



Title	2005年度 情報理論講義ノート
Author(s)	井上, 純一
Issue Date	2005-11-18T09:19:52Z
Doc URL	http://hdl.handle.net/2115/772
Rights(URL)	http://creativecommons.org/licenses/by-nc-sa/2.1/jp/
Type	learningobject
Note(URL)	http://www005.upp.so-net.ne.jp/j_inoue/index.html ; http://chaosweb.complex.eng.hokudai.ac.jp/~j_inoue/
Additional Information	There are other files related to this item in HUSCAP. Check the above URL.
File Information	InfoTheory05_slide6.pdf (第6回講義スライド)



[Instructions for use](#)



情報理論 #6

第6回講義 5月30日

情報科学研究科 井上純一

http://chaosweb.complex.eng.hokudai.ac.jp/~j_inoue/

情報源符号の平均符号長

(先週の復習)

$$A = \{a_1, a_2, \dots, a_K\}, f: A \mapsto B^+ = \{b_1, b_2, \dots, b_M\}$$

記号系列

符号語

l_i : 符号語 $f(a_i)$ の長さ

p_i : 記号 a_i の出現確率 $p(a_i)$

$$L = \sum_{i=1}^K p_i l_i$$

平均符号長

(例)

aa	00	1/4
ab	10	1/4
ba	01	1/4
bb	110	1/4

出現確率

$$L = 2 \times (1/4) + 2 \times (1/4) + 2 \times (1/4) + 3 \times (1/4) = 9/4$$

平均符号長

平均符号長の下限

(先週の復習)

$$\begin{aligned}
 L - H_M(X) &= -\sum_{i=1}^K p_i \log_M M^{-l_i} + \sum_{i=1}^K p_i \log_M p_i \\
 &= -\sum_{i=1}^K p_i \log_M cr_i + \sum_{i=1}^K p_i \log_M p_i \\
 &= \sum_{i=1}^K p_i \log \left\{ \frac{p_i}{r_i} \right\} - \log_M c \geq 0
 \end{aligned}$$

平均符号長

情報源のエントロピー

確率の規格化条件

$$\sum_{i=1}^K r_i = c^{-1} \sum_{i=1}^K M^{-l_i} = 1 \quad \text{より}$$

KL情報量で正の値を持つ

$$c = \sum_{i=1}^K M^{-l_i} \leq 1$$

クラフト不等式より

$$L \geq H_M(X)$$

一意復号可能な場合には、平均符号長をエントロピーよりも小さくすることができない

ハフマン符号

ハフマン符号：平均符号長が最短 (最適) な符号

$$L \geq H_M(X)$$

エントロピーの下限 (復習)

情報源アルファベット $A = \{A, B, C, D, E, F\}$, $B = \{0, 1\}$ 符号化に用いる記号

$$\{P_A, P_B, P_C, P_D, P_E, P_F\} = \{0.4, 0.3, 0.11, 0.09, 0.08, 0.02\}$$

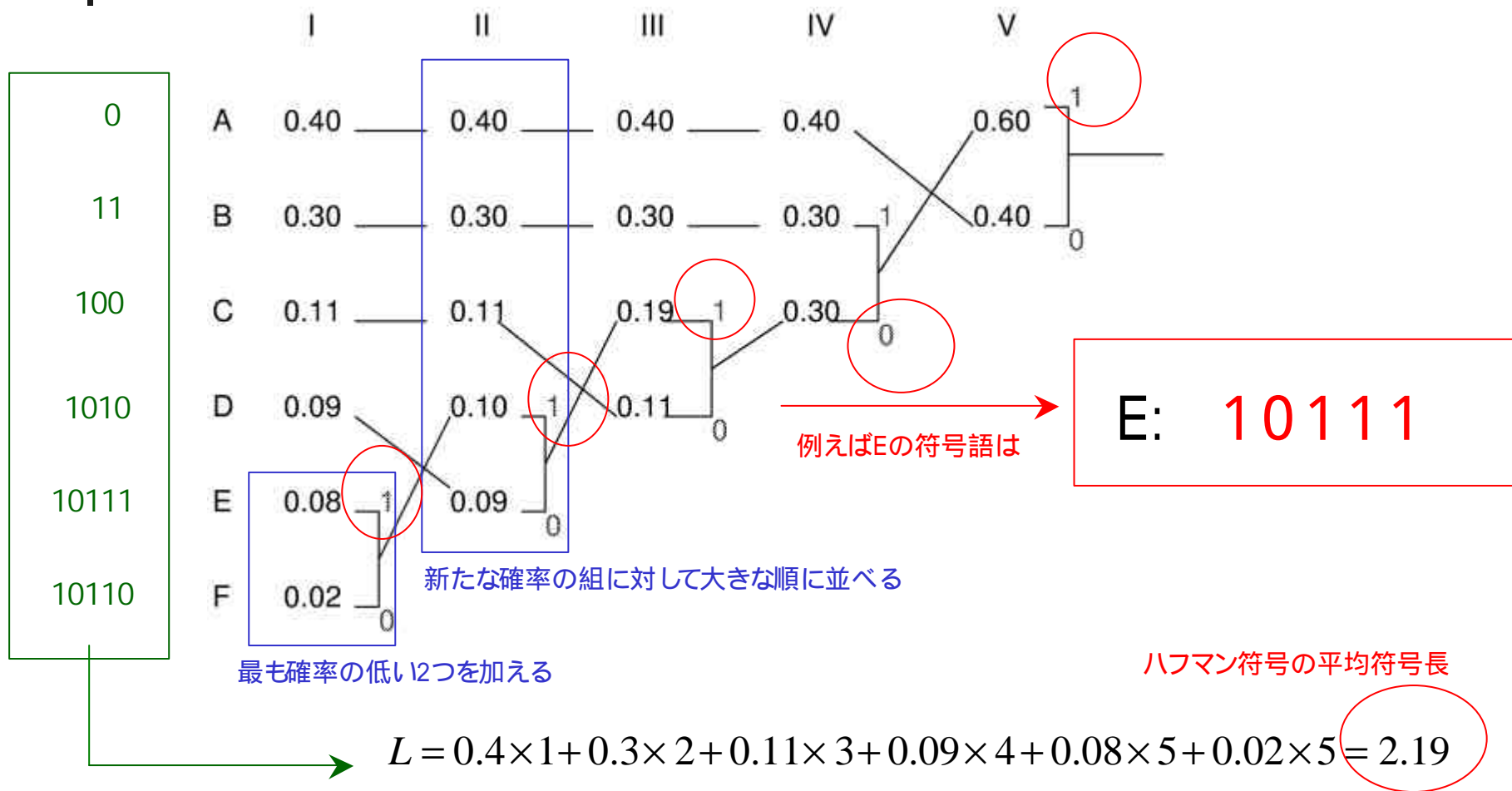
各アルファベットの出現確率が与えられているものとする

	l
A	0.40
B	0.30
C	0.11
D	0.09
E	0.08
F	0.02

直観的には出現確率が高い記号には長さが短い符号を割り当て、逆に出現確率が低い記号には長い符号を割り当てればよい

符号構成法の第1段階ではアルファベットを出現確率の大きな順に並べる

ハフマン符号の構成法



ハフマン符号の最適性

(準備)

補題

与えられた情報源に対し、以下の2つの条件を満たし、かつ平均符号長が最短の符号が存在する

- (1) 確率が最も小さな2つの符号は同じ節点から出ている2つの葉に割り当てられる
- (2) その節点レベルは木の中で最高である

(証明)

(2) a_M に対する符号が最高レベルの節ではないとする

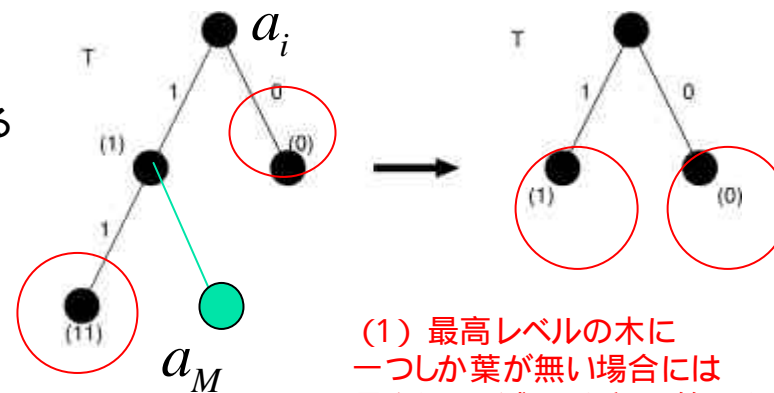
a_i と a_M を交換した新しい符号の木 T'

$$\begin{aligned} L(T) - L(T') &= p_i l_i + p_M l_M - p_i l_M - p_M l_i \\ &= (p_i - p_M)(l_i - l_M) \geq 0 \end{aligned}$$

T の最適性より $L(T) \leq L(T')$ と合わせると

$$L(T) = L(T')$$

上記の交換により、平均符号長は変化しない



(1) 最高レベルの木に一つしか葉が無い場合には長さを一つ減らした語頭符号を作れる
 T の平均符号長最短に反する

ハフマン符号の最適性

定理 ハフマン符号は平均符号長が最も短い符号である

(証明)

T : 平均符号長最短な木

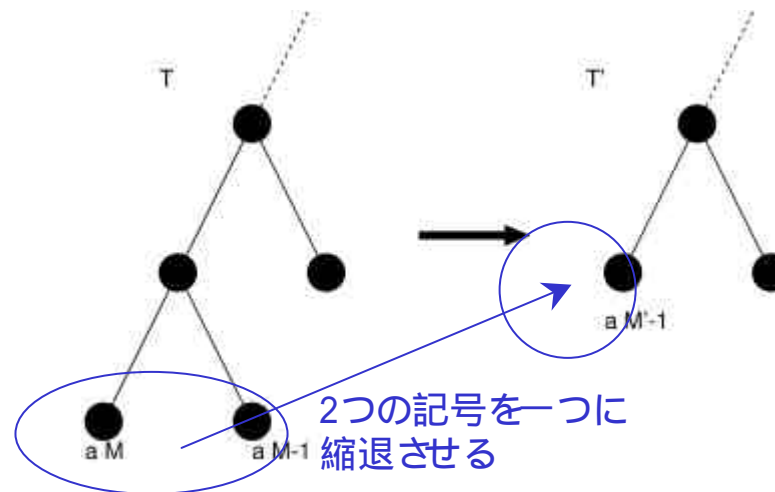
最高レベルには2つの葉

$$L(T) = p_1 l_1 + p_2 l_2 + \dots + p_{M-1} l_{M-1} + p_M l_{M-1}$$

T' : T を縮退させた木

2つの記号を縮退させた

$$L(T') = p_1 l_1 + p_2 l_2 + \dots + p_{M-2} l_{M-2} + (p_M + p_{M-1}) l_{M-2}$$



$$L(T) - L(T') = p_M \underbrace{(l_{M-1} - l_{M-2})}_1 + p_{M-1} \underbrace{(l_{M-1} - l_{M-2})}_1 = p_M + p_{M-1}$$

ハフマン符号の最適性

(続き)

C' に関する最適な符号の木を S' とする

$$L(S') = p_1 l_1 + p_2 l_2 + \cdots + p_{M-2} l_{M-2} + p_{M'-1} l_{M-2}$$

S' を展開して木 S を作る C c の記号数は c' と比べて一つだけ多い

$$L(S) = p_1 l_1 + p_2 l_2 + \cdots + p_{M-2} l_{M-2} + p_{M-1} l_{M-1} + p_M l_{M-1}$$

差をとると $L(S) - L(S') = p_M + p_{M-1}$

従って $0 \leq L(T') - L(S') = L(T) - L(S) \leq 0$

$$L(T) = L(S)$$

この事実から、記号数 n に関する帰納法より、
任意の記号数に対してハフマン符号の平均符号長は最短であることが示せる