6.4 ハミングの球充填限界式

最小距離 d の符号 C は $t = \lfloor \frac{d-1}{2} \rfloor$ 個の誤りを訂正することを見た。"球"[1]

$$S_t(\mathbf{u}) = \{\mathbf{v} \in \mathcal{V} \mid d((u), \mathbf{v}) \leq t\} \ (\mathbf{u} \in C) \tag{6.5}$$

は互いに素で、かつそれぞれの $S_t(\mathbf{u})$ は完全に $\mathbf{u}$ に復号される $\mathbf{v}$ なるベクトルからなる (けれどもそのような $\mathbf{v}$ を全て含む必要はない)。良い誤り訂正のためにはそれらの共通の t が大きくなると良い。しかしながら、良い伝送速度

$$R = \frac{\log_q M}{n}$$

のためには、それらの球の数 M が大きくなると良い。q と h をしたとき、それぞれの球は素なので、これら 2 つの目的は互いに競合する。$\mathcal{V}$ を互いに交わらない大きな球を充填しようとする $q \times q \times \cdots \times q$ の大きさで固定された n 次元の箱と考える。どの程度これを達成できるかというのは明らかに上限があり、次の結果はハミングの球充填限界式 [Ha50] と言って、この上限を正確に示す。

定理 6.15

C を符号長 n の q 元 t 誤り訂正符号とおく。このとき

$$M\Big(1 + \binom{n}{1}(q-1) + \binom{n}{2}(q-1)^2 + \cdots + \binom{n}{t}(q-1)^t\Big) \leq q^n$$

証明　それぞれの符号語 $\mathbf{u} \in C$ に対して、M 個の球 $S_t(\mathbf{u}) \subseteq \mathcal{V}$ がある。練習 5.4 のように、それぞれの $\mathbf{u} \in C$ と各 i に対して、$d(\mathbf{u}, \mathbf{v}) = i$ なるベクトル $\mathbf{v} \in \mathcal{V}$ の数は $\binom{n}{i}(q-1)^i$ 個である。このようなベクトル $\mathbf{v}$ は $\mathbf{u}$ と比べてちょうど i 個の位置のシンボルが異なっている必要がある。この位置は $\binom{n}{i}$ 通り選べ、そしてそれぞれの選択に対して、$\mathbf{v}$ のそれぞれ i 個の位置が $\mathbf{u}$ の対応する位置と違うように選ぶ方法が $q-1$ 通り存在する。$i = 0, 1, \ldots, t$ に対してこの数を足すことによって、 (6.5) から、各 $\mathbf{u} \in C$ に対して次が成り立つ。

$$\mid S_t(\mathbf{u}) \mid = 1 + \binom{n}{1}(q-1) + \binom{n}{2}(q-1)^2 + \cdots + \binom{n}{t}(q-1)^t \tag{6.6}$$

今これらの M 個の球は $2t < d$ より素であり、かつそれらは全ては q^n 個の要素を持つ $\mathcal{V}$ に含まれている。よって、$M \mid S_t(\mathbf{u}) \mid \leq q^n$ で、求められている結果を得る。

例 6.16

$q = 2$ かつ $t = 1$ のとき、定理 6.15 は $M \leq 2^n/(1+n)$ を与える。よって、M は整数なので $M \leq \lfloor 2^n/(1+n) \rfloor$ である。このようにして、各 $n = 1, 2, 3, 4, 5, 6, 7, \ldots$ に対して $M \leq 1, 1, 2, 3, 5, 9, 16, \ldots$ が得られる。

系 6.17

全ての F_q 上の t 誤り訂正線形 $[n, k]$-符号 C は以下を満たす。

$$\sum_{i=0}^{t} \binom{n}{i}(q-1)^i \leq q^{n-k}$$

[1] 厳密には、$d(\mathbf{u}, \mathbf{v}) = t$ ではなく $d(\mathbf{u}, \mathbf{v}) \leq t$ で定義されるボール (ball) あるいは球体 (solid sphere) であるが、これらを球と呼ぶ符号理論の習慣に従う

証明　$\dim(C) = k$ なので、$M = q^k$ となる。定理 6.15 の両辺を q^k で割ることによって上式を得る。

線形 $[n, k]$-符号 C では、それぞれの符号語は n 桁で、その k 桁を情報桁、$n-k$ 桁を検査桁とみなすことが出来る。系 6.17 は t 個の誤りを訂正するために必要な検査桁の数の下限を与える。

$$n - k \geq \log_q \big(\sum_{i=0}^{t} \binom{n}{i} (q-1)^i \big)$$

符号 C が完全であるとは、定理 6.15 の等号を満たすことを言う (線形符号の場合系 6.17 の等式を満たすことと同値である)。これは互いに素な球 $S_t(\mathbf{u}) \quad \in C$ が、全ての $\mathbf{v} \in \mathcal{V}$ がただ 1 つの符号語 $\mathbf{u}$ と最大でも t だけ離れた位置にあるように、$\mathcal{V}$ を完全に満たすことと同値である。(そのような完全な球充填は n 次元 $(n > 1)$ のユークリッド空間 $\mathbf{R}^n$ では不可能である。なぜならば、常に球同士の間に常に埋められていない隙間があるからである。平面での最も良い充填の方法はよく知られ、明らかであるが、$\mathbf{R}^3$ での対応する問題は解決されていなかったが、1998 年に Thomas Hales によって解決された。www.math.lsa.umich.edu/~hales を見よ。ユークリッド空間での球充填と符号理論の関係については [CS92] と [Th83] を見よ。)

練習 6.5　符号が完全であることは何らかの t に対して、最近傍復号が全ての重みが t 以下の誤りパターンを復号し、かつ重みが t より大きいときは一切訂正できないことの必要十分条件であることを示せ。

例 6.18

C を n が奇数の二元反復符号 $\mathcal{R}_n$ とする。これは $k = 1, q = 2$ かつ $t = \lfloor \frac{n-1}{2} \rfloor = \frac{n-1}{2}$ の線形符号で、よって系 6.17 で $n - k = n - 1$ である。今、$q - 1 = 1$ でかつ全ての i に対して、$\binom{n}{1} = \binom{n}{n-1}$ なので、

$$\sum_{i=0}^{t} \binom{n}{i} (q-1)^i = \sum_{i=0}^{t} \binom{n}{i} = \frac{1}{2} \sum_{i=0}^{n} \binom{n}{i} = \frac{1}{2} \cdot 2^n = 2^{n-1}$$

このようにして系 6.17 の境界に達するため、この符号は完全である。しかしながら n が偶数または、$q > 2$ のとき、$\mathcal{R}_n$ は完全ではない。図 6.4 はなぜ $\mathcal{R}_3$ が完全であるかを、どのように $\mathcal{V} = F_2^3$ の 8 つの点が 2 つの集合 $S_1(\mathbf{u})$ で分割されているかを $\mathbf{u} = 000$ を黒、111 を白として表すことで示している。全ての奇数 n に対して、F_2^n 上での似たような 2 つの集合の分割が存在する。

例 6.19

二元ハミング符号 $\mathcal{H}_7$ は線形 $[7, 4]$-符号で、つまり、$n = 7$ かつ $k = 4$ である。これは $q = 2$ かつ $t = 1$ なので、

$$\sum_{i=0}^{t} \binom{n}{i} (q-1)^i = 1 + \binom{n}{1} = 8 = q^{n-k}$$

で、符号は完全である。7 章ではこれは二元ハミング符号 $\mathcal{H}_n$ $(n = 2^c - 1)$ の仲間で、それらの全てが完全であることを見る。

練習 6.6　二元ハミング符号 $\mathcal{H}_7$ が使われ、通信路 Γ が $P > \frac{1}{2}$ かつ Δ が最近傍復号である BSC とする。誤り確率 $\Pr_E$ を求めよ。$Q = \bar{P}$ が小さい時、$\Pr_E \approx 21Q^2$ であることを示せ。

練習 6.7　C を拡大ハミング符号 $\bar{\mathcal{H}}_7$ とする。(練習 6.4 を見よ) $\mathbf{u} \in C$ として、いくつのベクトル $\mathbf{v} \in \mathcal{V} = F_2^8$ を $S_t(\mathbf{u})$ が覆うかを求め、この符号が完全でないことを示せ。

C がどんな二元符号であっても、定理 6.15 は以下を与える。

$$2^n \geq M\binom{n}{t} = 2^{nR}\binom{n}{t}$$

このように $2^{n(1-R)} \geq \binom{n}{t}$ なので、対数を取ると以下を得る。

$$1 - R \geq \frac{1}{n}\log_2 \binom{n}{t}$$

スターリングの近似式 $n! \sim (n/e)^n\sqrt{2\pi n}$ (たとえば [Fi83] または [La83] を見よ)、を $\binom{n}{t} = n!/n!(n-t)!$ の 3 つの階乗項に適用すると、右辺は t/n を定数として $n \to \infty$ のとき $H_2(t/n)$ に近づくことがわかる。このとき H_2 は §3.1 で定義した二元エントロピー関数である (練習 6.8 を見よ) 極限から得られる

$$H_2(\frac{t}{n}) \leq 1 - R$$

は $n \to \infty$ のときの伝送速度 R の二元符号が訂正する誤りの比率 t/n に対するハミングの限界式である。図 6.5 はこの不等式の示す領域を表している。$d \leq n$ かつ定理 6.10 が $t = \lfloor (d-1)/2 \rfloor$ を与えるので $t/n < 1/2$ である。

この領域のどの点が反復符号やパリティ検査符号、ハミング符号などの様々な二元符号のどれに対応しているかを決定するのはよい練習問題である。

練習 6.8　t/n を定数として $n \to \infty$ のとき $\frac{1}{2}\log_2 \binom{n}{t} \to H_2(t/n)$ を証明せよ。