

2016年度 通信理論

第3回 データ圧縮

(情報源符号化, 語頭符号, ハフマン符号)

2016. 4.14

植之原 裕行 uenohara.h.aa@m.titech.ac.jp

講義スケジュール(前半)

	日付	教科書	内容
第1回	4/7	第1章・第2章	通信システム概論: 通信伝送のモデルと具体例
第2回	4/11	第3章	情報源符号化 (情報量, エントロピー, 情報源符号化定理)
第3回	4/14	第4章2節	データ圧縮 (情報源符号化, 語頭符号, ハフマン符号)
第4回	4/18	第4章3節～4節	データ圧縮(イライアス符号)
第5回	4/21	第4章5節～6節	データ圧縮(ZL符号, ランレンジス符号)
第6回	4/25	第5章	通信路容量 (通信路の確率モデル, 相互情報量, 通信路容量)
第7回	4/28	第6章1節～2節	通信路符号化 (通信路符号化, 通信路符号化定理, 繰返し符号)
第8回	5/2	第1章～ 第6章2節	通信理論に関する演習

講義スケジュール(後半)

	日付	教科書	内容
第9回	5/ 9	第6章3、4節	誤り訂正符号(線形ブロック符号)
第10回	5/12	第6章5節	誤り訂正符号(巡回符号, 畳み込み符号)
第11回	5/16	第6章6節	誤り訂正符号(LDPC符号)
第12回	5/19	第7章	連続情報と連続通信路
第13回	5/23	第8章1節～2節	暗号の基礎理論
第14回	5/26	第8章3節～7節	暗号の例(共通鍵暗号、ハッシュ関数)
第15回	5/30	第8章3節～7節	暗号の例(公開鍵暗号)
	6/ 2		期末試験(予定)

符号化

- K 個のシンボルからなる情報源(アルファベット)を考える。
無記憶情報源(Memoryless Source)

$$\zeta = \{s_0, s_1, \dots, s_{K-1}\}$$

- 発生確率を以下に定める。

$$P(S = s_k) = p_k \quad \text{ただし} \quad \sum_{k=0}^{K-1} p_k = 1$$
$$k = 0, 1, 2, \dots, K-1$$

- 符号語

$$s_k \rightarrow (m_{k_1}, m_{k_2}, \dots, m_{k_n}) \quad \text{符号長: } l_k = n$$

符号の例

	C_1	C_2	C_3	C_4	C_5
s_0	00	0	0	0	0
s_1	01	10	01	00	00
s_2	10	110	011	11	1
s_3	11	1110	0111	00	11

一意 判定
でない できない

一意に復号可能な符号

C_1 : 等長符号 C_2, C_3 : 不等長符号

C_1, C_2 : 瞬時符号 C_3 : 非瞬時符号 符号と呼ばない

語頭符号 (Prefix Code)

- どの符号語も他の符号の語頭とならない符号

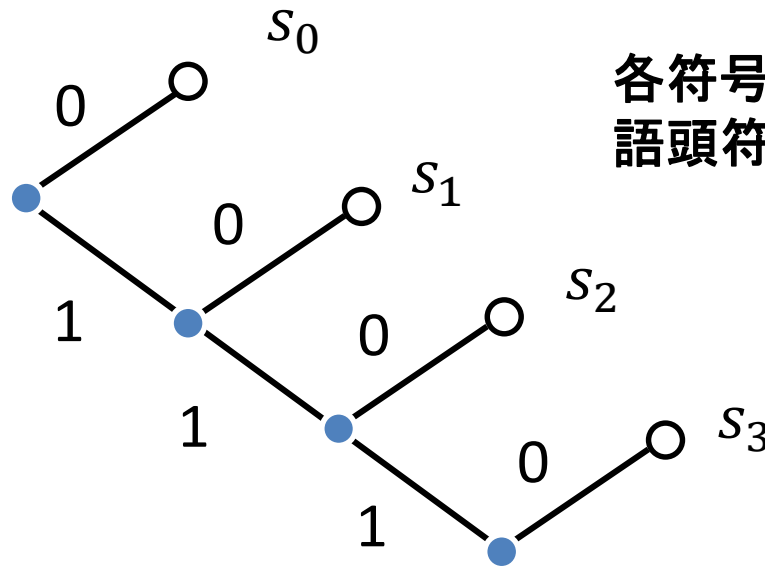
	C_1	C_2	C_3	発生確率
s_0	00	0	0	$\frac{1}{2}$
s_1	01	10	01	$\frac{1}{4}$
s_2	10	110	011	$\frac{1}{8}$
s_3	11	1110	0111	$\frac{1}{8}$

C_1, C_2 : 語頭符号 = 瞬時符号

C_3 : 語頭符号でない

語頭符号の復号(符号の木)

C_2



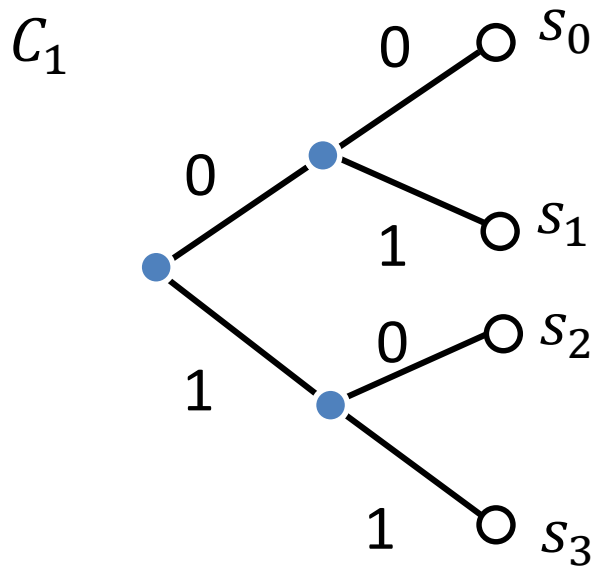
各符号が端点(葉)に位置する場合、
語頭符号(= 瞬時に復号可能)

s_3 を111としても語頭符号である

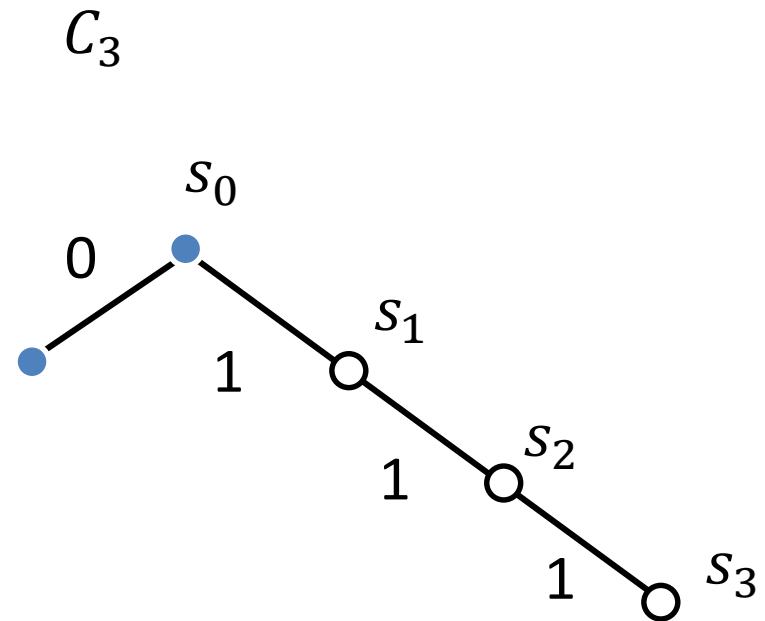
例: 1 0 1 1 1 1 1 0 0 0

s_2 s_4 s_3 s_1

語頭符号の復号(2)



語頭符号



各符号が端点(葉)に位置せず、
語頭符号でない

クラフト・マクミランの不等式

- 符号が一意的に復号可能であるための必要十分条件

$$\sum_{k=0}^{K-1} 2^{-l_k} \leq 1$$

l_k : 符号 s_k の符号長

最適符号長(1)

- クラフト・マクミランの不等式を満たす最小の符号長

$$\sum_{k=0}^{K-1} p_k = 1$$

ここで符号 s_k の発生確率として

$$p_k = 2^{-l_k}$$

が成り立つとすると、

$$l_k = -\log_2 p_k$$

なので、

$$\bar{L} = \sum_{k=0}^{K-1} p_k l_k = - \sum_{k=0}^{K-1} p_k \log_2 p_k = H(\zeta)$$

最適符号長(2)

前ページの符号において、 $l_k = -\log_2 p_k$ より大きい l_k はクラフト・マクミランの不等式を満たす。

また $-\log_2 p_k + 1$ までに整数が存在するので、以下の符号長を選ぶことができる。

$$l_k < -\log_2 p_k + 1$$

最適平均符号長：

$$\bar{L} = \sum_{k=0}^{K-1} p_k l_k < \sum_{k=0}^{K-1} (-p_k \log_2 p_k + p_k) = H(\zeta) + 1$$

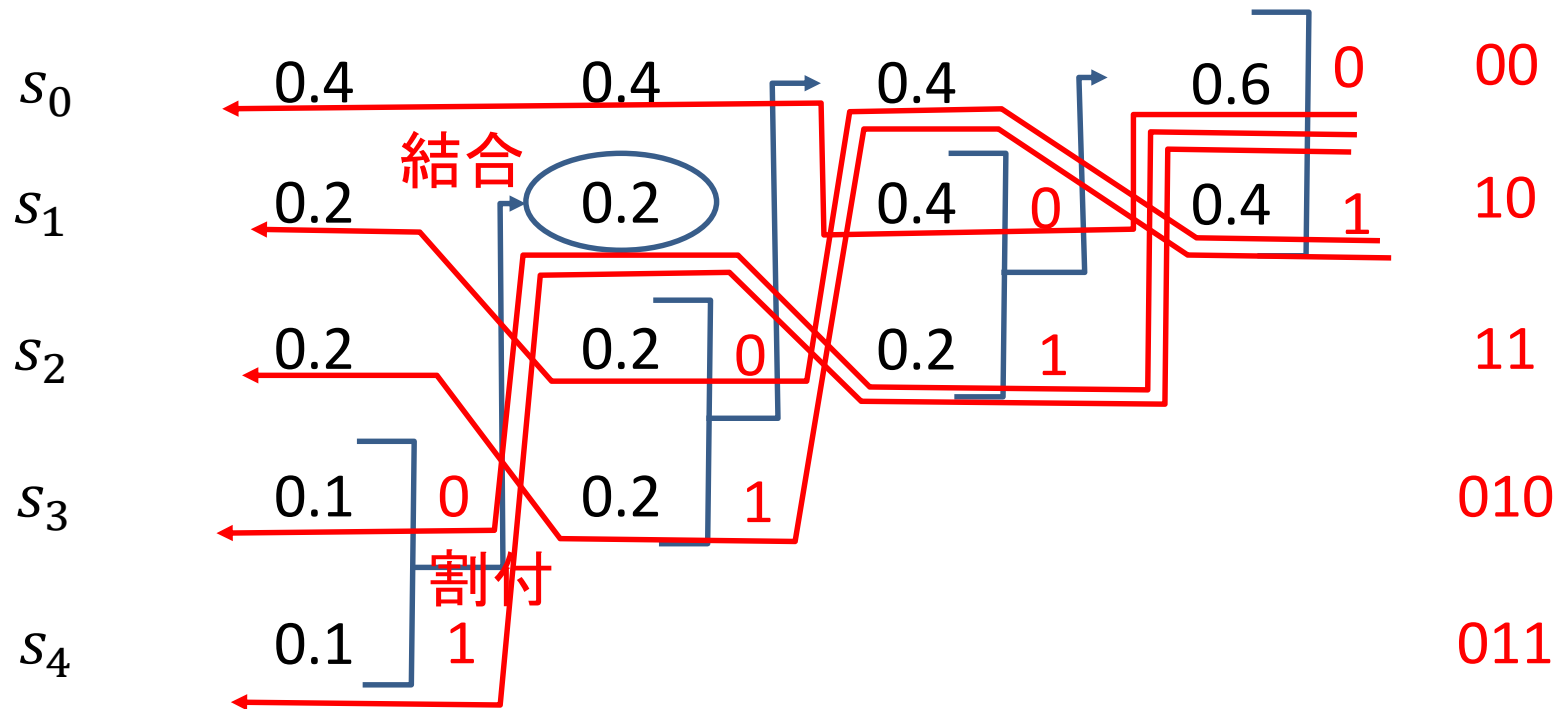
$$\therefore H(\zeta) \leq \bar{L} < H(\zeta) + 1$$

ハフマン符号

- 平均符号長が最小の瞬時符号が構成できる
- 符号化アルゴリズム
 - ① 情報源シンボルを、確率が下がる順に並べる。
-> 最も低い確率の2つのシンボルに0, 1を割り当てる。
 - ② ①の2つの確率の合計を新しいシンボルに結合されたものとみなす。
 - ③ 最後の2つになるまで繰り返す。
-> 元に戻りながら0,1の系列をたどって符号を形成

ハフマン・アルゴリズム

シンボル	確率	ステージI	ステージII	ステージIII	符号
------	----	-------	--------	---------	----



順序付

ハフマン符号の特性

	符号	発生確率
s_0	00	0.4
s_1	01	0.2
s_2	11	0.2
s_3	010	0.1
s_4	011	0.1

● エントロピー

$$\begin{aligned} H(\zeta) &= 0.4 \times \log_2 \left(\frac{1}{0.4} \right) + 0.2 \times \log_2 \left(\frac{1}{0.2} \right) \\ &\quad + 0.2 \times \log_2 \left(\frac{1}{0.2} \right) + 0.1 \times \log_2 \left(\frac{1}{0.1} \right) \\ &\quad + 0.1 \times \log_2 \left(\frac{1}{0.1} \right) = 2.12 \text{ [bit]} \end{aligned}$$

● 平均符号長

$$\begin{aligned} \bar{L} &= 0.4 \times 2 + 0.2 \times 2 + 0.2 \times 2 \\ &\quad + 0.1 \times 3 + 0.1 \times 3 = 2.2 \end{aligned}$$



① $\bar{L}$ は $H(\zeta)$ より3.67%だけ大きい。

② $\bar{L}$ は $H(\zeta) \leq \bar{L} < H(\zeta) + 1$ を満たす。

ハフマン符号生成の任意性

● 符号生成の任意性の要因

- 0, 1の割り当ての任意性
 - 同確率の符号の順位付け
 - 結合した符号の順位付け
- } 最適化への影響小
- ↓

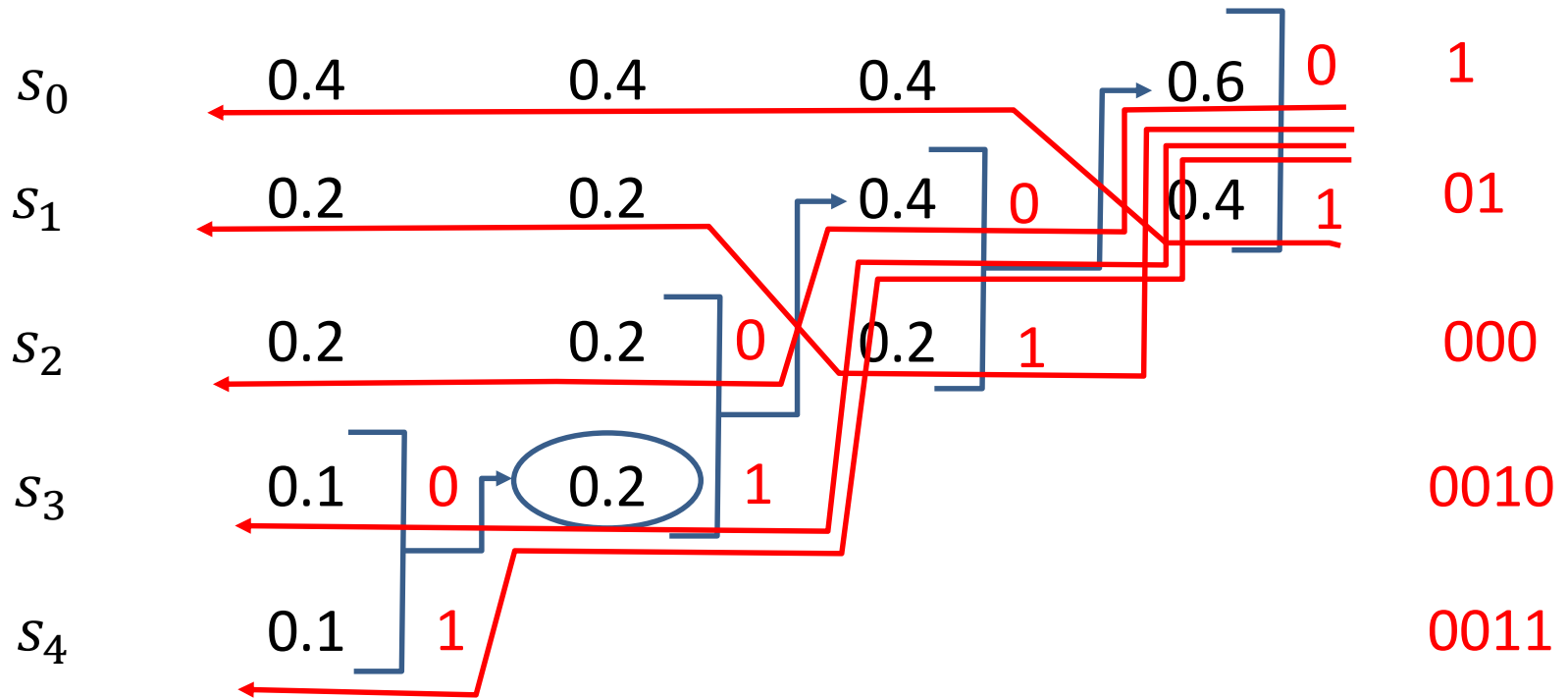
平均符号長の分散への影響

$$\sigma^2 = \sum_{k=0}^{K-1} p_k (l_k - \bar{L})^2$$

結合シンボルを高い位置に持っていくほど
分散は小さくなる(適す)

ハフマン符号の結合符号順位

シンボル	確率	ステージI	ステージII	ステージIII	符号
------	----	-------	--------	---------	----



結合符号順位による平均符号長・分散

- 平均符号長

$$\bar{L} = 0.4 \times 1 + 0.2 \times 2 + 0.2 \times 3 + 0.1 \times 4 + 0.1 \times 4 = 2.2$$

 13ページの平均符号長に等しい

- 符号長の分散

12ページの符号:

$$\begin{aligned}\sigma^2 &= 0.4 \times (2 - 2.2)^2 + 0.2 \times (2 - 2.2)^2 + 0.2 \times (2 - 2.2)^2 \\ &\quad + 0.1 \times (3 - 2.2)^2 + 0.1 \times (3 - 2.2)^2 = 0.16\end{aligned}$$

15ページの符号:

$$\begin{aligned}\sigma^2 &= 0.4 \times (1 - 2.2)^2 + 0.2 \times (2 - 2.2)^2 + 0.2 \times (3 - 2.2)^2 \\ &\quad + 0.1 \times (4 - 2.2)^2 + 0.1 \times (4 - 2.2)^2 = 1.36\end{aligned}$$

まとめ

- 語頭符号

どの符号語も他の符号の語頭とならない符号

- クラフト・マクミランの不等式

符号が一意に復号可能であるための必要十分条件

$$\sum_{k=0}^{K-1} 2^{-l_k} \leq 1$$

- ハフマン符号

平均符号長が最小の瞬時符号が構成できる