# 情報理論

第8回 講義 情報源の符号化(ハフマン符号)

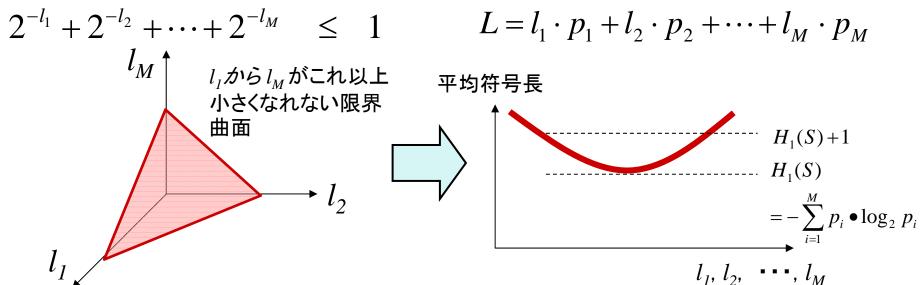
> 2015. 6. 10 植松 芳彦

## 前回分かったこと(1)

クラフトの不等式を制約条件として与えた時、平均符号長の下限は一次エントロピーで与えられる

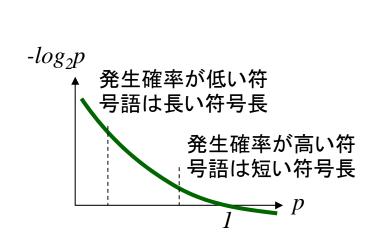
求めたいもの

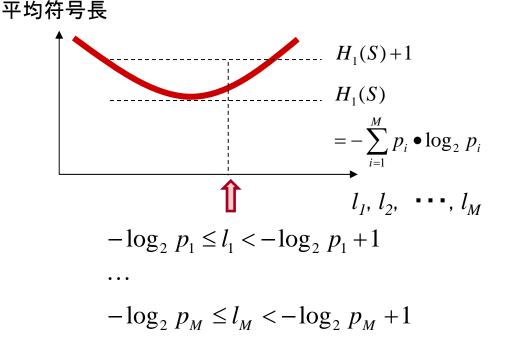
#### クラフトの不等式(制約条件)



## 前回分かったこと(2)

- 発生確率  $p_i$  の符号語には符号長として  $-\log_2 p_i$  を超える最小整数を割り当てることで、下限に近い平均符号長を達成
  - 発生確率が高い符号語には短い符号長
  - 発生確率が低い符号語には長い符号長



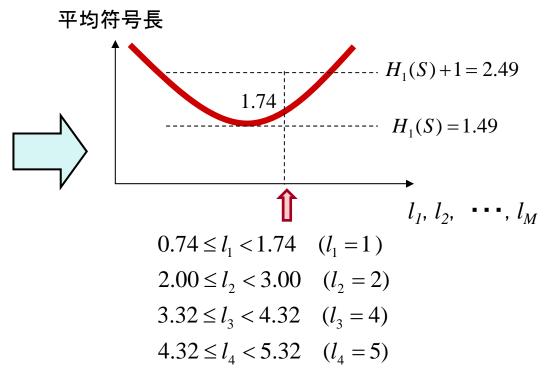


## さらなる平均符号長の短縮にむけて

- それなりに下限値に近い平均符号長は得られるが、必ずし も最小値になるとは限らない。
- より効率のよい符号の構成法を探索する.

#### 教科書表4.1の情報源記号列

情報源記号	発生確率		
Α	0.6		
В	0.25		
С	0.1		
D	0.05		



#### コンパクト符号

- コンパクト符号:情報源記号をひとつづつ一意復号可能な符号に符号化するとき,平均符号長を最小とする符号
- ハフマン符号: Huffmanが考案した構成法によるコンパクト 符号

表 4.1 情報原符号化の例 / 直接求めたい

情記	報	游号		碓	率	<i>C</i> ,	$c_{i_1}$	<i>c</i> ,	C,	C <sub>v</sub>	Cv
	A		T	0.	. 6	00	0	0	0	0	0
	$\boldsymbol{B}$			0.	. 25	01	10	10	01	10	10
	C			0.	.1	10	110	110	011	11	11
	D			0.	05	11	1110	111	111	01	0
平	均	符	号	長	£	2. 00	1. 60	1. 55	1. 55	1.40	1. 35
										4	

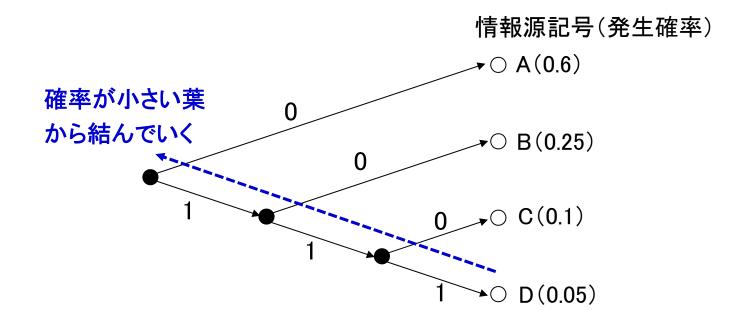
一意復号可能(瞬時)

一意復号可能

一意復号不可能

(非瞬時)

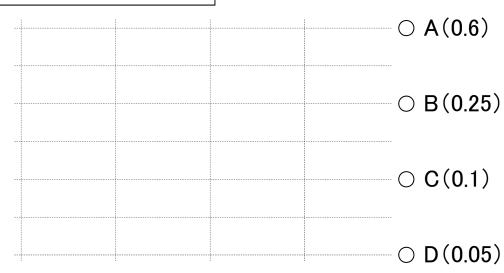
- 符号の木を情報源記号に対応した「葉」から作る.
- 発生確率が高い符号には短い符号語を、低い符号には 長い符号語を割り付けることを考慮し、確率が小さい葉 から結んでいく。



1. 各情報源記号に対応する葉を作成発生確率を付記

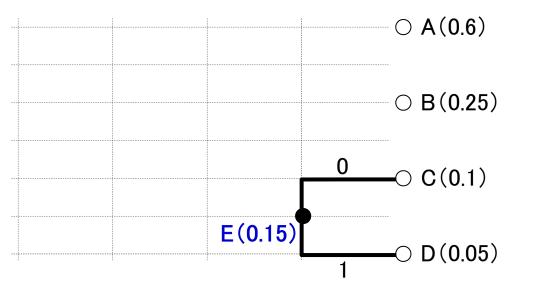
枝の書き方も変える 斜め線 ⇒ 水平+垂直線

情報源記号(発生確率)

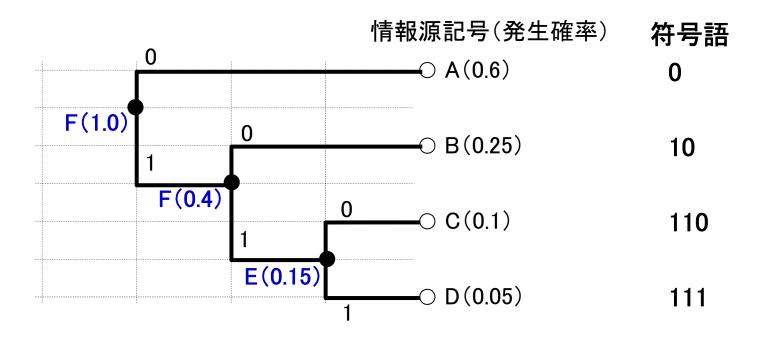


2. 確率最小の2枚の葉に対して接点を作成, 枝で連結 枝の一方に0, 他方に1を割り付け 接点に2枚の葉の確率の和を記入 接点を新たな葉と見做す





- 3. 葉が1枚になったら終了, そうでなければ2. の繰り返し
- 4. 符号語の確定

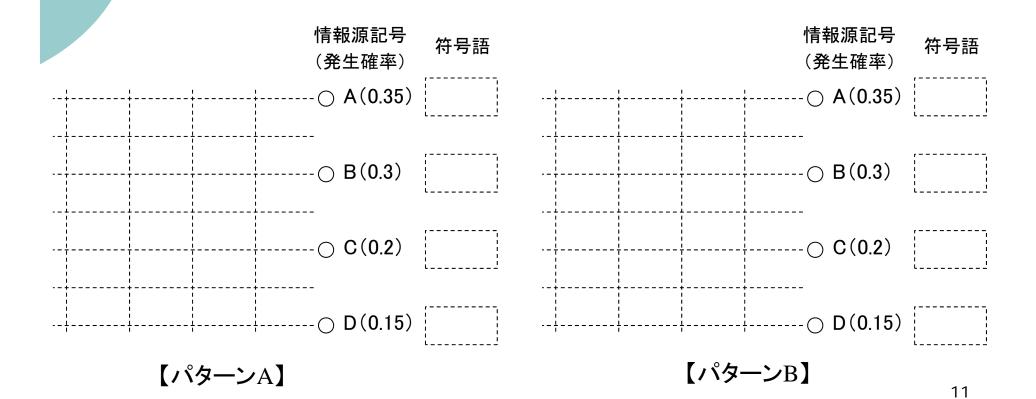


#### ハフマン符号の特徴

- 瞬時符号
  - 符号語が全て符号の木の葉にだけ割り当てられる
- 一意に決まらない
  - 各節点から出る枝のどちらにO, 1を割り当ててもよい
  - 確率が等しい葉(節)がある場合組み合わせ方は任意

## 【演習1】ハフマン符号の作り方

情報源記号A, B, C, Dを0.35, 0.3, 0.2, 0.15の確率で発生する情報源に対するハフマン符号の構成法は2通り存在する. 2通りの符号の木, 符号語を書いてみよう.



## 【演習1】ハフマン符号の作り方

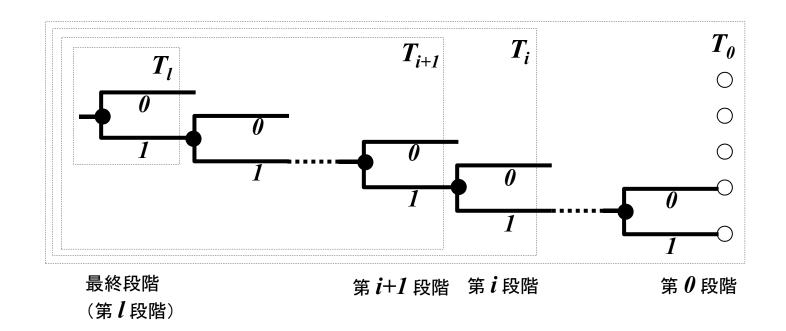
前ページの2通りの符号の木,符号語について,平均符号長を求めてみよう。

平均符号長 
$$L = l_1 \cdot p_1 + l_2 \cdot p_2 + \dots + l_M \cdot p_M$$
 (今の場合 $M = 4$ )

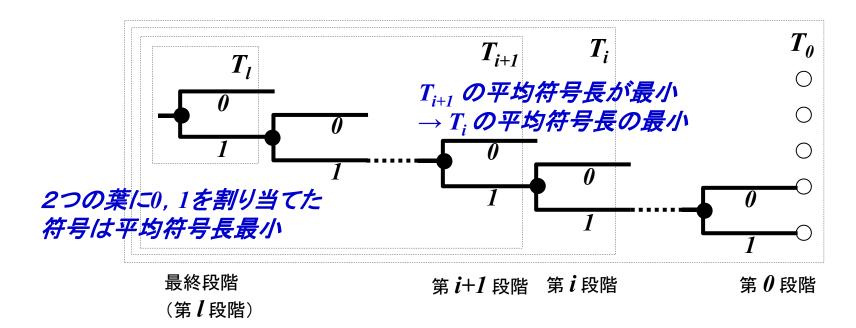
符号の木や符号語の形はちがうが、平均符号長は同じ、 かつ一次エントロピーにかなり近い値が得られている。

ー次エントロピー 
$$H_1(S) = -\sum_{i=1}^4 p_i \bullet \log_2 p_i \cong 1.93$$

- ハフマン符号の平均符号長は本当に最小値か検証
- ハフマン符号化は、下図の各段階毎に、確率が最も小さい2本の枝を節で結ぶ作業の繰り返し。
- 各 *i* 段階~根までの部分的な符号の木を *T<sub>i</sub>と*する.



- ただ2つの葉(とみなす節点)からなる最終段階の符号 の木 T,はコンパクト符号の木.
- $T_{i+1}$  がコンパクトの時  $T_i$  もコンパクトと言えれば、数学的帰納法により、 $T_{\varrho}$ がコンパクト符号の木であることを証明される.

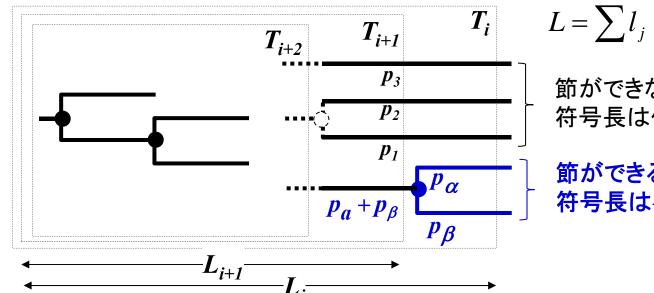


ullet  $T_{i+1}$ と $T_i$ の平均符号長の間には以下の関係がある.

$$L_i = L_{i+1} + 1 \cdot p_{\alpha} + 1 \cdot p_{\beta}$$

•  $T_i$ から  $T_{i+1}$ に移る過程で確率最小の枝を選んだはずな ので. 以下が成り立つ

$$p_{\alpha}, p_{\beta} \leq p_1, p_2, p_3$$

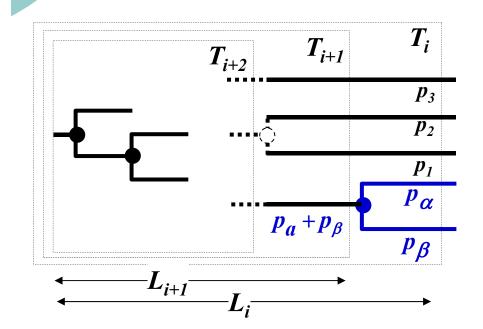


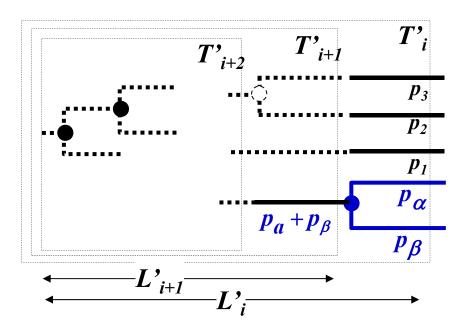
 $L = \sum l_i \cdot p_i$  の変化

節ができないので 符号長は伸びない

節ができるので 符号長は+1伸びる

- $T_{i+1}$  はコンパクトだが  $T_i$  はコンパクトでないとすると、より平均符号長が短い符号の木 $T'_i$ が存在するはず.
- また  $T'_i$ の確率最小の枝2本を接点で結ぶことで $T'_{i+1}$ ができるはず.





- $T_{i+1}$  もコンパクトでないことになり、当初の前提に矛盾.
- よって $T_{i+1}$  がコンパクトなら $T_i$ もコンパクト符号の木.

$$L_{i} = L_{i+1} + p_{\alpha} + p_{\beta}$$

$$T_{i+2} \qquad T_{i+1} \qquad T_{i}$$

$$p_{3}$$

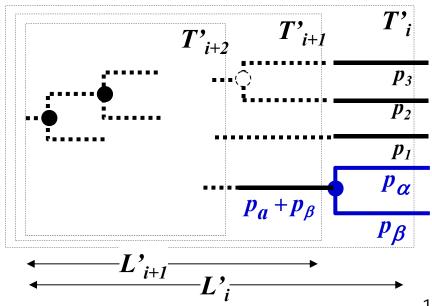
$$p_{2}$$

$$p_{1}$$

$$p_{\alpha} + p_{\beta}$$

$$p_{\beta}$$

$$L'_{i} = L'_{i+1} + p_{\alpha} + p_{\beta}$$



## 【演習2】ハフマン符号の作り方

情報源記号A, B, C, D, E, F, Gをそれぞれ0.34, 0.30, 0.12, 0.10, 0.08, 0.05, 0.01で出力する情報源をハフマン符号化しよう.

					(発	(生確率)	付亏諾
 , , , , , , ,	! ! !	1 1	L			A(0.34)	
 ;	;	; ! ! ! !	;       	;   		B(0.3)	
   	 	 				C(0.12)	
   	   	; 	 	; ;			[
 ; ! ! ! !	 	; ! ! !	 	 			
 ;	 	; ! ! !				E(0.08)	[]
   	         	 	,         	, 		F(0.05)	<u> </u>
 ! ! ! !	! ! ! <del>!</del>	! ! ! +	! ! ! <b></b>	    		G(0.01)	

## 【演習2】ハフマン符号の作り方

- 前ページのハフマン符号の平均符号長を求めよう.
- 一次エントロピーにかなり近い値が得られるはず。

一次エントロピー 
$$H_1(S) = -\sum_{i=A}^G p_i \bullet \log_2 p_i \cong 2.32$$

平均符号長 
$$L = l_A \cdot p_A + l_B \cdot p_B + \dots + l_G \cdot p_G$$

#### 本日のまとめ

- コンパクト符号(平均符号長を最小とする符号)のひとつとして、ハフマン符号の構成法を学んだ。
  - 情報源記号をひとつづつ一意復号可能な符号に符号化する前提
- ハフマン符号のコンパクト性を検証.
  - 符号化則のなかで、平均符号長が最小になるか。
  - 一次エントロピーにどこまで近づくか