情報科学類　情報理論テキスト　２０２２年度版

担当教員　システム情報系情報工学域　教授　伊藤誠 itoh.makoto.ge@u.tsukuba.ac.jp
岡山大学 助教（非常勤）　LEE Jieun

本テキストにおける問題を解くにあたっては，必要に応じて関数電卓などを利用してよい．この講義では細かい数値にはあまり興味がない．

0. 情報理論の目的 Purpose of Information Theory

(1) データをどこまで効率よく符号化（圧縮）できるか・・・情報源符号化定理
(2) データの伝送における誤りをどこまで減らせるか・・・・通信路符号化定理

1. 情報源符号化 Coding of messages

情報伝送系モデルは，次のように表現される．

情報源	→	符号器	→	通信路	→	復号器	→	受信者
Information Source		Encoder		Channel		Decoder		Information Sink

いくつかの重要な用語の定義をしておく．

符号化（coding）：記録媒体や通信路に適した形にするなどの目的で，データを変換する

符号語 (codeword): 符号化された記号列．符号語の長さは固定とは限らない（情報源の記号によって長さが異なるようにしてもよい）ことに注意．

符号(code): 情報源から出てきうる記号（情報源記号）の集合（これを、「アルファベット」という）A から，符号語の集合 B への写像 f: A→B.

r 元符号：符号語に用いる記号が r 種類の符号．
二元符号(binary code)，とくに符号語につかわれる記号の集合が{0, 1}のものがよく使われる

符号は，その性質によって表 1.1 のように分類されうる．

表 1.1　符号の分類

符号 code	正　則　(non-singular)	一意復号可能 (uniquely decodable)	瞬時復号可能 (instantaneously decodable)
			瞬時復号不可能
		一意復号不可能	
	特異(singular)		

- 正則：情報源の全ての相異なる記号に別々の符号語が割り当てられている
- 特異：情報源の相異なる記号の組の中に，同一の符号語が割当てられているものがある
- 一意復号可能（一意分節可能ともいう）：（有限長の）任意の符号語列から，もとの情報源記号列を一意に復元できる
- 瞬時復号可能：一つの符号語の末尾記号を読み込んだとき，それより先を読まなくてもその符号語を切り出すことができる

符号の例を，表 1.2 に示す.

表 1.2　符号の分類の例

情報源記号	符号					
	I	II	III	IV	V	VI
a1	00	0	0	0	0	0
a2	01	10	10	01	10	10
a3	10	110	110	011	11	11
a4	11	1110	111	111	01	0
	瞬時	瞬時	瞬時	一意	正則だが，一意でない	特異

表 1.2 の例において，つぎのことがわかる.

- 符号 VI は特異．0 という符号語が，a1 と a4 に割り当てられている.
- 符号 V は正則だが，一意復号可能ではない．たとえば，010 という符号語列があるとき，これは，a1a2(0-10)とも a4a1(01-0)とも復号されうる
- 符号 IV は一意復号可能ではあるが，瞬時復号可能ではない．たとえば，0111 という符号語列を考えよう. 最初の 0 を読み込んだところで, これを a1 と解釈してよいか，あるいは，01(a2), 011(a3)の最初と見なすべきかは判断つかない．ただし，0111 と最後まで読み込むと，これは，0-111 という風に読まざるを得ないため，一意に復号される

問題 1-1: 次の符号(1) – (6)はそれぞれ何符号だろうか.

情報源記号	符号					
	(1)	(2)	(3)	(4)	(5)	(6)
a1	00	0	11	00	0	0
a2	10	0	10	01	10	01
a3	110	1	01	10	11	10
a4	111	1	010			

符号語長(codeword length): 一つの符号語に含まれる記号の数を符号語長という．すべての符号語の符号語長が一定である符号を固定長符号(fixed-length code)といい，固定長符号でないものは可変長符号(variable-length code)という.

表 1.2 において，符号 I は固定長符号，それ以外は可変長符号である.

平均符号語長 $\overline{L}$: 情報源記号 $a_i,\ i=1,...,M$ (M は，情報源の記号の種類)の符号語長を l_i, 出現確率を p_i とするとき，平均符号語長 $\overline{L}$ を次のように定義する.

$$\overline{L}=\sum_{i=1}^{M} l_i p_i \tag{1.1}$$

平均符号語長は，符号語長の期待値である.

例題 1.1. 下の符号の場合，平均符号語長 $\overline{L}$ はいくらか.

情報源記号	符号	p_i
	II	
a1	0	1/2
a2	10	1/6
a3	110	1/6
a4	1110	1/6

解答.

$$\overline{L} = 1 \times \frac{1}{2} + 2 \times \frac{1}{6} + 3 \times \frac{1}{6} + 4 \times \frac{1}{6} = 2$$

問題 1-2: 表 1.2 の符号 II 以外，すなわち，符号 I,III,IV,V,VI の平均符号語長を求めてみよう．ただし，p_i は上の例と同じものとする．

情報源の符号化には，どのような符号が望ましいだろうか．一般には，次の二つのことが望まれる．

(1) 一意復号可能である（瞬時復号可能ならなおよい）
(2) 平均符号語長ができるだけ短い

瞬時復号可能であるためには，「語頭条件」(prefix condition)が要求される．

語頭（prefix）: 符号語の先頭部分のこと．先頭から 0 文字，1 文字，2 文字,. . . のすべてが語頭である．その意味では，符号語自身も語頭である．符号語自身を除いた語頭を，真の語頭ということがある．

例 1-1．符号語 101011 の場合，空系列（λ），1,10,101,1010,10101 が（真の）語頭である．

語頭条件：どの符号語も，他の符号語の語頭と一致しない．

語頭条件がなりたつと，各符号語の区切りまできたとき，そこが符号語の区切りであることが，先を読込まなくてもわかる．したがって，語頭条件が成り立つ符号は瞬時復号可能である．
なお，瞬時複号可能であることは，語頭条件できれいに決まるのであるが，一意復号可能であるための条件を求めるのは実は大変難しい（興味のある人は，「サーディナス・パターソンの定理」を調べるとよい）．

例 1-2．表 1.3 の符号は，符号語の 0, 01 が，他の符号語の語頭に一致するものがあることから，語頭条件を満たさない．

表 1.3 語頭条件を満たさない符号の例

符号	語頭
0	λ，0
01	λ，0, 01
011	λ，0, 01, 011
111	λ，1, 11, 111

例 1-3．表 1.4 の符号は，どの符号語も，他の符号語の語頭に一致するものがないことから，語頭条件を満たす．

表 1.4 語頭条件を満たす符号の例

符号	語頭
0	λ，0
10	λ，1, 10
110	λ，1, 11, 110
111	λ，1, 11, 111

符号の木 (tree)：r 元符号を，r 分木で表現する．根 (root)から，各ノード(node)への枝（branch）で，符号語を表現する．

例 1-4．00, 01, 10, 11 という符号語を持つ符号は，次のような符号の木で表すことができる．木の末端は、葉（leave）と呼ばれる。この例では、すべての符号語の終端が葉になっている。

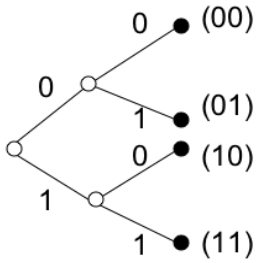


図 1.1

例 1-5．0, 01,011,111 の場合の符号の木．この例では、0 などの符号語の終端は葉にはなく、途中のノードにある。

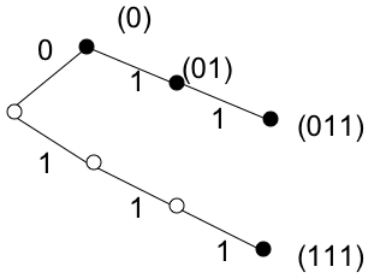


図 1.2

例 1-6．0, 10, 110, 111 の場合の符号の木．

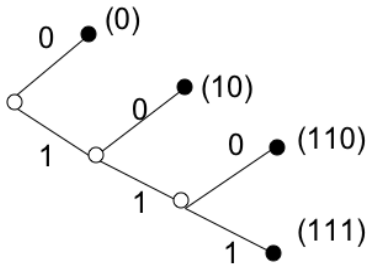


図 1.3

語頭条件を満たす符号は，全ての符号語が，葉で表されている．上の例 1.5 では，0, 01, 011, 111 の場合，0,01 が，葉でないところにあることからわかるように，この符号は瞬時復号可能ではない．実際，0111 という符号語列にたいして，0 まで読み込んだところでは，それが区切りかどうかがわからない．他方，例 1.6 の 0,10,110,111 の場合，全ての符号語が葉に割り当てられていることからわかるように，これは瞬時復号可能である．

クラフトの不等式 (Kraft inequality)：M 個の記号を持つ情報源を r 元符号化する場合，符号語長が $l_i \ \ (i=1,...,M)$ であるような瞬時復号可能符号は，下記の不等式を満たす．逆に，この不等式を満たす符号語長を持つように，瞬時復号可能な符号をつくることができる．

$$\sum_{i=1}^{M} r^{-l_i} \le 1 \tag{1.2}$$

これをクラフトの不等式という．なお，一意復号可能な符号でも上式を満たす．この意味では，上の不等式はマクミランの不等式という．

（証明のアイデア）

（語頭符号⇒クラフトの不等式）

根から出発して，r 個のノードをつくり，1/r ずつ重みを与えることを l_i の最大値まで再起的に行うこと考えよう．木の高さを h とすると，こうしてできる r 分木の葉には，それぞれ r^{-h} の重みがつく（葉は全部で r^h 枚）．一般に，r 分木の一部を刈り取った木の全ての葉（子ノードを持たないノード）の重みの和は 1 を超えない．

つぎに，符号語長が $l_i \ \ (i=1,...,M)$ の瞬時復号可能な符号を考えよう．最大符号語長を h とする．この符号の木は，その r 分木の一部を刈り取ったものに相当する（すべての符号語は葉につく）．したがって，瞬時複合可能な符号の木の葉の重み（符号語につく重み）は 1 を超えない．

（クラフトの不等式⇒語頭符号）

クラフトの不等式を満たす $l_i \ \ (i=1,...,M)$ について，最も小さな l_i の高さを持つ木を考える．その木の一つの葉を選び，対応する符号語とする．符号語に選ばれなかった葉を，次に最も小さい l_i を選び，その高さまで枝を伸ばす．このことにより 1 つ以上の葉が存在するので，そのうちの一つを選び，対応する符号語とする．以下，最も大きな l_i になるまで，枝を伸ばすことを繰り返す．その都度，一つ以上の葉が存在することから，対応する符号語を割り当てる．こうして得られた符号語の組は，明らかに語頭符号となる．

クラフトの不等式のもう少し厳密な証明を示しておこう．ただし，簡単のため，ここでは 2 分木で考えることにする．

（語頭符号⇒クラフトの不等式）

M 個の記号について，符号語の長さの順に並べ替える．すなわち，

$$l_1 \leq l_2 \leq \cdots \leq l_M \tag{1.3}$$

この語頭符号の木を T_1 とよぶことにする．ここで，T_1 の高さはl_Mであることに注意する．

高さl_Mの符号の木で，すべての葉が高さl_Mにあるものを T_2 とよぶことにする．T_2 において，T_1 におけるl_1の高さを持つ葉と同じ場所から先をすべて刈り取る．このとき刈り取られる部分は，$l_M - l_1$の高さを持つ木であるから，刈り取られる葉は，$2^{l_M - l_1}$だけある．同様に刈り取りを続けると，刈り取られる葉の総数 S は，

$$S = \sum_{i=1}^{M} 2^{l_M - l_i} \tag{1.4}$$

となるが，この S は T_2 の葉の総数2^{l_M}を超えることはない．すなわち，次式が成立する．

$$S = \sum_{i=1}^{M} 2^{l_M - l_i} \leq 2^{l_M} \tag{1.5}$$

この両辺を2^{l_M}で割れば，次式を得る．

$$\sum_{i=1}^{M} 2^{-l_i} \leq 1 \tag{1.6}$$

これは，クラフトの不等式に他ならない．

（クラフトの不等式⇒語頭符号）

ある符号語長の組み合わせ$l_1, l_2, \cdots, l_M$ $(l_1 \leq l_2 \leq \cdots \leq l_M)$が，クラフトの不等式

$$\sum_{i=1}^{M} 2^{-l_i} \leq 1 \tag{1.7}$$

を満たすとする．すべての葉が高さl_1の木 T を考えよう．この T の一つの葉を選び，l_1に該当する記号の符号語とする．符号語としなかった葉について，$l_2 - l_1$だけ枝を伸ばすとする．このとき，符号語としたことによって伸ばせなかった枝には，$2^{l_2 - l_1}$だけ葉がつけられたはずであるから，枝を伸ばしたことによってできる葉の総数は

$$2^{l_2} - 2^{l_2 - l_1} \tag{1.8}$$

となる．$2^{l_2} - 2^{l_2 - l_1} > 0$（すなわち，1 以上）が必ず成り立つことから，高さl_2の葉を一つとってl_2に該当する符号語に割り当てることができる．以下同様のことを繰り返していく．

符号語とならなかった葉から，さらに枝を伸ばす（ここまでの段階では，枝を伸ばすことにより葉は複数できるので，符号語に割り当てた葉以外にも葉はある．したがって，そこから枝を伸ばすことは可能）と，高さl_3の葉は，

$$2^{l_3} - 2^{l_3 - l_2} - 2^{l_3 - l_2} \tag{1.9}$$

となる．ここで，符号語長l_1, l_2, l_3がクラフトの不等式を満たすことから，

$$2^{-l_1} + 2^{-l_2} < \sum_{i=1}^{3} 2^{-l_i} \leq 1 \tag{1.10}$$

である．このことから，

$$2^{l_3} - 2^{l_3 - l_2} - 2^{l_3 - l_2} = 2^{l_3}(1 - 2^{-l_1} - 2^{-l_2}) > 0 \tag{1.11}$$

となり，高さl_3の葉を一つとってl_3に該当する符号語に割り当てることができる．同様に枝を伸ばしていっても，符号語を順次割り当てていくことが可能である．

高さl_Mに到達したとき，高さl_Mの葉は，

$$2^{l_M} - 2^{l_M - l_1} - \cdots - 2^{l_M - l_{M-1}} = 2^{l_M}\left(1 - \sum_{i=1}^{M-1} 2^{-l_i}\right) > 0 \qquad (1.12)$$

となり，一つ以上存在するので，符号語を割り当てることができる．

以上によって作成された符号は，明らかに語頭符号である．

クラフトの不等式の注意点：

瞬時復号不可能な符号でも，クラフトの不等式を満たすことがある．クラフトの不等式は，「そのような符号語長の組で，瞬時復号可能な符号が存在しうる」ことを言っているにすぎない．

例 1-7．次の表 1.5 の符号では，$\sum_{i=1}^{M} 2^{-l_i} = 2^{-1} + 2^{-2} + 2^{-3} + 2^{-3} = 1$となっていることからわかるように，クラフトの不等式を満たす．ただし，この符号は瞬時復号可能ではない．クラフトの不等式は，$[l_1, l_2, l_3, l_4] = [1,2,3,3]$となる瞬時復号可能な符号が存在することを主張する（表 1.2 の例では，符号 III が同じ符号長の組を持ち，瞬時復号可能な符号の例である）．

表 1.5 クラフトの不等式を満たすが瞬時復号可能ではない符号の例

符号 IV	l_i
0	1
01	2
011	3
111	3

完全な符号(complete code)：全ての符号語が完全な木（complete tree）の全ての葉を用いて割り当てられている符号．ここで，完全な木とは，「全てのノードが，r ヶの子ノードを持つか，子ノードを持たないかのいずれかである」木のことをいう．

例 1-8．完全でない符号

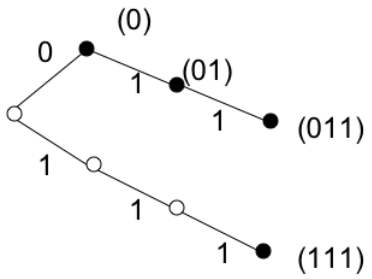


図 1.4

例 1-9. 完全な符号

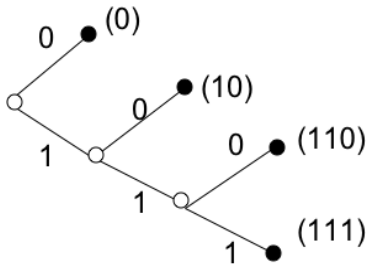


図 1.5

完全な符号の場合，クラフトの不等式において等号が成立する.

問題 1-3: 5 つの情報源記号(a1, a2, a3, a4, a5)を発生させる情報源について，瞬時複号可能な（二元）符号の例を示し，それがクラフトの不等式を満たすことを確認せよ.

瞬時復号可能な符号を作るには，クラフトの不等式を満たすことが重要であることが分かったが，平均符号語長を短くすることはどのようにして可能だろうか.

最短符号 : 語頭符号の中で平均符号語長が最小のものを最短符号という. 最小冗長符号とか，コンパクト符号などともいう.

記憶のない情報源(memoryless source of information)に対し，最短符号を構成する方法は，ハフマン符号という名で知られている. 2 元符号の場合，ハフマン符号の構成の方法はつぎのようである.

ハフマン符号化 (Huffman code) :
(1) 大きさ M のアルファベット A の記号の出現確率の小さいものから二つをまとめて一つの新たな記号とみなし，もとの二つの記号の出現確率の和を新たな記号の出現確率とする.
(2) すべての記号が一つの記号と見なされるまで，(1)の操作を繰り返す（M-1 回行うことになる）.
(3) 以上の結果を表す符号の木の各枝に，0 と 1 を割り当てる.
(4) 根から葉までの各枝に割り当てられた記号列が，その葉に対応する記号の符号語となる.

例 1-10. 次の表のように出現確率が与えられている情報源 A を考えよう.

表 1.6 情報源 A のアルファベットと出現確率

記号	出現確率 p_i
a_1	0.4
a_2	0.25
a_3	0.2
a_4	0.1
a_5	0.05

ハフマン符号化によって，つぎの図 1.6 のような符号の木を得る．

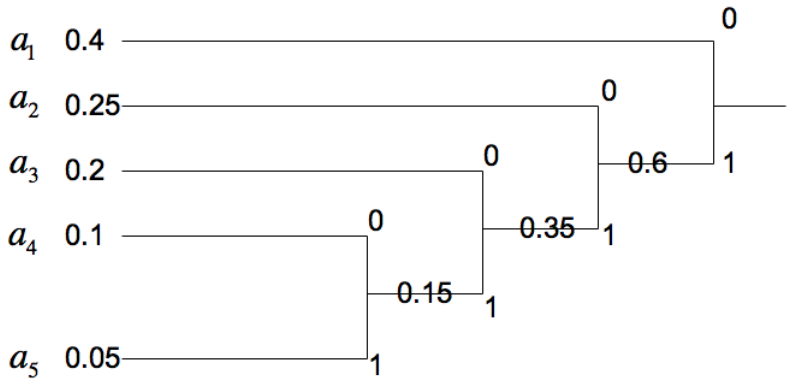


図 1.6 ハフマン符号化の例

すなわち，符号語は下記のとおりである．

表 1.7 情報源 A のハフマン符号

記号	符号語
a_1	0
a_2	10
a_3	110
a_4	1110
a_5	1111

一つの情報源に対して，できあがるハフマン符号は一意に定まるとは限らない場合があるので注意せよ．

ハフマン符号が最短符号となるのはなぜだろうか．証明の詳細は横尾(2004)等に譲ることとして，ここでは基本的な考え方を示しておく．図 1.7 は上の例で構築されたハフマン符号を再掲したものである．ここで S_2 は， a_4,a_5 を一つにまとめて 4 つの記号からなる情報源を意味しており，縮退情報源という． S_3,S_4 も同様である．

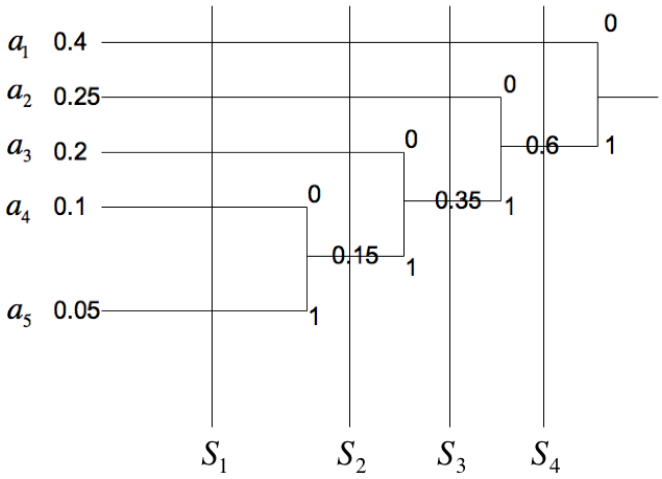


図 1.7　縮退情報源

明らかに，上図における S_4 の符号は最短符号である．一方，ハフマン符号化でつくられる S_{i+1} の符号が最短符号なら S_i の符号も最短符号となる．その理由は以下の通り． S_{i+1} の平均符号語長を $\overline{L_{i+1}}$ とすると，$\overline{L_{i+1}} = \overline{L_i} - p - q$ がなりたつ．ここで， p,q は S_i のなかで最も出現確率の小さい２つの記号の出現確率を意味する．S_i が最短符号でなければ $\overline{L'_i} < \overline{L_i}$ を満たすような $\overline{L'_i}$ を平均符号語長として持つ i 番目の縮退情報源の符号が存在することになるが，このとき， $\overline{L'_i}$ に基づいて改めて i+1 番目の縮退情報源を構築し，その平均符号語長を $\overline{L'_{i+1}}$ とすると

$$\overline{L'_{i+1}} = \overline{L'_i} - p - q < \overline{L_i} - p - q = \overline{L_{i+1}} \tag{1.13}$$

となり， $\overline{L_{i+1}}$ が最短符号の平均符号語長であることと矛盾する．以上のことから数学的帰納法によって図 1.7 の S_1 の符号は最短符号であることが保証される．

この，縮退情報源を右から左に広げていく方法は，クラフトの不等式の証明において，枝を徐々に伸ばしていったことと対応している．

なお，ハフマン符号は，ファックスや MP3 等でも利用されている重要な方法であるので，しっかりと理解しておくことが必要である．

問題 1-4: ４つの情報源記号(a1, a2, a3, a4)を発生させる情報源がある．それぞれの記号の生起確率は，以下の通りとなっている．

$$p_1 = \frac{1}{8}, p_2 = \frac{1}{8}, p_3 = \frac{1}{4}, p_4 = \frac{1}{2}$$

この情報源を，二元符号としてハフマン符号化せよ．

問題 1-5: 次の 5 つの情報源記号を発生させる情報源を二元符号としてハフマン符号化し，平均符号語長を求めよ．

記号	出現確率 p_i
a_1	0.3
a_2	0.23
a_3	0.2
a_4	0.15
a_5	0.12

問題 1-6: 6 つの情報源記号(a1, a2, a3, a4, a5, a6)を発生させる情報源がある．それぞれの記号の生起確率は，以下の通りとなっている．

$$p_1 = \frac{1}{9}, p_2 = \frac{1}{9}, p_3 = \frac{1}{9}, p_4 = \frac{1}{6}, p_5 = \frac{1}{6}, p_6 = \frac{1}{3}$$

この情報源をハフマン符号化せよ．

問題 1-7: 2 つの情報源記号(a1, a2)を発生させる情報源がある．それぞれの記号の生起確率は，以下の通りとなっている．

$$p_1 = \frac{1}{3}, p_2 = \frac{2}{3}$$

このとき，平均符号語長が 1 未満の瞬時復号可能な（二元）符号は作成できるだろうか．（ヒント：発想を柔軟に）

（付録 1-1．ユニバーサル符号）

ハフマン符号化は，情報源がどのような確率分布に従うかがわかっている場合には適用できるのであるが，そうでない場合には最短符号を構成することはできない．

現実には，情報源がどのような確率分布に従うかはわからないこともある．情報源の確率分布によらず，平均符号語長を短くすることのできる方法は存在する．そのような方法によって得られる符号は，ユニバーサル符号と呼ばれる．ユニバーサル符号の代表的なものに，LZ 符号と呼ばれるものがある．LZ 符号では，情報源の確率分布をしらなくても，一文字あたりのエントロピーを最小化することができる．符号化には，増分分解という手法を用いる．詳しくは，植松(2012)を参照されたい．

2. 情報量とエントロピー Information and Entropy

情報量(information): 事象 (event) A の生起確率 (probability)を $P(A)$ とあらわすとき，つぎの $I(A)$ を、情報量(information)という．

$$I(A) = \log_2 \frac{1}{P(A)} = -\log_2 P(A) \quad \text{（単位はビット（bit））} \tag{2.1}$$

この情報量は，自己情報量 (self information)とも呼ばれる．

情報量の性質：

- 確率の小さな事象の生起を知ったとき：情報量は大きい
- 確率の大きな事象の生起を知ったとき：情報量は小さい

$$P(A) \le P(B) \text{ ならば，} \quad I(A) = -\log_2 P(A) \ge -\log_2 P(B) = I(B) \tag{2.2}$$

- 二つの独立事象 A, B がともに生起をしたことを知ったとき：

$$I(A,B) = I(A) + I(B) \text{ (情報量の加法性)} \tag{2.3}$$

なお，事象 A の生起確率を p，事象 B の生起確率を q としたとき，A,B が独立であるならば，二つの事象の積事象（ここでは(A,B)と表すことにする）の生起確率は pq である．このとき，(A,B)を直接知ることによって得られる情報量 $I(A,B)$が，$I(A)$と I(B)の和になるためには，情報量は(2.1)式のように必ず対数の形をとらざるを得ないことが証明できる（付録 2.1 参照）．

問題 2-1: $\log_2 3 = 1.58$として，次の問いに答えよ
(1) fair なさいころの目が'1'だと知ったときの情報量を求めよ
(2) fair なさいころの目が偶数だと知ったときの情報量を求めよ

エントロピー(entropy):（離散）確率変数 (discrete random variable) X に対し，下記の $H(X)$を X のエントロピー (entropy)と呼ぶ．

$$H(X) = -\sum_X P(X = x) \log_2 P(X = x) \tag{2.4}$$

ただし，$0 \log_2 0 = 0$と定義する．

エントロピーは，X の情報量 $I(X)$の期待値の形をとっていることから，情報量の期待値（平均情報量）とみなしてよい．

エントロピーの意味： エントロピーは，情報源の不確実性の度合（どの情報源記号が出てくるかわからない度合）を表す．何が出てくるかわからないときの方が，情報を得たときの情報量は大きい（情報量の期待値が大きい）．

注意：確率変数 X が連続的である場合もエントロピーを定義することはできるのであるが，その場合エントロピーは絶対値としての意味を持たなくなる．ただし，その場合でも，相対的にエントロピーの大小を考えることには意味がある．この点について，詳しくは甘利(1970)を参照

のこと．

エントロピーの意味を，2 元エントロピー関数 (binary entropy function)を用いて具体的に考えよう．

いま, 確率変数 X は, 0 か 1 の値をとりうるものとする. また, $P(X=0)=p, P(X=1)=1-p$ とする．このとき，つぎの $H(X)$を 2 元エントロピー関数という．p の関数だという意味で $H(p)$ と書くこともある．

$$H(X) = -p\log_2 p - (1-p)\log_2(1-p) \tag{2.5}$$

2 元エントロピー関数は，上に凸であり，p=0.5 で最大値(1)，p=0, 1 で最小値(0)をとる．

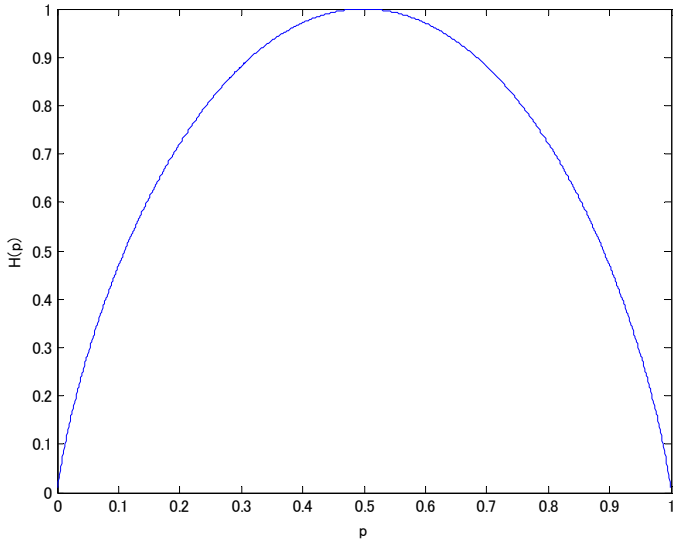


図 2.1　2 元エントロピー

同様に，一般の場合でもつぎのことがいえる．

- エントロピーが最大となるのは，一様分布の場合（場合の数が M ならエントロピーは $\log_2 M$）
- エントロピーが最小となるのは，ある基本事象の生起確率が 1（他の基本事象の生起確率は全て 0）である場合．エントロピーの最小値はゼロである．すなわち，エントロピーは非負である（問題 2-3）．

エントロピーの最大値は，ラグランジュの未定乗数法で求めることができる（ラグランジュの未定乗数法については，ビショップ(2007)などにわかりやすく書かれている）．ラグランジュの未定乗数法は，制約条件のもとで極値を得るためにしばしば用いられる方法である．ある凸集合において上に凸な関数に対しては，ラグランジュの未定乗数法によって，その凸集合の範囲内で最大値が得られる．エントロピーは，上に凸な関数であることが知られている（付録 2.2 参照）．

具体的にエントロピーの最大値を求めるには，次のように行う．問題は次のように与えられる．

$$\begin{cases} maximize \quad H(X) = -\sum_{i=1}^{M} p_i \log_2 p_i \\ s.t. \qquad \sum_{i=1}^{M} p_i = 1 \end{cases} \tag{2.6}$$

この問題の解は，つぎの連立方程式の解によって与えられる（ラグランジュの未定乗数法）．

$$\begin{cases} \frac{\partial}{\partial p_i} \left(H(X) - \lambda \left(\sum_{i=1}^{M} p_i - 1 \right) \right) = 0 \\ \sum_{i=1}^{M} p_i = 1 \end{cases} \tag{2.7}$$

これをとくと， $p_i = \frac{1}{M}$ が得られる．したがって，エントロピーは $\log_2 M$ になる．

問題 2-2: 上の連立方程式を解いてみよ．（重要）

問題 2-3: エントロピーが非負であることを証明せよ．また，ある基本事象の生起確率が 1（他の基本事象の生起確率は全て０）である場合にのみエントロピーがゼロとなることを確認せよ．（重要）

問題 2-4: 数値の計算には電卓を用いてよい．
(1) 2 つの情報源記号(a1, a2)を発生させる情報源がある．それぞれの記号の生起確率は，以下の通りとなっている．このとき，この記号のエントロピーを求めよ．

$$p_1 = \frac{1}{4}, p_2 = \frac{3}{4}$$

(2) (1) の２つの記号をまとめて新たな記号を４つ作る（独立に 2 回記号が出現すると考えよ），すなわち，a1a1,a1a2,a2a1,a2a2．このとき，この記号のエントロピーを求めよ．

問題 2-5: 数値の計算には電卓を用いてよい．
英語の文字が以下の確率で出現する．1 文字のエントロピーを求めよ．
E, T: 1/8
A,I,N,O,S,H,R,D: 1/16
L,U,C,M: 1/32
R,W,Y,G,P,B,V,K,Q,J,X,Z: 1/96

問題 2-6: 二つの独立な実験 A,B から得られる記号（結果）が次のようになっている．
実験 A. 記号（結果）は，a1, a2. $p_1 = \frac{1}{2}, p_2 = \frac{1}{2}$
実験 B. 記号（結果）は，b1, b2. $p_1 = \frac{1}{3}, p_2 = \frac{2}{3}$
(1) それぞれの実験のエントロピーを求めよ．
(2) 実験 A,B を組み合わせた実験 AB（記号は a1b1,a1b2,a2b1,a2b2）のエントロピーを求めよ．

情報量の定義（再訪）： 情報をえることによって，エントロピーが H から H'に変化したならば，その情報が持つ情報量 I は，次の通りである．

$$I = H - H' \tag{2.8}$$

これは、最初に定義した自己情報量とどんな関係があるだろうか.
たとえば，正しく作られたサイコロを考える．一番最初は，１から６のうち，どの目も「同様に確からしい」と考えられるので，その時のエントロピー*H*は，

$$H = -\sum_{i=1}^{6} \frac{1}{6} \log_2 \frac{1}{6} = 1 + \log_2 3$$

となる．つぎに，「奇数の目が出た」という情報 *I* を得たとしよう．これによって，残る可能性としては{1,3,5}となるが, それのうちいずれの目が出やすいかということについては何の知識も持っていないとすると，そのときのエントロピー*H'*は,

$$H = -\sum_{i=1}^{3} \frac{1}{3} \log_2 \frac{1}{3} = \log_2 3$$

となる．したがって，情報量 *I* は,

$$I = 1 + \log_2 3 - \log_2 3 = 1$$

となる．奇数の目が出るという確率は，$P(\{1,3,5\}) = \frac{3}{6} = \frac{1}{2}$であるから，奇数の目が出ることの自己情報量は

$$I(\{1,3,5\}) = -\log_2 \frac{1}{2} = 1$$

であるから，ちょうどエントロピーの差分と一致することが確認できた.

今後，通信路に伝送される情報量などを議論するにあたり，エントロピーに関して抑えておくべき重要な性質がいくつかあるので，まとめてみておくことにする．ここでは，条件付きエントロピーと，結合エントロピーが重要である.

条件付きエントロピー (conditional entropy) : 条件付きエントロピー $H(X|Y)$ は，次のように定義される.

$$H(X|Y) = \sum_{Y} P(Y=y) H(X|Y=y) \tag{2.9}$$

すなわち，確率変数 *Y* がある実現値 *y* をとるもとでのエントロピー $H(X|Y=y)$ の期待値である．なお， $H(X|Y=y)$ は，次式で定義される.

$$H(X|Y=y) = -\sum_{X} P(X=x|Y=y) \log_2 P(X=x|Y=y) \tag{2.10}$$

(2.9), (2.10)式から，以下を得ることができる.

$$H(X|Y) = -\sum_X \sum_Y P(X=x, Y=y) \log_2 P(X=x|Y=y) \tag{2.11}$$

いくつかの教科書などでは，後者の書き方で条件付きエントロピーを定義している．ただし，(2.11)の書き方だとなぜこれを条件付きエントロピーと呼ぶのかがわかりづらいので，条件付きエントロピーを(2.9)式に基づいて理解することを勧める．

問題 2-7: 条件付きエントロピーの非負性，すなわち $H(Y \mid X) \geq 0$ を証明せよ．等号が成り立つのはどのような場合であるか.

結合エントロピー (joint entropy)：二つの確率変数 X,Y について，つぎのものを結合エントロピーとよぶ

$$H(X,Y) = -\sum_{X}\sum_{Y} P(X=x,Y=y)\log_2 P(X=x,Y=y) \tag{2.12}$$

エントロピーのチェイン則 (chain rule)：一般に，次式がなりたつ．これを，エントロピーのチェイン則という．

$$H(X,Y) = H(X) + H(Y \mid X) \tag{2.13}$$

（証明） X,Y の実現値を x,y とする．

$$\begin{aligned}
H(X,Y) &= -\sum_{x}\sum_{y} p(x,y)\log_2 p(x,y) \\
&= -\sum_{x}\sum_{y} p(x)p(y \mid x)\log_2 p(x)p(y \mid x) \\
&= -\sum_{x}\sum_{y} p(x)p(y \mid x)\{\log_2 p(x) + \log_2 p(y \mid x)\} \\
&= -\sum_{x}\left(p(x)\log_2 p(x)\sum_{y} p(y \mid x)\right) - \sum_{x}\left(p(x)\sum_{y} p(y \mid x)\log_2 p(y \mid x)\right) \\
&= -\sum_{x} p(x)\log_2 p(x) - \sum_{x}\big(p(x)H(Y \mid X=x)\big) \\
&= H(X) + H(Y \mid X)
\end{aligned} \tag{2.14}$$

問題 2-8: $H(X,Y,Z) = H(X) + H(Y \mid X) + H(Z \mid X,Y)$ を証明せよ.

結合エントロピーの性質：　つぎの不等式が成り立つ.

$$H(X,Y) \leq H(X) + H(Y) \tag{2.15}$$

上式において，等号が成立するのは，X,Y が独立の場合である．

（証明） X,Y の実現値を x,y とする．

$$\begin{aligned} H(X)+H(Y) &= -\sum_x p(x)\log_2 p(x) - \sum_y p(y)\log_2 p(y) \\ &= -\sum_x\sum_y p(x,y)\log_2 p(x) - \sum_x\sum_y p(x,y)\log_2 p(y) \\ &= -\sum_x\sum_y p(x,y)\{\log_2 p(x)+\log_2 p(y)\} \\ &= -\sum_x\sum_y p(x,y)\log_2 p(x)p(y) \\ &\geq -\sum_x\sum_y p(x,y)\log_2 p(x,y) \\ &= H(X,Y) \end{aligned} \tag{2.16}$$

なお，上の証明においては，次の関係を利用している．$\sum_i p_i = 1$，$\sum_i q_i = 1$ のとき，

$$-\sum_i q_i \log_2 q_i \leq -\sum_i q_i \log_2 p_i \tag{2.17}$$

問題 2-9: $\log_e x \leq x-1$（ただし，$x>0$）なる関係を用いて，(2.16)式を示せ．

問題 2-10: $H(Y|X) \leq H(Y)$ を証明せよ．

問題 2-11: $H(X) \leq H(X,Y)$ を証明せよ．

付録 2-1．なぜ情報量を、対数を使って定義するのか？

確率 p で起こる事象の情報量を $f(p)$ とかく．この f は，微分可能であると仮定する．

いま，二つの独立な事象 A,B があり，それぞれの生起確率が p,q であるとする．この A,B の積事象 $A\cap B$ を考えよう．この積事象の発生を知ることによって得られる情報量 $f(pq)$ は，それぞれの事象の情報量 $f(p)$, $f(q)$ の和となることを仮定する．すなわち，

$$f(pq)=f(p)+f(q)$$

この f がどのように表現されるかを考えよう．通常の関数のように考えるために，p の代わりに x を用いて $f(x)$ の微分を考える．ごく小さな ε に対し， $f(x+\varepsilon x)$ を考えると，

$$f(x+\varepsilon x)=f((1+\varepsilon)x)=f(1+\varepsilon)+f(x)$$

となることから，次式が成り立つ．

$$\frac{f(x+\varepsilon x)-f(x)}{\varepsilon x}=\frac{f(1+\varepsilon)}{\varepsilon x}=\frac{1}{x}\cdot\frac{f(1+\varepsilon)}{\varepsilon}$$

ここで， $\varepsilon\to 0$ （すなわち $\varepsilon x\to 0$）として， $c=\lim_{\varepsilon\to 0} f(1+\varepsilon)/\varepsilon$ とおけば，次式を得る．

$$\lim_{\varepsilon x\to 0}\frac{f(x+\varepsilon x)-f(x)}{\varepsilon x}=f'(x)=\frac{c}{x}$$

したがって， $f(x)$ は次式の形とならざるを得ない．

$$f(x)=c\log_e x+d \quad (\text{ただし，}\ d\ \text{は定数})$$

あとは，定数である c,d を適当に決めればよい．ここで， $x=1(p=1)$ とするとき，ある事象の生起が確実であることは情報量がまったくないことを意味すると考えれば， $f(x)=f(1)=0$ としてよいので，

$$f(1)=c\log_e 1+d=0$$

より， $d=0$ が得られる．つぎに，2 者択一のときの情報量を 1（ビット）とみなすものと仮定すれば，

$$f\left(\tfrac{1}{2}\right)=c\log_e\tfrac{1}{2}=1$$

より， $c=-\log_2 e$ を得る． $f(x)$ を，底を 2 とする対数で書きなおせば，最終的に次式が得られる．

$$f(x)=-\log_2 x$$

付録 2.2　凸関数

上に凸な関数：開区間 (a,b) で定義された実関数 $f(x)$ が任意の $x_1, x_2 \in (a,b)$ と任意の $\lambda\ (0 \le \lambda \le 1)$ に対して

$$\lambda f(x_1) + (1-\lambda) f(x_2) \le f(\lambda x_1 + (1-\lambda) x_2)$$

を満たすとき，この $f(x)$ は (a,b) において上に凸である．

$f(x) = \log_2 x$ とすると，この $f(x)$ は上に凸である．

同じように，n 次元空間の凸集合 I 上で定義された n 変数実関数 $f(\mathbf{x})$ が，n 次元ベクトルの $\mathbf{x}, \mathbf{y}$ に対して，

$$\lambda f(\mathbf{x}_1) + (1-\lambda) f(\mathbf{x}_2) \le f(\lambda \mathbf{x}_1 + (1-\lambda) \mathbf{x}_2)$$

を満たすとき，この $f(\mathbf{x})$ は I において上に凸である．

ここでは証明は省略するが，$H(X) = f(p_1, ..., p_M) = -\sum_{i=1} p_i \log_2 p_i$ は，上に凸である．このことにより，ラグランジュの未定乗数法を用いて $H(X)$ の最大値を求めることができる．

Jenzen の不等式：有限個の実数値 $x_1, ..., x_M$ をとりうる離散確率変数 X，上に凸な関数 $f(x)$ に対して，次の不等式が成り立つ．これを，Jenzen の不等式という．

$$\sum_{i=1}^{M} p_i f(x_i) \le f\left(\sum_{i=1}^{M} p_i x_i \right)$$

ただし，$p_i = P(X = x_i)$ である．Jenzen の不等式に基づいて，ダイバージェンス（付録 6-1 参照）の非負性が得られ，さらに平均相互情報量（６章参照）の非負性，エントロピーの上界などを得ることができる．

3.　マルコフ情報源

ここまで何気なく使ってきた「情報源」という言葉を，改めてもう少し厳密に定義する．

情報源(information source): 出力する記号列の集合（アルファベットという）の要素を実現値とする確率変数が，ある確率分布に従って「次々に」記号を出力する

記号の出てくるパターンによっては，エントロピーが低い（何が出てくるか予想できる）場合もあるかもしれない．

つぎの例を考えよう．2 種類の記号 *H*, *T* を出す情報源 *X* があるとする． 1 回だけ情報が出てくるという前提で考えるならば，*P*(*H*)=*P*(*T*)=1/2 と仮定すると，*X* のエントロピー*H*(*X*)は，$H(X) = -0.5\log_2 0.5 - 0.5\log_2 0.5 = 1$（これは二元エントロピーの最大値でもある）が，次々に出てくる記号が

HTHTHTHTHTHTHTHTHTHTHTHTHTHTH ・・・

と続くとするならば，出現確率（頻度）は *P*(*H*)=*P*(*T*)=1/2 であっても，エントロピーは 1 だと考えるのはおかしいだろう．では，どのように考えたらよいだろうか．このことを考えるために，出力される記号のまとまり（ブロック）という考え方を導入する．

ブロック：情報源から出力される記号列を *n* ヶまとめた記号列を，長さ *n* のブロックという

例 3-1: 長さ３のブロック

$$\underbrace{x_1,x_2,x_3}_{X_1^{n=3}},\underbrace{x_4,x_5,x_6}_{X_2^{n=3}},\underbrace{x_7,x_8,x_9}_{X_3^{n=3}},x_{10},x_{11},.....$$

長さ *n* のブロックのエントロピー（*n* 次エントロピー）をつぎのように定義する．

$$H(X^n) = -\sum_{X^n} P(X^n)\log_2 P(X^n) \tag{3.1}$$

問題 2-4 では，2 次のエントロピーを考えたことになることに注意する．

情報源の（一文字あたりの）エントロピーは，次のように定義される．

$$H(X) = \lim_{n\to\infty}\frac{1}{n}H(X^n) \tag{3.2}$$

これは，エントロピーレート (entropy rate)とも呼ばれる．ただし，この *H*(*X*)は実際には求まらないので，近似が必要である．

マルコフ情報源 (Markov information source) : 現在の記号の出現確率が，有限の過去（直前の *m* ヶの記号）にのみ依存するとするとき，これを *m* 重マルコフ情報源という

記号系列確率のマルコフモデルによる近似： 情報源記号列 $X^n = x_1x_2x_3x_4x_5x_6\cdots x_n$ を考える．

$$P(X^n) = p(x_1x_2x_3x_4x_5x_6\cdots x_n) \tag{3.3}$$

これを，以下のような方法で近似する．

0 重マルコフモデルの場合（各 x_i が互いに独立）：

$$P(X^n) = p(x_1)p(x_2)p(x_3)p(x_4)p(x_5)p(x_6)\cdots p(x_n) \tag{3.4}$$

これは，$P(x_i) = p(x_i \mid x_1 \cdots x_{i-1})$ ということを意味することでもあり，このような情報源を記憶のない（独立生起）情報源(memoryless source)ということがある．

なお，$x_1,...,x_n$ が互いに独立に同一の分布に従うとき，i.i.d. (independently identically distributed) な情報源という．

1 重マルコフモデル（単純マルコフモデル） (simple Markov model)：

$$P(X^n) = p(x_1)p(x_2 \mid x_1)p(x_3 \mid x_2)p(x_4 \mid x_3)\cdots p(x_n \mid x_{n-1}) \tag{3.5}$$

1 重マルコフモデルでは，一つ前の状態にのみ依存して現在の出現確率が決まる．

2 重マルコフモデル (2nd-order Markov model)：

$$P(X^n) = p(x_1)p(x_2 \mid x_1)p(x_3 \mid x_1x_2)p(x_4 \mid x_2x_3)\cdots p(x_n \mid x_{n-2}x_{n-1}) \tag{3.6}$$

なお，$m > 2$ の m 重以上のマルコフ情報源については，長さ m の記号集合をつくることによって，1 重マルコフモデルで表すことができる（例 3-3 参照）．

マルコフモデルによるエントロピーレートの計算

記憶のない情報源 X におけるエントロピーレート $H(X)$は，どのようなものであるだろうか．結合エントロピーの性質(2.14)式においてすべての確率変数が独立である場合により，ｎ次エントロピーは次のようになる．

$$\begin{aligned} H(X^n) &= H(X_1 \cdots X_n) \\ &= H(X_1) + \cdots + H(X_n) \\ &= nH(X^1) \end{aligned} \tag{3.7}$$

したがって，

$$H(X) = \lim_{n\to\infty} \frac{1}{n} H(X^n) = \lim_{n\to\infty} \frac{1}{n} nH(X^1) = H(X^1) \tag{3.8}$$

要するに，記憶のない情報源のエントロピーレートは，1 回限りの試行が持つエントロピーそのものである．これは，前後の記号の出現に依存しない情報源なのであるから，ごく自然な結果であるといえる．

他方，記憶のある情報源の場合はどうだろうか．この場合，エントロピーのチェイン則と，$H(X \mid Y) \le H(X)$ （問題 2-10 参照）により，ｎ次エントロピーは次に示すように

$nH(X^1) \geq H(X^n)$である．

$$\begin{aligned} H(X^n) &= H(X_1 \cdots X_n) \\ &= H(X_1) + H(X_2 \mid X_1) + \cdots + H(X_n \mid X_1 \cdots X_{n-1}) \\ &\leq nH(X^1) \end{aligned} \tag{3.9}$$

では，記憶のある情報源の場合，エントロピーレートはどのような形となるだろうか．単純マルコフの場合，$H(X^n) = H(X_1 \cdots X_n) = H(X_1) + H(X_2 \mid X_1) \cdots + H(X_n \mid X_{n-1})$であることから，次式が成り立つ．

$$H(X) = \lim_{n\to\infty} \frac{1}{n} H(X^n) = \lim_{n\to\infty} \frac{1}{n} \big(H(X_1) + (n-1)H(X_i \mid X_{i-1})\big) = H(X_i \mid X_{i-1}) \tag{3.10}$$

つまり，単純マルコフ性を持つ情報源のエントロピーレートは，そのマルコフ性に由来する，直前に発せられた記号とその次に発生する記号との間の条件付きエントロピーである．これも，記憶のない情報源と同様，納得しやすい結果であるといえる．ただし，この条件付きエントロピーを具体的にどう計算したらよいのかはまだ見通しが得られていない．条件付きエントロピーを計算するには，

$$H(X_i|X_{i-1}) = \sum_{X_{i-1}} P(X_{i-1} = x_{i-1}) H(X_i|X_{i-1} = x_{i-1})$$

における$P(X_{i-1} = x_{i-1})$の値を必要とするのであるが，これは時刻に依らないで，X_{i-1}がある特定の実現値を取る確率を意味するものであり，それが具体的にいくらなのかということはマルコフモデルでは与えられていない．そこで，この確率をどうやって求めたらよいかについて考えなければならない．

時刻によらない$P(X_{i-1} = x_{i-1})$は，定常確率と呼ばれる．定常確率を求める方法を考えよう．このことを見通しよく検討していくために，状態遷移図という考え方を導入する．

状態遷移図（state transition diagram），シャノン線図： 次の例で説明する．

例 3-2: $p(w \mid w) = \frac{3}{4}, p(b \mid w) = \frac{1}{4}, p(w \mid b) = \frac{1}{4}, p(b \mid b) = \frac{3}{4}$ となるような 1 重マルコフ情報源を考える．これを，次の図のように表す．○が状態をあらわし，エッジの上の *a*/*p* は，*a* という記号が確率 *p* で発生することを表す．

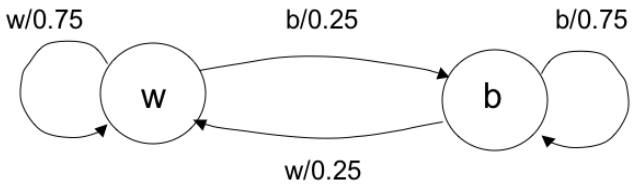


図 3.1 状態遷移図の例

例 3-3: $p(w \mid ww) = 0.2, p(w \mid wb) = 0.6, p(w \mid bw) = 0.5, p(w \mid bb) = 0.9$ であるような 2 重マルコフ情報源を考える．この場合，次のようにすると，あたかも 1 重マルコフモデルのように扱えることがわかるだろう．

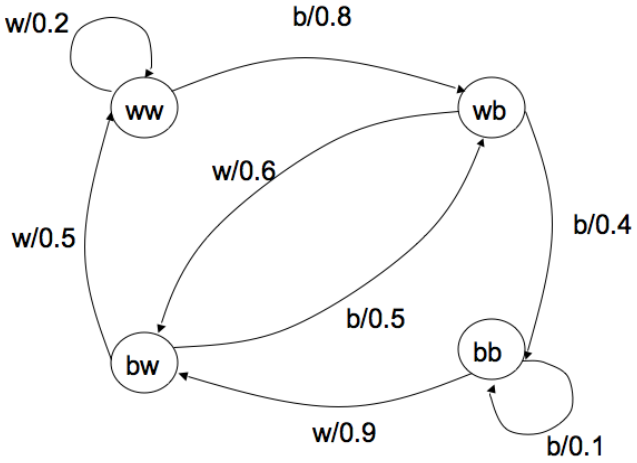


図 3.2　２重マルコフ情報源の１重マルコフ的表現

遷移確率 (transition probability)：

マルコフモデルにおいて，状態 $s_i(i=1,...,M)$ にある確率を知りたい．$s_i \to s_j$ への遷移確率を p_{ij} とかくとき，次の行列を遷移行列という．

$$\Pi = \begin{bmatrix} p_{11} & \cdots & p_{1M} \\ \vdots & \ddots & \vdots \\ p_{M1} & \cdots & p_{MM} \end{bmatrix} \tag{3.11}$$

例 3-2 における遷移行列は，つぎのようになる．

$$\Pi = \begin{bmatrix} \frac{3}{4} & \frac{1}{4} \\ \frac{1}{4} & \frac{3}{4} \end{bmatrix}$$

ある時刻において，状態 $s_i(i=1,...,M)$ にある確率を w_i とする．次の時刻において，状態 s_i にある確率 $w_i^{'}$ をかんがえると，つぎのようになる．

$$\begin{aligned} w_i &= w_1 p_{1i} + w_2 p_{2i} + w_3 p_{3i} \cdots w_M p_{Mi} \\ &= \begin{bmatrix} w_1 & \cdots & w_M \end{bmatrix} \begin{bmatrix} p_{1i} \\ \vdots \\ p_{Mi} \end{bmatrix} \end{aligned} \tag{3.12}$$

このことから，次の時刻における状態は，

$$
\begin{aligned}
\mathbf{w}' &= \begin{bmatrix} w_1' & \cdots & w_M' \end{bmatrix} \\
&= \begin{bmatrix} w_1 & \cdots & w_M \end{bmatrix} \begin{bmatrix} p_{11} & \cdots & p_{1M} \\ \vdots & \ddots & \vdots \\ p_{M1} & \cdots & p_{MM} \end{bmatrix} \\
&= \mathbf{w}\Pi
\end{aligned}
\tag{3.13}
$$

さらに，次の時刻では $\mathbf{w}'' = \mathbf{w}'\Pi = \mathbf{w}\Pi^2$，n 回の状態遷移によって，$\mathbf{w}\Pi^n$ となる．はじめのころは，状態遷移の前後で，状態 $s_i(i=1,...,M)$ のとる確率はことなるであろうが，何度も繰り返していけばいずれ落ち着くことが考えられる．実際，後で述べるエルゴード性を持つマルコフモデル(simple Markov model)では，Π^n がある行列 Π^∞ に収束する．この Π^∞ は，$\mathbf{w}^\infty = \mathbf{w}^\infty\Pi$ を満たすベクトル $\mathbf{w}^\infty$ を M 個並べたものとなる（$\Pi^\infty = [\mathbf{w}^\infty, \mathbf{w}^\infty, \cdots \mathbf{w}^\infty]^T$）ことが知られている．この $\mathbf{w}^\infty$ こそが，定常確率である．

定常確率を改めて $\mathbf{w}$ と書くと，$\mathbf{w}$ は，次の連立方程式の解として求まる．

$$
\begin{cases} \mathbf{w} = \mathbf{w}\Pi \\ \sum_{i=1}^{M} w_i = 1 \end{cases}
\tag{3.14}
$$

例 3-4: つぎのような状態遷移図をもつマルコフモデルの定常確率を求めよう．

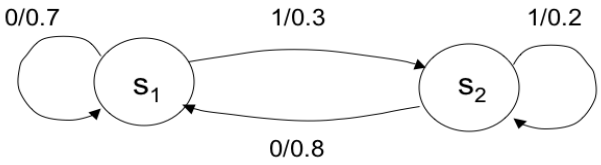


図 3.3 定常確率を持つマルコフモデルの例

$\mathbf{w} = \begin{bmatrix} w_1, w_2 \end{bmatrix}, \Pi = \begin{bmatrix} 0.7 & 0.3 \\ 0.8 & 0.2 \end{bmatrix}$ とおき，つぎの連立方程式の解を求める．

$$
\begin{cases} \mathbf{w} = \mathbf{w}\Pi \\ w_1 + w_2 = 1 \end{cases}
\tag{3.15}
$$

上式を書き下してみると，

$$
\begin{cases} w_1 = 0.7w_1 + 0.8w_2 \\ w_2 = 0.3w_1 + 0.2w_2 \\ w_1 + w_2 = 1 \end{cases}
$$

(一つ目と二つ目の式は従属であることに注意)

これを解くと，

$$
\mathbf{w} = \begin{bmatrix} \frac{8}{11}, \frac{3}{11} \end{bmatrix}
$$

定常確率が存在する条件＝エルゴード性 : エルゴード性とは，次の条件を満たすことをいう．

(1) 既約である（すべての状態間に遷移するパスが存在する）
(2) 周期的でない(周期的であるとは，ある状態からぐるっと回って元に戻るループを考えたとき，各ループのエッジの本数に公約数があるものをいう．詳しくは，甘利(1970)を参照のこと)

例 3-5: 既約でないもの

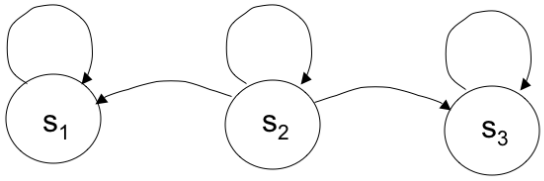


図 3.4 既約でない状態遷移図の例

この例では，S3 から S1 へのパスが存在しない．このような例では，落ち着き先がどこになるかが初期状態に依存する．

例 3-6: 周期的なもの（周期的である場合，Π^nが収束しない）

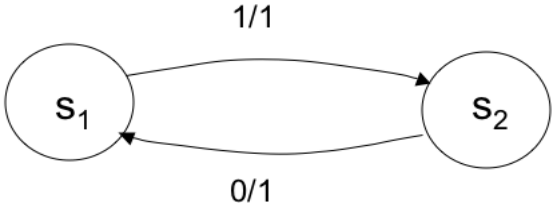


図 3.5 周期的な状態遷移図の例

この例では，(3.14)式の連立方程式は解けてしまうし，直観的にも定常確率が存在しそうに見えなくもない．この例は，本章で最初に示した病的な例（HTHTHTHTH・・・をくりかえすもの）であり，奇数回目，偶数回目にどちらの状態にあるかが確定している．このため，「時刻に依らずに」状態 i にある確率が定まらない．このようなものは定常確率が存在するとは言えないことになる．

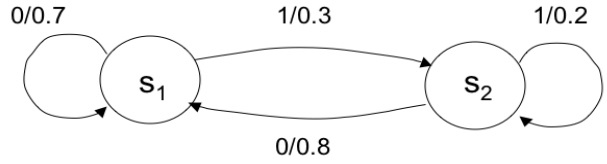
マルコフモデルのエントロピーレートの具体的な計算

マルコフモデル X において，情報源のエントロピーレートを求めてみよう．(3.10)で見たように，$H(X)=H(X_i \mid X_{i-1})$ であった．ここで，一般に，条件付きエントロピー $H(X \mid Y)$ は，次のように定義されることを思い出そう．

$$H(X \mid Y) = \sum_{Y=y} P(Y=y) H(X \mid Y=y) \tag{3.16}$$

この場合エントロピーレートは，条件付きエントロピーとして求まる．具体的に，マルコフ情報源のエントロピーレートの求め方を，次の例を用いて説明する．

例 3-7:　つぎの状態遷移図



で表現されるマルコフモデルは，定常確率がつぎのようであった．

$$\mathbf{w} = \left[\tfrac{8}{11}, \tfrac{3}{11}\right]$$

ここで，「ある特定の状態」にあるときに，その状態でどの程度のエントロピーがあるかを求めてみよう．状態 s_1 では 0,1 が出る確率がそれぞれ 0.7, 0.3 だから $p(0|s_1)=0.7, p(1|s_2)=0.3$ と書くことができるので，　s_1 におけるエントロピー $H(X|s_1)$ $H(X|S_1)$はつぎのようになる．

$$\begin{aligned} H(X|s_1) &= -p(0|s_1)\log_2 p(0|s_1) - p(1|s_1)\log_2 p(1|s_1) \\ &= -0.7\log_2 0.7 - 0.3\log_2 0.3 \\ &= 0.881[bit] \end{aligned}$$

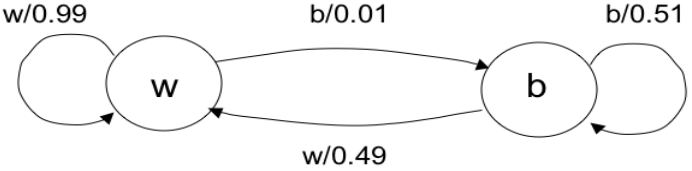
同様に，

$$\begin{aligned} H(X|s_2) &= -p(0|s_2)\log_2 p(0|s_2) - p(1|s_2)\log_2 p(1|s_2) \\ &= -0.8\log_2 0.8 - 0.2\log_2 0.2 \\ &= 0.722[bit] \end{aligned}$$

全体としてのエントロピーレートは，次のように求めることができる．

$$\begin{aligned} H(X) &= p(s_1)H(X|s_1) + P(s_2)H(X|s_2) \\ &= \frac{8}{11}\cdot 0.881 + \frac{3}{11}\cdot 0.722 \\ &= 0.838[bit] \end{aligned}$$

問題 3-1: 以下のマルコフ情報源 M を考える．このときつぎの問いに答えよ．



(1) この状態遷移図に対応する遷移行列を書け
(2) このマルコフモデルの定常確率を求めよ
(3) この情報源のエントロピーレートを求めよ．

問題 3-2: エルゴード性の成り立たない情報源を一つ考えてみよ.

4. 情報源符号化定理（シャノンの第一定理）

情報源符号化定理（シャノンの第一定理）

ブロックの長さnを十分大きくとれば，1 記号あたりの平均符号語長$\frac{\overline{L}}{n}$を，エントロピーレート$H(X)$へいくらでもちかづけることができる．すなわち，任意の$\varepsilon > 0$に対し，

$$H(X) \le \frac{\overline{L}}{n} < H(X) + \varepsilon \tag{4.1}$$

以下では，この情報源符号化定理の証明を示す．ただし，本稿では，二元符号の場合に限定して議論する．

まず，つぎの補題を示す．

補題： 独立生起情報源U（ここでは記号が 1 つだけ出てくるものとする）に対し，次式を満たすような符号語長の組を持つ瞬時復号可能な符号が存在する．

$$H(U) \le \overline{L} < H(U) + 1 \tag{4.2}$$

補題の証明．

最初に，$\overline{L}$の下限について，$H(U) \le \overline{L}$を示す．記号の種類をMとし，i番目の記号の出現確率をp_i，符号語長をl_iとする．Uのエントロピー$H(U)$と平均符号語長$\overline{L}$はつぎのようになる．

$$H(U) = \sum_{i=1}^{M} p_i \log_2 \frac{1}{p_i}, \quad \overline{L} = \sum_{i=1}^{M} p_i l_i \tag{4.3}$$

したがって，次を示すことができればよいことになる．

$$H(U) - \overline{L} = \sum_{i=1}^{M} p_i \log_2 \frac{1}{p_i} - \sum_{i=1}^{M} p_i l_i \le 0 \tag{4.4}$$

いま，考察の対象としている符号がクラフトの不等式を満たすものとする．すなわち，

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

ここで，$q_i = 2^{-l_i}$とおくと，$l_i = \log_2 \frac{1}{q_i}$であるから，

$$
\begin{aligned}
H(U)-\overline{L} &= \sum_{i=1}^{M} p_i \log_2 \frac{1}{p_i} - \sum_{i=1}^{M} p_i \log_2 \frac{1}{q_i} \\
&= \sum_{i=1}^{M} p_i \log_2 \frac{q_i}{p_i} \\
&\le (\log_2 e) \sum_{i=1}^{M} p_i \left(\frac{q_i}{p_i} - 1 \right) \\
&= (\log_2 e) \sum_{i=1}^{M} (q_i - p_i) \\
&= (\log_2 e) \left(\sum_{i=1}^{M} q_i - \sum_{i=1}^{M} p_i \right) \le 0
\end{aligned}
\tag{4.5}
$$

つぎに，$\overline{L} < H(U)+1$を示す．$\overline{L}$の下限についての証明から，$p_i = q_i = 2^{-l_i}$のとき$H(U) = \overline{L}$を満たし，このとき$l_i = \log_2 \frac{1}{p_i}$である．しかし符号長は整数でなければならないから，つぎの関係を満たす整数l_i^*を符号長としよう．

$$
\log_2 \tfrac{1}{p_i} \le l_i^* < \log_2 \tfrac{1}{p_i} + 1 \tag{4.6}
$$

このとき，$l_i = \log_2 \frac{1}{p_i} \le l_i^*$であることから，$2^{-l_i^*} \le p_i$である．これをすべての$i$について足し合わせると，

$$
\sum_{i=1}^{M} 2^{-l_i^*} \le \sum_{i=1}^{M} p_i = 1 \tag{4.7}
$$

上式は，$\{l_1^*,...,l_M^*\}$がクラフトの不等式を満たすことを意味しているから，$\{l_1^*,...,l_M^*\}$で瞬時復号可能な符号をつくることができる．さらに，

$$
l_i^* < \log_2 \tfrac{1}{p_i} + 1 \tag{4.8}
$$

に対し，両辺にp_iを乗じ，すべてのiについて足し合わせると，

$$
\overline{L} = l_1^* p_1 + \cdots l_M^* p_M < \sum_{i=1}^{M} p_i \log_2 \tfrac{1}{p_i} + \sum_{i=1}^{M} p_i = H(U) + 1 \tag{4.9}
$$

すなわち，

$$
\overline{L} < H(U) + 1 \tag{4.10}
$$

（補題証明終わり）

上の補題を用いて，シャノンの第一定理を示そう．

まず，記憶のない情報源 X の場合を考える．ブロックの長さを n とするとき，ブロック化された新しい記号X^nのエントロピーは$H(X^n) = nH(X^1) = nH(X)$であることから，上の補題に$H(X^n)$を代入すると，次式をえる．

$$nH(X) \le \bar{L} < nH(X) + 1 \tag{4.11}$$

両辺を n で割れば,

$$\frac{nH(X)}{n} \le \frac{\bar{L}}{n} < \frac{nH(X)}{n} + \frac{1}{n} \tag{4.12}$$

すなわち,

$$H(X) \le \frac{\bar{L}}{n} < H(X) + \frac{1}{n} \tag{4.13}$$

ここで，n はいくらでも大きくできるので，定理が証明できたことになる．

つぎに，マルコフ情報源の場合を考えよう．$H(X^n)$を補題の不等式に代入して(注：定常確率が存在するなら，補題は適用可能)両辺を n で割ると，次式をえる．

$$\frac{H(X^n)}{n} \le \frac{\bar{L}}{n} < \frac{H(X^n)}{n} + \frac{1}{n} \tag{4.14}$$

ここで，$H(X) = \lim_{n\to\infty} \frac{1}{n} H(X^n)$に注意して，n の極限をとると、次式を得る．

$$H(X) \le \frac{\bar{L}}{n} < H(X) + \frac{1}{n} \tag{4.15}$$

（定理証明終わり）

(4.6)式で符号語長を与えた語頭符号は，シャノン・ファノ符号と呼ばれている．

ちなみに，2 元符号ではなく，r 元符号の場合は，情報源符号化第一定理は次の形となるが，本質的には 2 元符号の場合と違いはない．

$$\frac{H(X)}{\log_2 r} \le \frac{\bar{L}}{n} < \frac{H(X)}{\log_2 r} + \frac{1}{n} \tag{4.16}$$

補足：

ブロック長を長くとると，一般には符号化が困難になる．M 種類の記号を持つ情報源を考えよう．ブロック長を n とすると，その結果あらたに作られる記号の種類は M^n になる．英数字（若干の記号も含め，全１００種類あるとしよう）を符号化するとき，ブロック長を 10 にするだけで，記号の種類は$100^{10} = 100{,}000{,}000{,}000{,}000{,}000{,}000$にものぼる．

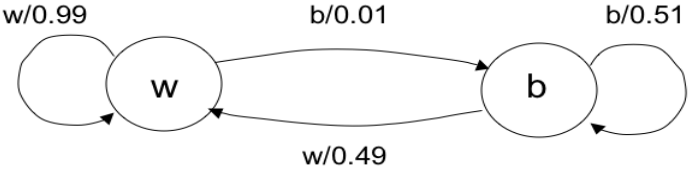
したがって，実際には，ブロック長を大きくとることはできない．

問題 4-1: 記号{0,1}を出力する記憶のない情報源 A を考える．記号の出現確率は p(0)=2/3, p(1)=1/3 であり，情報源 A のエントロピーは H(X)=0.918 [bit]である．このとき，次の問いに答えよ．

(1) 情報源 A をブロック符号化せずに最短 2 元符号化し，平均符号語長を求めよ．このとき，補題の式を満たしているかどうか確認せよ．

(2) 情報源 A をブロック符号化した場合（n=2,3）の最短二元符号を求め，平均符号語長を求めよ．

問題 4-2: つぎのマルコフ情報源 M を考える．



(1) ブロック符号を最短符号化し（n=1,2,3），平均符号語長を計算せよ．

(2) 先に求めた H(X)を用いて，情報源符号化定理が成立していることを確認せよ．

5. 通信路符号化 Channel coding

通信路(channel)：データの送信（「記憶」として考えてもいい）のための通る通路（記憶媒体）．多くの場合，「雑音」によって，データに変化が生じうる．

雑音(noise)：通信路の入力に加わりうる変化

通信路符号化の目的：通信路に雑音があっても正しく送信できるようにする．通信路を通る符号語を，通信路符号語とよぶ．

パリティ(parity)：
偶数パリティ：符号語の１の数を偶数個になるようにすること（０の数としても同じこと）
奇数パリティ：符号語の１の数を奇数個になるようにすること

パリティ検査 (parity check)：符号語中の１の数を偶（奇）数個に固定する．そのために，パリティ検査ビットとして，何ビットか本来の符号語に付加する．このことによって，送信された符号語の１の数が奇（偶）数個となっていたら，送信の途中でどこかに誤りが生じたということがわかる．

パリティ検査の方法
送信データを，$X = x_1x_2x_3x_4x_5x_6\cdots x_k$とする．パリティ検査用の冗長ビット$x_{k+1}$を追加することによって，通信路符号語を$X = x_1x_2x_3x_4x_5x_6\cdots x_kx_{k+1}$のようにする．送信側での$x_{k+1}$の値の決定や，受信側での検査は，排他的論理和を用いる．
排他的論理和とは，２を法とする演算で，次の性質を持つ．

$$\begin{aligned} &0 \oplus 0 = 0,\ \ 1 \oplus 1 = 0 \\ &0 \oplus 1 = 1,\ \ \ 1 \oplus 0 = 1 \end{aligned} \tag{5.1}$$

送信側でx_{k+1}の値は次のように定める．

$$x_{k+1} = x_1 \oplus x_2 \oplus \cdots \oplus x_k \tag{5.2}$$

そのようなx_{k+1}は，つねに$x_1 \oplus x_2 \oplus \cdots \oplus x_k \oplus x_{k+1} = 0$を満たす．
受信側では，受け取った記号列$X = x_1x_2x_3x_4x_5x_6\cdots x_kx_{k+1}$から，次のパリティ検査方程式

$$S = x_1 \oplus x_2 \oplus \cdots \oplus x_k \oplus x_{k+1} \tag{5.3}$$

を計算し，$S = 0$なら検査合格，$S = 1$なら誤り発生，と判断する（ただし，ここでは，誤るとしても高々１文字まで，と仮定した場合）．

例 5-1: 通信路に送ろうとする情報記号が 00, 01, 10, 11 のとき，偶数パリティとして検査ビットを次のように加える．

$$00 \ \rightarrow\ 000,01 \ \rightarrow\ 011,10 \ \rightarrow 101,11 \ \rightarrow\ 110$$

送信データは３つの記号からなるので，パリティ検査方程式は

$$S = x_1 \oplus x_2 \oplus x_3 \tag{5.4}$$

となる．

検査と訂正:

(１ビットの)パリティ検査では,（どこかに）誤りがあるということまではわかっても，どこが誤っているかはわからない．したがって，訂正はできない．

たとえば，つぎのような「受信空間」を考えよう．送信する情報は，000, 011, 101, 110 のうちのいずれかである．通信路において，高々1 ビットの誤りが発生するとしよう．もし，100 という記号列をうけとったとき，送信すべき情報のリストにないので，誤りがあるということまではわかるが，もとの符号語は，110, 000, 101 のいずれでもありうる．

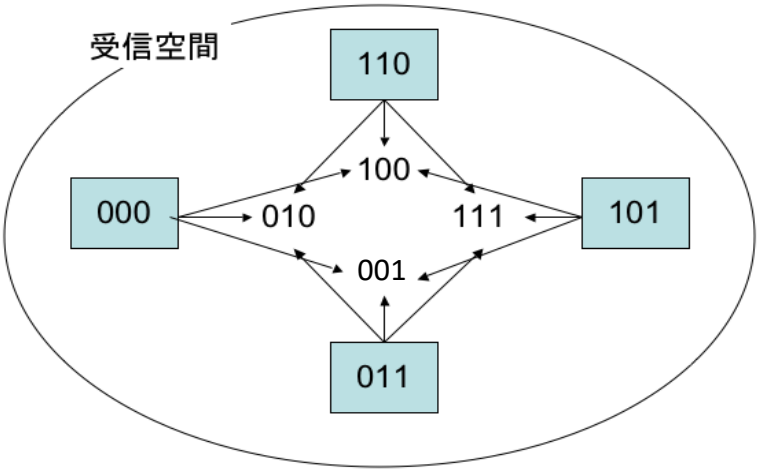


図 5.1 受信空間

誤りの訂正をできるようにするにはどうしたらよいだろうか．例えば，同じ記号を３回繰り返し送信する，というような方法がありうる．すなわち，0 を送りたいときは 000, 1 を送りたいときは 111 とする．この場合，1 符号語あたり 1 ビット以下でしか誤りが発生しないとすれば，000 から誤りうるパターンの集合

$$\{001, 010, 100\}$$

と，111 から誤りうるパターンの集合

$$\{110, 101, 011\}$$

とでは，共通部分が空であることから，誤りを検出し，さらに訂正まで可能である．ただし，情報の伝送の効率は悪い（１つの記号を送るのに３ビットを要している）．

符号のレート（情報伝送速度）：記号の種類を M とし，n ビットの記号列で送信するとき，次で表される量 R を符号のレート（あるいは，情報伝送速度）という．

$$R = \frac{\log_2 M}{n} \tag{5.5}$$

上の例では，符号のレートは　0 もしくは 1 という 2 種類の記号を３ビットで送信するので，$\frac{\log_2 2}{3} = \frac{1}{3}$ である．

符号の最小距離:
すでに見たように，符号語同士が著しく似ていなければ，少しの誤りがあっても訂正可能である．どれくらい誤りが発生したら他の符号語と一致してしまうかを，符号語間の距離を定義することによって論ずる．

ハミング距離(Hamming distance): 通信路符号語 $X = x_1 x_2 \cdots x_n$ と $Y = y_1 y_2 \cdots y_n$ のハミング距離 $hamming(X,Y)$ を，次のように定義する．

$$hamming(X,Y) = \sum_{i=1}^{n} \delta(x_i, y_i) \tag{5.6}$$

ただし，

$$\delta(x,y) = \begin{cases} 0 & \text{if } x = y \\ 1 & \text{if } x \neq y \end{cases} \tag{5.7}$$

なお，2 元符号の場合は，$\delta(x,y) = x \oplus y$（排他的論理和）で実現できることに注意せよ．

例 5-2:

$X = 010101, Y = 010100$ のとき，$hamming(X,Y) = 1$
$X = 001, Y = 010$ のとき，$hamming(X,Y) = 2$
$X = 000, Y = 000$ のとき，$hamming(X,Y) = 0$

二元符号の空間を考える．たとえば，長さ３の二元符号 $X = x_1 x_2 x_3$ の空間を，次の図のように表現しよう．

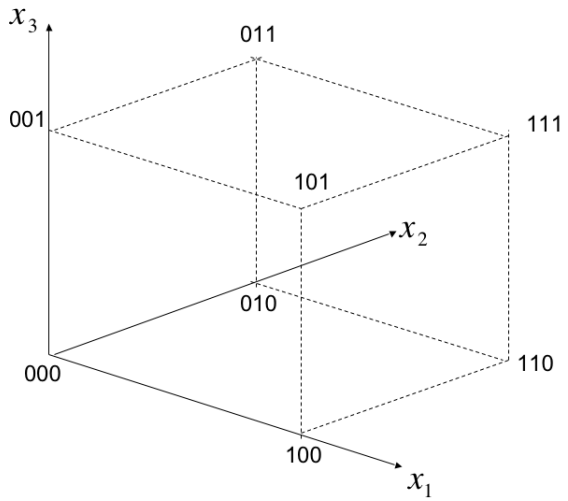


図 5.2　二元符号の空間

ハミング球 : ハミング距離で距離が測られる空間において，中心からハミング距離 *r* 以内の点をすべて集めた集合を，ハミング球とよぶ.

例 5-3: 長さ３の二元符号の空間における，中心 000 ,半径 1 のハミング球C_1は下図左のようになる．

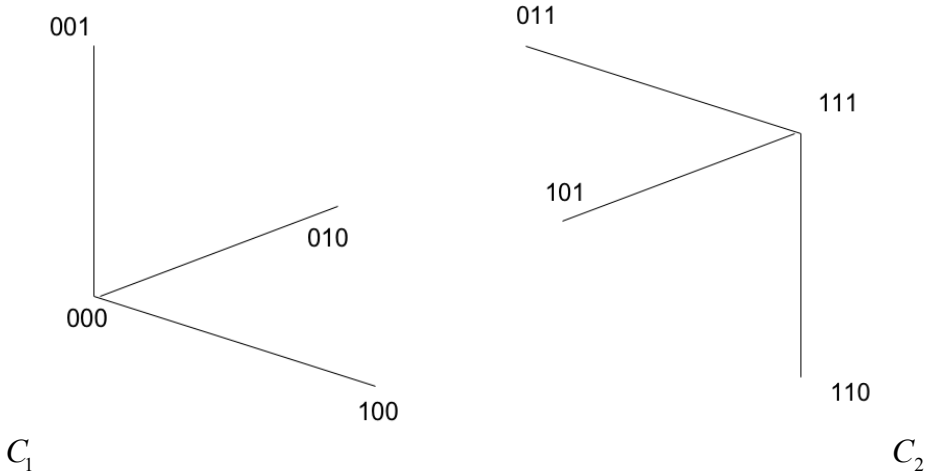


図 5.3　ハミング球

同様に，長さ３の二元符号の空間における，中心 111 ,半径 1 のハミング球C_2は上図右のようになる．両者は排反（共通部分を持たない）であることに注意せよ．

排反な２つのハミング球があるとき，それぞれの中心を符号語として考えると，ハミング球の半径以内の誤りならば訂正が可能である．

最小距離（minimum distance）： 通信路符号$\{X_1,...,X_p\}$（符号語の集合のこと）について，この通信路符号の最小距離 *d* を次のように定義する．

$$d = \min_{i \neq j} hamming(X_i, X_j) \tag{5.7}$$

すなわち，*d* は，符号語間のハミング距離の最小値を意味する．

例 5-4: 最小距離が２であるような符号の例を示す．

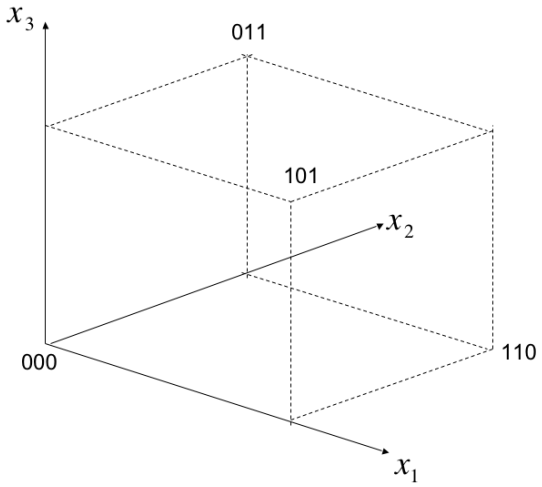


図 5.4　最小距離が 2 である符号の例

誤り検知と訂正:

最小距離 d が２:1 ビット誤っても他の符号語にはならないので, 1 ビットの誤り検出は可能. ただし，誤りを共有しうるので，訂正はできない.

最小距離 d が３：2 ビットまで誤っても他の符号にはならないので，2 ビットまでの誤り検出が可能．1 ビットまでの誤りなら，訂正が可能.

最小距離 d が４：３ビットまで誤りの検出が可能．1 ビットの誤りは訂正できる.

より一般には,

最小距離 d が s+1 以上なら，s 重誤り（s ビットの誤り）の検知が可能.

最小距離 d が，2t+1 以上なら，t 重誤り（t ビットの誤り）の訂正が可能.

問題 5-1: 送信する情報が４種類あるとする．このとき，符号のレート（情報伝送速度）が 1/3 よりも大きくなるように，1 重誤り訂正符号を作成せよ.

水平垂直パリティ検査符号 : 複数のパリティ検査ビットを加えて, 誤り位置がわかるようにしたい（誤り訂正）.

例 5-5: ４ビットの情報記号： $x_1x_2x_3x_4$ に対し，４ビットの検査記号： $c_1c_2c_3c_4$ とする
ここで， $c_1 = x_1 \oplus x_2, c_2 = x_3 \oplus x_4, c_3 = x_1 \oplus x_3, c_4 = x_2 \oplus x_4$ とする．検査式は，次の通りである.

$$S_1 = x_1 \oplus x_2 \oplus c_1, S_2 = x_3 \oplus x_4 \oplus c_2, S_3 = x_1 \oplus x_3 \oplus c_3, S_4 = x_2 \oplus x_4 \oplus c_4$$

検査の結果,

- $S_1 = S_2 = S_3 = S_4 = 0$ なら，誤りなしと見なしていい
- S_1, S_2, S_3, S_4 のいずれか一つのみが 1 なら，対応する c_i に誤りが生じている
- S_1, S_2, S_3, S_4 のうち２つが 1 なら，つぎの所に誤りがある
  - $S_1, S_3 : x_1$
  - $S_1, S_4 : x_2$
  - $S_2, S_3 : x_3$
  - $S_2, S_4 : x_4$
  - 上記以外，どこかに間違いがあるがどれかはわからない

たとえば，0011 という記号列を送信したいものとしよう．このとき，検査記号は 0011 となるとなるので，送られるのは 00110011 となる．受信側で，00010011 という記号列を受け取ったとしよう．このとき,

$$S_1 = x_1 \oplus x_2 \oplus c_1 = 0, S_2 = x_3 \oplus x_4 \oplus c_2 = 1, S_3 = x_1 \oplus x_3 \oplus c_3 = 1, S_4 = x_2 \oplus x_4 \oplus c_4 = 0$$

となる．したがって， $S_2 = S_3 = 1$ となっていることから， x_3 に誤りが生じているということがわかる.

上記のような水平垂直パリティ符号は，元々送りたい記号は４ビットであるのに，検査記号に４ビットも使っていて，効率が悪い．効率がよく，誤りの訂正が可能な符号はどのようにして作ることができるだろうか．その答えの一つが，ハミング符号と呼ばれるものであるが，ハミング符号を説明するために，まず，パリティ検査ブロック符号というものを考える.

パリティ検査ブロック符号：

最初に，上の例について，送信前の符号語には下記が成立するので，これを検査方程式として利用する．

$$\begin{array}{ccccccccccccc} x_1 & \oplus & x_2 & & & & & \oplus & c_1 & & & & = 0 \\ & & & & x_3 & \oplus & x_4 & & & \oplus & c_2 & & = 0 \\ x_1 & & & & x_3 & & & & & & & \oplus & c_3 & = 0 \\ & & x_2 & & & \oplus & x_4 & & & & & \oplus & c_4 & = 0 \end{array} \tag{5.8}$$

一般に，検査方程式は，次のような形で表される．

$$\begin{array}{ccccccccccccc} h_{11}x_1 & \oplus & h_{12}x_2 & \oplus & h_{13}x_3 & \oplus & \cdots & \oplus & h_{1n}x_n & = & 0 \\ h_{21}x_1 & \oplus & h_{22}x_2 & \oplus & h_{23}x_3 & \oplus & \cdots & \oplus & h_{2n}x_n & = & 0 \\ \vdots & & \vdots & & \vdots & & \vdots & & \vdots & \vdots & \vdots \\ h_{m1}x_1 & \oplus & h_{m2}x_2 & \oplus & h_{m3}x_3 & \oplus & \cdots & \oplus & h_{mn}x_m & = & 0 \end{array} \tag{5.9}$$

ここで，$h_{ij} \in \{0,1\}$．これを行列表記すると，つぎのようにできる．

$$H\mathbf{x} = O \tag{5.10}$$

ただし，H(パリティ検査行列)，$\mathbf{x}$（通信路符号語）は次のように表される．[1]

$$, \quad H = \begin{bmatrix} h_{11} & h_{12} & h_{13} & \cdots & h_{1n} \\ h_{21} & h_{22} & h_{23} & \cdots & h_{2n} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ h_{m1} & h_{m2} & h_{m3} & \cdots & h_{mn} \end{bmatrix} \quad \mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} \qquad O = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix} \tag{5.11}$$

受信した記号列 $\mathbf{y}$ について $H\mathbf{y} = O$ がみたされれば問題ないが，そうでない場合はどう考えればよいか．これには，シンドローム(syndrome)という考え方を利用する．受信符号語 $\mathbf{y}$ について，

$$H\mathbf{y} = \mathbf{s} = \begin{bmatrix} s_1 \\ s_2 \\ \vdots \\ s_m \end{bmatrix} \tag{5.12}$$

をシンドロームという．このシンドローム $\mathbf{s}$ が持つ性質を調べてみよう．いま，

---

[1] マルコフ性の議論では，符号語を確率変数の積の形であらわしていた．ここでは，説明の都合上，符号語をベクトルとして表現する．このように，議論の内容に応じて，符号語の表現の仕方はいくつかの方式がある．

$$\mathbf{e} = \begin{bmatrix} e_1 \\ e_2 \\ \vdots \\ e_n \end{bmatrix}, \quad e_i \in \{0,1\} \tag{5.13}$$

を, 誤りが発生する箇所が 1, そうでない箇所が 0 となるようなベクトルとして考える. これを, 誤りベクトルとか, 誤りパターンなどとよぶ. 受信符号語 $\mathbf{y}$ は, もとの通信路符号語 $\mathbf{x}$ とこの誤りベクトル $\mathbf{e}$ によって, $\mathbf{y} = \mathbf{x} \oplus \mathbf{e}$ として得られることになる. ここで, シンドローム $\mathbf{s}$ は,

$$\begin{aligned} \mathbf{s} = H\mathbf{y} = H(\mathbf{x} \oplus \mathbf{e}) &= H\mathbf{x} \oplus H\mathbf{e} \\ &= O \oplus H\mathbf{e} \\ &= H\mathbf{e} \end{aligned} \tag{5.14}$$

すなわち, シンドローム $\mathbf{s}$ は,「送信された符号語が何であるかによらず」誤りパターン $\mathbf{e}$ できまってしまう. したがって, 個々の誤りパターン $\mathbf{e}$ と, シンドローム $\mathbf{s} = H\mathbf{e}$ との関係をあらかじめ求めておけば, どのような誤り $\mathbf{e}$ が受信符号語 $\mathbf{y}$ に含まれているかがシンドロームによってわかってしまうことになる. 誤りパターン $\mathbf{e}$ がわかれば,

$$\begin{aligned} \mathbf{y} \oplus \mathbf{e} = (\mathbf{x} \oplus \mathbf{e}) \oplus \mathbf{e} &= \mathbf{x} \oplus (\mathbf{e} \oplus \mathbf{e}) \\ &= \mathbf{x} \oplus O \\ &= \mathbf{x} \end{aligned} \tag{5.15}$$

によって, 正しい符号語 $\mathbf{x}$ に戻すことができる. これをシンドローム復号という.

以前にみた水平垂直パリティ符号のシンドロームについて考えよう. 情報記号４桁, 検査記号４桁であるから, 符号語長は８となる. 高々１つの誤りが発生しうるとすると, 誤りなしもふくめて誤りパターンは９種類あることになる. ところで, シンドロームは４桁であることから, ２の４乗, すなわち１６種類の誤りパターンを区別できる. すなわち, この場合, シンドロームの桁数が多すぎるという意味で, 無駄があることがわかる（この点に関し, 次に述べるハミング符号では無駄がない）.

問題 5-2: つぎのパリティ検査行列 H について, 次の問いに答えよ

$$H = \begin{bmatrix} 1 & 1 & 0 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 & 1 & 0 \\ 0 & 1 & 1 & 0 & 0 & 1 \end{bmatrix}$$

(1) ６種類の誤りパターン $[1\ 0\ 0\ 0\ 0\ 0]^T$,...,$[0\ 0\ 0\ 0\ 0\ 1]^T$（ただし, T は転置を表す）それぞれに対しシンドロームを求めよ.

(2) 100011 という受信語が得られたとする. 誤りパターンと, 正しい送信語を求めよ.

ハミング符号(Hamming code)：

2 以上の整数 m について，1 から $n=2^m-1$ までを 2 進数表示した列ベクトルをすべて並べて得られる検査行列をもつ符号.

このとき，符号語長が $n=2^m-1$，検査記号の長さは m 桁（したがって，情報記号の桁数は $k=2^m-m-1$）となり，後でも述べるように 1 個の誤りを訂正可能である.

検査記号が m 桁であるから，パリティ検査方程式が m 個である. したがってシンドロームは，m 次元列ベクトルとなり，全部で 2^m 種類あることになる. そのうち「誤りなし」を確保すると，誤りパターンとしては 2^m-1 個シンドロームで対応させうる．通信路に送る記号列は全部で $n=2^m-1$ 桁であるから，n 桁のうち 1 個までの誤りならば，全ての誤りパターンを対応するシンドロームでカバーできる（したがって，1 個までの誤りならば訂正可能である）．また，ハミング符号では，誤りパターンが誤り位置と対応する.

ハミング符号の基本的な考え方について，もう少し詳しく見てみよう．$m=3$ の場合を例にとる．このとき，$n=2^m-1=7, k=2^m-m-1=4$ となる．10 進数の 1–7 を，2 進数で表し，それを縦に書いた行列を考える（ゼロベクトルでない全ての列ベクトルを並べたもの）.

$$H = \begin{bmatrix} 0 & 0 & 0 & 1 & 1 & 1 & 1 \\ 0 & 1 & 1 & 0 & 0 & 1 & 1 \\ 1 & 0 & 1 & 0 & 1 & 0 & 1 \end{bmatrix} \tag{5.16}$$

この行列 H に対し，$H\mathbf{x}=O$ を満たす記号系列 $\mathbf{x}$ を符号語として採用するというのがハミング符号である.

このように定義すると，最小距離は必ず 3 となる（問題 5-3）ので，単一誤り訂正可能である. また，ハミング符号における符号語は，半径を 1 とするハミング球の中心にあり，それらのハミング球によって，可能な信号のパターンが覆い尽くされている(問題 5-4)．すなわち，無駄がない.

ちなみに，ハミング符号の任意の二つの符号語 $\mathbf{x}_1$, $\mathbf{x}_2$ に対し $\mathbf{x}=\mathbf{x}_1\oplus\mathbf{x}_2$ も符号語となる．なぜなら，$H\mathbf{x}=H(\mathbf{x}_1\oplus\mathbf{x}_2)=H\mathbf{x}_1\oplus H\mathbf{x}_2=O\oplus O=O$ だからである．ゼロベクトルは明らかに符号語となるので，ハミング符号の符号語は線形部分空間をなす（ここで挙げた例では，符号語は 7 次元ベクトルであるが，符号語の集合は，4 次元線形部分空間をなす）．このような符号を，線形符号という.

問題 5-3: 上記のハミング符号の最小距離が 3 であることを証明せよ.
（ヒント：線形符号の最小距離は，符号語のハミング重み（符号語中の 1 の個数）の最小値に等しいという性質がある．この性質を利用し，$H\mathbf{x}=O$ を満たす，ゼロベクトルでない $\mathbf{x}$ が 3 つ以上の 1 を持つことを示せ.）

問題 5-4: ハミング符号における符号語は，半径を 1 とするハミング球の中心にあり，それらのハミング球によって，可能な信号のパターンが覆い尽くされていることを示せ.

組織符号と生成行列：

ハミング符号において，具体的な符号語を求めるためには $H\mathbf{x}=O$ を解く必要があるが，これには手間がかかるので，もう少し楽に符号語を得られるとよい．このためには，組織符号という考え方と，生成行列というものを利用すればよい．

情報記号のベクトルを $\mathbf{u}=[u_1,\cdots,u_k]^T$ に対し，符号語 $\mathbf{x}=[x_1,\cdots,x_n]^T$ を与える行列を $G=[g_{ij}]$ とする．すなわち，$\mathbf{x}=G\mathbf{u}$ で

$$\begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix} = \begin{bmatrix} g_{11} & \cdots & g_{1k} \\ \vdots & \ddots & \vdots \\ g_{n1} & \cdots & g_{nk} \end{bmatrix} \begin{bmatrix} u_1 \\ \vdots \\ u_k \end{bmatrix} \tag{5.17}$$

である．この G は生成行列と呼ばれる．G がきまっていれば，u を与えることによって符号語 $\mathbf{x}$ は容易に定まる．このような G はどうやって得られるだろうか．パリティ検査行列 H に対し，列を適当に並べ替えることによって，次式を得るものとする．

$$H' = \begin{bmatrix} h_{11} & \cdots & h_{1k} & 1 & \cdots & 0 \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ h_{m1} & \cdots & h_{mk} & 0 & \cdots & 1 \end{bmatrix} \tag{5.18}$$

このような H' は標準形とよばれる．この H' では，左側が情報記号に対応し，右側が検査記号に対応している．このように情報記号と検査記号とが見た目で区別できる符号は，組織符号と呼ばれる．組織符号のパリティ検査行列 H' に対しては，生成行列は次のように容易に求まる．いま，$H'=[G_R \quad I_m]$ とし，I_m は m 次元の単位行列で，G_R は，

$$G_R = \begin{bmatrix} h_{11} & \cdots & h_{1k} \\ \vdots & \ddots & \vdots \\ h_{m1} & \cdots & h_{mk} \end{bmatrix} \tag{5.19}$$

である．ここで，

$$G = \begin{bmatrix} I_k \\ G_R \end{bmatrix} \tag{5.20}$$

とする．$H'\mathbf{x}=H'G\mathbf{u}=O$ が任意の $\mathbf{u}$ に対して成立するためには，$H'G=O$ でなければならない．ところが，

$$H'G=[G_R \quad I_m]\begin{bmatrix} I_k \\ G_R \end{bmatrix} = G_R \oplus G_R = O \tag{5.21}$$

である．したがって，上のように，標準形のパリティ検査行列があれば G は簡単に求まる．

なお，生成行列の考え方を使えるのは，ハミング符号に限らない．線形符号と呼ばれるクラスの符号は，ハミング符号と同様に生成行列によって符号語を得ることができる（ハミング符号は線形符号）．

例 5-5: 上で与えたパリティ検査行列Hから，生成行列を求めてみよう．まず，Hを標準化すると，つぎのようになる．

$$H' = \begin{bmatrix} 0 & 1 & 1 & 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 1 & 0 & 1 & 0 \\ 1 & 1 & 0 & 1 & 0 & 0 & 1 \end{bmatrix} \tag{5.22}$$

このとき，

$$G_R = \begin{bmatrix} 0 & 1 & 1 & 1 \\ 1 & 0 & 1 & 1 \\ 1 & 1 & 0 & 1 \end{bmatrix} \tag{5.23}$$

であるから，

$$G = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 1 & 1 & 1 \\ 1 & 0 & 1 & 1 \\ 1 & 1 & 0 & 1 \end{bmatrix} \tag{5.24}$$

となる．

H'について，シンドロームを求めてみよう．$\mathbf{e}_2 = [0,1,0,0,0,0,0]^T$（2 ビット目に誤りが生じる）のとき，

$$H'\mathbf{e}_2 = \begin{bmatrix} 0 & 1 & 1 & 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 1 & 0 & 1 & 0 \\ 1 & 1 & 0 & 1 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 0 \\ 1 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix} = \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix} \tag{5.25}$$

となる．すなわち，i ビット目に誤りが生じるときは，パリティ検査行列の i 列目がシンドロームとして現れる．

送信したい記号列として，$\mathbf{u} = [1,1,0,0]^T$ を考えよう．このとき，符号語 $\mathbf{x}$ は，

$$\mathbf{x} = G\mathbf{u} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 1 & 1 & 1 \\ 1 & 0 & 1 & 1 \\ 1 & 1 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 \\ 1 \\ 0 \\ 0 \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \\ 0 \\ 0 \\ 1 \\ 1 \\ 0 \end{bmatrix} \tag{5.26}$$

となる.

問題 5-5: 上の例で $H'\mathbf{x}$ を計算してみると零ベクトルとなることを確認せよ．また，$\mathbf{y} = \mathbf{x} \oplus \mathbf{e}_2$ となるとき，シンドロームはどのようになるか.

問題 5-6: 検査記号の桁数が 3 のハミング符号を作成せよ（すべての符号語を列挙せよ）.

問題 5-7: 問題 5-6 の符号における符号のレートはいくらか.

(付録 5-1)　線形符号の一般理論（の基礎）

線形符号の議論では，説明の都合上，符号語をベクトルとして表現する．別の文脈では，確率変数の積として表現することもあることに注意せよ．

線形符号：n 次元ベクトル $\mathbf{x} = [x_1, \cdots x_n]^T$ （ただし， $x_i \in \{0,1\}$ ）が作る空間 S の線形部分空間 $W = \{w_1, ..., w_{2^k}\}$ の要素を符号語とする符号 W

W は線形部分空間であるから，任意の $\mathbf{w}_i, \mathbf{w}_j \in W$ について， $\mathbf{w}_i + \mathbf{w}_j \in W$ が成り立つ

$\mathbf{w} \in W$ は，k 本の一次独立なベクトル $\mathbf{g}_1, \cdots \mathbf{g}_k$ （符号の生成ベクトルという）の線形結合としてあらわされる．すなわち，

$$\mathbf{w} = \sum_{i=1}^{k} a_i \mathbf{g}_i \text{，ただし，} a_i \in \{0,1\}$$

このとき， $\mathbf{a} = [a_1, \cdots, a_k]^T$ が，送信したい情報そのものである．送信したい情報から符号語を作り出すという意味で， $\mathbf{g}_1, \cdots \mathbf{g}_k$ を符号の生成ベクトルといい，行列 $G = [\mathbf{g}_1, \cdots \mathbf{g}_k]$ を生成行列という．

w の性質をもう少し見てみよう．w をベクトル表記してみると，

$$\mathbf{w} = G\mathbf{a}$$

となる．生成行列 G に対し， $HG = O$ を満たす H をとり，右側から a をかけてみると，

$$HG\mathbf{a} = H(G\mathbf{a}) = H\mathbf{w} = O$$

となる．すなわち，線形符号における符号語は，ある行列 H（これをパリティ検査行列という）にかけるとゼロベクトルになるものでもある．

線形符号の復号：

線形符号を復号する場合は，基本的にはシンドローム復号を行えばよい．誤りベクトルを e とするとき，受信信号 **y** は，

$$\mathbf{y} = \mathbf{w} \oplus \mathbf{e}$$

であり，シンドロームは次の性質を持つ．

$$\begin{aligned} S = H\mathbf{y} = H(\mathbf{w} \oplus \mathbf{e}) &= H\mathbf{w} \oplus H\mathbf{e} \\ &= O \oplus H\mathbf{e} \\ &= H\mathbf{e} \end{aligned}$$

したがって，誤りベクトルとシンドロームの関係をあらかじめ調べておけば，シンドロームから誤りベクトルが特定できる．誤りベクトルを受信信号に加えれば，

$$\mathbf{y} \oplus \mathbf{e} = (\mathbf{w} \oplus \mathbf{e}) \oplus \mathbf{e} = \mathbf{w}$$

によって正しく復号ができる．

しかし，複数個所の誤りが生じる可能性を考慮に入れると，あるシンドローム S をもたらす誤りベクトルは複数存在しうる．その場合は，限界距離復号といって，シンドローム S をもたらす誤りベクトルのうち，1 の数（誤りの数）が最も小さいもの（言い換えると，受信信号から最もハミング距離の近いもの）を，送信信号とみなせばよいだろう．

線形符号の誤り訂正能力：

線形符号の最小距離（もちろん，ここではハミング距離の意味である）が 2d+1 以上ならば，d 個の誤りを訂正できる．

なお，線形符号における最小距離は，

「0 以外の符号の重みの最小値」（＊）

に等しい．ここで，符号の重みとは，符号語に現れる 1 の個数を意味する．この（＊）が意味することは意外なことに思われるが，線形符号では一般に成り立つ性質である．なぜなら，符号 **u**,**v** との距離が最小距離だとする．このとき，符号間の距離を $hamming(u,v)$ であらわすと，距離は平行移動しても変わらないので，次式が成り立つ．

$$hamming(u,v) = hamming(u+u, v+u) = hamming(0,w)$$

なお，$w = v+u$ もまた符号語となることは，この符号が線形符号ならば必ず保証される．したがって，この符号の最小距離は符号 0 と w の距離，すなわち，w における 1 の個数として決まる．この w は，すべての符号語のなかで 1 の個数が最小である．もしそうでないとすると，もっと 1 の個数が少ない w’が存在することになり，0 と w’との距離 $hamming(0,w')$ は $hamming(u,v)$ よりも小さくなり，$hamming(u,v)$ が最小距離だとの仮定に反することになる．

6.　通信路符号化定理 Channel Coding Theorem

雑音のある通信路のモデル (noisy channel)：雑音(noise)によってゆがめられる可能性のある通信路(channel)の場合，次の図 6.1 のようにモデル化することができる．なお，入力は記号が２種類しかなくても，消失を含めて出力は３種類，などということもありうる（図 6.2）．

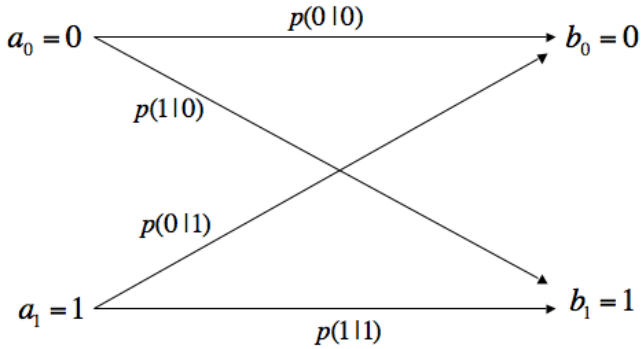


図 6.1 雑音のある通信路（二元通信路）

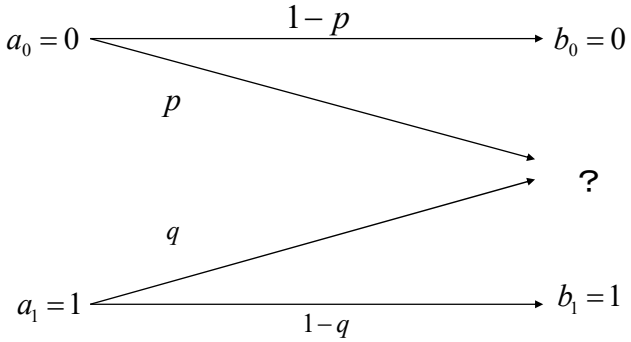


図 6.2　雑音のある通信路（二元消失通信路）

このような通信路が，どれだけの情報量を運ぶ能力があるかを知りたい．

情報量：事象（記号）a_iの生起についての情報量は次式で定義されるのだった．

$$I(a_i) = -\log_2 p(a_i) \tag{6.1}$$

記号b_jを受信した条件のもとでのa_i，すなわち$(a_i \mid b_j)$の情報量は，つぎのようになる．

$$I(a_i \mid b_j) = -\log_2 p(a_i \mid b_j) \tag{6.2}$$

$p(a_i \mid b_j)$の値は，どのように得られるだろうか．これには，Bayes の定理 (Bayes' theorem)を使うのであった．上の図のように，通信路の情報 $p(b_j \mid a_i)$ は所与のものとする．また，$p(a_i)$ 自身も通信路への入力の性質としてわかっているものとする．この場合，Bayes の定理によってつぎのようになる．

$$p(a_i \mid b_j) = \frac{p(a_i, b_j)}{p(b_j)} = \frac{p(a_i)p(b_j \mid a_i)}{\sum_i p(a_i)p(b_j \mid a_i)} \tag{6.3}$$

通信路がもたらす情報量は，つぎの「相互情報量」として与えられる．

相互情報量 (mutual information)： 受信前の状態から，b_jを受信することによって得られる情報量を相互情報量$I(a_i;b_j)$という．最終的に得られる情報量は図 6.3 の双方において等しいと考えれば，相互情報量は，次のようになる．

$$I(a_i;b_j) = I(a_i) - I(a_i \mid b_j) \tag{6.4}$$

受信前の状態 ----------> a_iを知る
（情報量$I(a_i)$）

受信前の状態 ---------- b_jを受信 ----------> a_iを知る
相互情報量　　情報量$I(a_i \mid b_j)$

図 6.3　相互情報量

相互情報量は，どのように書き下すことができるだろうか.

$$\begin{aligned} I(a_i;b_j) &= I(a_i) - I(a_i \mid b_j) \\ &= -\log_2 p(a_i) + \log_2 p(a_i \mid b_j) \\ &= \log_2 \frac{p(a_i \mid b_j)}{p(a_i)} \\ &= \log_2 \frac{p(a_i,b_j)}{p(a_i)p(b_j)} \end{aligned} \tag{6.5}$$

上式からわかるように，a_iとb_jとが独立である場合，$I(a_i;b_j)=0$である．

平均相互情報量 (average mutual information)：相互情報量$I(a_i;b_j)$の期待値$\bar{I}(X;Y)$を，平均相互情報量とよぶ.

次の(6.7)式の関係（条件付きエントロピー）を思い出すと，平均相互情報量は(6.8)式のように求められる．

$$\begin{aligned} H(X \mid Y) &= \sum_Y P(Y=y) H(X \mid Y=y) \\ &= -\sum_Y P(Y=y) \sum_X P(X=x \mid Y=y) \log_2 P(X=x \mid Y=y) \\ &= -\sum_X \sum_Y P(Y=y) P(X=x \mid Y=y) \log_2 P(X=x \mid Y=y) \\ &= -\sum_X \sum_Y P(X=x, Y=y) \log_2 P(X=x \mid Y=y) \end{aligned} \tag{6.6}$$

$$
\begin{aligned}
\bar{I}(X;Y) &= E[I(X;Y)] \\
&= E\left[\log_2 \frac{p(a_i \mid b_j)}{p(a_i)}\right] \\
&= \sum_i \sum_j p(a_i,b_j)\log_2 \frac{p(a_i \mid b_j)}{p(a_i)} \\
&= -\sum_i \sum_j p(a_i,b_j)\log_2 p(a_i) + \sum_i \sum_j p(a_i,b_j)\log_2 p(a_i \mid b_j) \\
&= -\sum_i \log_2 p(a_i) \sum_j p(a_i,b_j) + \sum_i \sum_j p(a_i,b_j)\log_2 p(a_i \mid b_j) \\
&= -\sum_i p(a_i)\log_2 p(a_i) + \sum_i \sum_j p(a_i,b_j)\log_2 p(a_i \mid b_j) \\
&= H(X) - H(X \mid Y)
\end{aligned}
\tag{6.7}
$$

雑音のない通信路の場合，$p(b_j \mid a_i) = 0, (i \neq j)$ であることから，

$$\bar{I}(X;Y) = H(X) \tag{6.8}$$

エントロピー $H(X)$ と，平均相互情報量 $\bar{I}(X;Y)$ との関係を整理してみよう．

エントロピー $H(X)$
- 情報を知ることによって平均的に得られる情報量
- 情報源の不確かさ

平均相互情報量 $\bar{I}(X;Y)$
- 受信記号を知ることによって平均的に得られる情報量
- 受信記号を知ることによって減少する情報源の不確かさの平均

なお，$\bar{I}(X;Y) = \bar{I}(Y;X)$ が成立する．このことは，エントロピーのチェイン則から明らかである．なぜなら，エントロピーのチェイン則から，

$$H(X,Y) = H(X) + H(Y \mid X) = H(Y) + H(X \mid Y) \tag{6.9}$$

が成立するが，上式を変形すると，

$$H(X) - H(X \mid Y) = H(Y) - H(Y \mid X) \tag{6.10}$$

となるが，この式は，まさに $\bar{I}(X;Y) = \bar{I}(Y;X)$ を示している．

二元通信路の平均相互情報量を計算してみよう．

例 6-1: $p(a_i) = 0.5$ とし，$p(0 \mid 0) = 0.99, p(1 \mid 1) = 0.7$ とすると，

$$\bar{I}(X;Y) = H(X) - H(X \mid Y) = 1 - 0.543 = 0.457$$

となるはずである（確かめてみよう）．

通信路容量 (channel capacity) : 平均相互情報量は，通信路への入力の確率分布（$p(a_i)$）に依存し，これを変更することによって大きくしたり小さくしたりできるが，通信路の特性（$p(a_i \mid b_j)$）は変えられない．通信路への入力をどんなに工夫しても，これ以上大きくできないという平均相互情報量を，通信路容量といい，次式で定義される．

$$C = \max_{p(a_i)} \bar{I}(X;Y) \tag{6.11}$$

問題 6-1:
二元通信路の平均相互情報量を計算してみよう．
上の例で，$p(a_i) = 0.6$ としたら，平均相互情報量はどれほどになるか．0.7 ではどうか．

通信路容量を，図 6.4 の二元対称通信路の場合について考えよう．

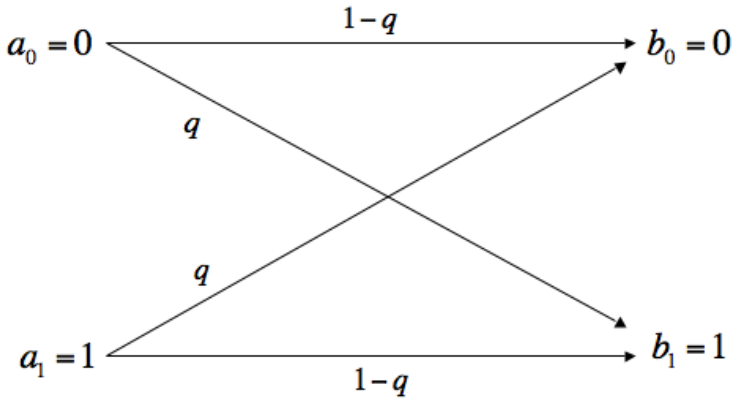


図 6.4 二元対称通信路

$p(a_0) = p_0$ とおく．さきにみたように，$\bar{I}(X;Y) = H(X) - H(X \mid Y) = H(Y) - H(Y \mid X)$ である．$H(Y \mid X)$ は次のように求まる．

$$\begin{aligned} H(Y \mid X) &= -\sum_i \sum_j p(a_i, b_j) \log_2 p(b_i \mid a_j) \\ &= -q \log_2 q - (1-q) \log_2 (1-q) \\ &= H(q) \end{aligned} \tag{6.12}$$

ここで，$H(Y \mid X) = H(q)$ は通信路への入力（$p(a_i)$）に依存しないことに注意しよう．したがって，

$$C = \max_{p(a_i)} H(Y) - H(q) \tag{6.13}$$

である．ここで，$H(Y)$ は

$$H(Y) = -\sum_j p(b_j) \log_2 p(b_j) \tag{6.14}$$

つまるところこれは二元エントロピーであって，[0,1]の値をとりうる．ここで，

$$p(b_0) = p_0(1-q) + (1-p_0)q$$
$$p(b_1) = p_0 q + (1-p_0)(1-q) \tag{6.15}$$

は，$p_0 = \frac{1}{2}$とすればそれぞれ 0.5 の値をとることから，

$$C = \max_{p(a_i)} H(Y) - H(q) = 1 - H(q) \tag{6.16}$$

通信路容量の意味：

通信路容量は，１送信記号あたり，最大で C [bit]送信できることを意味する．いいかえると，1 bit の情報を送信するために，1/C [送信記号/bit]を送る必要があることになる．これに関し，「1/C [送信記号/bit]以上で送信すれば，任意の小さな誤り率で送信できる符号が存在する」ことを主張するのが，通信路符号化定理（シャノンの第２定理）である．

問題 6-2: ある天気予想のサイトでは，過去の事例にもとづくと，予想と実際の天気との同時確率が次のように与えられている．このとき，的中率と天気の予想と実際の天気との相互情報量を求めよ．

	実際に晴れ	実際は雨
「晴れ」と予想	0.1	0.2
「雨」と予想	0.3	0.4

メッセージ(message)：通信路符号器への入力をメッセージという．

例 6-2: 以前の例で，「0」を送りたいときは 000 とし，「1」を送りたいときは 111 とする，というようなことを考えた．この場合，もとの，「0」，「1」がメッセージにあたる．

送信前のメッセージの記号は，あるアルファベットBの要素であるが，受信後，復号器からの出力は一般にはBの要素でないこともありうる（復号できない，という場合）．

復号誤り率 (probability of error)： 送信メッセージ（確率変数 M で表す）が$M = m \in B$とし，復号器からの出力$\hat{M} \neq m$のときは復号誤りが生じたといい，復号誤りの発生する確率（復号誤り率）P_eは，つぎのようになる．

$$P_e = \sum_{m \in B} P(M = m, \hat{M} \neq m) = \sum_{m \in B} P(M = m) P(\hat{M} \neq m \mid M = m) \tag{6.17}$$

例 6-3: {0,1}の２種類の記号（それぞれの記号の出現確率は 0.5 とする）を 1 ビットあたり確率 p で誤りが生じる二元対称通信路に通す．

それぞれの記号をそのまま送信する場合，情報伝送速度（符号のレート）$R = \frac{\log_2 2}{1} = 1$である．また，この場合，明らかに$P(\hat{M} \neq 0 \mid M = 0) = P(\hat{M} \neq 1 \mid M = 1) = p$ であることから，

$$P_e = P(M=0)P(\hat{M} \neq 0 \mid M = 0) + P(M=1)P(\hat{M} \neq 1 \mid M = 1) = \tfrac{1}{2} \cdot p + \tfrac{1}{2} \cdot p = p$$

これに対し，「0」に対しては 000,「1」に対しては 111 と通信路符号語を与えるものとする．

以前みたように，情報伝送速度 $R=\frac{\log_2 2}{3}=\frac{1}{3}$である．いま，受信後の記号列が{000,100,010,001}なら $\hat{M}=0$とし，{011,101,110,111}なら $\hat{M}=1$とする．このときの復号誤り率は，どのようになるだろうか.

$m=0$ のとき， $\hat{M}=1$となるのは，２カ所に誤りが生じる場合と，３カ所すべてに誤りが生じる場合がある．ビットごとの誤りの発生は互いに独立であると考えると，

- ３カ所すべてに誤りが生じる(111 を受信する)確率は， p^3
- ２カ所に誤りが生じるのは，011,101,110 のいずれかを受信する場合で，個々の生起確率は $p^2(1-p)$，場合の数は ${}_3C_2=3$通りある．これらは排反だから，あわせて $3p^2(1-p)$

であることから， $P(\hat{M}\neq 0\mid M=0)=3p^2(1-p)+p^3=p^2(3-2p)$

同様に，m=1 のとき， $P(\hat{M}\neq 1\mid M=1)=p^2(3-2p)$

したがって，

$$P_e=P(M=0)P(\hat{M}\neq 0\mid M=0)+P(M=1)P(\hat{M}\neq 1\mid M=1)=p^2(3-2p) \quad (6.18)$$

0,1 をそのまま送信する方式と，3 回繰り返して送信する方式とで，情報伝送速度と復号誤り率をまとめるとつぎのようになる．

表 6.1 復号誤り率

方式	情報伝送速度	復号誤り率
0,1 そのまま	1	p
３回繰り返す	1/3	$p^2(3-2p)$

$p\le\frac{1}{2}$のとき， $p\ge p^2(3-2p)$がなりたつので，３回繰り返す方式の方が，情報伝送速度は遅いが，より正確に情報を伝えることができる．

問題 6-3: 二元対称通信路において，１ビットあたりの誤りの確率を 0.1 とするとき，この通信路の容量を求めよ．また，この通信路に，(7,4)ハミング符号（情報ビットを４ビット，検査ビットを３ビットとすることによって合計７ビットで符号化するハミング符号）を送信する場合の復号誤り率を求めよ．（ヒント：ハミング符号は１重誤り訂正符号であるから，１つの符号語内に誤りが１個以下であれば訂正できる．２個以上だと復号誤りが生じる）.

上の例で観察したこと，すなわち，「通信路符号の冗長さを増すことによって，情報伝送速度は低下するが，復号誤り率も低下する」ということは，より一般に成り立ちそうであると直観的には思える．つまり，情報伝送速度をよくしようとおもうと，復号誤り率が悪くなり，逆に復号誤り率をよくしようと思うと，情報伝送速度が悪化する，という風に，情報伝送速度と復号誤り率との間にはトレードオフの関係が成立しそうに思われる．

ところが，通信路符号化定理は，（ある条件のもとでは）情報伝送速度と復号誤り率との関係が必ずしもトレードオフの関係になるとは限らないことを示している．

通信路符号化定理（シャノンの第 2 定理）：

通信路容量 C　[bit/送信記号]の通信路に対し，情報伝送速度が R [bit/送信記号]のとき，つぎの二つのことが成り立つ

(1) $R<C$ならば，任意に小さい復号誤り率で送信できる通信路符号が存在する
(2) $R>C$ならば，$R-C+\varepsilon$(εは任意の正定数)の復号誤り率で送信できる通信路符号が存在する．ただし，$R-C$より小さな誤り率で送信できる通信路符号は存在しない

つまり，情報伝送速度を通信路容量以下に抑える限り，復号誤り率は任意に小さい値にすることができる．この意味では，情報伝送速度と復号誤り率との間にはトレードオフの関係はない．（ただし，情報伝送速度を上げすぎると，復号誤り率の下限に$R-C$というカベが出てくるので，この意味では情報伝送速度と復号誤り率との間のトレードオフの関係は残っている．）

なお，通信路符号化定理は，そのような符号が「どこかに存在する」ことを述べているだけであって，どのように構成できるかについては述べていない．具体的な符号の構成法に基づく証明を，植松(2012)で確認できる．

通信路符号化定理で存在が示されている符号は，「十分に長い記号系列」を持つことを前提とする．実際に符号語長を長くしてしまうと，符号化や復号化に大きな計算コストを強いることとなる．

ここでは，通信路符号化定理の（１）の部分のみ証明の概観を示してみよう．

まず，情報源の記号列における代表系列 （あるいは標準系列）と呼ばれるものを示す．面白いことに，すべての代表系列は互いに生起確率がほぼ等しく，それらを足すとほぼ 1 になる．代表系列とは，おおざっぱに言うと，大数の（弱）法則に従う記号列のことと思っておけばよい．情報源記号が M 種類（$a_1,...,a_M$）ある[2]とし，それぞれの生起確率が $p_1,...,p_M$ であるような独立生起情報源 X を考えよう．なお，この情報源 X は，$C=H(X)-H(X|Y)$ を満たすものであるとする．

いま，長さ n の記号列($X^n=x_1x_2\cdots x_n$)を考えることとし，n が十分に大きいとすると，大数の法則によって，それぞれの記号は $np_1,...,np_M$ 個ずつ現れると考えてよい（本当は np_i は一般には整数ではないが，n が大きければ，i 番目の記号は近似的に np_i 個に近い数の個数をとる）．

代表系列（典型系列）　(typical sequence)：

それぞれの記号をほぼ $np_1,...,np_M$ 個ずつもつものを代表系列とよぶ（あるいは，典型系列ともいう）．

一つの記号の個数は同じでも、並びがいろいろありうるため，代表系列は実際にはたくさんある．

代表系列の生起確率と個数を考えてみよう．一つの代表系列の生起確率を p とすると，

$$p=p_1^{np_1}\times\cdots\times p_M^{np_M} \tag{6.19}$$

となる．両辺の対数をとると，つぎのようになる．

---

[2] ここでの M は定数である．説明の都合上、別のところで，M を確率変数として使っている部分があるので注意．

$$\log p = \sum_i np_i \log p_i = -n\sum_i -p_i \log p_i = -nH(X) \tag{6.20}$$

したがって，

$$p = 2^{-nH(X)} \tag{6.21}$$

となる．これは，どの代表系列であっても生起確率は等しく p となることを意味している（漸近等分割性）．

代表系列の個数を Nとすると，p=1/N と考えてよいから，つぎのようになる．

$$N = \frac{1}{p} = 2^{nH(X)} \tag{6.22}$$

ちなみに，長さ n の記号列において，記号の種類が M 種類あるとすると，あり得る記号列は全部で M^n 個である．$H(X) \le \log_2 M$ であることを考えると，N は M^n に比べてごく小さい値であることに注意する．

問題 6-4: メッセージが{0, 1}の 2 種類がありうる(M=2)として，n=10 とする．このとき，$N = 2^{nH(X)}$ と M^n とはどれほど異なるだろうか． $p(0) = 0.4, p(1) = 0.6$ の場合を例にとって，$N : M^n$ を計算してみよ．

さて，受信側を考えよう．受信される記号 Y の記号列 $Y^n = y_1 y_2 \cdots y_n$ についても代表系列を考えることができて，一つの代表系列の生起確率と代表系列の個数はそれぞれ $2^{-nH(Y)}$，$2^{nH(Y)}$ となる．送信された一つの代表系列は，受信側の代表系列のどれか一つになる．X^n を送信して Y^n になった場合の「条件付き代表系列」 $Y^n \mid X^n$ の生起確率と個数を求めてみよう．入力の記号 X がわかっているときの出力 Y のエントロピーは $H(Y \mid X)$ であるから，これまでと同様に考えると，長さ n の記号列を送った場合の出力記号列は，その生起確率がそれぞれ $2^{-nH(Y|X)}$ で，その個数はほぼ $2^{nH(Y|X)}$ であると考えてよい．すなわち，n を十分に大きくすると，一つの送信記号列に対応する受信記号列は $2^{nH(Y|X)}$ 個のどれかであると思ってよく，それ以外のものが出て来ないと見てよい．このことは，次の図に示すようにイメージすればよい．うまく送信側の異なる二つの代表系列を選べば，それらに対応する二つの出力側の代表系列はほとんど重ならないことに注意する（甘利は，一つの入力記号系列に対する出力記号系列の代表系列を，「勢力圏」と呼んでいる）．

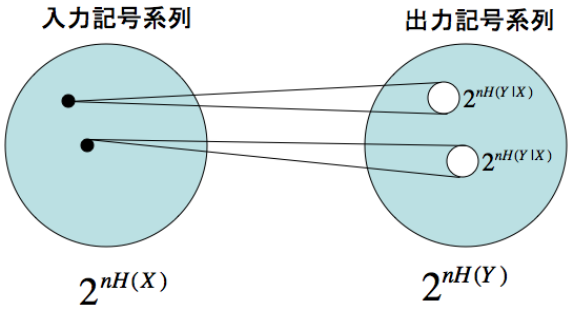


図 6.5 勢力圏

受信側の記号系列（勢力圏）が互いに重ならないようにするためには，送信側をどのように選べばよいだろうか．いま，そのような送信側の代表系列が2^{nR}個とれたとする．このとき，2^{nR}種類の記号をnビットの記号列で送信することになるので，符号のレート（情報伝送速度）は，

$$\frac{\log_2 2^{nR}}{n} = R \tag{6.23}$$

となる．通信路符号化定理は，このRを通信路容量C以下に設定すれば，復号誤り率を限りなく小さくできることを主張する．そのようにできる理由を見てみよう．

まず，Rの上限を考えよう．実際に受信される記号系列は全体でほぼ$2^{nH(Y)}$個であるが，一つの送信信号系列について$2^{nH(Y|X)}$個の可能性が生じることから，

$$2^{nR} \le 2^{nH(Y)} \div 2^{nH(Y|X)} = 2^{n\{H(Y)-H(Y|X)\}} \tag{6.24}$$

となる．ここで，$H(Y)-H(Y\,|\,X)$は高々通信路容量Cで抑えられることから，

$$2^{nR} \le 2^{n\{H(Y)-H(Y|X)\}} \le 2^{nC} \tag{6.25}$$

すなわち，異なる「勢力圏」が重ならないようにしようと思ったら，

$$R \le C \tag{6.26}$$

にせざるを得ない．この点が，通信路符号化定理の本質的に重要な部分といえる（通信路符号化定理は，$R<C$ならば「勢力圏」がほとんど重ならないようにできることを述べている）．

いま，$R<C$としよう．このとき，ある$\varepsilon>0$をもちいて，$R=C-\varepsilon$と表すことができる．送信記号の代表系列から，2^{nR}個をランダムに選んで符号語とする（ランダム符号）ことにする．このとき，一つの記号系列が符号語に選ばれる確率は，つぎのようになる．

$$\frac{2^{nR}}{2^{nH(X)}} = 2^{n\{R-H(X)\}} \tag{6.27}$$

逆に，符号語として選ばれない確率は，つぎのとおりである．

$$1-2^{n\{R-H(X)\}} \tag{6.28}$$

ひとつの受信記号系列(これをyとよぶ)が与えられたとするとしよう．一つの記号を受信したときに残るエントロピーは$H(X\,|\,Y)$であるから，yを与える可能性のある送信記号系列は$2^{nH(X|Y)}$個とおもってよい．その中に，符号語が一つしかなければよい．そのような確率P（正しく復号できる確率）は，次の通りである．

$$P = \left\{1-2^{n\{R-H(X)\}}\right\}^{2^{nH(X|Y)-1}} \tag{6.29}$$

ここで，$2^{nH(X|Y)-1} \cong 2^{nH(X|Y)}$とし，xが十分小さければ$(1-x)^y \cong 1-xy$となる関係を用いて，次のようにできる．

$$
\begin{aligned}
P &\cong \left\{1-2^{n\{R-H(X)\}}\right\}^{2^{nH(X|Y)}} \\
&\cong 1-2^{n\{H(X)-H(X|Y)-\varepsilon-H(X)\}}2^{nH(X|Y)} \\
&= 1-2^{-n\{H(X|Y)+\varepsilon\}}2^{nH(X|Y)} \\
&= 1-2^{-\varepsilon n}
\end{aligned}
\tag{6.30}
$$

したがって，$n\to\infty$とすれば，$P\to 1$となる．これによって，通信路符号化定理の(1)が示されたことになる．

以上によって，通信路符号化定理の（１）の部分が証明できたことになる．

(付録 6-1) 相対エントロピー，ダイバージェンス，カルバック・ライブラー情報量：

二つの確率分布 $P(X_1 = x) = p(x)$， $P(X_2 = x) = q(x)$ があるとき，次のように定義される量を，相対エントロピーもしくはダイバージェンスと呼ぶ．

$$D(p \| q) = \sum_x p(x) \log_2 \frac{p(x)}{q(x)}$$

これは，パターン認識などでしばしば使われる量であり，提案者の名前を取ってカルバック・ライブラー情報量と呼ばれたりする．

この相対エントロピーは，かならず非負であり，二つの確率分布が一致する場合に限り 0 となる．このため，二つの確率分布の「距離」のようなもの，すなわち二つの確率分布がどの程度似通っているかを表すのに使われることがある．ただし，一般には，

$$D(p \| q) \neq D(q \| p)$$

であることから，相対エントロピーは厳密には「距離」ではない（距離の公理を満たさない）．

平均相互情報量は，その定義から明らかなように，相対エントロピーの一種である．

次ページ以降は，今後順次追加していく

７．様々な符号

7.1 巡回符号

巡回符号：
$\mathbf{x} = [x_1, x_2, \ldots x_{n-1}, x_n]$が符号語であるとき，$\mathbf{x}' = [x_n, x_1, x_2 \ldots x_{n-2}, x_{n-1}]$も符号語であるような符号を，巡回符号という．

巡回符号は，線形符号の一種である．巡回符号は，これを計算する回路を構成するのが容易であり，よく利用される．

巡回符号の議論では，説明の都合上，符号語をベクトルではなく，多項式としてあらわすことがある．つまり，符号語 $\mathbf{x} = [x_1, x_2, \ldots x_{n-1}, x_n]$を

$$x(t) = x_1 + x_2 t + x_3 t^2 + \cdots + x_n t^{n-1}$$

とかく．たとえば，n=4 として，$\mathbf{x} = [1,1,0,1]$という符号語の場合，この符号語は

$$x(t) = 1 + t + 0t^2 + 1t^3 = 1 + t + t^3$$

と表される．

こうした符号について，$t^n - 1$を法とする多項式環 S を考える．この場合，t^nは1に，t^{n+1}は t に，t^{n+2}はt^2に見なされる，という具合となる．
符号語$\mathbf{x} = [x_1, x_2, \ldots x_{n-1}, x_n]$を巡回的にずらした$\mathbf{x}' = [x_n, x_1, x_2 \ldots x_{n-2}, x_{n-1}]$は，

$$x(t) = x_1 + x_2 t + x_3 t^2 + \cdots + x_n t^{n-1}$$

に対して，$tx(t)$が相当する．つまり，

$$tx(t) = x_1 t + x_2 t^2 + x_3 t^3 + \cdots x_{n-1} t^{n-1} + x_n t^n = x_n + x_1 t + x_2 t^2 + x_3 t^3 + \cdots + x_{n-1} t^{n-1}$$

となり，これが$\mathbf{x}' = [x_n, x_1, x_2 \ldots x_{n-2}, x_{n-1}]$に相当する．

巡回符号を表す多項式$x(t)$の全体は，$t^n - 1$を法とする多項式環のイデアルをなす．すなわち，$x(t)$，$y(t)$が巡回符号の符号語であるとき，

$$x(t) + y(t)$$

$$x(t)y(t)$$

のいずれも符号語となる．

ある多項式$g(t)$が，$t^n - 1$を割り切ることができるとき（ある$h(t)$が存在して$g(t)h(t) = t^n - 1$），$g(t)$に多項式をかけて得られる多項式の全体，すなわち

$$\{z(t)\mathrm{g}(t)\}$$

は，一つのイデアルをなし，これによって一つの巡回符号が形成される．この巡回符号は，$g(t)$をもとにして，以下の一次独立な多項式

$$g(t), tg(t), t^2 g(t), \cdots, t^{k-1} g(t)$$

の線形結合としてあらわすことができる(ただし$g(t)$は m 次の多項式とし, k=n-m). この場合, 符号語は2^k個とることができる.

このような$g(t)$に対して, ある符号語$w(t) = z(t)g(t)$をとると$w(t)g(t)$はイデアルの元であるから符号語であり, $g(t)h(t) = t^n - 1 = 0$に注意すると,

$$w(t)h(t) = z(t)g(t)h(t) = z(t) \cdot 0 = 0$$

となる. 同様に,

$$w(t) \cdot th(t) = 0, w(t) \cdot t^2 h(t) = 0, \cdots$$

といった具合に, 符号語$w(t)$は, $h(t)$ (この$h(t)$は$g(t)h(t) = t^n - 1 = 0$を満たすものとして得られるもの)から作られるイデアルの元を乗ずると 0 になるようなものであるという性質を持つ.

$g(t)$が m 次の多項式であることから, $h(t)$は k=n-m 次の多項式であり,

$$h(t) = h_1 + h_2 t + x_3 t^2 + \cdots + h_k t^{k-1}$$

とかける. 巡回符号の符号語は, $w(t)h(t) = 0$をみたすが, これはベクトルで表現すると

$$H = \begin{bmatrix} h_k & \cdots & h_1 & 0 & \cdots & 0 \\ 0 & h_k & \cdots & h_1 & \ddots & \vdots \\ 0 & 0 & h_k & \cdots & h_1 & \\ \vdots & \ddots & \ddots & & & \\ 0 & \cdots & 0 & h_k & \cdots & h_1 \end{bmatrix}$$

$$\mathbf{w} = [w_1, \cdots w_n]^T$$

なる H,w について

$$\mathrm{Hw} = 0$$

が成り立つ. この H が, この巡回符号のパリティ検査行列となっている.

巡回符号の例を一つ挙げよう. n=7 とする. このとき,

$$t^n - 1 = (1 + t^2 + t^3)(1 + t^2 + t^3 + t^4)$$

とできる. $g(t) = 1 + t^2 + t^3$とすると, $h(t) = 1 + t^2 + t^3 + t^4$であり,

$$\mathrm{H} = \begin{bmatrix} 1 & 1 & 1 & 0 & 1 & 0 & 0 \\ 0 & 1 & 1 & 1 & 0 & 1 & 0 \\ 0 & 0 & 1 & 1 & 1 & 0 & 1 \end{bmatrix}$$

となる. 実は, これはハミング符号のパリティ検査行列と同じものである.

ほかの巡回符号の例:

・BCH (Bose-Chaudhuri-Hocquenghem) 符号

・リード・ソロモン符号

参考文献

瀧保夫(1978). 情報論 I，岩波書店
⇒ この講義資料の下敷きになっている．古典.

植松友彦(2011). イラストで学ぶ情報理論の考え方，講談社.
⇒ イラストで学ぶ となっているが，イラストはこの本にとって本質的でない.
しかし，わかりやすく書かれている．レベルも低すぎない.

横尾英俊（2004）. 情報理論の基礎，共立出版.
⇒ わかりやすく書かれているが，ややレベルを落としすぎている感あり.

甘利俊一(1970). 情報理論，ダイヤモンド社.
⇒ 名著．わかりやすく書かれている．今でも入手可能だろう.

小林欣五，森田啓義(2008). 情報理論講義，培風館.
⇒ 難しい．しかし厳密にしっかり書かれている.

Slepian, D. (1973). Key papers in the development of information theory, IEEE Press.
⇒ 情報理論発展初期の重要な論文が集められている.

J.グリック（2013）. インフォメーション，新潮社.
⇒ 「情報理論」の本ではないが，「情報」の歴史を幅広く，具体的事例の積み重ねによって示した異色作．情報理論の諸概念（符号化，エントロピー，等）が基調をなしている

C.M.ビショップ (2007). パターン認識と機械学習（上），シュプリンガージャパン.
⇒ パターン認識の本であるが，わかりやすい．ビショップは天才．ラグランジュ乗数はあくまでも付録であるが，わかりやすく書かれている.

Moser, S.M., Chen, Pn-Ning (2012). A student Guide to Coding and Information Theory, Cambridge University Press.
⇒ 情報理論の英語の教科書。シンプルでわかりやすい。