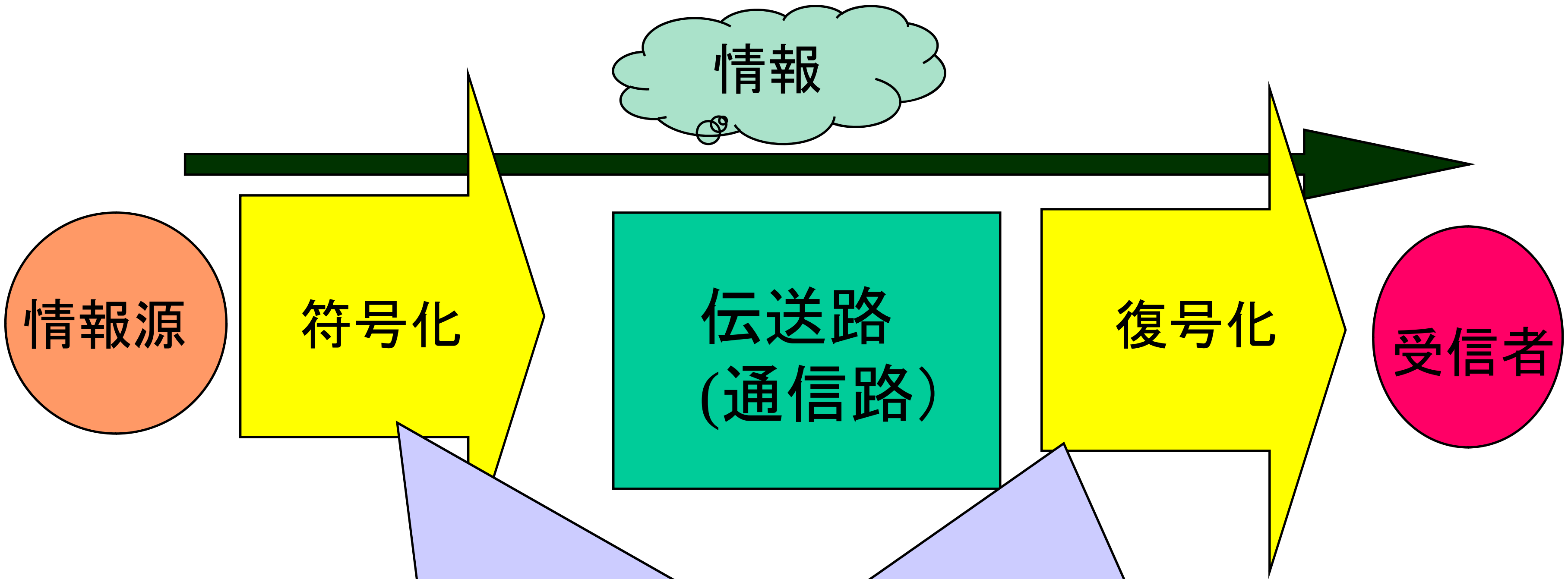
5.情報源符号化（5章）

情報源符号化の役割



今回扱う。
☆雑音のない理想的な場合に、情報源アルファベットを符号に変換する。

符号化の形式化1

符号:符号語 $\boldsymbol{c}_k$ の集合

$$S=\{s_1,s_2,\cdots,s_n\}$$

情報源アルファベット

符号化

$$\phi: S \to \boldsymbol{C}$$

$$\boldsymbol{C}=\{\boldsymbol{c}_1,\boldsymbol{c}_2,\cdots,\boldsymbol{c}_n\}$$

ここで

$$\begin{aligned}\boldsymbol{c}_1 &= c_{11}c_{12}\cdots c_{1l_1},\\ \boldsymbol{c}_2 &= c_{21}c_{22}\cdots c_{2l_2},\\ &\vdots\\ \boldsymbol{c}_n &= c_{n1}c_{n2}\cdots c_{nl_n}\end{aligned}$$

復号化

$$\phi^{-1}: \boldsymbol{C} \to S$$

$$c_{ij}\in X=\{x_1,x_2,\cdots,x_r\}$$

符号アルファベット:符号に用いられる記号 x_l の集合。
この個数が r のとき、 r 元符号という。
符号記号:符号に用いられる記号 x_l

符号化の形式化2

符号化は、一種の関数とみなせる。

符号化

$$|\boldsymbol{c}_j| = l_j$$

$$\phi : s_j \mapsto \boldsymbol{c}_j = c_{j1}c_{j2}\cdots c_{jl_j}$$

$$\phi(s_j) = \boldsymbol{c}_j = c_{j1}c_{j2}\cdots c_{jl_j}$$

符号語長

全単射が多い

復号化

$$\phi^{-1} : c_{j1}c_{j2}\cdots c_{jl_j} \mapsto s_j$$

復号化は、符号化の逆関数とみなせる。

$$\phi^{-1}(\boldsymbol{c}_j) = s_j$$

符号化の例

情報源

$$S = \begin{Bmatrix} a & , & b & , & c & , & d \\ 0.4 & , & 0.3 & , & 0.2 & , & 0.1 \end{Bmatrix}$$

を、符号アルファベットを $X = \{0, 1\}$ とする2元符号化する。

$$\phi_1 = \{a \mapsto 0, b \mapsto 10, c \mapsto 110, d \mapsto 1110\}$$

符号化の関数を求めることを符号化ということもある。

このとき、符号長の集合 L は以下で与えられる。

$$L = \{l_a, l_b, l_c, l_d\}$$

$$= \{1, 2, 3, 4\}$$

練習

$$\phi_1 = \{a \mapsto 0, b \mapsto 10, c \mapsto 110, d \mapsto 1110\}$$

次の文字列を符号 ϕ_1 に従って符号化せよ。

（1）

$$cab$$

（2）

$$abaadccbba$$

練習2

$$\phi_1 = \{a \mapsto 0, b \mapsto 10, c \mapsto 110, d \mapsto 1110\}$$

次の文字列を符号${\phi_1}^{-1}$に従って復号化せよ。

（1）

$$01011001000010 1110$$

（2）

$$11001001011 00110$$

平均符号長

定義：平均符号長

$$S = \begin{Bmatrix} s_1 & , & \cdots & , & s_n \\ P(s_1) & , & \cdots & , & P(s_n) \end{Bmatrix}$$

の符号長の集合を $L = \{l_1, \cdots, l_n\}$ とする。

このとき、平均符号長 $\overline{L}$ は次式で定義される。

$$\overline{L} = \sum_{s_k \in S} P(s_k) l_k$$

平均：確率を乗じて総和を取る。

平均符号長例1

情報源 $S=\begin{Bmatrix} a & , & b & , & c & , & d \\ 0.4 & , & 0.3 & , & 0.2 & , & 0.1 \end{Bmatrix}$ の符号を

$$\phi_1 = \{a \mapsto 0, b \mapsto 10, c \mapsto 110, d \mapsto 1110\}$$

とする。

このとき、符号長の集合は、

$$L_1 = \{l_1(a), l_1(b), l_1(c), l_1(d)\} = \{1, 2, 3, 4\}$$

であるので、平均符号長 $\overline{L_1}$ は、次式で求められる。

$$\overline{L_1} = P(a)l_1(a) + P(b)l_1(b) + P(c)l_1(c) + P(d)l_1(d)$$

$$= 0.4 \times 1 + 0.3 \times 2 + 0.2 \times 3 + 0.1 \times 4$$

$$= 2$$

確率の大きい記号には短い符号を、
確率の小さい記号には長い符号を用
いた方が効率が良い。

平均符号長例2

情報源 $S=\begin{Bmatrix} a & , & b & , & c & , & d \\ 0.4 & , & 0.3 & , & 0.2 & , & 0.1 \end{Bmatrix}$ の符号を

$$\phi_2=\{a\mapsto 1110, b\mapsto 110, c\mapsto 10, d\mapsto 0\}$$ とする。

このとき、符号長の集合は、

$$L_2=\{l_2(a),l_2(b),l_2(c),l_2(d)\}=\{4,3,2,1\}$$

であるので、平均符号長 $\overline{L}_2$ は、次式で求められる。

$$\overline{L}_2=P(a)l_2(a)+P(b)l_2(b)+P(c)l_2(c)+P(d)l_2(d)$$

$$=0.4\times 4+0.3\times 3+0.2\times 2+0.1\times 1$$

$$=3$$

確率の大きい記号に長く、確率の小さい記号に短ると平均符号長が長くなる。

平均符号長の効果

$$S = \begin{Bmatrix} a & , & b & , & c & , & d \\ 0.4 & , & 0.3 & , & 0.2 & , & 0.1 \end{Bmatrix}$$

約40文字
（$\overline{L_1} \times 20$ ）

$$00101101010011100000101110010110110101100$$

$$\phi_1 = \{a \mapsto 0, b \mapsto 10, c \mapsto 110, d \mapsto 1110\}$$

$$aabcbbadaaabdabccbca$$

情報源からの20文字

$$\phi_2 = \{a \mapsto 1110, b \mapsto 110, c \mapsto 10, d \mapsto 0\}$$

$$111011101101010110110111001110111011101100$$

$$11101101010110101110$$

約60文字
（$\overline{L_2} \times 20$ ）

練習

S : スペード(Spade)
H:ハート(Heart)
D:ダイヤ(Diamond)
C : クラブ(Club)

（1）情報源

$$\text{トランプ} = \begin{Bmatrix} S & , & H & , & D & , & C \\ \frac{1}{2} & , & \frac{1}{4} & , & \frac{1}{8} & , & \frac{1}{8} \end{Bmatrix}$$

に対する符号を

$$\phi = \{S \mapsto 110, H \mapsto 111, D \mapsto 10, C \mapsto 0\}$$

とする。このとき、平均符号長を求めよ。
（2）同じ情報源に対して、別の符号を割り当て、（1）より平均符号長を短くせよ。また、そのときの平均符号長を求めよ。

等長符号と可変長符号

定義：等長符号と可変長符号

ある符号を$C=\{\boldsymbol{c}_1,\boldsymbol{c}_2,\cdots,\boldsymbol{c}_n\}$とし、その符号長集合を
$L=\{l_1,l_2,\cdots,l_n\}$ とする。

符号長 $l_i=|c_i|,1\leq i\leq n$ が全て等しいとき、
$\boldsymbol{C}$ を等長符号をいう。
長さの異なる符号 $l_i\neq l_j,0\leq i,j\leq n,$ が存在するとき、
$\boldsymbol{C}$ を可変長符号という。

$$\begin{aligned}
\boldsymbol{c}_1 &= c_{11}c_{12}\cdots c_{1l},\\
\boldsymbol{c}_2 &= c_{21}c_{22}\cdots c_{2l},\\
&\vdots\\
\boldsymbol{c}_n &= c_{n1}c_{n2}\cdots c_{nl}
\end{aligned}$$

等長符号

可変長符号

$$\begin{aligned}
\boldsymbol{c}_1 &= c_{11}\cdots c_{1l_1},\\
\boldsymbol{c}_2 &= c_{21}\cdots\cdots\cdots c_{2l_2},\\
&\vdots\\
\boldsymbol{c}_n &= c_{n1}\cdots\cdots c_{nl_n}
\end{aligned}$$

等長符号の平均符号長

性質:等長符号の平均符号長

等長符号の平均符号長は、ある一つの記号の符号長と等しい。

$$
\begin{aligned}
\boldsymbol{c}_1 &= c_{11}c_{12}\cdots c_{1l}, \\
\boldsymbol{c}_2 &= c_{21}c_{22}\cdots c_{2l}, \\
&\vdots \\
\boldsymbol{c}_n &= c_{n1}c_{n2}\cdots c_{nl}
\end{aligned}
$$

$$
\begin{aligned}
\overline{L} &= \sum_{s_j \in S} P(s_j) l_j \\
&= l \underbrace{\sum_{s_j \in S} P(s_j)}_{1} \\
&= l
\end{aligned}
$$

等長符号例

$$S = \begin{Bmatrix} a & , & b & , & c & , & d \\ 0.4 & , & 0.3 & , & 0.2 & , & 0.1 \end{Bmatrix}$$

$$\phi_3 = \{a \mapsto 00, b \mapsto 01, c \mapsto 10, d \mapsto 11\}$$

平均符号長　$\overline{L_3}$　は、次式で求められる。

$$\overline{L_3} = 0.4 \times 2 + 0.3 \times 2 + 0.2 \times 2 + 0.1 \times 2$$

$$= 1 \times 2$$

$$= 2$$

符号例一覧

$$S = \begin{Bmatrix} a & , & b & , & c & , & d \\ 0.4 & , & 0.3 & , & 0.2 & , & 0.1 \end{Bmatrix}$$

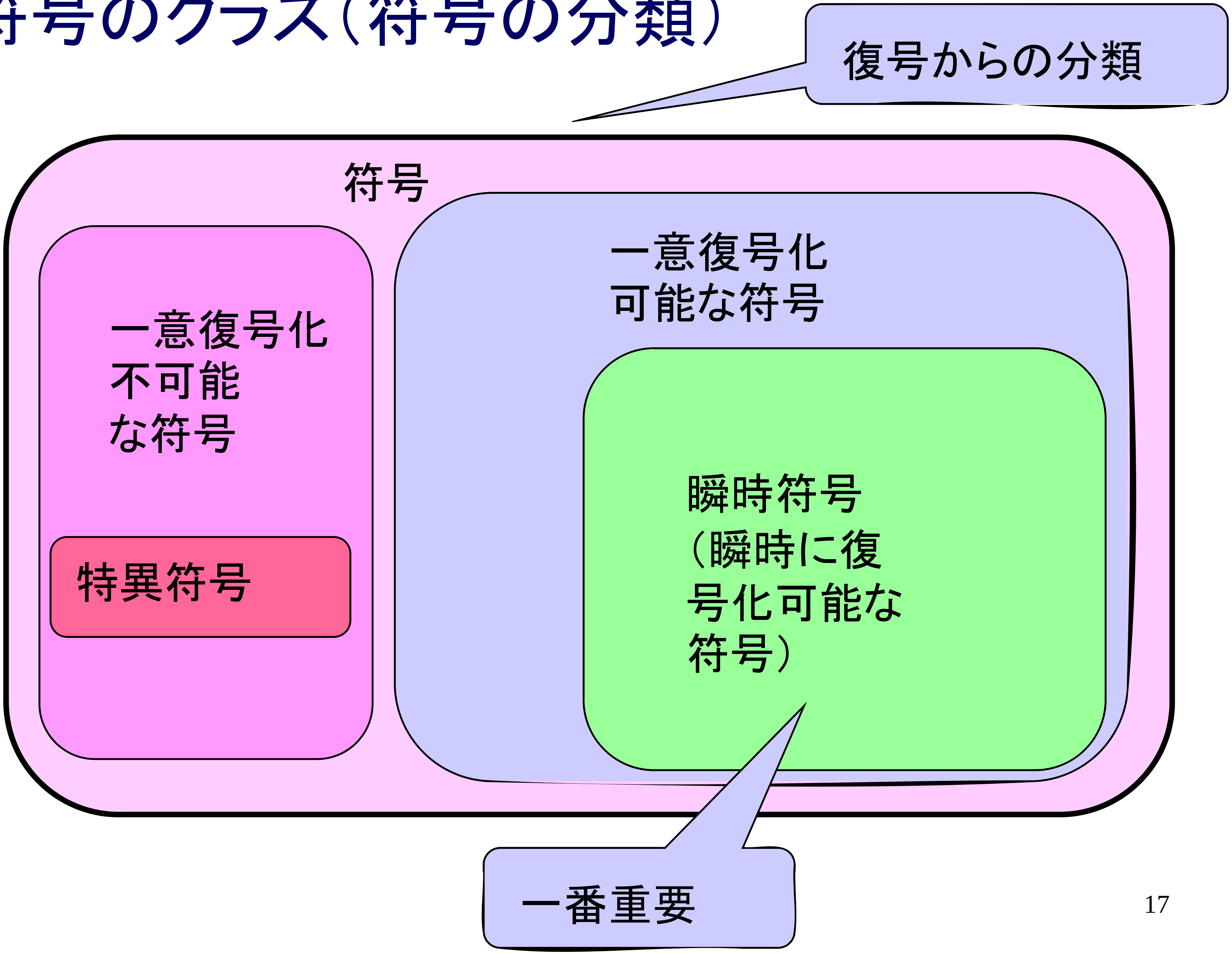
平均符号長

可変長符号

	a	b	c	d	$\overline{L}$
ϕ_1	0	10	110	1110	2
ϕ_2	1110	110	10	0	3
ϕ_3	00	01	10	11	2

等長符号

符号のクラス（符号の分類）



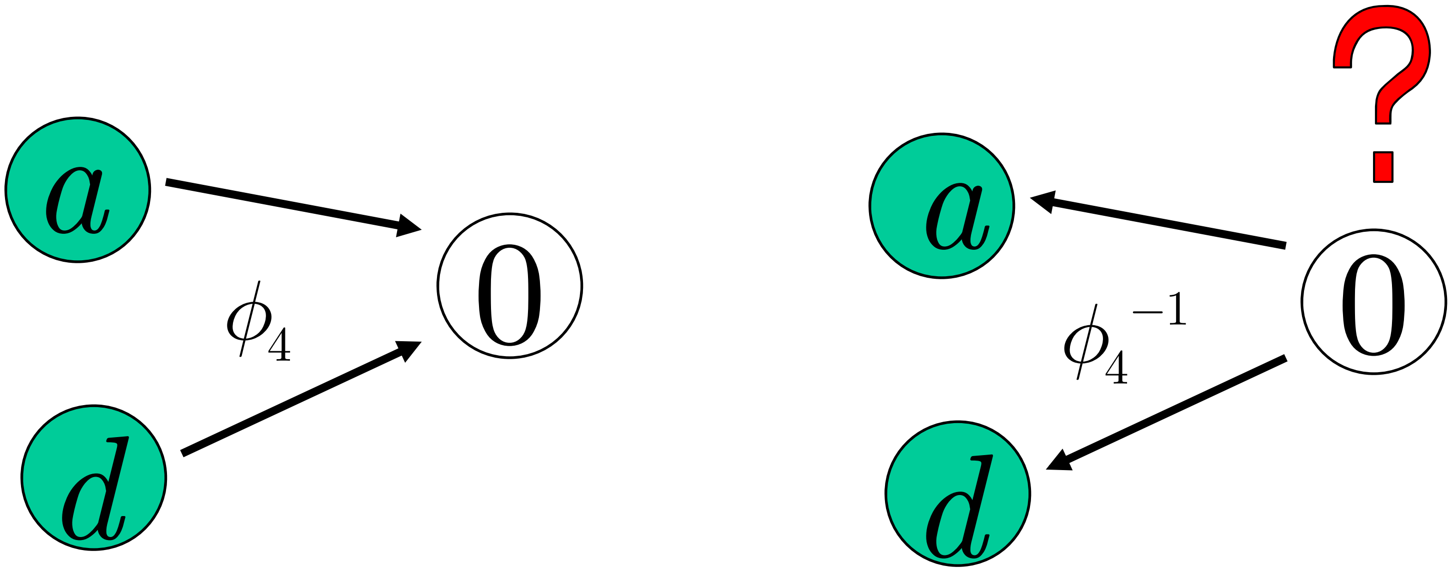
特異符号

定義:特異符号

2つ以上の情報源記号に、一つの符号語を対応させた符号を**特異符号**という。

特異符号例

$$\phi_4 = \{a \mapsto 0, b \mapsto 1, c \mapsto 10, d \mapsto 0\}$$



復号化の一意性

定義：一意複合可能

符号記号の系列から、対応する情報源の系列を一意に求められ符号を一意に復号可能な符号といい、一意に求められない符号を一意に復号不可能な符号という。

特異符号のように符号記号毎に一意に復号不可能な場合だけでなく、系列として一意に復号不可能場合も含む。もちろん次の命題は成り立つ。

性質：特異符号の一意複合不可能性

特異符号は一意に復号不可能な符号である。

一意に復号不可能な符号例

$$\phi_5 = \{a \mapsto 0, b \mapsto 1, c \mapsto 01, d \mapsto 10\}$$

特異符号
ではない

$$abbacd$$

$$\phi_5$$

$$01100110$$

$$\phi_5^{-1}$$

$$abbaabba$$

$$abbaabd$$

$$abbacd$$

$$cdcd$$

?

瞬時符号

定義:瞬時符号

情報源記号の系列を符号化したものが、時系列で送られるとする。このとき、符号記号の系列から情報源記号の区切りが瞬時に判断できる符号を**瞬時符号**という。
(ここで、瞬時とは、次の情報源記号の符号語が送られてこなくても、符号語の終わりが判別できることを指す。)

$$\boldsymbol{s}_{t_1}\,\boldsymbol{s}_{t_2}\,\cdots\,\boldsymbol{s}_{t_n}$$

$$\downarrow$$

$$\boldsymbol{c}_{t_1}\,\boldsymbol{c}_{t_2}\,\cdots\,\boldsymbol{c}_{t_n}$$

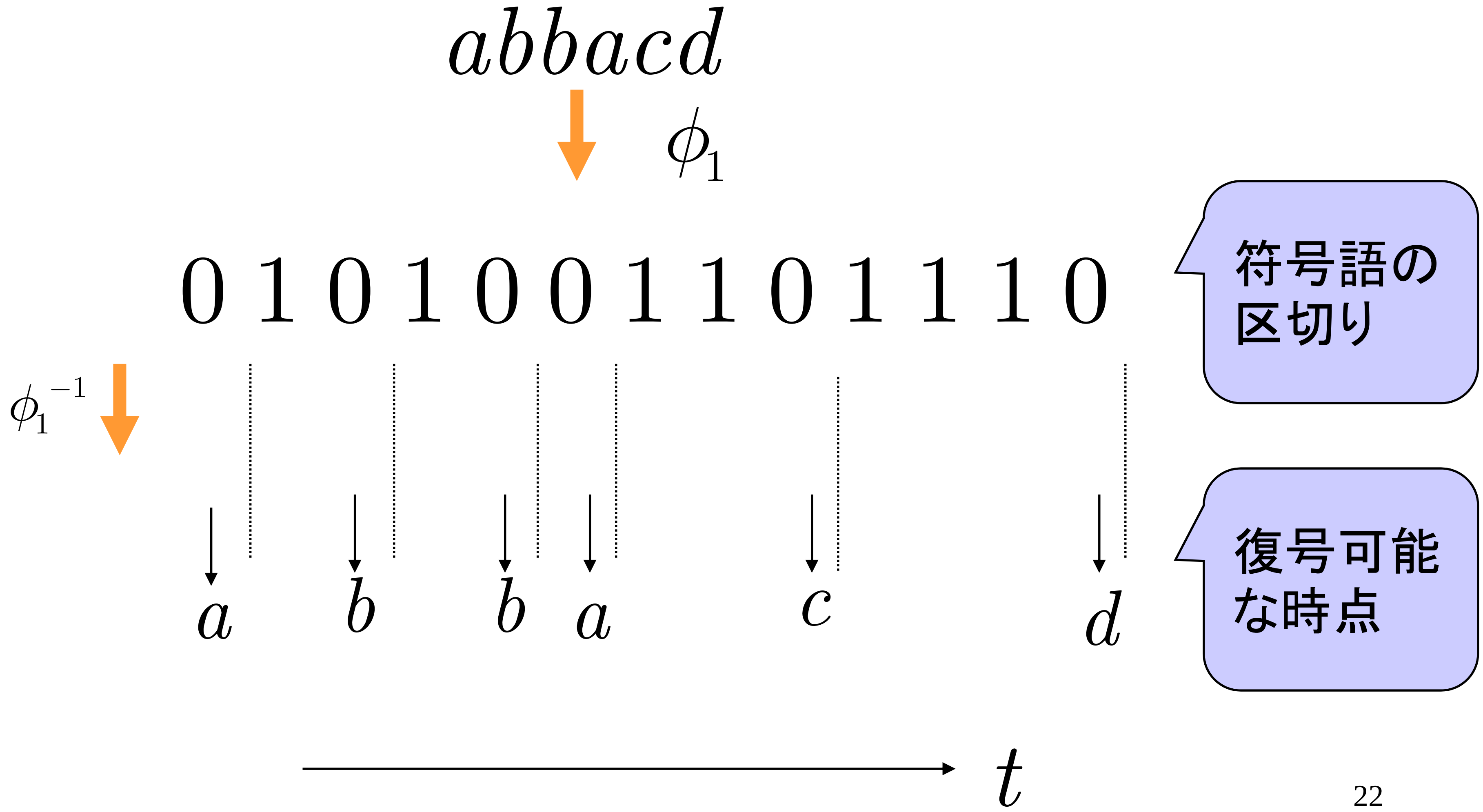
$$x_1^{t_1}\cdots x_{l_1}^{t_1}\,\big|\,x_1^{t_2}\,x_2^{t_2}\cdots x_{l_2}^{t_2}\,\big|\,x_1^{t_3}\,x_2^{t_3}\cdots x_{l_3}^{t_3}\cdots x_1^{t_n}\cdots x_{l_n}^{t_n}\quad t$$

この時点で
1記号復号できる。

瞬時符号例

瞬時符号

$$\phi_1 = \{a \mapsto 0, b \mapsto 10, c \mapsto 110, d \mapsto 1110\}$$



非瞬時符号例

非瞬時符号

$$\phi_6 = \{a \mapsto 0, b \mapsto 01, c \mapsto 011, d \mapsto 0111\}$$

$$abbacd$$

$$\phi_6$$

0 0 1 0 1 0 0 1 1 0 1 1 1

ϕ_6^{-1}

a b b a c d ?

t

符号語の区切り

復号可能な時点

ϕ_6 は一意復号可能な符号であるが、
瞬時復号可能な符号ではない。

練習

次の２つの符号に対して、与えられた通報を符号化し、さらに復号化せよ。

$$\phi_1 = \{a \mapsto 0, b \mapsto 10, c \mapsto 110, d \mapsto 1110\}$$

$$\phi_6 = \{a \mapsto 0, b \mapsto 01, c \mapsto 011, d \mapsto 0111\}$$

（１）

$$caabacdbba$$

（２）

$$accabbbadaab$$

符号の木

符号を一つの木として捉えると、符号の性質を理解しやすい。
ここでは、2元符号についての符号の木を示す。

定義：符号の木

接点：符号記号の区切り

枝：符号記号

として、各接点から2分岐（r元の場合はr分岐）させた木を符号の木という。各符号語は根から対応する接点までの経路上の符号記号の系列として求められる。

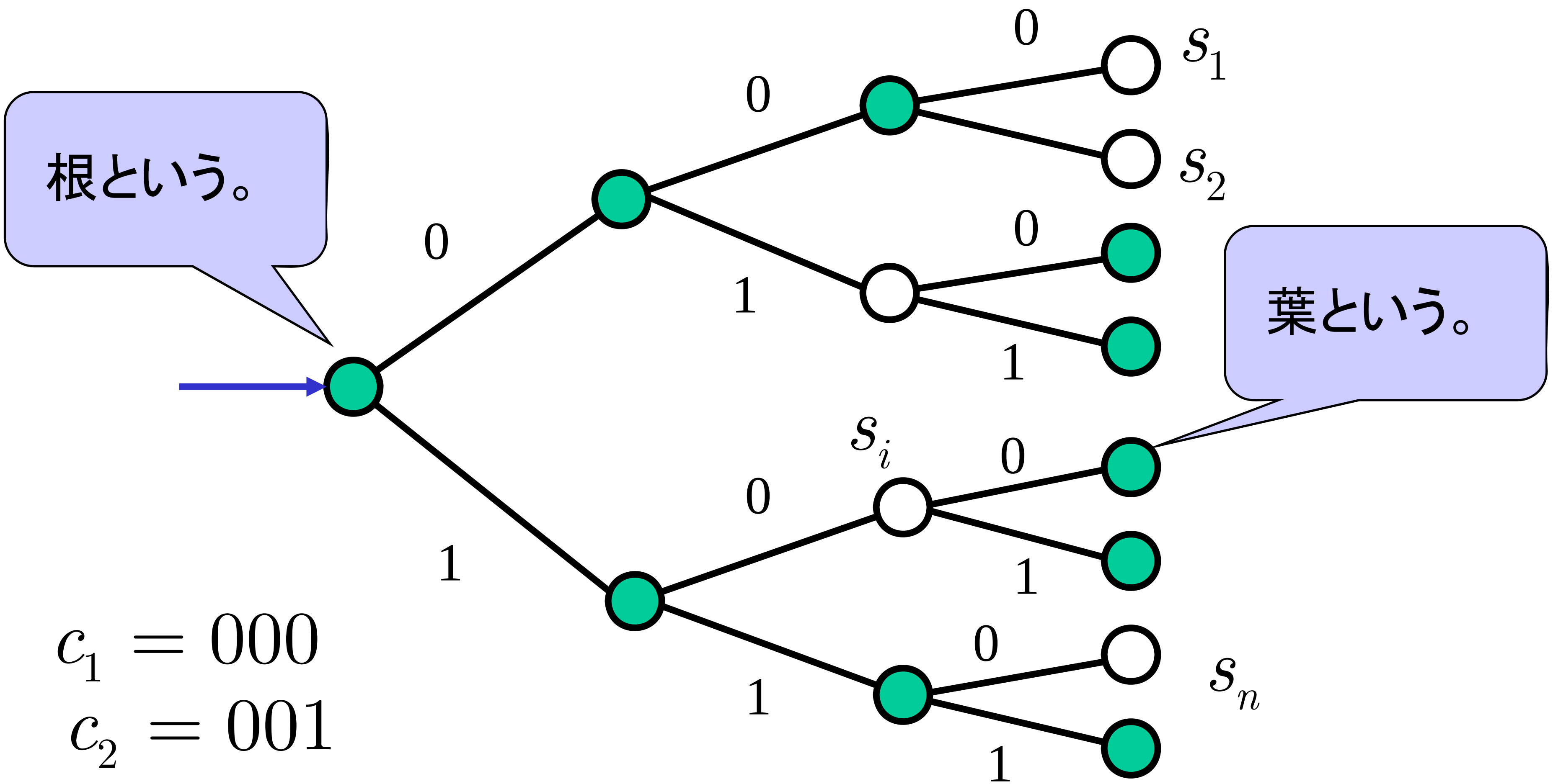
●:区切り

○ :符号語に対応

0:上の枝
1:下の枝

$$S = \{s_1, s_2, \cdots, s_n\}$$

$$\boldsymbol{C} = \{\boldsymbol{c}_1, \boldsymbol{c}_2, \cdots, \boldsymbol{c}_n\}$$



$$c_1 = 000$$

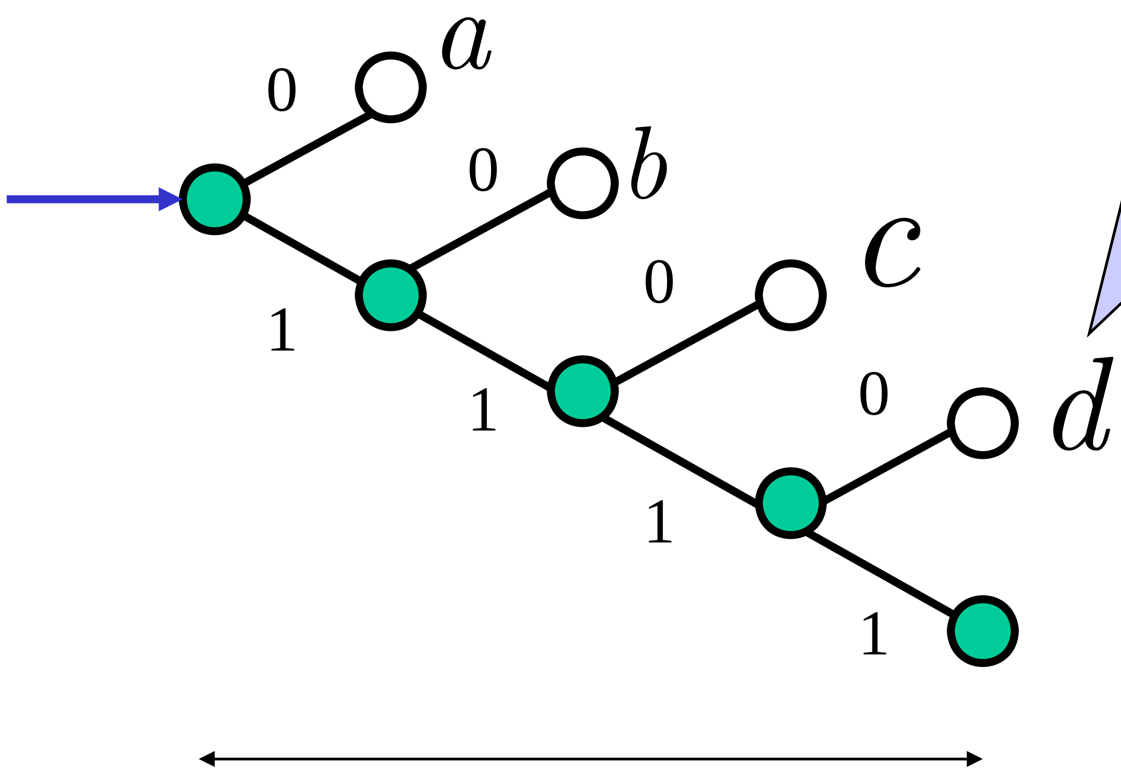
$$c_2 = 001$$

$$c_i = 10$$

符号の木の例

$$\phi_1 = \{a \mapsto 0, b \mapsto 10, c \mapsto 110, d \mapsto 1110\}$$

$$\boldsymbol{C}_1 = \{0, 10, 110, 1110\}$$



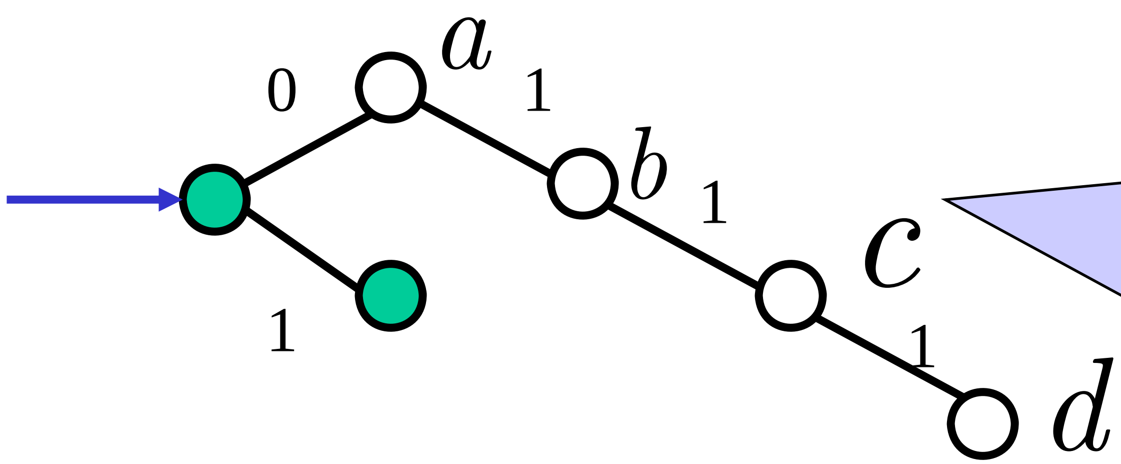
瞬時符号は、
符号語が葉に
しか割り当てら
れない。

符号長に対応（深さという）

符号の木の例2

$$\phi_6 = \{a \mapsto 0, b \mapsto 01, c \mapsto 011, d \mapsto 0111\}$$

$$\boldsymbol{C}_6 = \{0, 01, 011, 0111\}$$



非瞬時符号は、符号語が葉以外にも割り当てられる。

練習

以下の符号に対して、符号の木を作成せよ。

（1）

$$\phi_2 = \{a \mapsto 1110, b \mapsto 110, c \mapsto 10, d \mapsto 0\}$$

（2）

$$\phi_3 = \{a \mapsto 00, b \mapsto 01, c \mapsto 10, d \mapsto 11\}$$

（3）

$$\phi_5 = \{a \mapsto 0, b \mapsto 1, c \mapsto 01, d \mapsto 10\}$$

瞬時符号の性質1

性質:符号の木を用いた瞬時符号の判別

瞬時符号であるための必要十分条件は、
符号の木として表現したとき全ての符号語が葉に割り当てられていることである。

語頭

定義：語頭

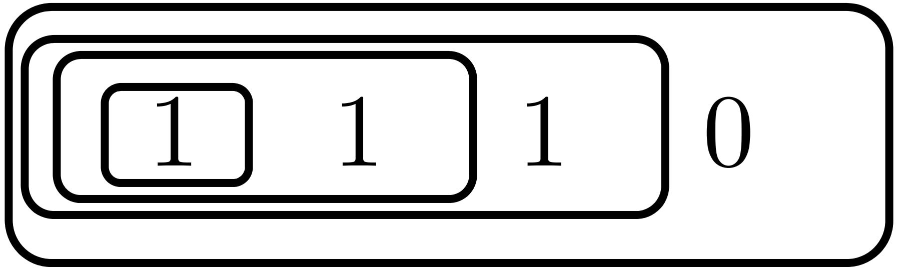
符号語$\boldsymbol{c}_i = c_{i1}c_{i2}\cdots c_{il_i} \in \boldsymbol{C}$　に対して、

$c_{i1}c_{i2}\cdots c_{ij}$　　$1 \le j < l_i$

を（符号語 $\boldsymbol{c}_i$ の）語頭(prefix)という。

$$\phi_1 = \{a \mapsto 0, b \mapsto 10, c \mapsto 110, d \mapsto 1110\}$$

$\phi_1(d) = 1110$　の語頭



1

11

111

（瞬時符号の）語頭条件

性質：語頭を用いた瞬時符号の判別

瞬時符号であるための必要十分条件は、
各符号が他の符号の語頭になっていない。

$$\phi_1 = \{a \mapsto 0, b \mapsto 10, c \mapsto 110, d \mapsto 1110\}$$

語頭

$\phi_1(a) = 0$ $\phi_1(b) = 10$ $\phi_1(c) = 110$ $\phi_1(d) = 1110$

ϕ 1 1 1

11 11

111

空集合

これらが符号に含まれない。

クラフトの不等式

符号長で瞬時符号を特徴づけることができる。

性質:クラフトの不等式

符号語長の集合が $L=\{l_1,l_2,\cdots,l_n\}$ であるような
r元瞬時符号 $C=\{\boldsymbol{c}_1,\boldsymbol{c}_2,\cdots,\boldsymbol{c}_n\}$ が存在するための
必要十分条件は、次式が成り立つことである。

$$\sum_{i=1}^{n} r^{-l_i} \le 1$$

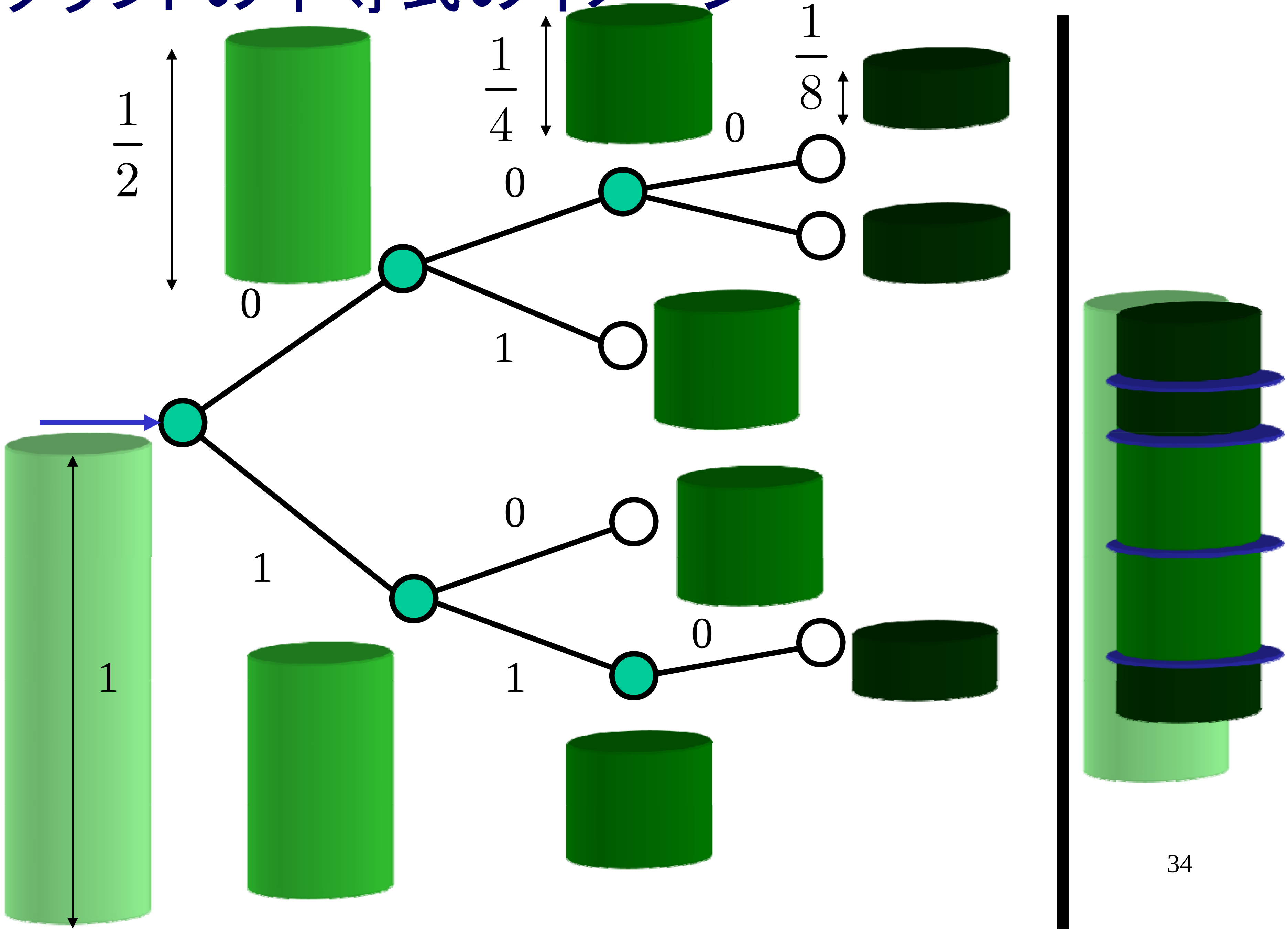
この不等式を**クラフトの不等式**という。

性質:2元符号のクラフトの不等式

２元符号（{0,1}への符号化）の場合のクラフトの不等式

$$\sum_{i=1}^{n} 2^{-l_i} \le 1$$

クラフトの不等式のイメージ



クラフトの不等式の確認

$$\phi_1 = \{a \mapsto 0, b \mapsto 10, c \mapsto 110, d \mapsto 1110\}$$
$$L = \{1, 2, 3, 4\}$$

$$\sum_{i=1}^{4} 2^{-l_i} = \frac{1}{2} + \frac{1}{4} + \frac{1}{8} + \frac{1}{16}$$
$$= \frac{15}{16}$$
$$\leq 1$$

クラフトの不等式の利用

クラフトの不等式はあくまでも、瞬時符号の符号長に関する条件である。したがって、以下の命題しか成り立たない。

瞬時符号である。 $\xrightarrow{\quad}$ / $\not\longleftarrow$ クラフトの不等式を満足する。

例えば、

$$\phi_6 = \{a \mapsto 0, b \mapsto 01, c \mapsto 011, d \mapsto 0111\}$$

の符号はクラフトの不等式を満足するが、瞬時符号ではない。

練習

以下の符号に対して、クラフトの不等式を満たすか調べよ。
また、瞬時符号かどうかを答えよ。

（1）

$$\phi_2 = \{a \mapsto 1110, b \mapsto 110, c \mapsto 10, d \mapsto 0\}$$

（2）

$$\phi_3 = \{a \mapsto 00, b \mapsto 01, c \mapsto 10, d \mapsto 11\}$$

（3）

$$\phi_5 = \{a \mapsto 0, b \mapsto 1, c \mapsto 01, d \mapsto 10\}$$

情報源符号化定理(平均符号長の下限)

性質:情報源符号化定理(重要)

無記憶情報源 S の N 次拡大情報源 S^N に対して、次式を満たす平均符号長 $\overline{L}$ を持つr元瞬時符号が構成できる。

$$\forall \varepsilon > 0$$

$$\frac{H(S)}{\log r} \le \overline{L} < \frac{H(S)}{\log r} + \varepsilon$$

この定理の理解が当面の目標

性質:2元符号版の情報源符号化定理

$$\forall \varepsilon > 0$$

$$H(S) \le \overline{L} < H(S) + \varepsilon$$

エントロピーの重要性の再確認。
平均符号長の下限がエントロピーである。

拡大情報源

ここでは、情報源について再考する。

定義：拡大情報源

情報源 $S=\{s_1,s_2,\cdots,s_n\}$ に対して、 S の情報源記号を N 個並べた順列すべてを情報源アルファベットとする情報源を S （元の情報源）の N 次拡大情報源といい S^N と表す。すなわち、

$$S^N=\left\{s_1',s_2',\cdots,s_{n^N}'\right\}$$

$$\forall k \quad s_k' = s_{k_1}s_{k_2}\cdots s_{k_N}$$

$$\forall k,i \quad s_{k_i}\in S$$

Sの記号をN個ならべて新たな記号とする。

拡大情報源例

$$S=\begin{Bmatrix} a & , & b \\ 1/3 & , & 2/3 \end{Bmatrix}$$ の2次拡大情報源を求める。

まず、2次拡大情報源アルファベットは以下のようになる。

$$S^2=\{aa,ab,ba,bb\}=\{A,B,C,D\}$$

2記号で、
1情報源記号扱いに注意する。

次に、各確率を求める。

$$P(A)=P(aa)=P(a)\times P(a)=1/9$$
$$P(B)=P(ab)=P(a)\times P(b)=2/9$$
$$P(C)=P(ba)=P(b)\times P(a)=2/9$$
$$P(D)=P(bb)=P(b)\times P(b)=4/9$$

$$\therefore S^2=\begin{Bmatrix} aa & , & ab & , & ba & , & bb \\ 1/9 & , & 2/9 & , & 2/9 & , & 4/9 \end{Bmatrix}$$

$$S = \begin{Bmatrix} a & , & b \\ 1/3 & , & 2/3 \end{Bmatrix}$$

の3次拡大情報源は以下となる。

$$S^3 = \left\{ \begin{matrix} aaa & , & aab & , & aba & , & abb & , \\ \dfrac{1}{27} & , & \dfrac{2}{27} & , & \dfrac{2}{27} & , & \dfrac{4}{27} & , \end{matrix} \right.$$

$$\left. \begin{matrix} baa & , & bab & , & bba & , & bbb \\ \dfrac{2}{27} & , & \dfrac{4}{27} & , & \dfrac{4}{27} & , & \dfrac{8}{27} \end{matrix} \right\}$$

練習

次の拡大情報源を求めよ。

（１）

$$S_1 = \begin{Bmatrix} \alpha & , & \beta & , & \gamma \\ \dfrac{1}{6} & , & \dfrac{1}{3} & , & \dfrac{1}{2} \end{Bmatrix}$$ の2次拡大情報源 S_1^2 。

（２）

$$S_2 = \begin{Bmatrix} a & , & b \\ \dfrac{1}{4} & , & \dfrac{3}{4} \end{Bmatrix}$$ の3次拡大情報源 S_2^3 。

拡大情報源のエントロピー

性質:拡大情報源の平均符号長

無記憶情報源 S の平均符号長を $\overline{L}$ とし、
S の N 次拡大情報源 S^N の平均符合長を $\overline{L_N}$ とする。このとき、次が成り立つ。

$$H(S^N) = N \times H(S)$$

$$\overline{L_N} = N \times \overline{L}$$

エントロピーはN倍。エントロピーが1記号あたりの情報量であることから、妥当といえる。

拡大情報源の1記号には、元の情報源記号がN個含まれる。

練習

次のエントロピーをそれぞれ求めよ。

（1）

$$S_1 = \begin{Bmatrix} \alpha & , & \beta & , & \gamma \\ \dfrac{1}{6} & , & \dfrac{1}{3} & , & \dfrac{1}{2} \end{Bmatrix}$$

に対して、$H(S_1)$ および $H(S_1{}^2)$

（2）

$$S_2 = \begin{Bmatrix} a & , & b \\ \dfrac{1}{4} & , & \dfrac{3}{4} \end{Bmatrix}$$

に対して、$H(S_2)$ および $H(S_2{}^3)$

平均符号長の性質

性質:瞬時符号の平均符号長

無記憶情報源 S に対して、次式を満たす平均符号長 $\overline{L}$ を持つr元瞬時符号が構成できる。

$$\frac{H(S)}{\log r} \leq \overline{L} < \frac{H(S)}{\log r} + 1$$

証明

証明方針:

（1）クラフトの不等式を満たす符号長集合を持つ符号を構成する。（瞬時符号では必ず存在する。）
（2）構成した符号が命題の式を満たすことを示す。

（1） $1 \le k \le n$ に対して

$$\frac{-\log P(s_k)}{\log r} \le l_k < \frac{-\log P(s_k)}{\log r} + 1$$

この式を満たす自然数はいつも一つだけ存在する。

を満たす符号長集合 $L = \{l_1, l_2, \cdots, l_n\}$ を持つ符号 $\boldsymbol{C} = \{\boldsymbol{c}_1, \boldsymbol{c}_2, \cdots, \boldsymbol{c}_n\}$ を構成する。

この符号がクラフトの不等式を満たすことを示す。

左の不等号より、

$$-\log P(s_k) \le l_k \log r$$

$$\therefore -l_k \log r \le \log P(s_k)$$

$$\therefore r^{-l_k} \le P(s_k)$$

符号全ての和をとる。

$$\sum_{k=1}^{n} r^{-l_k} \le \sum_{k=1}^{n} P(s_k) = 1$$

クラフトの不等式

よって、クラフトの不等式を満たす。(したがって、
前のスライドの条件を満たす符号が存在する。)

（2） 各項に $P(s_k)$ を乗じる。

$$\frac{-P(s_k)\log P(s_k)}{\log r} \le P(s_k) l_k < \frac{-P(s_k)\log P(s_k)}{\log r} + P(s_k)$$

辺々総和をとる。

$$\sum_{k=1}^{n}\frac{-P(s_k)\log P(s_k)}{\log r} \le \sum_{k=1}^{n} P(s_k) l_k < \sum_{k=1}^{n}\frac{-P(s_k)\log P(s_k)}{\log r} + \sum_{k=1}^{n} P(s_k)$$

分子はエントロピー

平均符号長

確率の和

したがって、次式が成り立つ。

$$\frac{H(S)}{\log r} \le \overline{L} < \frac{H(S)}{\log r} + 1$$

QED.

情報源符号化定理の証明

性質:情報源符号化定理(シャノンの第1定理)

$$\frac{H(S)}{\log r} \le \overline{L} < \frac{H(S)}{\log r} + \varepsilon$$

証明　N次拡大情報源 S^Nに対して瞬時情報源の平均符号長の性質を適用する。

$$\frac{H(S^N)}{\log r} \le \overline{L_N} < \frac{H(S^N)}{\log r} + 1$$

さらに、拡大情報源の平均符号長の性質を適用する。

$$\frac{NH(S)}{\log r} \le N\overline{L_N} < \frac{NH(S^N)}{\log r} + 1$$

$$\frac{H(S)}{\log r} \le \overline{L} < \frac{H(S)}{\log r} + \frac{1}{N}$$

QED.

符号の効率と冗長度

定義:符号の効率

次式で定められる e を符号の効率という。

$$e \equiv \frac{H(S)}{\overline{L}} \qquad (0 \leq e \leq 1)$$

効率的

非効率的
(符号が極端に長い)

定義:符号の冗長度

次式で定められる r を符号の冗長度という。

$$r \equiv 1 - e = \frac{\overline{L} - H(S)}{\overline{L}} \qquad (0 \leq r \leq 1)$$

冗長的
(符号が極端に長い)

冗長性なし

符号の効率の計算

情報源 $S=\begin{Bmatrix} a & , & b & , & c & , & d \\ 0.4 & , & 0.3 & , & 0.2 & , & 0.1 \end{Bmatrix}$ の符号

$$\phi_1=\{a\mapsto 0, b\mapsto 10, c\mapsto 110, d\mapsto 1110\}$$

の効率を求める。

$$H(S)=\frac{4}{10}\log\frac{10}{4}+\frac{3}{10}\log\frac{10}{3}+\frac{2}{10}\log\frac{10}{2}+\frac{1}{10}\log 10$$

$$=\log 10-\left(\frac{4}{10}\log 4+\frac{3}{10}\log 3+\frac{2}{10}\log 2\right)$$

$$\simeq 3.322-(0.8+0.476+0.2)$$

$$\simeq 1.846$$

$$\overline{L}=0.4\times 1+0.3\times 2+0.2\times 3+0.1\times 4=2$$

$$e=\frac{H(S)}{\overline{L}}=\frac{1.846}{2}=0.923$$

練習

次の情報源に対する符号の効率を求めよ。

$$S = \begin{Bmatrix} a & , & b & , & c & , & d \\ 0.4 & , & 0.3 & , & 0.2 & , & 0.1 \end{Bmatrix}$$

（1）

$$\phi_2 = \{a \mapsto 1110, b \mapsto 110, c \mapsto 10, d \mapsto 0\}$$

（2）

$$\phi_7 = \{a \mapsto 0, b \mapsto 10, c \mapsto 110, d \mapsto 111\}$$