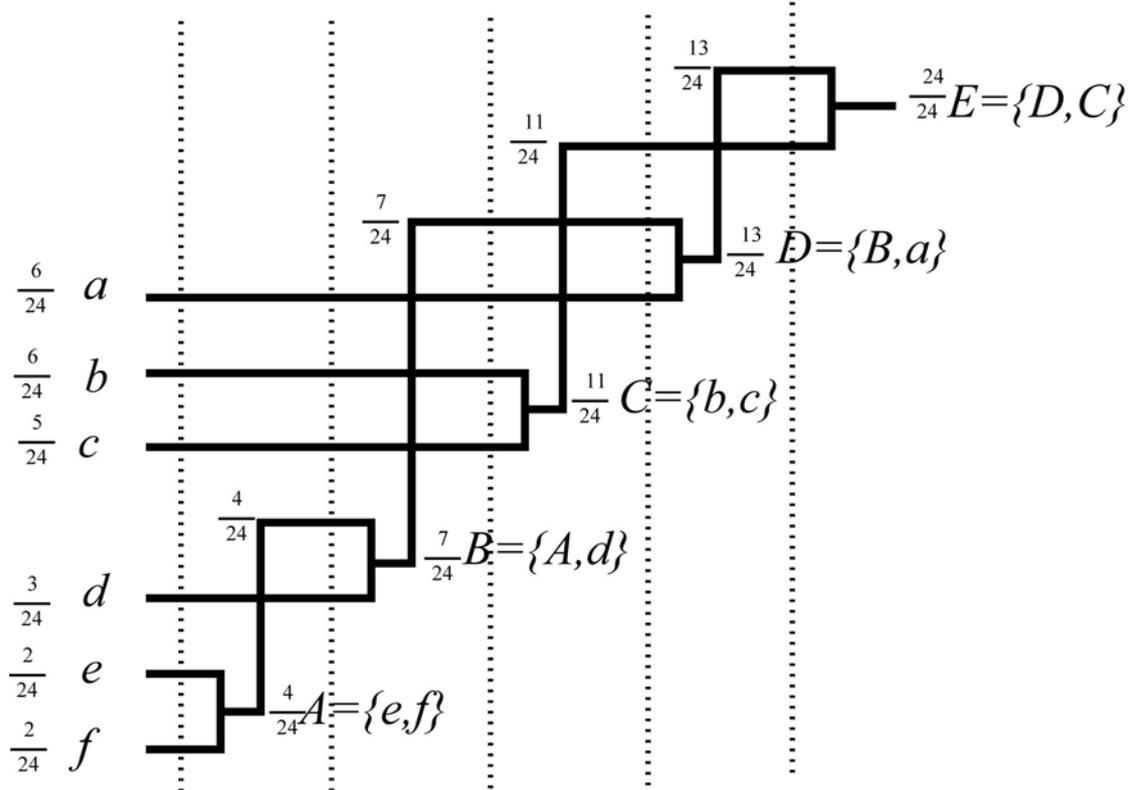
$$e(\phi_S) = \frac{H(S)}{\bar{L}_S} \simeq \frac{2.44}{2.67} \simeq 0.914$$

2.ハフマン符号化(コンパクト符号化)

(1) 情報源 S に対して、ハフマン符号 ϕ_H を求めよ。

次の図のように、確率の小さい2つの情報源記号を縮退し、縮退情報源を繰り返し求める。その際、符号の木の葉より順に符号記号を割り当てる。ただし、各垂直な点線において、縮退情報源が確率の降順に並ぶようにする。

辺：符号記号(上に伸びる辺が0、下が1)



上図により、次のように符号が求められる。

$$\phi_H = \{a \mapsto 01, b \mapsto 10, c \mapsto 11, d \mapsto 001, e \mapsto 0000, f \mapsto 0001\}$$

(2) 符号 ϕ_H に対して符号の木を示し、瞬時符号可能であるかを判別せよ。

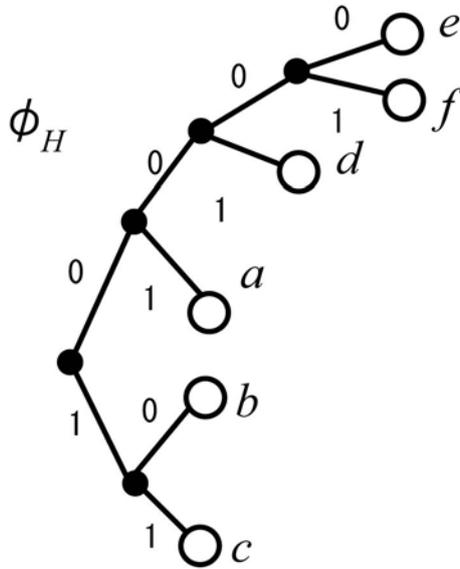
符号の木が次のように得られる。

節点：符号記号の切れ目

●：符号記号の切れ目

○：符号語

辺：符号記号（上に伸びる辺が0、下が1）



全ての符号語が葉に割り当てられているので、瞬時複号可能な符号（瞬時符号）である。

(3) 符号 ϕ_H に対して、クラフトの不等式が成立するか判別せよ。

符号 ϕ_H の符号語長の集合は $L_H = \{l_H(a), l_H(b), l_H(c), l_H(d), l_H(e), l_H(f)\} = \{2, 2, 2, 3, 4, 4\}$

$\sum_{\alpha \in S} 2^{-l_H(\alpha)}$ を計算し、1 との大小比較を行う。

$$\begin{aligned} \sum_{\alpha \in S} 2^{-l_H(\alpha)} &= \frac{1}{2^2} + \frac{1}{2^2} + \frac{1}{2^2} + \frac{1}{2^3} + \frac{1}{2^4} + \frac{1}{2^4} \\ &= \frac{4 + 4 + 4 + 2 + 1 + 1}{16} \\ &= \frac{16}{16} \leq 1 \end{aligned}$$

よって、クラフトの不等式は成立する。

(4) 符号 ϕ_H の平均符号長と効率を求めよ。

平均符号長 \overline{L}_H を求める。

$$\begin{aligned}
\overline{L}_H &= \sum_{\alpha \in S} P(\alpha) l_H(\alpha) \\
&= \frac{6}{24} \times 2 + \frac{6}{24} \times 2 + \frac{5}{24} \times 2 + \frac{3}{24} \times 3 + \frac{2}{24} \times 4 + \frac{2}{24} \times 4 \\
&= \frac{12 + 12 + 10 + 9 + 8 + 8}{24} \\
&= \frac{59}{24} \\
&\simeq 2.46 \quad [\text{bit} / \text{記号}]
\end{aligned}$$

よって、効率 $e(\phi_H)$ は次のように求められる。

$$e(\phi_H) = \frac{H(S)}{\overline{L}_H} \simeq \frac{2.44}{2.46} \simeq 0.992$$