The Hamming Codes

Mitsuru Takigahira

TL;DR

ハミング $[7, 4]$ 符号 $\mathcal{H}_7$ は 1 重誤り訂正完全 2 元符号で、伝送速度 $\frac{4}{7}$ である。
実際これは 1 重誤り訂正 2 元符号の無限列のうちの 1 つであり、
この列は符号長 n が増えるほど伝送速度 R は 1 に近づく。
この符号はハミングによって 1950 年に考案された[1]が、
ゴレイも同時期に独自にこれらを発見した
(どちらが先に発見されたかの議論については [Th83] を見よ。)

[1] [Ha50]

ハミング符号の構成

1 重誤り訂正 2 元符号に対して、ハミングの球充填限界式 (系 6.17) において $t = 1$ 及び $q = 2$ とおくため、 完全符号に対して以下の条件が成立する。

$$2^{n-k} = 1 + \binom{n}{1} = n + 1$$

$c = n - k$ (検査桁の桁数) とおけば、この条件は以下の式と同値である

$$n = 2^c - 1 \tag{7.4}$$

$k = n - c = 2^c - 1 - c$ なので、n 及び k のとりうる値は次のようになる。

c	$=$	1	2	3	4	5	$\cdots$
n	$=$	1	3	7	15	31	$\cdots$
k	$=$	0	1	4	11	26	$\cdots$

ハミング符号の構成

このようなパラメータに対して符号を構成しよう。
$t = 1$ とおき、系 7.31 からそのような符号 $\mathcal{C}$ が存在することは

- 階数 c
- すべての 2 つの列の組が線形独立となっている

なる $c \times n$ 行列 H が $\mathbf{F}_2$ 上に存在することと同値で、$\mathbf{F} = \mathbf{F}_2 = \{0, 1\}$ より、

- H の任意の列 $\mathbf{c}_i$ が非零
- H のすべての列が互いに異なっている

つまり H は互いに異なる $n = 2^c - 1$ 個の非零な長さ c の列ベクトルからなる。
このとき、2^c 個の相異なる長さ c の 2 元ベクトルのみが存在し、列 $\mathbf{c}_i$ の選択は存在しない。これらは何らかの順の非零な長さ c の列すべて ($2^c - 1$ 個) になる。
($c = 3$ の場合については例 7.3 を見よ)
これらの列ベクトルは c 個の標準基底ベクトルを含み、それらは線形独立である。
よってこの行列 H は階数 c を持つ。ここから $\mathcal{C}$ の存在が示され、
これらのパラメータを持つどんな 2 つの線形符号も (列の交換の下) 同値となる。
$\mathcal{C}$ を符号長 $n = 2^c - 1$ の 2 元ハミング符号 $\mathcal{H}_n$ と呼ぶ。
(厳密にはここでは符号長 n の符号の集合を構成したが、それらはすべて同値で、単一の符号 $\mathcal{H}_n$ とみなされる傾向がある。)

例

例 7.32

$\mathcal{H}_1$ は符号長 1 の単一の符号語 0 からなるつまらないケースなので、
$c = 1$ については無視する。 $c = 2$ のとき、$n = 3$ で、

$$H = \begin{pmatrix} 0 & 1 & 1 \\ 1 & 0 & 1 \end{pmatrix};$$

なので、$\mathcal{H}_3$ は $\mathbf{v} = v_1 v_2 v_3$ に対して、$v_1 + v_3 = v_2 + v_3 = 0$
つまり $v_1 = v_2 = v_3$ なる 2 元の符号語からなり、これは 2 元反復符号 $\mathcal{R}_3$ である。
$c = 3$ の場合、既に考えたハミング符号 $\mathcal{H}_7$ を得る。
$c \geq 4$ に対して、符号長 $n = 2^c - 1$ の新たな完全符号 $\mathcal{H}_n$ を無限に得る。
これらの符号は $c \to \infty$ としたとき、伝送速度

$$R = \frac{k}{n} = \frac{2^c - 1 - c}{2^c - 1} \to 1$$

となるが、それらはただ 1 つの誤りしか訂正しない。よって $\mathrm{Pr}_E \not\to 0$ である
(練習 7.4 を見よ)

練習

練習 7.4

2 元ハミング符号 $\mathcal{H}_7$ に対して、Γ を $P > \frac{1}{2}$ の BSC、Δ を最近傍復号とおくとき、Pr_E を示せ。$n \to \infty$ のとき Pr_E はどうなるか？

解答

$$
\begin{aligned}
\mathrm{Pr}_E &= 1 - \sum_{i=0}^{1} \binom{n}{i} P^{n-i}(1-P)^i \\
&= 1 - \left(P^n(1-P)^0 + nP^{n-1}(1-P)^1\right) \\
&= 1 - P^n - nP^{n-1}(1-P)
\end{aligned}
$$

となり、$P < 1$ のとき、$\mathrm{Pr}_E \to 1\ (n \to \infty)$、$P = 1$ のとき、$\mathrm{Pr}_E \to 0\ (n \to \infty)$

ハミング符号の最近傍復号

ハミング符号の最近傍復号はとても簡単である。

$\mathcal{H}_n$ は $t = 1$ の完全符号で、すべての重み 1 までの誤りパターン $\mathbf{e}$ を訂正する。
$\mathbf{u} \in \mathcal{H}_n$ が伝送され、$\mathrm{wt}(\mathbf{e}) = 1$ として $\mathbf{v} = \mathbf{u} + \mathbf{e}$ が受信されたと仮定する。
よって $\mathbf{e} = \mathbf{0}$ あるいは $\mathbf{e}$ は $\mathbf{e}_i \in \mathcal{V}$ の標準基底ベクトル $\mathbf{e}_i$ となる。

受信者は $\mathbf{s} = \mathbf{v}H^T$ を計算し、これを $\mathbf{v}$ のシンドローム[2]と呼ぶ。今、
$\mathbf{v}H^T = (\mathbf{u} + \mathbf{e})H^T = \mathbf{u}H^T + \mathbf{e}H^T = \mathbf{e}H^T$ ($\mathbf{u}H^T = \mathbf{0}$ なので) で、
かつこれは $\mathbf{e} = \mathbf{0}$ 及び $\mathbf{e} = \mathbf{e}_i$ に対してそれぞれ $\mathbf{0}$ または $\mathbf{c}_i^T$ となる。

もし $\mathbf{s} = \mathbf{0}$ ならば、受信者は $\mathbf{v}$ を $\Delta(\mathbf{v}) = \mathbf{v} = \mathbf{u}$ として復号する
$\mathbf{s} = \mathbf{c}_i^T$ ならば、$\Delta(\mathbf{v}) = \mathbf{v} - \mathbf{e}_i$ となり $\mathbf{v}$ の i 番目を変更することで得られる。
この方法は $\mathrm{wt}(\mathbf{e}) \leq 1$ のときは常に正しく復号するが、
$\mathrm{wt}(\mathbf{e}) > 1$ の場合一切正しく復号しない。
何らかの $\mathbf{u}' \in \mathcal{H}_n$ $(\mathbf{u}' \neq \mathbf{u})$ に対し、$\mathrm{wt}(\mathbf{e}') \leq 1$ のもと $\mathbf{v} = \mathbf{u}' + \mathbf{e}'$ のとき、
上記のアルゴリズムは $\mathbf{v}$ を $\Delta(\mathbf{v}) = \mathbf{u}'$ として誤って復号してしまう。

[2]医学用語で、シンドロームとは患者のどこが悪いかの兆候のこと

例

例 7.33

以下の組織符号形式のパリティ検査行列を持つ $\mathcal{H}_7$ を用いる。

$$\begin{pmatrix} 0 & 1 & 1 & 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 1 & 0 & 1 & 0 \\ 1 & 1 & 0 & 1 & 0 & 0 & 1 \end{pmatrix}$$

$\mathbf{u} = 1101001 \in \mathcal{H}_7$ が送信され、$\mathbf{v} = 1101101 \in \mathcal{V}$ が受信されたとする。
つまり誤りパターンは $\mathbf{e} = \mathbf{e}_5$ であり、シンドロームは $\mathbf{s} = \mathbf{v}H^T = 100$ で、
H の 5 番目の列の転置 $\mathbf{c}_5^T$ である。これは 5 番目に誤りがあることを示していて、
$\Delta(\mathbf{v}) = 1101001 = \mathbf{u}$ を得る。
一方で、$\mathbf{v}' = 1001101$ が受信された場合、誤りパターンは $\mathbf{e}' = \mathbf{e}_2 + \mathbf{e}_5$ で、
シンドロームは $\mathbf{s}' = 001 = \mathbf{c}_7^T$ となり、誤りの位置が 7 番目にあることを示し、
$\Delta(\mathbf{v}) = 1001100 \neq \mathbf{u}$ となる。
このように、2 つの誤りを訂正する代わりに符号は 3 番目の誤りを作ってしまう。

練習

練習 7.5

例 7.33 のパリティ検査行列を用いて $\mathbf{u} = 1100110$ が $\mathcal{H}_7$ の符号語かを検査せよ。
この符号語 $\mathbf{u}$ が送信され、$\mathbf{v} = 1000110$ が受信されたと仮定し、
シンドロームと $\Delta(\mathbf{v})$ を求めよ。
$\mathbf{v}' = 0000110$ が受信された場合何が起こるかを調べ、説明せよ。

解答

板書する。

ハミング符号の復号

H の各列 $\mathbf{c}_i$ の順番 $\mathbf{c}_1^T, \ldots \mathbf{c}_n^T$ が $i = 1, \ldots, n$ の 2 進数表現となっている場合、
$\mathcal{H}_7$ の復号は特に簡単で、$n = 7$ の場合 H は例 7.13 にあるような以下の形になる

$$\begin{pmatrix} 0 & 0 & 0 & 1 & 1 & 1 & 1 \\ 0 & 1 & 1 & 0 & 0 & 1 & 1 \\ 1 & 0 & 1 & 0 & 1 & 0 & 1 \end{pmatrix}$$

これは置換 (1362547) を例 7.33 で用いられたパリティ検査行列の列に対して
適用したものと同値である。
シンドローム $\mathbf{s} = \mathbf{0}$ のとき $\mathbf{e} = \mathbf{0}$ と解釈され、つまり誤りは無く、
単一の誤り $\mathbf{e}_i$ が発生した場合、非零のシンドローム $\mathbf{s}$ は誤りの位置 i の
2 進数表現となっている。

例

例 7.34

先程挙げたパリティ検査行列で定義される $\mathcal{H}_7$ と同値な符号を用いる。
置換 (1362547) を例 7.33 で用いられた符号語 1101001 に適用すると、
$\mathbf{u} = 1010101 \in \mathcal{H}_7$ を得る。
これが送信され $\mathbf{v} = 1010001$ が受信された場合、
シンドロームは $\mathbf{s} = \mathbf{v}H^T = 101$ で、 これは 5 の 2 進数表現であり、
よって $\mathbf{v}$ の 5 番目を変更することで $\Delta(\mathbf{v}) = 1010101 = \mathbf{u}$ を得る。

練習

練習 7.6

例 7.34 で用いた $\mathcal{H}_7$ の同値な符号を用いる。
$\mathbf{u} = 0111100$ が送信され $\mathbf{v} = 0011100$ が送信された場合、どうなるかを説明せよ。

解答

板書する。

$q > 2$ の場合の完全符号

素数の類乗数 $q > 2$ に対しても似たような完全 1 重誤り訂正線形符号の構成が存在する。 H の列を

$$n = \frac{q^c - 1}{q - 1} = 1 + q + q^2 + \cdots + q^{c-1} \tag{7.5}$$

対の線形独立な長さ c の $\mathbf{F}_q$ 上のベクトルとなるようにとる。
(これは可能な最大数である。練習 7.7 を見よ)
結果得られた線形符号は符号長 n で次元 $k = n - c$ となり、
最小距離 $d = 3$ よって $t = 1$ となる。 (再び練習 7.7 を見よ)
2 元の場合のように、$R \to 1$ $(c \to \infty)$ だが、$\mathrm{Pr}_E \not\to 0$

練習

練習 7.7

$\mathcal{W} = \mathbf{F}_q^c$ のとき、$\mathcal{W}$ のどんな 2 つのベクトルの組も線形従属とならないようなベクトルの組の最大個数が $(q^c - 1)/(q - 1)$ であることを示せ。
任意のそのようなベクトルの集合がパリティ検査行列 H の列を形成する場合、結果得られる $\mathbf{F}_q$ 上の線形符号は完全で 1 重誤り訂正となることを示せ。

解答

題意を満たすベクトルの集合は任意の 2 組のベクトルが従属でないため
$i \neq j \Rightarrow \mathbf{c}_i \neq \alpha \mathbf{c}_j$ for all $\alpha \in \mathbf{F}_q$ で各々のベクトルは 1 次元部分空間の基底となる。
つまり、$\mathcal{W}$ 上にある全ての 1 次元部分空間から非零元を 1 つずつとって、
任意の 2 つのベクトルの組が独立となる最大個数の元を持つ集合を作れる。
$\mathcal{W}$ 上の 1 次元部分空間の個数は不明だが、1 次元部分空間にあるベクトルは、
基底 $\mathbf{e}$ に対し $\alpha\mathbf{e}$ $(\alpha \in \mathbf{F}_q)$ の形で、零ベクトルを除けば $q - 1$ 個存在する。
全ての 1 次元部分空間の直和が $\mathcal{W}$ かつ $\mathbf{0}$ のみが各 1 次元部分空間の共通元で、
この 1 次元部分空間の個数は $\mathcal{W}$ の非零元の個数を $q - 1$ で割ることで得られ、
これは $(q^c - 1)/(q - 1)$ となる。

練習

解答 (続き)

$\mathcal{W}$ の非零かつ任意の 2 組のベクトルが線形独立となる最大個数のベクトルの組が形成するパリティ検査行列を H とおくと、
符号 $\mathcal{C} = \{\mathbf{v} \in \mathbf{F}_q^n \mid \mathbf{v}H^T = \mathbf{0}\}$ を定義できる。また、H は c 行を持つ。
補題 7.17 から H は階数 c を持ち、$\mathcal{C}$ の次元は $k = n - c$ となる。
H の列 $\mathbf{c}_i, \mathbf{c}_j$ を取ると、$\mathbf{c}' = \mathbf{c}_i + \mathbf{c}_j$ は $\mathbf{c}_i, \mathbf{c}_j$ のどちらとも線形独立である。
H の列は題意を満たすベクトルの最大個数ある為、$\mathbf{c}'$ は H の列として含まれる。
(厳密には必ず $\mathbf{c}'$ の定数倍が H の列に含まれる)
よって、H の線形従属となる最小の列の個数は 3 個となり、定理 7.27 より
この符号の最小距離は $d = 3$ で、よって $t = 1$ である。
系 6.17 からハミングの球充填限界式を考えると
$\sum_{i=0}^{1} \binom{n}{i}(q-1)^i = 1 + n(q-1) = 1 + \frac{q^c-1}{q-1}(q-1) = 1 + q^c - 1 = q^c = q^{n-k}$
よってこの符号 $\mathcal{C}$ は完全である。
以上から、得られた符号 $\mathcal{C}$ は 1 重誤り訂正完全 $[n, n-c]$ 符号となる

例

例 7.35

$q = 3, c = 2$ のとき $n = 4, k = 2$ となり、以下のパリティ検査行列を得られ、

$$H = \begin{pmatrix} 1 & 1 & 1 & 0 \\ 1 & 2 & 0 & 1 \end{pmatrix}$$

これは $\mathbf{F}_3$ 上の 1 重誤り訂正線形 $[4, 2]$ 符号を与える。