10. 通信路符号化法 ― 単一誤りの検出と訂正

通信路符号化の枠組みを導入し，ノイズのある離散通信路を使って情報を伝送する符号化の基本的な手法を紹介する．

10.1 通信路符号化の枠組み

通信路の入出力アルファベットは$\{0,1\}$とし，通信路符号化の枠組みを図 1 に示す．送ろうとする情報語$(x_1,\cdots,x_k)$ $(x_i \in \{0,1\})$を適宜符号化して符号語$(w_1,\cdots,w_n)$ $(w_i \in \{0,1\})$を作り出して，通信路に送り込む．通常は，情報語から計算される検査ビットと呼ばれる$n-k$個のビットを情報語に追加してnビットにする．このようなタイプの符号を(n,k)符号と呼ぶ．

ここで使用している通信路モデルでは，通信路で誤り$e=(e_1,\cdots,e_n)$が混入して，通信路からは$y=(y_1,\cdots,y_n)=(w_1+e_1,\cdots,w_n+e_n)$が出力される（＋の演算は後述する GF(2)による．ここでは，単に符号語の各ビットにe_iが「加算」されるという理解でよい）．シンドローム計算では通信路から出力された受信語から「シンドローム」という値を算出する．

誤り検出符号の場合は，誤りがあればシンドロームの値は0なければ1となるようにシンドロームの計算法が設計される．誤り訂正符号の場合は，シンドロームの値が0のときは誤りなし，0以外の値のときは，そこから誤っているビットの位置が計算できるように設計される．2元符号が使われているときは，出力アルファベットは$\{0,1\}$しかないので，誤り位置として指定された位置のビットの値が0ならば，入力側の情報ビットの値は1であり，その位置のビットの値が1であればその情報ビットの値は 0 であったということになる．

m誤り訂正符号では，実際に生じた誤りの個数がm個以下であったときは，このようにして訂正することによって正しい情報語が復元されることが保証されている．

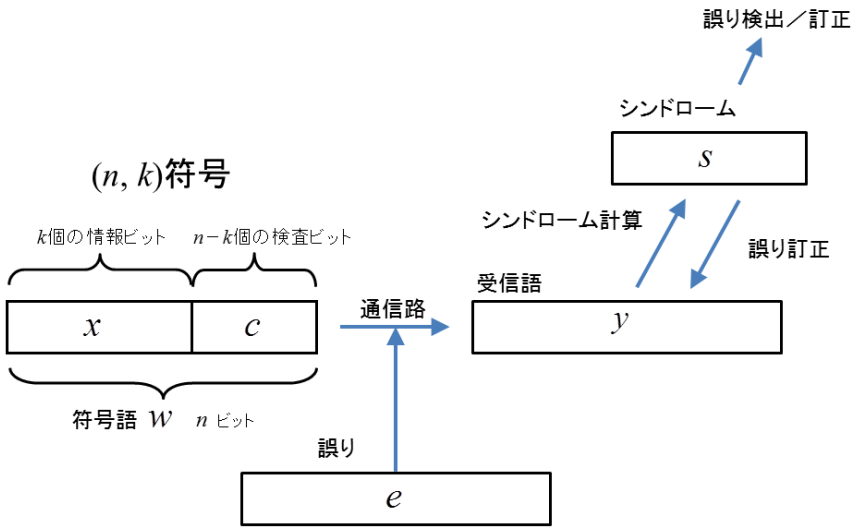

図 1．通信路符号化の枠組み

10.2 単一パリティ検査

最も原始的な符号化法である単一パリティ検査について紹介する.

① 符号化：情報語$(x_1,\cdots,x_k)$が与えられたとき,

$$c=\begin{cases}0 & ;\text{値が 1 の情報ビットの個数が偶数のとき}\\ 1 & ;\text{値が 1 の情報ビットの個数が奇数のとき}\end{cases}$$

となるようにcを求め，$w=(w_1,\cdots,w_{n-1},w_n)=(x_1,\cdots,x_k,c)$とする．ただし，$k=n-1$.

$c=x_1+\cdots+x_k=\sum_{i=1}^{k}x_i$とも書ける．ここで，+は$\mathrm{GF}(2)=\{0,1\}$の上での加算であり,

$$0+0=1+1=0$$
$$0+1=1+0=1$$

によって定義される.

これによって長さが$k+1$で，1の個数が偶数であるような全ての2元系列からなる符号長$n=k+1$，符号語数 $M=2^k$の符号語の集まりができる.

例えば，$k=2$のとき，$C=\{000,011,101,110\}$である.

② 復号：通信路からの出力$y=(y_1,\cdots,y_k,y_n)$に対して$s=y_1+\cdots+y_n$を計算する.

$$\begin{aligned}s&=y_1+\cdots+y_n\\&=(w_1+e_1)+\cdots+(w_n+e_n)\\&=(y_1+\cdots+y_n)+(e_1+\cdots+e_n)\\&=(x_1+\cdots+x_k+c)+(e_1+\cdots+e_n)\\&=(x_1+\cdots+x_k+x_1+\cdots+x_k)+(e_1+\cdots+e_n)\\&=e_1+\cdots+e_n\end{aligned}$$

であるから，誤りが高々1個しか生じていないと仮定できる時は，誤りが生じたか否かを検出できる．すなわち，単一パリティ誤り検査符号において単一誤りが検出可能である．しかし，そのときどこに誤りが起きていたかまでは導けないので，誤り訂正は可能ではない．例えば，$k=2$のとき100が受信されたと，誤りが高々一つであると仮定したとしても元の符号語が，000，110，101のいずれであったのかを特定することはできない.

10.3 水平垂直パリティ検査符号

水平垂直パリティ検査は，二つの正整数s,tに対して，情報ビット数$s\times t$に以下のようにして作られる$s+t+1$個の検査ビットを加えて構成される$((s+1)(t+1),st)$符号であり，1個の誤りを訂正できる.

$s=3,t=4$の場合を例にとって説明しよう．送信する情報ビット数は$3\times4=12$ビットであり，例えば，(0,1,1,0,1,0,1,1,0,0,0,1)．これを3×4に折りたたんで図 2(a)のように表示する.

$x_1: 0$	$x_2: 1$	$x_3: 1$	$x_4: 0$
$x_5: 1$	$x_5: 0$	$x_7: 1$	$x_8: 1$
$x_9: 0$	$x_{10}: 0$	$x_{11}: 0$	$x_{12}: 1$

(a) 情報ビットの例

$w_1 = x_1: 0$	$w_2 = x_2: 1$	$w_3 = x_3: 1$	$w_4 = x_4: 0$	$w_{13} = c_1: 0$
$w_5 = x_5: 1$	$w_6 = x_6: 0$	$w_7 = x_7: 1$	$w_8 = x_8: 1$	$w_{14} = c_2: 1$
$w_9 = x_9: 1$	$w_{10} = x_{10}: 0$	$w_{11} = x_{11}: 0$	$w_{12} = x_{12}: 1$	$w_{15} = c_3: 1$
$w_{16} = c_4: 1$	$w_{17} = c_5: 1$	$w_{18} = c_6: 0$	$w_{19} = c_7: 1$	$w_{20} = c_8: 0$

(b)　パリティの与え方

y_1	y_2	y_3	y_4	y_{13}
y_5	y_6	y_7	y_8	y_{14}
y_9	y_{10}	y_{11}	y_{12}	y_{15}
y_{16}	y_{17}	y_{18}	y_{19}	y_{20}

(c) 受信語の例

図 2．水平垂直パリティ検査符号

符号化を行うときは，図 2(b)のように縦，横それぞれにパリティビットを付け，さらに全体についてもパリティビットをつける．

復号時には，縦横についてシンドロームを計算する．すなわち図 2(c)を受け取ったとき，シンドロームとして

$$s_1 = y_1 + y_2 + y_3 + y_4 + y_{13}$$
$$s_2 = y_5 + y_6 + y_7 + y_8 + y_{14}$$
$$s_3 = y_9 + y_{10} + y_{11} + y_{12} + y_{15}$$
$$s_4 = y_{16} + y_{17} + y_{18} + y_{19} + y_{20}$$
$$s_5 = y_1 + y_5 + y_9 + y_{16}$$
$$s_6 = y_2 + y_6 + y_{10} + y_{17}$$
$$s_7 = y_3 + y_7 + y_{11} + y_{18}$$
$$s_8 = y_4 + y_8 + y_{12} + y_{19}$$
$$s_9 = y_{13} + y_{14} + y_{15} + y_{20}$$

を計算する．

通信路で，誤りが生じていなければ，$k = 1, \cdots, 9$に対して，$s_k = 0$である．誤りが1個であれば，$\{s_1, \cdots, s_9\}$のなかのちょうど2個が1になり，誤りが生じたこととどこでそれが生じたかを一意に同定できる．例えば，y_{11}に誤りが起きて，$y_{11} = 1$になったとしたら，s_3とs_7の値が1になり，両者の計算に含まれるy_{11}に誤りが起きたことがわかる．

以上を一般化しよう．水平垂直パリティ検査符号では，s,t個の情報ビットを$s\times t$個の配列に並べ，すべての行および列の 1 の数が偶数になるように$s+t+1$個の検査ビットを作る（図 3）．

① 符号化：情報語$(x_1,\cdots,x_{st})$が与えられたとき，

$$c_k=\begin{cases}\displaystyle\sum_{i=1}^{t} x_{(k-1)t+i} & (1\le k\le s)\\ \displaystyle\sum_{i=1}^{t} x_{(i-1)t+k-s} & (s+1\le k\le s+t)\\ \displaystyle\sum_{i=1}^{s} c_i=\sum_{i=1}^{t} c_{s+i} & (k=s+t+1)\end{cases}$$

となるように$c_k\ (k=1,\cdots,s+t+1)$を求め，$w=(x_1,\cdots,x_{st},c_1,\cdots,c_{s+t+1})$とする．

②　復号：通信路からの出力$y=(y_1,\cdots,y_{st},y_{st+1},\cdots,y_{st+s+t+1})$に対して次のようにシンドロームを計算する．

$$s_k=\begin{cases}y_{st+k}+\displaystyle\sum_{i=1}^{t} y_{(k-1)t+i} & (1\le k\le s)\\ \displaystyle\sum_{i=1}^{t+1} y_{s(t+1)+i} & (k=s+1)\\ y_{st+k-1}+\displaystyle\sum_{i=1}^{s} y_{(i-1)t+k-s-1} & (s+2\le k\le s+t+1)\\ y_{st+s+t+1}+\displaystyle\sum_{i=1}^{s} y_{st+i} & (k=s+t+2)\end{cases}$$

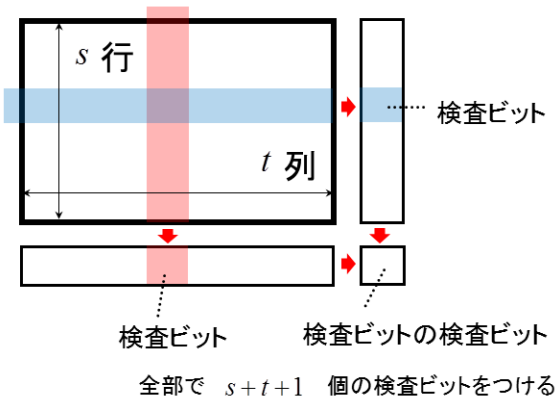

図 3. 水平垂直パリティ検査符号の原理

【問題】

水平垂直パリティ検査符号によって，2誤り検出ができることを示せ．また，2誤り訂正しようとしてもできない場合があることを示せ．

10.4 ハミング符号

ハミング符号は，水平垂直パリティ検査符号より効率のよい単一誤り訂正符号である．ここで，(15,11)ハミング符号を例にとって，その構成法を説明する．

① 符号化：与えられた情報語を$(x_1,x_2,x_3,x_4,x_5,x_6,x_7,x_8,x_9,x_{10},x_{11})$としよう．検査ビット$(c_1,c_2,c_3,c_4)$を次のように構成する．

$$c_1 = x_1 + x_2 + x_3 + x_4 \qquad + x_9 + x_{10} + x_{11}$$
$$c_2 = x_1 + x_2 + x_3 \quad + x_5 \quad + x_7 + x_8 \qquad + x_{11}$$
$$c_3 = x_1 + x_2 \quad + x_4 + x_5 + x_6 \quad + x_8 \quad + x_{10}$$
$$c_4 = x_1 \quad + x_3 + x_4 + x_5 + x_6 + x_7 \quad + x_9$$

により検査ビットをつくり，

$$w = (x_1,x_2,x_3,x_4,x_5,x_6,x_7,x_8,x_9,x_{10},x_{11},c_1,c_2,c_3,c_4)$$

とする．

このようにして作られた符号語

$$w = (w_1,w_2,w_3,w_4,w_5,w_6,w_7,w_8,w_9,w_{10},w_{11},w_{12},w_{13},w_{14},w_{15})$$

は，

$$w_1 + w_2 + w_3 + w_4 \qquad + w_9 + w_{10} + w_{11} + w_{12} \qquad = 0$$
$$w_1 + w_2 + w_3 \quad + w_5 \quad + w_7 + w_8 \qquad + w_{11} \quad + w_{13} \qquad = 0$$
$$w_1 + w_2 \quad + w_4 + w_5 + w_6 \quad + w_8 \quad + w_{10} \qquad + w_{14} \quad = 0$$
$$w_1 \quad + w_3 + w_4 + w_5 + w_6 + w_7 \quad + w_9 \qquad + w_{15} = 0$$

を満足する．

② 復号：

受信語$y=(y_1,y_2,y_3,y_4,y_5,y_6,y_7,y_8,y_9,y_{10},y_{11},y_{12},y_{13},y_{14},y_{15})$に対して，

$$s_1 = y_1 + y_2 + y_3 + y_4 \qquad + y_9 + y_{10} + y_{11} + y_{12}$$
$$s_2 = y_1 + y_2 + y_3 \quad + y_5 \quad + y_7 + y_8 \qquad + y_{11} \quad + y_{13}$$
$$s_3 = y_1 + y_2 \quad + y_4 + y_5 + y_6 \quad + y_8 \quad + y_{10} \qquad + y_{14}$$
$$s_4 = y_1 \quad + y_3 + y_4 + y_5 + y_6 + y_7 \quad + y_9 \qquad + y_{15}$$

により，シンドロームを計算する．
通信路で誤りが生じていなければ，s_1=s_2=s_3 = s_4 = 0となる．

誤りが1個w_kのところに生じた場合，すなわち，$y_i = w_i + e_i, e_k = 1, e_i = 0\ (i \neq k)$のときは，$y_k$を算入しているシンドロームのみその値が1になる．例えば，誤りがw_7のところに生じた場合は，y_7を算入しているシンドロームs_2,s_4の値が1になり，そうでないs_1,s_3の値は0である．

実は，誤りの生じ得る位置1～15に対してシンドローム(s_1,s_2,s_3,s_4)の値のパターンがすべて異なるように，シンドロームの算出式を作っていた．そのため，生じる誤りの個数が高々1個である場合は，シンドロームの値のパターンから誤りが生じたか否か，1個の誤りが生じた場合には，どの位置に生じたか特定できる．ここでは，2元符号を用いる仮定していたので，誤りの生じた位置が特定できれば，通信路に入力された符号語を復元できる．

生成行列と検査行列

以上述べたことを GF(2)上の行列演算として表現できる．上に述べた(15,11)ハミング符号を例にとって説明する．与えられた情報語を

$$(x_1,x_2,x_3,x_4,x_5,x_6,x_7,x_8,x_9,x_{10},x_{11})$$

としよう．すると，

$$\begin{aligned} w = (&x_1, x_2, x_3, x_4, x_5, x_6, x_7, x_8, x_9, x_{10}, x_{11},\\ &x_1 + x_2 + x_3 + x_4 + x_9 + x_{10} + x_{11},\\ &x_1 + x_2 + x_3 + x_5 + x_7 + x_8 + x_{11},\\ &x_1 + x_2 + x_4 + x_5 + x_6 + x_8 + x_{10},\\ &x_1 + x_3 + x_4 + x_5 + x_6 + x_7 + x_9) \end{aligned}$$

と符号語を構成するという演算を，生成行列

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

という行列を使って，情報語を表す横ベクトル$x=(x_1,x_2,x_3,x_4,x_5,x_6,x_7,x_8,x_9,x_{10},x_{11})$から，符号語$w=(w_1,w_2,w_3,w_4,w_5,w_6,w_7,w_8,w_9,w_{10},w_{11},w_{12},w_{13},w_{14},w_{15})$を表す横ベクトルを作り出す演算：

$$w = xG$$

で作り出せるものとして表現できる．

同様に，受信語$y=(y_1,y_2,y_3,y_4,y_5,y_6,y_7,y_8,y_9,y_{10},y_{11},y_{12},y_{13},y_{14},y_{15})$に対して，シンドロームのベクトル$(s_1,s_2,s_3,s_4)$:

$$\begin{aligned}
s_1 &= y_1 + y_2 + y_3 + y_4 \qquad\quad + y_9 + y_{10} + y_{11} + y_{12} \\
s_2 &= y_1 + y_2 + y_3 \quad + y_5 \quad + y_7 + y_8 \quad\; + y_{11} \quad + y_{13} \\
s_3 &= y_1 + y_2 \quad + y_4 + y_5 + y_6 \quad + y_8 \quad + y_{10} \qquad + y_{14} \\
s_4 &= y_1 \quad + y_3 + y_4 + y_5 + y_6 + y_7 \quad + y_9 \qquad\quad + y_{15}
\end{aligned}$$

を作り出す演算については，検査行列

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

を用いて

$$s = yH^{\mathrm{T}} = (y_1, \cdots, y_{15}) \begin{bmatrix}
1 & 1 & 1 & 1 \\
1 & 1 & 1 & 0 \\
1 & 1 & 0 & 1 \\
1 & 0 & 1 & 1 \\
0 & 1 & 1 & 1 \\
0 & 0 & 1 & 1 \\
0 & 1 & 0 & 1 \\
0 & 1 & 1 & 0 \\
1 & 0 & 0 & 1 \\
1 & 0 & 1 & 0 \\
1 & 1 & 0 & 0 \\
1 & 0 & 0 & 0 \\
0 & 1 & 0 & 0 \\
0 & 0 & 1 & 0 \\
0 & 0 & 0 & 1
\end{bmatrix}$$

と表現できる．

一般のハミング符号

単一誤りが訂正できるためには，検査行列の全ての列が互いに異なり，全零でなければよい（図 4）．全零の列があると，そこで生じた誤りと，誤りなしの場合が区別できないからである．

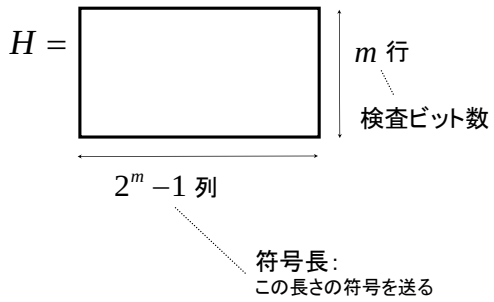

図 4．ハミング符号の検査行列

以上から，ハミング符号は，

符号長　$n = 2^m - 1$

情報ビット数　$k = 2^m - 1 - m$

検査ビット数　m

によって規定される，$(2^m - 1, 2^m - 1 - m)$符号となる．

演習 1

通信路行列$\begin{bmatrix} 1-p & p \\ p & 1-p \end{bmatrix}$ $\left(0 < p < \frac{1}{2}\right)$で規定される無記憶 2 元対称通信路がある．与えられた長さ5の情報ビット列（ただし，m=(m_1,m_2,m_3,m_4,m_5) $(m_i \in \{0,1\}, 1 \le i \le 5)$に対して，行列

$$A = \begin{bmatrix} a_{11} & a_{12} & a_{13} & a_{14} & a_{15} \\ a_{21} & a_{22} & a_{23} & a_{24} & a_{25} \\ a_{31} & a_{32} & a_{33} & a_{34} & a_{35} \\ a_{41} & a_{42} & a_{43} & a_{44} & a_{45} \end{bmatrix} \left(a_{ij} \in \{0,1\}, 1 \le i \le 4, 1 \le j \le 5\right)$$

を用いて，$1 \le i \le 5$に対しては，$x_i = m_i$，$6 \le i \le 9$に対しては，$x_i = \sum_{j=1}^{5} a_{i-5,j} m_j$（演算はブール代数に基づいて行うものとする）によって定まる通信路符号語 x=$(x_1,x_2,x_3,x_4,x_5,x_6,x_7,x_8,x_9)$を伝送する．

復号(decoding)は，Aを次のように拡張した行列H

$$H=\begin{bmatrix} a_{11} & a_{12} & a_{13} & a_{14} & a_{15} & 1 & 0 & 0 & 0 \\ a_{21} & a_{22} & a_{23} & a_{24} & a_{25} & 0 & 1 & 0 & 0 \\ a_{31} & a_{32} & a_{33} & a_{34} & a_{35} & 0 & 0 & 1 & 0 \\ a_{41} & a_{42} & a_{43} & a_{44} & a_{45} & 0 & 0 & 0 & 1 \end{bmatrix}$$

を用いて，通信路からの出力$y=(y_1,y_2,y_3,y_4,y_5,y_6,y_7,y_8,y_9)$から計算したシンドローム$g=(g_1,g_2,g_3,g_4)^{\mathrm{T}}=Hy^{\mathrm{T}}$を用いて行うものとする．以下では，さらに行列$A$は次のような行列$B$に等しいとする．

$$A=B=\begin{bmatrix} b_4 & b_5 & b_6 & b_7 & b_8 \\ b_3 & b_4 & b_5 & b_6 & b_7 \\ b_2 & b_3 & b_4 & b_5 & b_6 \\ b_1 & b_2 & b_3 & b_4 & b_5 \end{bmatrix} (b_i \in \{0,1\}, 1 \le i \le 8)$$

このとき，次の問いに答えよ．

1. 最尤復号法による復号を行うにはどうすればよいか説明せよ．また，それによって，$(b_1,b_2,b_3,b_4,b_5,b_6,b_7,b_8)=(1,0,1,0,1,1,0,1)$のとき，出力語$(0,1,0,0,1,1,0,0,1)$と$(0,0,1,0,1,0,0,0,1)$がそれぞれどのように復号されるかを示せ．ただし，通信路への入力となる情報ビット列$m=(m_1,m_2,m_3,m_4,m_5)$の生起確率は一様に等しいとする．

2. $(b_1,b_2,b_3,b_4,b_5,b_6,b_7,b_8)=(1,0,1,0,1,1,0,1)$とするとき，この通信路符号の誤り訂正能力をその根拠とともに示せ．

3. $b_1,b_2,b_3,b_4,b_5,b_6,b_7,b_8$の値がそれぞれ独立に$\{0,1\}$のなかから$\frac{1}{2}$の確率で決められるとき，相異なる$(s_1,s_2,s_3,s_4,s_5,s_6,s_7,s_8,s_9)$と$(s_1',s_2',s_3',s_4',s_5',s_6',s_7',s_8',s_9')$（ただし，$s_i,s_i' \in \{0,1\}, 1 \le i \le 9$）が同じシンドローム$g=(g_1,g_2,g_3,g_4)^{\mathrm{T}}=Hy^{\mathrm{T}}$を与える確率のアンサンブル平均を求めよ．

4. $b_1,b_2,b_3,b_4,b_5,b_6,b_7,b_8$の値がそれぞれ独立に$\{0,1\}$のなかから$\frac{1}{2}$の確率で決められるとするとき，最尤復号法を用いた復号で誤りが発生する確率のアンサンブル平均$\overline{P}(e)$の上限を求めよ．