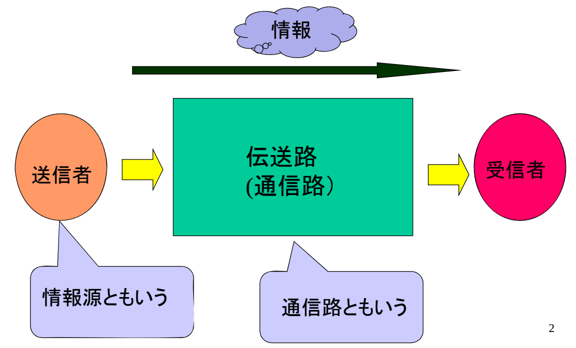
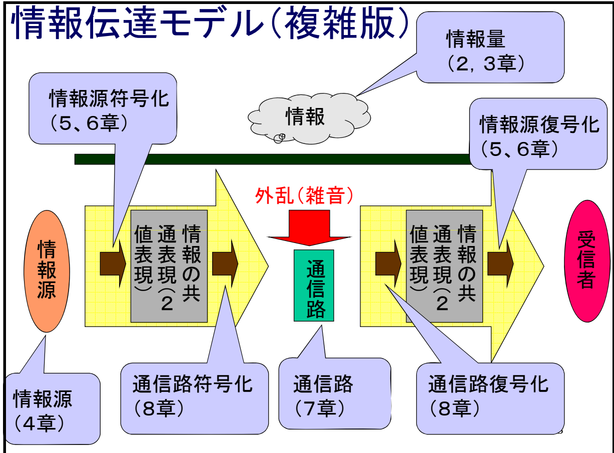
通信路符号化(8章)

情報伝達モデル(簡易版)



情報伝達モデル(複雑版)



情報源符号化と通信路符号化の比較

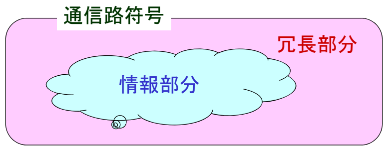
符号化	目標	実現状況	符号長	定理
情報源符号化	効率化	保存領域・通信時間の節約	最短符号の実現(平均符号長で評価)	情報源符号化定理(シャノンの第1基本定理)
通信路符号化	信頼性向上	誤りの検出・訂正	冗長性を付加(情報速度で評価)	通信路符号化定理(シャノンの第2基本定理)

通信路符号化の基礎概念

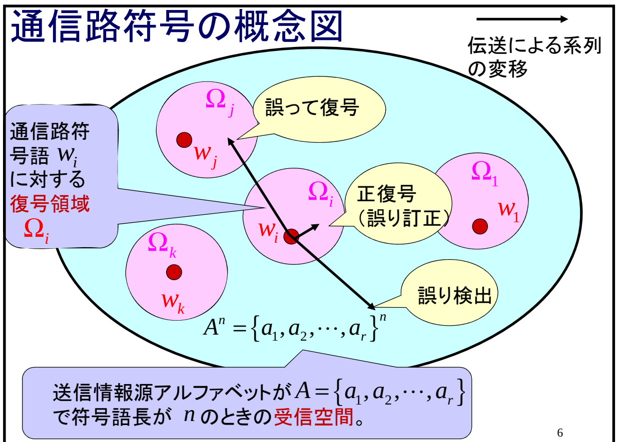
通信路符号化は信頼性の向上を達成するのが目的であり、そのために情報部分の他に冗長部分を加える。すなわち、以下のように表せる。

通信路符号＝情報部分＋冗長部分

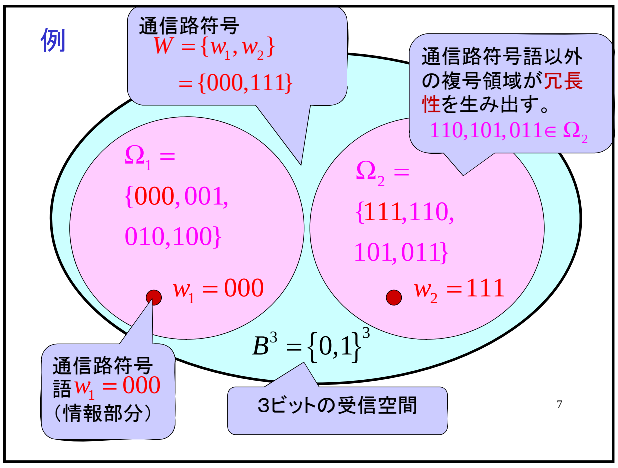
通信路符号構成の際には、いかにして冗長部分を作り出すかが重要となる。



通信路符号の概念図



送信情報源アルファベットが $A=\{a_1,a_2,\cdots,a_r\}$ で符号語長が n のときの受信空間。



（通信路符号における）記号系列の個数

（送信）情報源アルファベットを $A=\{a_1,a_2,\cdots,a_r\}$ とする。

A の元からなる長さ n の系列すべての集合を A^n と書く。

このとき、 A^n には $|A|^n=r^n$ 個の要素が含まれる。

例

$$B=\{0,1\}$$
$$B^3=\{000,001,010,011,100,101,110,111\}$$
$$|B|=2$$
$$|B^3|=|B|^3=2^3=8$$

8

通信路符号化では、送信記号の系列すべてを符号として用いるのではなくて、その中の特定の系列だけを符号として扱う。このことが冗長性を生み出す。

例

$$B^3=\{000,001,010,011,100,101,110,111\}$$

の部分集合である $W=\{000,111\}$ の要素だけを通信路符号として用いる。

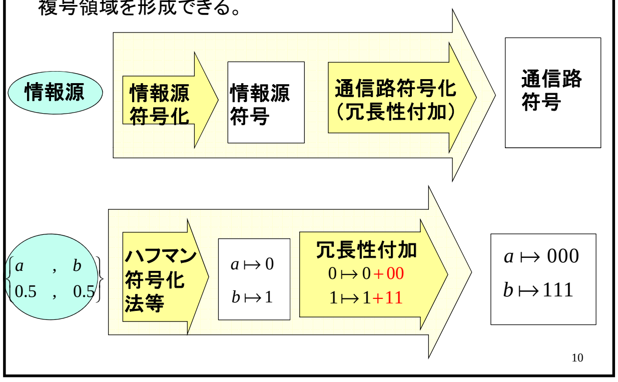
定義：（通信路符号）

このように、送信記号系列の部分集合 $W \subseteq A^n$ を（通信路）符号という。また、通信路符号に含まれる各系列 $w \in W$ を（通信路）符号語という。なお、通信路符号としては等長符号が用いられることが多い。

9

情報の付加と通信路符号

情報源符号語に冗長性を表す記号列を付加することで、複号領域を形成できる。



10

代表的系列

定義：（代表的系列）

情報源 $A=\begin{Bmatrix} a_1 & , & a_2 & , & \cdots & , & a_r \\ p_1 & , & p_2 & , & \cdots & , & p_r \end{Bmatrix}$ の長さ n

の系列の中で、各記号 a_i の出現個数 n_i と n の比が生成確率 p_i に等しいような系列を代表的系列という。

11

例

情報源 $A=\begin{Bmatrix} a & , & b & , & c \\ 1/6 & , & 2/6 & , & 3/6 \end{Bmatrix}$ の長さ12の代表的系列。

$$p_a \simeq \frac{n_a}{n}, \therefore n_a=np_a=12\times\frac{1}{6}=2$$
$$p_b \simeq \frac{n_b}{n}, \therefore n_b=np_b=12\times\frac{2}{6}=4$$
$$p_c \simeq \frac{n_c}{n}, \therefore n_c=np_c=12\times\frac{3}{6}=6$$

bbaccccabccb
cbccbacbcabc
⋮

記号出現頻度が、同一。（順序は異なる。）十分な長さの通報は、ほぼ代表的系列。

12

練習

次の各情報源と系列の長さに対して、代表的系列をそれぞれ3個示せ。

（1）

$$B=\begin{Bmatrix} 0 & , & 1 \\ 1/3 & , & 2/3 \end{Bmatrix} \qquad n=24$$

（2）

$$S=\begin{Bmatrix} a & , & b & , & c & , & d \\ 1/2 & , & 1/4 & , & 1/6 & , & 1/12 \end{Bmatrix} \qquad n=24$$

13

代表的系列の個数

性質：（代表的系列の個数とエントロピー）

情報源 $A=\begin{Bmatrix} a_1 & , & a_2 & , & \cdots & , & a_r \\ p_1 & , & p_2 & , & \cdots & , & p_r \end{Bmatrix}$ の十分な長さn

の代表的系列の個数 N は、次式で表される。

$$N=2^{nH(A)}$$

ここで、$H(A)$は情報源 A のエントロピー。

14

証明

長さ n の代表的系列を $\mathbf{w}=w_1\cdots w_n$ とし、
その発生確率を $P(\mathbf{w})$ とする。

記号 a_i の発生確率が p_i であり、それが $\mathbf{w}$ に $n_i=np_i$ 個含まれているので次式が成り立つ。

$$P(\mathbf{w})=\prod_{a_i\in A} p_i^{\,n_i}=\prod_{i=1}^{r} p_i^{\,np_i}$$

底の変換と指数法則より、次のように計算できる。

$$P(\mathbf{w})=\prod_{i=1}^{r} 2^{np_i\log p_i}=2^{n\sum_{i=1}^{r} p_i\log p_i}$$

$$=2^{-nH(A)}$$

15

n が十分大きければ、代表的系列以外の発生確率は0に漸近する。したがって、代表的系列 $\mathbf{w}$ の個数Nは発生確率の逆数である。すなわち、次式が成り立つ。

$$N=\frac{1}{P(\mathbf{w})}=2^{nH(A)}$$

QED

16

情報（伝送）速度

定義：（通信路符号の情報（伝送）速度）

通信路符号 $W=\{w_1,\cdots,w_M\}\subseteq A^n$ に対して、
通信路記号１記号あたりに伝送される情報量を、
情報速度あるいは**情報伝送速度**という。
情報速度は主に
R [bit/(通信路）記号]
の記号が用いられ、

$$R=\frac{\log|W|}{n}=\frac{\log M}{n}$$

と表される。

速度を意味する
Rate
の頭文字。

$$情報量=-\log\frac{1}{M}=\log M$$

$$情報速度=\frac{情報量}{記号数}$$

17

情報伝送速度の物理的意味

1秒（単位時間）あたりに伝送可能な記号 k[記号/秒] とする。
このとき、1秒あたりで伝送される情報量 R^* [bit/秒] は、
次式で与えられる。

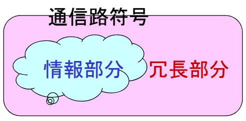
$$R^*[bit/秒]=k[記号/秒]\times R[bit/記号]$$

[bbs](Bit　Per　Second)

物理的な通信速度としてよく現れる。
１秒あたりに通信されるビット数（通信路符号を伝送する際の記号数）

18

情報（伝送）速度の直観的意味



$2^{\text{情報部分のビット長}}$ は送信系列の総数

$2^{\text{通信路符号のビット長}}$ は受信の可能性がある系列の総数

$$\text{情報速度} = \frac{\text{情報部分のビット長}}{\text{通信路符号のビット長}}$$

$$= \frac{\text{情報部分のビット長}}{\text{情報部分のビット長＋冗長部分のビット長}}$$

19

例

通信路符号 $W = \{000, 111\} \subseteq \{0,1\}^3$

の情報速度 R_W を求めよ。

解）

$$R_W = \frac{\log|W|}{n} = \frac{\log 2}{3} = \frac{1}{3} \quad [bit/(\text{通信路})\text{記号}]$$

情報源記号が2種類である。よって、送信情報源として、

$$S = \begin{Bmatrix} a & , & b \\ 0.5 & , & 0.5 \end{Bmatrix}$$

を接続すれば1ビット伝送できる。（相互情報量を1ビットにできる。）よって、

R_W ＝伝送される情報量/通信路符号長

また、例えば、$\phi = \{a \mapsto 0, b \mapsto 1\}$ のように情報源符号化も可能である。よって、

R_W＝情報源符号長/通信路符号長

20

練習

次の通信路符号の情報速度を求めよ。

（1） $W_1 = \{000, 011, 101, 110\} \subseteq \{0,1\}^3$

（2） $W_2 = \{000000, 011011, 101101, 111111\} \subseteq \{0,1\}^6$

（3） $W_3 = \begin{Bmatrix} 0000, 0011, 0101, 0110, \\ 1001, 1010, 1100, 1111 \end{Bmatrix}$

（4） $W_4 = \begin{Bmatrix} 0000\,00000, 0001\,01011, 0010\,01101, 0011\,00110, \\ 0100\,10011, 0101\,11000, 0110\,11110, 0111\,10100, \\ 1000\,10101, 1001\,11110, 1010\,11000, 1011\,10011, \\ 1100\,00110, 1101\,01101, 1110\,01011, 1111\,00000 \end{Bmatrix}$

21

情報速度と符号語数

情報源 A の長さ n の代表的系列の N 個中から M 個の系列を選んだとする。これらの系列（符号語）を代表的系列の中から等しい確率で用いるとする。

$$\left.\begin{array}{l} \mathbf{w}_1 = w_1^1 \cdots w_n^1 \\ \mathbf{w}_2 = w_1^2 \cdots w_n^2 \\ \vdots \\ \mathbf{w}_N = w_1^N \cdots w_n^N \end{array}\right\} N = 2^{nH(A)}$$

M

22

このとき、各符号を伝送したときの情報量 R_n は、次式である。

$$R_n = -\log_2 \frac{1}{M} = \log_2 M \ [bit]$$

自己情報量の式

この情報の伝送に n 記号用いているので、情報速度は次式で与えられる。

$$R = \frac{\log_2 M}{n} [bit/\text{記号}]$$

23

この式を逆に用いることにより、符号語数 M は情報速度 R と符号語長 n で表すことができる。すなわち、次式が成り立つ。

$$M = 2^{nR} \ [\text{個}]$$

この式は通信路符号化定理を導くときに用いられる。

24

通信路符号化定理（重要）

性質：（通信路符号化定理）

通信路容量 C の通信路 T において、通信路符号 Wを用いて情報を伝達する。
この通信路符号 Wの複号誤り率を $P_e(W)$、情報速度を $R(W)$ と表す。
任意の $\varepsilon > 0$ に対して、

$$R(W) < C$$

ならば

$$P_e(W) < \varepsilon$$

を満たす（通信路）符号 W が存在する。
逆に、$R(W) > C$ なら、$P_e(W) < \varepsilon$となる符号W は存在しない。

情報速度が通信路容量未満なら、複号誤り率を限りなく0にできるという意味。（通信路容量が情報速度の上限）

25

証明

A_0を通信路に接続すれば、伝送される情報量がC となる情報源

証明の方針

（1）通信路容量 C を達成する情報源をA_0 とする。
（2）A_0 から発生する長さ n の代表的系列の中から、$M = 2^{nR}$ 個の符号語をランダムに選び、その集合を符号

$$W = \{w_1, w_2, \cdots, w_M\}$$

とする。
（3）これらより、誤り確率を求める。

情報源 A_0 → 通信路T（通信路容量 C ）

26

受信される情報源を B_0とする。

このとき、A_0 は通信路容量を達成する情報源なので、通信路容量の定義より次式が成り立つ。

$$C = H(A_0) - H(A_0 \mid B_0)$$

一方、前に示したように、長さ n の代表的系列の数は

$$N = 2^{nH(A_0)}$$

ランダム符号化という。

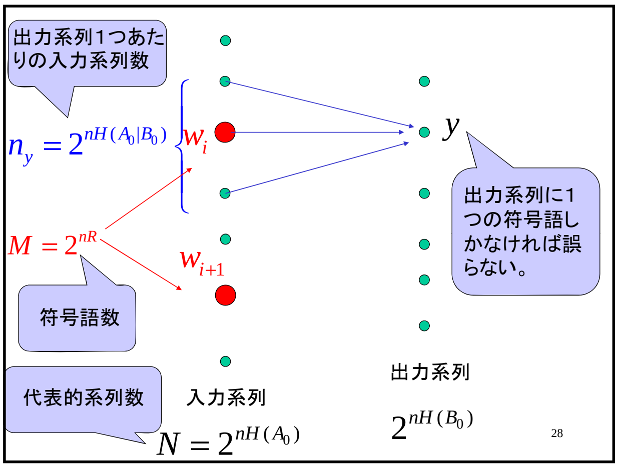
である。これらの中から代表的系列を、$M = 2^{nR}$ 個ランダムに選ぶ。（もちろん $R < H(A_0)$ である。）
したがって、ある代表的系列が符号語として選ばれる確率は、

$$p_s = \frac{M}{N} = \frac{2^{nR}}{2^{nH(A_0)}}$$

選択（Select）確率

である。

27



28

符号語 w_i を通信路 T を通して送り、受信系列 y が得られたとする。通信路には一般に誤りがあり、送信符号（送信系列）は確率的にしかわからない。
しかし、条件付エントロピー $H(A_0 \mid B_0)$ が平均情報量を意味することから、受信系列が y として受信される長さ n の代表的系列の個数 n_y は平均すれば次式表される。

$$n_y = 2^{nH(A_0|B_0)}$$

条件付エントロピーが1記号あたりの情報量であり、[bit/記号]の単位を持つことに注意する。

29

もし、$n_y = 2^{nH(A_0|B_0)}$ 個の代表的系列の中に w_i 以外に $W = \{w_1, \cdots, w_M\}$ の符号語がなければ、送られた符号語が w_i 受信語 y から一意に確定し、誤りなく復号できる。

n_y 個の代表的系列を選んだとき、w_i 以外の W の符号語を一つも選ばれない確率 p_c は、先ほど求めた p_s を用いて以下のように表せる。

$$p_c = (1 - p_s)^{n_y}$$

正しく（Correctly）複号できる確率。

連続して n_y 回符号語を選ばない確率。

これらまでの式を元にこの確率を計算する。

30

$$p_c = (1 - p_s)^{n_y}$$
$$= \left(1 - \frac{2^{nR}}{2^{nH(A_0)}}\right)^{2^{nH(A_0|B_0)}}$$
$$\simeq 1 - 2^{nH(A_0|B_0)} \bullet \left(\frac{2^{nR}}{2^{nH(A_0)}}\right)$$
$$= 1 - 2^{-n(H(A_0) - H(A_0|B_0))} \bullet 2^{nR}$$
$$= 1 - 2^{-n(C-R)}$$

テイラー展開の1次の項

$R < C$ なので $C - R > 0$

これは誤りなく復号できる確率なので、誤り確率 p_e は次式で表される。

$$p_e = 1 - p_c = 2^{-n(C-R)}$$

以上より、誤り確率は n を大きくすればいくらでも小さくできる。

（前半の証明終）

31

（後半の証明）

$R > C$ でしかも $p_e \to 0$ とできたとする。

このときには、誤りなく復号されるのだから、通信路を通じて実際に R [bit/記号]の情報量が伝送されたことになる。
しかし、これは通信路容量の定義に矛盾する。

QED

32

情報源符号化と通信路符号化のまとめ2

符号化	定理	意味
情報源符号化	シャノンの第一定理（情報源符号化定理）	平均符号長はエントロピー以上 $\frac{H(S)}{\log r} < \overline{L}$
通信路符号化	シャノンの第二定理（通信路符号化定理）	情報速度は通信路容量以下 $R < C = \max\{H(A) - H(A\|B)\}$

33

情報源符号化定理と通信路符号化定理の関係（シャノンの第1定理と第2定理の関係）

性質：雑音の無い通信路における通信路符号化定理

理想的な通信路（雑音の無い通信路）で情報伝送する場合、情報源符号化の定理と通信路符号化の定理は等価となる。

証明

送信者は $S = \begin{Bmatrix} s_1 & , & s_2 & , & \cdots & , & s_n \\ p_1 & , & p_2 & , & \cdots & , & p_n \end{Bmatrix}$ を、r元送信アルファベット

$A = \{a_1 \ , \ \cdots \ , \ a_r\}$ で符号化して送信し、受信アルファベット

$B = \{b_1 \ , \ \cdots \ , \ b_{r'}\}$ の記号列が受信されるものとする。

34

また、符号を

$\phi = \{s_1 \mapsto \boldsymbol{w}_1 \ , \ \cdots \ , \ s_i \mapsto \boldsymbol{w}_i \ , \ \cdots \ , \ s_n \mapsto \boldsymbol{w}_n\}$

とする。ここで、各符号語は、送信記号系列である。

$$\phi(s_i) = \boldsymbol{w}_i = a_1^i a_2^i \cdots a_j^i \cdots a_{l_i}^i, \quad a_j^i \in A$$

この符号の平均の情報速度 $R(\phi)[bit/送信記号]$ は、

情報源のエントロピー $H(S)[bit/情報源記号] = H(S)[bit/符号語]$ と、

平均符号長 $\overline{L(\phi)}$ [送信記号/符号語] を用いて次式で表わされる。

$$R(\phi) = \frac{H(S)}{\overline{L(\phi)}} \quad [bit/送信記号]$$

35

一方、雑音の無い通信路では、通信路容量は次式で表せる。

$$C = \max\{H(A) - H(A|B)\}$$
$$= \max H(A)$$
$$= \log r \quad [bit/送信記号]$$

雑音がないので、$H(A|B) = 0$

符号化されたものを送信アルファベット $A = \{a_1, \cdots, a_r\}$ の（新たな）情報源とみなす。全ての記号が均等に生成するときが、通信路容量を満足する。よって、以下のような通信路に対する送信情報源となる。

$$A = \begin{Bmatrix} a_1 & , & \cdots & , & a_r \\ 1/r & , & \cdots & , & 1/r \end{Bmatrix}$$

36

これらを、通信路符号化定理に代入する。

$$R(\phi) < C$$

$$\therefore \frac{H(S)}{\overline{L(\phi)}} < \log r$$

$$\therefore \frac{H(S)}{\log r} < \overline{L(\phi)}$$

これは、r　元記号を用いた情報源符号化定理である。
逆に、情報源符号化定理と、雑音の無い通信路における
通信路容量より、通信路定理を導ける。

QED

37