情報理論＜シラバスより＞

１．ハフマン符号の原理

ハフマン符号の原理について概説し，ハフマン符号の有効性を理解する．

２．拡大情報源とハフマン符号化

エントロピーH の情報源を n 次拡大した情報源では，エントロピーが n 倍となることを学ぶ．また，n 次拡大情報源に対してハフマン符号化を行った場合には，さらに効率の良い符号化となることを理解する．

３．ランレングス符号

世の中には，ある特定の情報が連続して生起しやすいような情報源がある．このような情報源においては生起する情報を 1 つずつ符号化して伝送するよりも，情報の連続数（ランレングス）を符号化した方が効率は良い．この符号化方法であるランレングス符号とその性質について理解する．

４．ランレングスハフマン符号

ランレングス符号化とハフマン符号化を併用することで，さらに効率の良い符合となることを理解する．

５．通信路符号化

これまで学んだ情報源符号化は誤りの無い通信路モデルを対象とした符合化であったのに対し，ここでは誤りのある通信路モデルを対象とする通信路符号化について学ぶ．一般に冗長を付加することで誤りに対して強くなることを理解する．

６．誤りのある通信路モデルにおける情報量

誤りのある通信路モデルを具体的に与え，誤り率と受信側で得られる相互情報量の関係を理解する．また通信の信頼性を向上させるための方法についても理解する．

７．各種の情報源と符号化方法の性質

これまでに学んできた各種の情報源の性質，並びに符号化方法の性質について再確認する．

1．ハフマン符号とランレングス符号

1.1　ハフマン符号

ハフマン符号とは，1952 年にデビット・ハフマンによって開発された符号である．ハフマン符号は平均符号長が最短（最適）な符号であり，JPEG や LHA などの圧縮 format で使用されている．

＜ハフマン符号化の手順＞

情報源 $X=\begin{pmatrix} a_1 & a_2 & \cdots\cdots & a_n \\ p_1 & p_2 & \cdots\cdots & p_n \end{pmatrix}$ を符号化する．

ただし，a_m：符号$(m=1,2,\cdots,n)$，p_i：生起確率$(i=1,2,\cdots,n)$，$\sum_{i=1}^{n} p_i = 1$ である．

①	生起確率 p_i が降順になるよう符号 a_m を縦に並べ，その横にその生起確率 p_i を記入しておく．
②	生起確率が最も低い符号と，その次に低い符号を選出する．
③	その二つの生起確率を足し合わせ，生起確率が降順になるよう符号を再び並べ替える．
④	②・③（縮退）を最後の一つになるまで繰り返す．
⑤	逆から辿っていき，生起確率が高い方を 0，低い方を 1 として値を並べたものを符号とする．

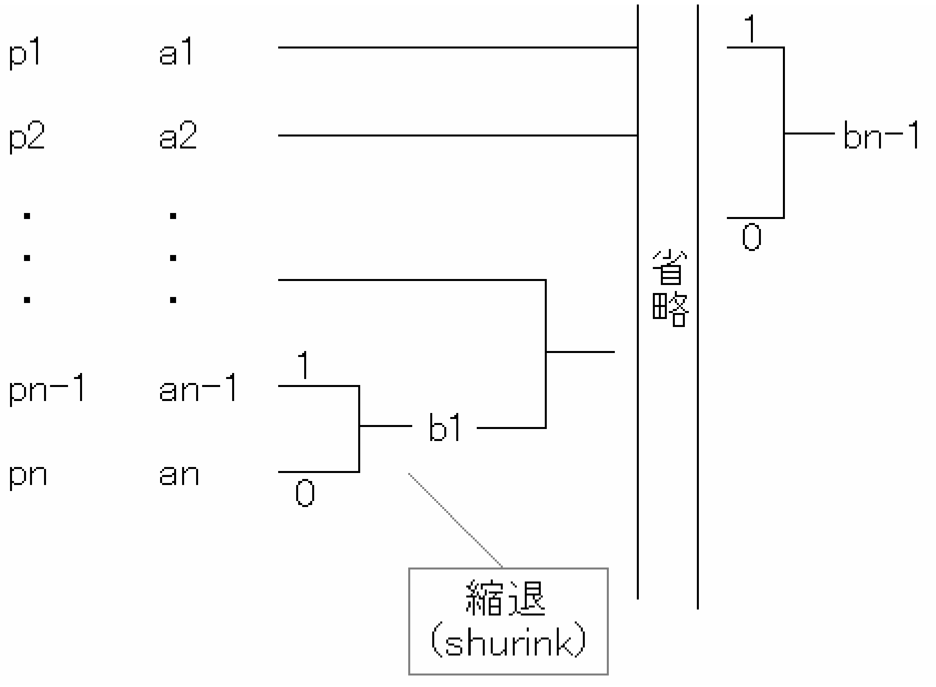


図 1　ハフマン符号の手順

[確認問題]

情報源 $X=\begin{pmatrix} a_1 & a_2 & a_3 & a_4 & a_5 & a_6 \\ 7/20 & 4/20 & 3/20 & 3/20 & 2/20 & 1/20 \end{pmatrix}$ をハフマン符号によって符号化しなさい．また，j 回縮退したときの情報源を Xj として，その符号と生起確率を示し，Xj における平均符号長 Lj を求めなさい．

[確認問題]

情報源 $X = \begin{pmatrix} a_1 & a_2 & a_3 & a_4 & a_5 & a_6 \\ 7/20 & 4/20 & 3/20 & 3/20 & 2/20 & 1/20 \end{pmatrix}$ をハフマン符号によって符号化しなさい．また，j 回縮退したときの情報源を Xj として，その符号と生起確率を示し，Xj における平均符号長 Lj を求めなさい．

[確認問題の解答]

符号化の解答例を図 2 に示す．

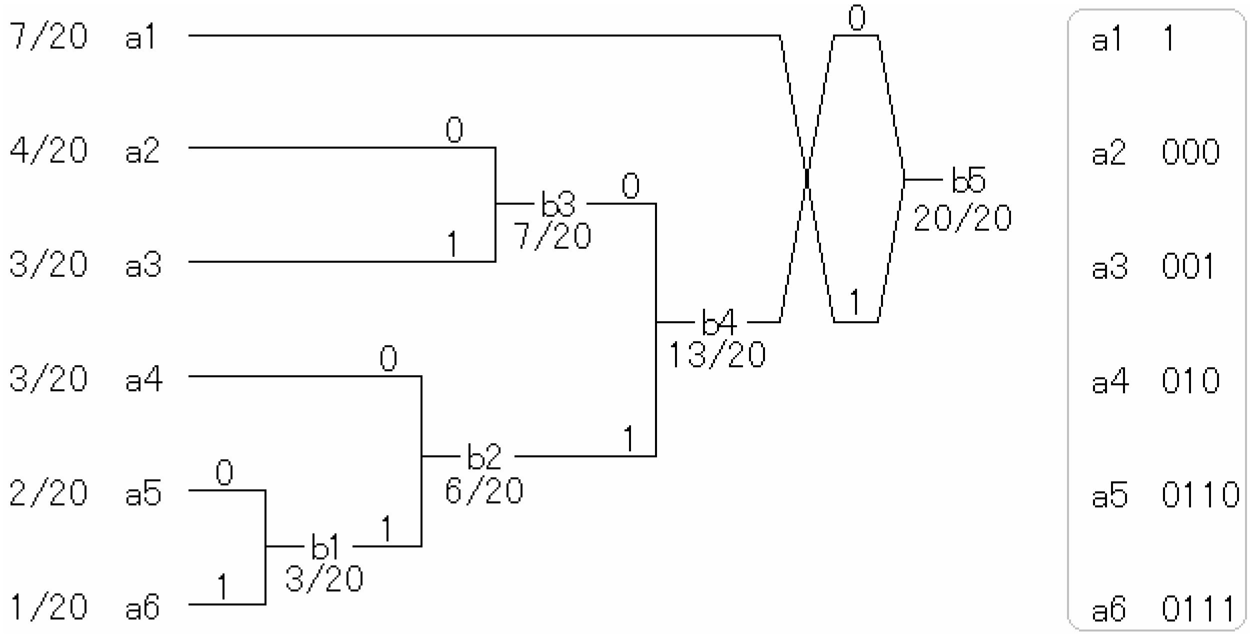


図 2　ハフマン符号確認問題の解答例（二回目の縮退，及び四回目の縮退にて別解あり）

図 2 に示す解答例を用いて，j 回縮退したときの情報源を Xj として，その符号と生起確率を示し，Xj における平均符号長 Lj を求める．平均符号長 Lj は，生起確率×符号長の総和によって求められる．

表 1　情報源 Xj 及び平均符号長 Lj

$X0 = \begin{pmatrix} a_1 & a_2 & a_3 & a_4 & a_5 & a_6 \\ 7/20 & 4/20 & 3/20 & 3/20 & 2/20 & 1/20 \end{pmatrix}$	L0 = (7×1＋4×3＋3×3＋3×3＋2×4＋1×4)/ 20 = 49/ 20　= 2.45
$X1 = \begin{pmatrix} a_1 & a_2 & a_3 & a_4 & b_1 \\ 7/20 & 4/20 & 3/20 & 3/20 & 3/20 \end{pmatrix}$	L1 = (7×1＋4×3＋3×3＋3×3＋3×3)/ 20 = 46/ 20
$X2 = \begin{pmatrix} a_1 & a_2 & a_3 & b_2 \\ 7/20 & 4/20 & 3/20 & 6/20 \end{pmatrix}$	L2 = (7×1＋4×3＋3×3＋6×2)/ 20 = 40/ 20
$X3 = \begin{pmatrix} a_1 & b_3 & b_2 \\ 7/20 & 7/20 & 6/20 \end{pmatrix}$	L3 = (7×1＋7×2＋6×2)/ 20 = 33/ 20
$X4 = \begin{pmatrix} a_1 & b_4 \\ 7/20 & 13/20 \end{pmatrix}$	L4 = (7×1＋13×1)/ 20 = 1

1.2　ハフマン符号最適性の証明

ハフマン符号は平均符号長 L が最小となる符号化である．その証明を以下で行う．

j 回縮退したときの平均符号長を Lj とすると，Lj および Lj-1 の間の関係は式(1)のようになる．

$$L_j = L_{j+1} + P_0(j) + P_1(j) \tag{1}$$

（ただし，$P_0(j)$・$P_1(j)$は Xj ⇒ Xj+1 に縮退されるときの確率）

この式より，j = k－1 のときを考え，背理法を用いて証明を行う．

仮定：「L_kが最短であるが，L_{k-1}が最短でない」

(1)より，j = k－1 のとき，

$$L_k = L_{k-1} - P_0(k-1) - P_1(k-1) \tag{2}$$

が成り立つ．ここで，L_kの最小値を L_k' と表す．これを用いると，仮定は「$L_k = L_k'$ であるが，$L_{k-1} > L_{k-1}'$ である」と表される．この仮定より，次式が成り立つ．

$$L_k = L_{k-1} - P_0(k-1) - P_1(k-1) > L_{k-1}' - P_0(k-1) - P_1(k-1) = L_k'$$

上式は，$L_{k-1} > L_{k-1}'$ なら，上式の L_k よりも小さい値 L_k' が存在することを示す．すなわち，$L_k > L_k'$となるので，仮定と矛盾する．

従って，「L_kが最短であるが，L_{k-1}が最短でない」という仮定は不合理であるから，「L_kが最短であれば，L_{k-1}も最短である」ということが成り立っていなければならない．

1.3　n 次拡大情報源とブロックハフマン符号

これまでは情報源 1 記号ごとの符号化を前提としてきた．しかし，一般に複数の情報源記号をまとめて符号化すると，さらに効率がよくなることが期待できる．

n 次拡大情報源とは，情報源の通報を n 個単位でブロック化し，それを一つの通報としたもの．また，ブロックハフマン符号とは，n 次拡大情報源に対してハフマン符号化を行うものである．

ブロックハフマン符号では，情報源の拡大次数を大きくすると（ブロックの長さを大きくすると）平均符号長 L は，限りなくエントロピーH に近づくことが知られている．

[確認問題]

情報源 $X = \begin{pmatrix} A & B \\ 5/6 & 1/6 \end{pmatrix}$ をブロック化し，2 次拡大情報源 X_2 及び 3 次拡大情報源 X_3 を求めなさい．また，それぞれをブロックハフマン符号化し，平均符号長 L，L_2，L_3 及びエントロピーH，H_2，H_3 を求めなさい．

[確認問題の解答]

ブロックハフマン符号化した結果を表 2 に示す．

表 2　n 次拡大情報源とブロックハフマン符号

$X = \begin{pmatrix} A & B \\ 5/6 & 1/6 \end{pmatrix}$	A：0　，　B：1
$X_2 = \begin{pmatrix} AA & AB & BA & BB \\ 25/36 & 5/36 & 5/36 & 1/36 \end{pmatrix}$	AA：0　，　AB：11 BA：100　，　BB：101
$X_3 = \begin{pmatrix} AAA & AAB & ABA & BAA & ABB & BAB & BBA & BBB \\ 125/216 & 25/216 & 25/216 & 25/216 & 5/216 & 5/216 & 5/216 & 1/36 \end{pmatrix}$	AAA：0　，　AAB：100 ABA：101　，　BAA：110 ABB：11100　，　BAB：11101 BBA：11110　，　BBB：11111

表 2 の結果を用いて，平均符号長 L，L_2，L_3 及びエントロピーH，H_2，H_3 を求める．

表 3　平均符号長とエントロピー

X	$L = 1$	$H = \log 3 - \frac{5}{6}\log 5 + 1$　　0.65	一文字当たり 0.65
X_2	$L_2 = (25\times1+5\times2+5\times3+1\times3)/36$ $= 53/36$　　1.47	$H_2 = 2(\log 3 + 1 - \frac{5}{6}\log 5)$　　1.30	一文字当たり 0.735
X_3	$L_3 = (125+25\times9+25\times3+5)/216$ $= 430/216$　　1.99	$H_3 = 3(\log 3 + 1 - \frac{5}{6}\log 5)$　　1.95	一文字当たり 0.664

⇒　拡大情報源の次数 n を大きくしていくことで，一文字当たりの情報量を減らすことができる！！

しかし，もともとの情報量である H より小さくすることはできない．

一文字あたりの情報量 ＞ H

（ 設問の場合：一文字あたりの情報量 ＞ 0.65 ）

1.3　n 次拡大情報源とブロックハフマン符号（続き　―生起確率を P とすると―）

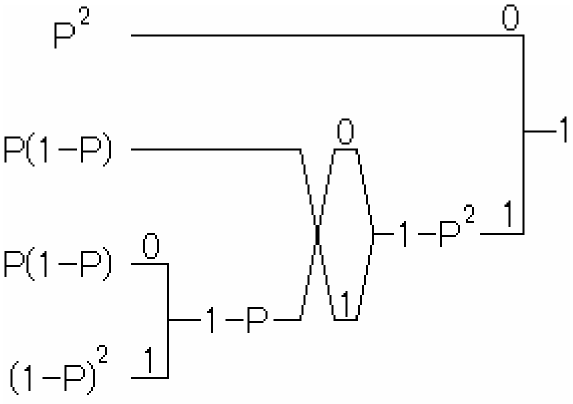
情報源 X を $X = \begin{pmatrix} A & B \\ P & 1-P \end{pmatrix}$ としたとき，n 次拡大情報源がどうなるかを計算する．（ $P > 1-P$　とする ）

符号　P：0　，　（1－P）：1

エントロピー　$H = -P\log P - (1-P)\log(1-P)$

$= -P\log P - \log(1-P) + P\log(1-P)$

2 次拡大情報源 $X_2 = \begin{pmatrix} AA & AB & BA & BB \\ P^2 & P(1-P) & P(1-P) & (1-P)^2 \end{pmatrix}$



AA：0
AB：11
BA：100
BB：101

エントロピー　$H_2 = 2\{-P\log P - (1-P)\log(1-P)\}$

平均符号長　$L_2 = -P^2 - P + 3$

＜拡大情報源について＞

元の情報源記号のいくつかをまとめて新しい情報源記号とし，この新しい情報源記号を持つ情報源を**拡大情報源**（extension source）と呼ぶ．一般に，n 次の拡大情報源を S^n で表す．

定理１：　ある情報源およびその n 次拡大情報源をそれぞれ S および S^n とするとき

$H(S^n) = nH(S)$　　が成立する．

定理２：　（クラフトの不等式）　M 元情報源 S を 2 進符号符号化したとき，その符号が瞬時符号となるための必要十分条件は，

$$\sum_{k=1}^{M} 2^{-l_k} \leqq 1$$

が成り立つことである．ここで，l_k は符号語の長さである．

定理３：　情報源 S のエントロピーを H(S)，その平均符号長を L とすると，

$H(S) \leqq L \leqq H(S)+1$　　が成り立つ．

1.4　ランレングス符号（run length code）

ランレングス符号とは，連なる(run)長さ(length)を用いて符号化するものである．

情報源 $X = \begin{pmatrix} B & W \\ 1/6 & 5/6 \end{pmatrix}$ から発生した文字列｛WWBWWWWWWBWBBWWWWWWWWWWB｝をランレングス符号化で符号化してみると以下のようになる．（もともとの文字列は 24 ビットである）

　　2　　　6　　　1　0　　　　10

→ | WWB | WWWWWWB | WB | B | WWWWWWWWWWB |

⇒　10　110　1　0　1010

上記のように，ランレングス符号をそのまま用いると，効率よく符号化できたように見えるが，実際にはこれは一意に復号可能ではない．そのため，固定長ランレングス符号がよく用いられる．

固定長ランレングス符号とは，まとめる長さを 2^n-1 までと最初に決め，全てを n ビット固定で符号する方法である．固定長ランレングス符号によって，上記の情報源を符号化すると，以下のようになる

　　2　　　6　　　1　0　　　7　　　　3

→ | WWB | WWWWWWB | WB | B | WWWWWWW | WWWB |

⇒　010　110　001　000　111　011　　　　（3 ビット固定の場合）

このとき，情報量は 18 ビットまで減少しているため，ランレングス符号は有効といえる．

固定長ランレングス符号の場合，平均符号長 L を最初に与えることになる（上記の場合は 3 ビット）．

このとき，1 情報源当たりの符号長は，$l_N = \dfrac{L}{n_N} = \dfrac{P \times L}{1-(1-P)^N}$ で表される．なお，n_N とは平均ブロック長であり，以下の式で求められる．

$$n_N = 1 \times P + 2 \times (1-P) + \cdots\cdots + N(1-P)^{N-1}P + N(1-P)_N$$

$$= \frac{1-(1-P)^N}{P}$$

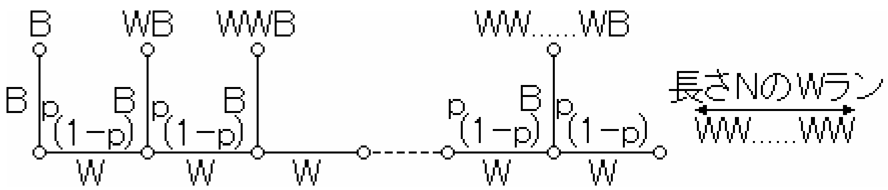


表 4　固定長ランレングス符号

符号	生起確率	長さ
B	P	1
WB	$(1-P)P$	2
WWB	$(1-P)^2P$	3
・	・	・
・	・	・
・	・	・
W……WB	$(1-P)^{N-1}P$	N-1
W……WW	$(1-P)^N$	N

1.5　固定長ランレングス＋ハフマン符号

固定長ランレングスによってまとめた文字列をハフマン符号化することで，より効率的な符号化を行うことができる．

情報源 $X = \begin{pmatrix} B & W \\ 1/6 & 5/6 \end{pmatrix}$ を固定長ランレングス符号で 2 ビットにまとめると，以下のようになる．

情報源 $X_2 = \begin{pmatrix} B & WB & WWB & WWW \\ 1/6 & 5/36 & 25/216 & 125/216 \end{pmatrix}$　→ この情報源 X_2 をハフマン符号化する．

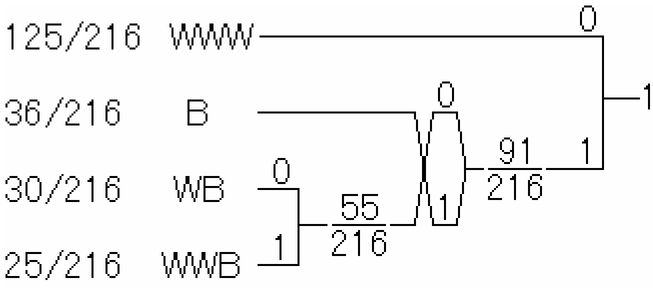


図 3　固定長ランレングス＋ハフマン符号

B：11　，　WB：100　，　WWB：101　，　WWW：0

→ 平均符号長 $L = (36\times2+30\times3+25\times3+125\times1)/216$

$= 362/216 \quad 1.67 < 2$

上記のように，平均符号長 L が固定長ランレングスそのままの場合よりも小さくなっているため，より効率のいい符号化ができている．

――

1．ハフマン符号とランレングス符号

1.1　ハフマン符号

ハフマン符号とは何か．ハフマン符号の構成方法．縮退とは．平均符号長＝∑生起確率×符号長．

1.2　ハフマン符号最適性の証明

$L_j = L_{j+1} + P_0(j) + P_1(j)$.「$L_k$ が最短であるが，L_{k-1} が最短でない」と仮定して，背理法によって証明する．

1.3　n 次拡大情報源とブロックハフマン符号

ブロックハフマン符号とは．一文字あたりの情報量＞もともとの情報量 H.

1.4　ランレングス符号

ランレングス符号とは．固定長ランレングス符号とは．1 情報源当たりの符号長 $l_N = \dfrac{L}{n_N} = \dfrac{P\times L}{1-(1-P)^N}$

固定長ランレングスのブロック長の上限 $N=2^L-1$

1.5　固定長ランレングス＋ハフマン符号

2．通信路符号化

2.1　情報源符号化と通信路符号化

これまでは，誤りのない場合についての符号化を行ってきたが，実際の通信時には誤りが生じる可能性がある．これまでの符号化は情報源符号化であり，誤りが生じる可能性を考慮して符号化するものを通信路符号化とよぶ．

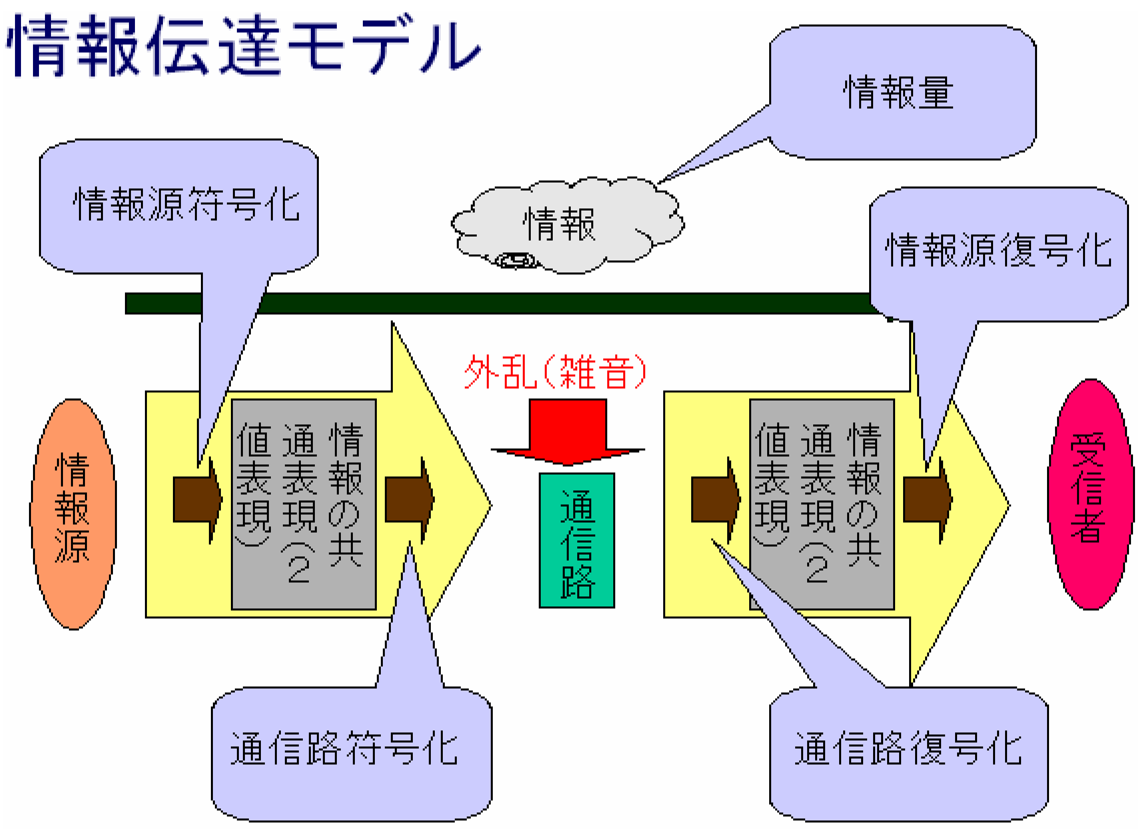


図 4　情報伝達モデル

表 5　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

符号化の際には，いかにして冗長部分を作り出すかが重要となる．

2.2　通信路行列と 2 元対称通信路

通信路では雑音などにより**誤り**が起きる．この誤りの頻度を**誤り確率**（error probability）という．
雑音レベル v は一般に

$$p(v) = \frac{1}{\sqrt{2\pi N}}\ \exp\left(-\frac{v^2}{2N}\right)$$

という確率で表される．

[通信路行列]

通信路に入力される記号を送信記号と呼び，通信路から出力される記号を受信記号と呼ぶことにする．ここで，通信路を下図のように，送信記号 $\{x_1, x_2, \cdots\cdots, x_S\}$ を入力とすると，受信記号 $\{y_1, y_2, \cdots\cdots, y_R\}$ が出力されるというモデルで表す．

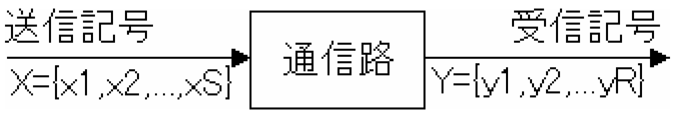


図 5　通信路

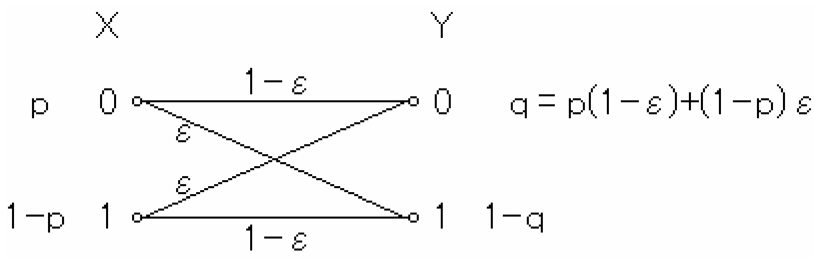
このとき，**通信路行列**（channel matrix）P は次式のようになる．

$$\mathbf{P} = \begin{pmatrix} p_{11} & p_{12} & \cdots & p_{1R} \\ p_{21} & p_{22} & \cdots & p_{2R} \\ \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot \\ p_{S1} & p_{S2} & \cdots & p_{SR} \end{pmatrix}$$

ただし，$p_{ij} = P(Y=y_j \mid X=x_i)$ とする

[2 元対称通信路]

2 元対称通信路とは下図に示すような，送信記号および受信記号とも{0,1}の 2 元記号であり，0→1 となる確率と 1→0 となる確率が同じである通信路である．この確率を反転確率（記号誤り確率）と呼び，ε で表す．



$$\mathbf{P} = \begin{pmatrix} 1-\varepsilon & \varepsilon \\ \varepsilon & 1-\varepsilon \end{pmatrix}$$ ：通信路行列

図 6　2 元対称通信路

$$X = \begin{pmatrix} 0 & 1 \\ p & 1-p \end{pmatrix} \qquad Y = \begin{pmatrix} 0 & 1 \\ q & 1-q \end{pmatrix}$$

[確認問題]

図 6 の場合において，H(Y)および H(Y | X)を求めなさい．
また，I(X;Y)=H(Y)－H(Y | X)を大きくするにはどうすればよいか答えなさい．

[確認問題]

図 6 の場合において，H(Y)および H(Y | X)を求めなさい．

また，I(X;Y)=H(Y)－H(Y | X)を大きくするにはどうすればよいか答えなさい．

[確認問題の解答]

$H(Y) = -q \log q - (1-q)\log(1-q)$

$H(Y|X) = p\{-\varepsilon \log \varepsilon - (1-\varepsilon)\log(1-\varepsilon)\} + (1-p)\{-(1-\varepsilon)\log(1-\varepsilon) - \varepsilon \log \varepsilon\}$

$\qquad = -\varepsilon \log \varepsilon - (1-\varepsilon)\log(1-\varepsilon)$

H(Y | X)は，送信側での文字の発生確率に無関係．従って，I(X;Y)を大きくするには H(Y)を大きくすればよい．H(Y)は $q = \frac{1}{2}$ のとき最大値“１”であり，$q = \frac{1}{2}$ のときの I(X;Y)の値は以下のようになる．

$I(X;Y) = 1 - \varepsilon \log \varepsilon - (1-\varepsilon)\log(1-\varepsilon)$

$\varepsilon=0$ または $\varepsilon=1$ で $I(X;Y)=1$ で最大

$\varepsilon=0.5$ で $I(X;Y)=0$ で最小

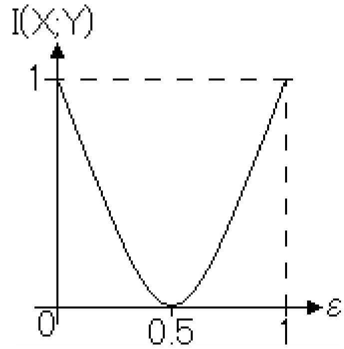
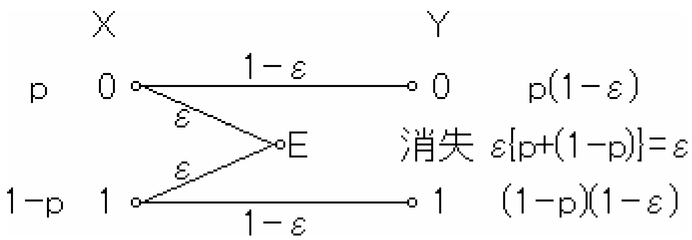


図 7　I(X;Y)と ε の関係

[2 元消失通信路]

2 元消失通信路とは，図 8 に示すような 0 と 1 の中間に消失を表す記号 E を加えた通信路である．このモデルでは 0→1，または 1→0 となる場合は考えていない．



$\mathbf{P} = \begin{pmatrix} 1-\varepsilon & \varepsilon & 0 \\ 0 & \varepsilon & 1-\varepsilon \end{pmatrix}$ ：通信路行列

図 8　2 元消失通信路

$X = \begin{pmatrix} 0 & 1 \\ p & 1-p \end{pmatrix}$

[確認問題]

図 8 の場合において，H(Y)および H(Y | X)を求めなさい．

また，I(X;Y)=H(Y)－H(Y | X)を大きくするにはどうすればよいか答えなさい．

[確認問題]

図 8 の場合において，H(Y)および H(Y | X)を求めなさい．

また，I(X;Y)=H(Y)−H(Y | X)を大きくするにはどうすればよいか答えなさい．

[確認問題の解答]

$$H(Y) = -p(1-\varepsilon)\log p(1-\varepsilon) - \varepsilon\log\varepsilon - (1-p)(1-\varepsilon)\log(1-p)(1-\varepsilon)$$

$$= -\varepsilon\log\varepsilon - (1-\varepsilon)\log(1-\varepsilon) + (1-\varepsilon)\{-p\log p - (1-p)\log(1-p)\}$$

$$H(Y|X) = p\{-(1-\varepsilon)\log(1-\varepsilon) - \varepsilon\log\varepsilon\} + (1-p)\{-\varepsilon\log\varepsilon - (1-\varepsilon)\log(1-\varepsilon)\}$$

$$= -\varepsilon\log\varepsilon - (1-\varepsilon)\log(1-\varepsilon)$$

$$I(X;Y) = (1-\varepsilon)\{-p\log p - (1-p)\log(1-p)\}$$

I(X;Y)の値は以下のようになる．

p=0.5 で I(X;Y)は最大値“ 1− ε ”

p=0 または p=1 で I(X;Y)は最小値“ 0 ”

↓

p=0.5 のとき，

ε =0 なら I(X;Y)=1

ε =1 なら I(X;Y)=0

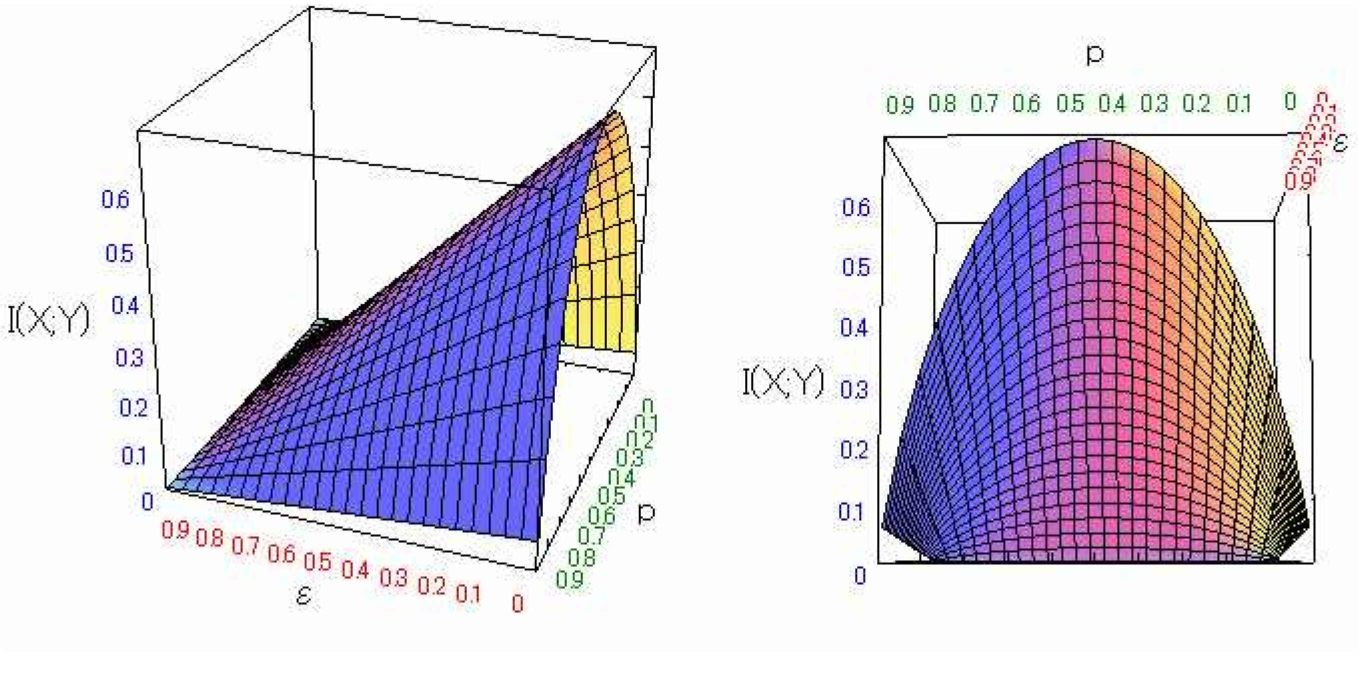


図 9　2 元消失通信路の I(X;Y)・ ε ・p の関係

＜演習問題＞

[1] 2 元対称通信路において，$P(X=0) = P(Y=0) = 0.5$，$P(Y=0|X=1) = P(Y=1|X=0) = 0.2$ なるときの相互情報量 $I(X;Y)$を求めよ．

[2] つぎのような情報源 S がある．

$$S = \begin{pmatrix} a_1 & a_2 & a_3 & a_4 & a_5 & a_6 \\ 0.05 & 0.40 & 0.10 & 0.12 & 0.18 & 0.15 \end{pmatrix}$$

(1) S のエントロピーH(S)を計算せよ．

(2) 各情報源記号をハフマン符号の符号語に符号化せよ．

(3) 平均符号長 L を求めよ．

[3] 52 枚のトランプをよく切って，任意に 1 枚取り出してハートのマークが出たら A，それ以外のマークがでたら B を出力する情報源 S がある．以下の問に答えよ．

(1) この情報源 S のエントロピーH(S)を計算せよ．

(2) この情報源から発生する情報源記号を長さ 2, および 3 の長さでブロック化してハフマン符号に符号化せよ

(3) それぞれの 1 情報源記号当たりの平均符号長を求めよ．

[4] つぎのような情報源 S を考える．

$$S = \begin{pmatrix} A & B \\ 0.1 & 0.9 \end{pmatrix}$$

この情報源 S をランレングス符号を用いて符号化する．この場合，A を区切りとして B のランレングスを用いる．以下の問に答えよ．

(1) この情報源 S のエントロピーH(S)を計算せよ．

(2) 符号語の長さを 3 とした固定長ランレングス符号で符号化した場合の 1 情報源記号当たりの平均符号長を計算せよ．

(3) N=4 とした場合，ランレングスハフマン符号で符号化せよ．また，このときに 1 情報源記号当たりの平均符号長を計算せよ．

[5] 誤り率 $\varepsilon=0.1$ の 2 元対称通信路の通信路容量を計算せよ．

[6] 「トン」と「ツー」の二つの音を用いて情報伝送を行う場合を考える．どちらの音も持続時間は 0.5s である．伝送過程において途中で雑音が発生して，それにかき消されてどちらの音か分からなくなることがある．このようなことは平均 5s に 1 回生じる．1s 当たりに受信側に届く情報量を計算せよ．

[7] つぎの通信路行列 T で与えられる通信路の通信路容量を計算せよ．

$$T = \begin{pmatrix} 0.8 & 0.1 & 0.1 \\ 0.1 & 0.8 & 0.1 \\ 0.1 & 0.1 & 0.8 \end{pmatrix}$$

[8] 今，二つの 2 元対称通信路を並列に用いて $X_1(x_{11},x_{12})$・$X_2(x_{21},x_{22})$が $Y(y_1,y_2,y_3,y_4)$に情報を送信する通信路を考えた場合，この通信路はそれぞれの通信路容量の和となることを確かめよ．それぞれの通信路容量は $C_1=1-H(\varepsilon_1)$，$C2=1-H(\varepsilon_2)$である．

＜演習問題の解答＞

[1] 2 元対称通信路において，$P(X=0) = P(Y=0) = 0.5$，$P(Y=0\,|\,X=1) = P(Y=1\,|\,X=0) = 0.2$ なるときの相互情報量 $I(X;Y)$を求めよ．

[1 の解答]

$P(Y=0) = P(Y=0\,|\,X=0)P(X=0) + P(Y=0\,|\,X=1)P(X=1) = 0.8\times 0.5 + 0.2\times 0.5 = 0.5$

$P(Y=1) = 0.5$

$H(Y) = -P(Y=0)\log_2 P(Y=0) - P(Y=1)\log_2 P(Y=1)$

$\quad = -0.5\log_2 0.5 - 0.5\log_2 0.5 = 1$

$H(Y\,|\,X) = -P(Y=0\,|\,X=0)P(X=0)\log_2 P(Y=0\,|\,X=0) - P(Y=0\,|\,X=1)P(X=1)\log_2 P(Y=0\,|\,X=1)$

$\quad - P(Y=1\,|\,X=0)P(X=0)\log_2 P(Y=1\,|\,X=0) - P(Y=1\,|\,X=1)P(X=1)\log_2 P(Y=1\,|\,X=1)$

$\quad = -0.8\times 0.5\log_2 0.8 - 0.2\times 0.5\log_2 0.2 - 0.2\times 0.5\log_2 0.2 - 0.8\times 0.5\log_2 0.8$

$\quad\quad 0.72$

$I(X;Y) = H(Y) - H(Y\,|\,X) \qquad 1 - 0.72 = 0.28$ 〔ビット／記号〕

[2] つぎのような情報源 S がある．

$$S = \begin{pmatrix} a_1 & a_2 & a_3 & a_4 & a_5 & a_6 \\ 0.05 & 0.40 & 0.10 & 0.12 & 0.18 & 0.15 \end{pmatrix}$$

(1) S のエントロピー$H(S)$を計算せよ．

(2) 各情報源記号をハフマン符号の符号語に符号化せよ．

(3) 平均符号長 L を求めよ．

[2 の解答]

(1) $H(S) = -0.05\log_2 0.05 - 0.4\log_2 0.4 - 0.1\log_2 0.1 - 0.12\log_2 0.12 - 0.18\log_2 0.18 - 0.15\log_2 0.15$

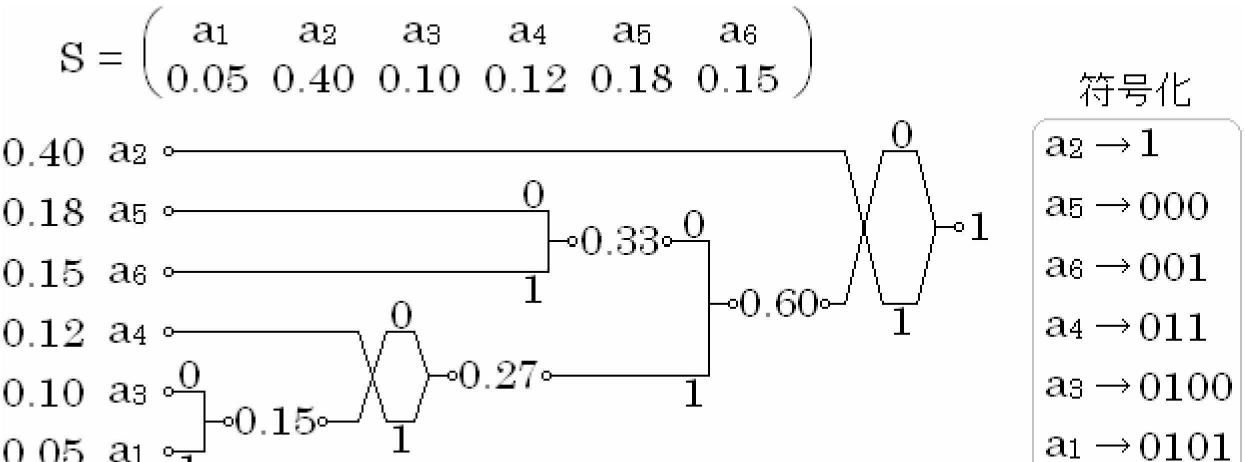
$\quad = 0.1 + 0.05\log_2 5 + 0.4\log_2 5 - 0.4 + 0.1\log_2 5 + 0.1 + 0.24\log_2 5 - 0.12\log_2 3$

$\quad + 0.36\log_2 5 + 0.18 - 0.36\log_2 3 + 0.3 + 0.15\log_2 5 - 0.15\log_2 3$

$\quad = 0.28 - 0.63\log_2 3 + 1.3\log_2 5$

$\quad\quad 2.30$ 〔ビット〕

(2) 解答を解図 1 に示す．



解図 1　各情報源記号をハフマン符号化する

(3) 平均符号長 $L = 0.40\times 1 + 0.18\times 3 + 0.15\times 3 + 0.12\times 3 + 0.10\times 4 + 0.05\times 4$

$\quad = 0.4 + 0.54 + 0.45 + 0.36 + 0.4 + 0.2 = 2.35$ 〔ビット／記号〕

[3] 52 枚のトランプをよく切って，任意に 1 枚取り出してハートのマークが出たら A，それ以外のマークがでたら B を出力する情報源 S がある．以下の問に答えよ．

(1) この情報源 S のエントロピーH(S)を計算せよ．

(2) この情報源から発生する情報源記号を長さ 2, および 3 の長さでブロック化してハフマン符号に符号化せよ

(3) それぞれの 1 情報源記号当たりの平均符号長を求めよ．

[3 の解答]

情報源 S は以下のような 2 元情報源で表される．

$$S = \begin{pmatrix} A & B \\ 0.25 & 0.75 \end{pmatrix}$$

(1) $H(S) = -0.25\log_2 0.25 - 0.75\log_2 0.75$

$= 0.5 + 1.5 - 0.75\log_2 3$

0.811 〔ビット〕

(2) 長さ 2 および長さ 3 でブロック化した情報源をそれぞれ S_2，S_3 とすると，S_2，S_3 は以下のようになる．

$$S_2 = \begin{pmatrix} AA & AB & BA & BB \\ 1/16 & 3/16 & 3/16 & 9/16 \end{pmatrix}$$

$$S_3 = \begin{pmatrix} AAA & AAB & ABA & BAA & ABB & BAB & BBA & BBB \\ 1/64 & 3/64 & 3/64 & 3/64 & 9/64 & 9/64 & 9/64 & 27/64 \end{pmatrix}$$

この二つの拡大情報源を符号化すると，解表 1 のようになる．

解表 1　ブロックハフマン符号化（長さ 2 および長さ 3 の場合）

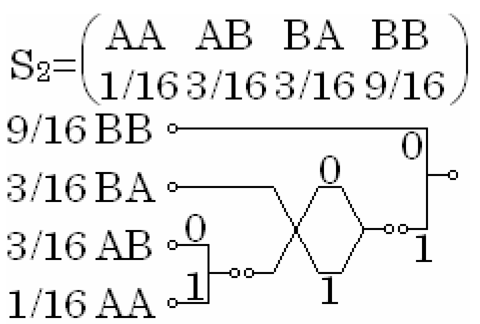
<table>
<tr><th colspan="2">長さ 2 の場合</th><th colspan="2">長さ 3 の場合</th></tr>
<tr><td rowspan="2">BB → 0</td><td rowspan="2">AB → 100</td><td>BBB → 1</td><td>AAB → 00000</td></tr>
<tr><td>BBA → 001</td><td>ABA → 00001</td></tr>
<tr><td rowspan="2">BA → 11</td><td rowspan="2">AA → 101</td><td>BAB → 010</td><td>BAA → 00010</td></tr>
<tr><td>ABB → 011</td><td>AAA → 00011</td></tr>
</table>

(3) 情報源ブロックの長さが 2 の場合および 3 の場合の平均符号長をそれぞれ L_2，L_3 とする．

$L_2 = \{ (9\times1+3\times2+3\times3+1\times3) / 16 \} / 2 = 27/32 = 0.84375$ 〔ビット／記号〕

$L_3 = \{ (27\times1+9\times3+9\times3+9\times3+3\times5+3\times5+3\times5+1\times5) / 64 \} / 3$

$= 79 / 96$　　0.823 〔ビット／記号〕



解図 2　長さが 2 の場合のブロックハフマン符号化

[4] つぎのような情報源 S を考える．

$$S = \begin{pmatrix} A & B \\ 0.1 & 0.9 \end{pmatrix}$$

この情報源 S をランレングス符号を用いて符号化する．この場合，A を区切りとして B のランレングスを用いる．以下の問に答えよ．

(1) この情報源 S のエントロピー H(S)を計算せよ．

(2) 符号語の長さを 3 とした固定長ランレングス符号で符号化した場合の 1 情報源記号当たりの平均符号長を計算せよ．

(3) N=4 とした場合，ランレングスハフマン符号で符号化せよ．また，このときに 1 情報源記号当たりの平均符号長を計算せよ．

[4 の解答]

(1) $H(S) = 0.1 \log_2 0.1 - 0.9 \log_2 0.9$

$= 0.1 + 0.1 \log_2 5 + 0.9 + 0.9 \log_2 5 - 1.8 \log_2 3$

0.469 〔ビット〕

(2) L=3 の場合 N=7 となるから，情報源ブロックの平均の長さ n_7 は以下のようになる．

$$n_7 = p + 2p(1-p) + \cdot\cdot\cdot + Np(1-p)^{N-1} + N(1-p)^N$$

$$= \frac{1-(1-p)^N}{p}$$

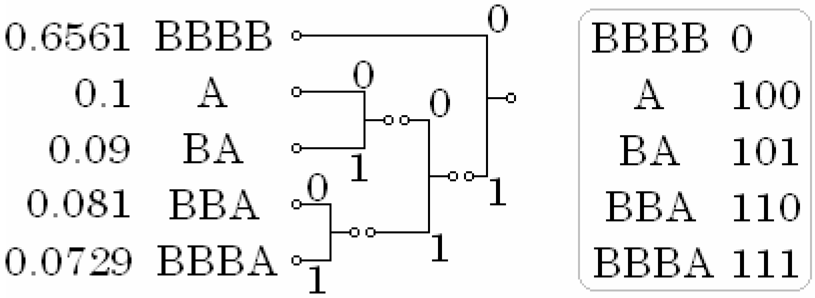
$$= \frac{1-0.9^7}{0.1} \quad 5.217$$ 〔記号／ブロック〕

従って，1 情報源記号当たりの平均符号長は以下のようになる．

$$l_7 = \frac{L}{n_7} \quad \frac{3}{5.217} \quad 0.575$$ 〔ビット／記号〕

(3) 解答を解図 3 に示す．

$$Sr = \begin{pmatrix} A & BA & BBA & BBBA & BBBB \\ 0.1 & 0.09 & 0.081 & 0.0729 & 0.6561 \end{pmatrix}$$



解図 3　ランレングスハフマン符号化

1 情報源ブロック当たりの平均符号長 L および情報源ブロックの平均の長さ n_4 は以下のようになる．

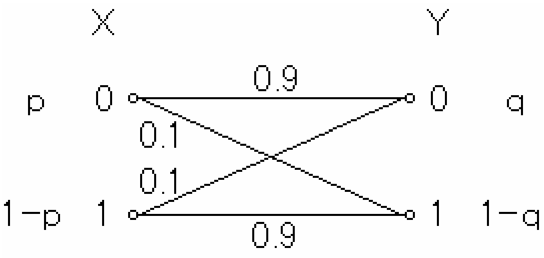
$L = 1\times0.6561 + 3\times0.1 + 3\times0.09 + 3\times0.081 + 3\times0.0729 = 1.6878$ 〔ビット／ブロック〕

$$n_4 = \frac{1-0.9^4}{0.1} = 3.439$$ 〔記号／ブロック〕

従って，1 情報源記号あたりの平均符号長は $l_4 = 1.6878/3.439$　0.491 〔ビット／記号〕となる．

[5] 誤り率 ε =0.1 の 2 元対称通信路の通信路容量を計算せよ.

[5 の解答]



解図 4　ε =0.1 のときの 2 元対称通信路

通信路容量 C = 1－H(Y | X)より，以下のようになる.

$C = 1 - H(Y|X) = 1 + 0.1\log_2 0.1 + 0.9\log_2 0.9 \quad 1 - 0.469 = 0.531$ 〔ビット／記号〕

[6] 「トン」と「ツー」の二つの音を用いて情報伝送を行う場合を考える．どちらの音も持続時間は 0.5s である. 伝送過程において途中で雑音が発生して, それにかき消されてどちらの音か分からなくなることがある. このようなことは平均 5s に 1 回生じる．1s 当たりに受信側に届く情報量を計算せよ.

[6 の解答]

これは 2 元消失通信路として表される．消失が生じる頻度は 10 回に 1 回となるから ε =0.1 である．これにより，通信路容量 C が求まる.

$C = 1 - \varepsilon = 1 - 0.1 = 0.9$ 〔ビット／記号〕

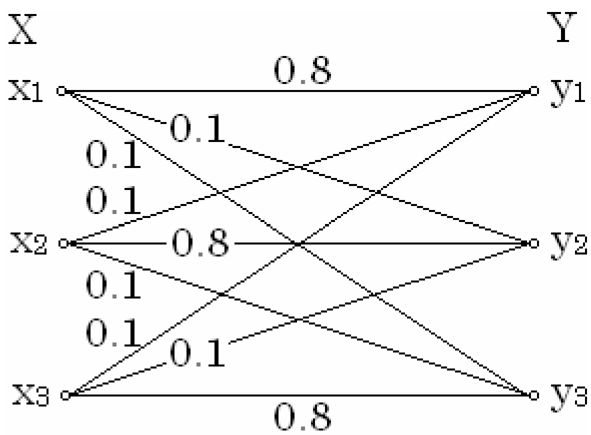
したがって，1 秒間に届く情報量は C/0.5 = 1.8 〔ビット／s〕となる.

[7] つぎの通信路行列 T で与えられる通信路の通信路容量を計算せよ.

$$T = \begin{pmatrix} 0.8 & 0.1 & 0.1 \\ 0.1 & 0.8 & 0.1 \\ 0.1 & 0.1 & 0.8 \end{pmatrix}$$

[7 の解答]

通信路に入力する記号を x_1，x_2，x_3，出力される記号を y_1，y_2，y_3 とし，確率変数をそれぞれ X，Y とする.



解図 5　3 元通信路

I(X;Y) = H(Y)－H(Y | X)であり，H(Y | X)は通信路行列 T を用いて求めることができ，入力される記号の生起確率に依存しない．入力される記号が等確率(1/3)で生起するときには，出力される記号も等確率(1/3)となるから，このとき

$H(X) = H(Y) = \log_2 3$

である．また，H(Y | X)は,

$H(Y|X) = -0.8\log_2 0.8 - 0.1\log_2 0.1 - 0.1\log_2 0.1$

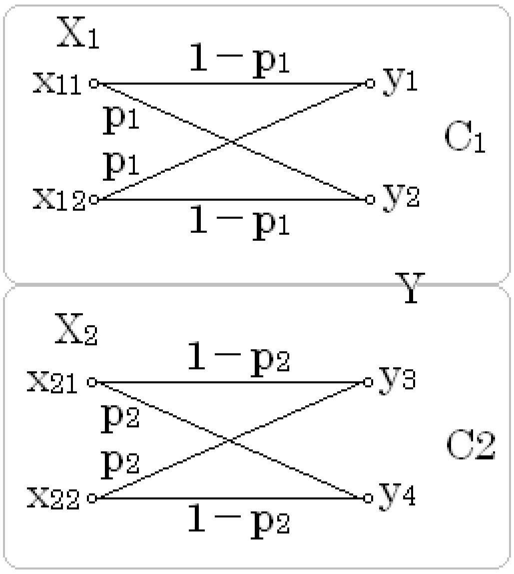
である．したがって，H(Y)が最大のときに I(X;Y)も最大になるから，通信路容量 C は,

$C = \log_2 3 + 0.8\log_2 0.8 + 0.2\log_2 0.1 \quad 0.663$〔ビット〕

となる.

[8] 今，二つの 2 元対称通信路を並列に用いて $X_1(x_{11},x_{12})$・$X_2(x_{21},x_{22})$が $Y(y_1,y_2,y_3,y_4)$に情報を送信する通信路を考えた場合，この通信路はそれぞれの通信路容量の和となることを確かめよ．それぞれの通信路容量は $C_1=1-H(\varepsilon_1)$，$C2=1-H(\varepsilon_2)$である．

[8 の解答]



解図 6　2 元対称通信路を並列で用いる

$H(Y)$を最大にすれば，$I(X;Y)$の最大値が得られる．

C_1 の出力を Y_1，C_2 の出力を Y_2 としたとき，

$$H(Y) = H(Y_1)+H(Y_2)$$

であり，

$$I(X;Y) = I(Y_1;X_1)+I(Y_2;X_2)$$
$$I(Y_1;X_1) = H(Y_1)-H(Y_1|X_1)$$
$$I(Y_2;X_2) = H(Y_2)-H(Y_2|X_2)$$

となる．また，$H(Y_1)$，$H(Y_2)$は X_1，X_2 が等確率で送信記号を発生するときに受信記号 Y も等確率となり，このときに，

$$H(Y) = 1+1 = 2$$

で最大値となる．$H(Y|X)$を計算すると，

$$H(Y|X) = -(1-p_1)\log_2(1-p_1)-p_1\log_2 p_1-(1-p_2)\log_2(1-p_2)-p_2\log_2 p_2$$
$$= H(Y_1|X_1)+H(Y_2|X_2)$$

となるから，通信路容量は以下のようになる．

$$\begin{aligned} C &= 2-H(Y|X) \\ &= 2-H(Y_1|X_1)-H(Y_2|X_2) \\ &= 1-H(Y_1|X_1)+1-H(Y_2|X_2) \\ &= C_1+C_2 \end{aligned}$$

$C = \max I(X;Y)$

→ $C = 1-H(Y|X)$　（2 元対称通信路の場合）

→ $C = 1-\varepsilon$　（2 元消失通信路の場合）