

1 誤り訂正符号と充填問題

通信路符号化定理の証明において、誤りを

1. $d(\mathbf{x}, bfy) > \rho$ なる場合
2. $d(\mathbf{x}, bfy) \leq \rho$ であるが、ある $\mathbf{x}' \in \mathcal{C} \setminus \{\mathbf{x}\}$ について $d(\mathbf{x}', \mathbf{y}) \leq d(\mathbf{x}, \mathbf{y})$ なる場合.

の二種類に分けて各々の確率が評価した.

二番目の誤りの確率の評価が通信路符号化定理の複雑な部分であった. 更にいえば証明の手法上, 二番目の誤りの確率については「少なくとも一つは誤り確率が低くなる $\mathcal{C}$ がある」という事をしめたのみにとどまり, 具体的にどの $\mathcal{C}$ を選べば良いかについては言及しなかった事に注意しよう.

しかるに, 二番目の誤りが起きるには符号 $\mathcal{C} \subset A^n$ が互いの距離が 2ρ 以下の 2 点を含まなければならない. 逆に言えば符号 $\mathcal{C}$ の 2 点間の最小距離が 2ρ より大きい場合, 二番目の誤りの確率を評価する困難が必要が無い.

その様な $\mathcal{C}$ は A^n に詰められた半径 ρ の“球”(の中心)の集まりと捉えることができる. 逆に「 A^n に半径 ρ のボールを幾つ詰め込めるか?」という問題は充填問題と呼ばれる.

以下, $\{0, 1\}^n$ 上に (正規化していない) ハミング距離 $d(\mathbf{x}, \mathbf{y}) = \sum_{k=1}^n 1 - \delta_{x_k y_k}$ を入れた空間上で議論する:

Definition 1. 符号 $\mathcal{C}$ は各符号語を中心とした半径 $t \in \mathbb{N}$ のボール

$$B(t, \mathbf{c}) = \{\mathbf{x} \in \{0, 1\}^n \mid d(\mathbf{x}, \mathbf{c}) \leq t\}$$

が互いに交わりを持たないとき, ハミング距離 t の誤りを訂正するという.

実際, ハミング距離 t の誤りを訂正する符号において, 最近接符号語へ訂正する決定則 $\Delta^{\mathcal{C}}$ が $d(\mathbf{X}, \mathbf{Y}) \leq t$ なる誤りを必ず正しく修正する.

1.1 充填の限界に関する議論

ハミング距離に関する半径 $r \in \mathbb{N}$ のボールの“体積”(含まれる元の数)が

$$1 + \binom{n}{1} + \binom{n}{2} + \cdots + \binom{n}{r}$$

である事に注意すれば, 符号語の総数は次で制限される;

Proposition 1 (ハミングの球充填限界式). $\mathcal{C} \subset \{0, 1\}^n$ のハミング距離 t の誤りを訂正するとき, 符号語の総数 $M = \#\mathcal{C}$ は

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

満たす.

特に等式を与える符号を**完全符号**という.

逆に次の不等式が成り立つハミング距離 t の誤りを訂正する符号の存在が保証できる.

Proposition 2.

$$M(1 + \binom{n}{1} + \binom{n}{2} + \cdots + \binom{n}{2t+1}) \geq 2^n$$

を満たす M についてハミング距離 t の誤りを訂正し, $M = \#\mathcal{C}$ なる符号 $\mathcal{C} \subset \{0, 1\}^n$ が存在する.

Proof. 最小距離 $2t+1$ を満たす部分集合の族において、包含関係で極大な $\mathcal{C} \subset \{0,1\}^n$ が求めるものである事を背理法を用いて示す。極大な $\mathcal{C}$ が

$$\bigcup_{\mathbf{a} \in \mathcal{C}} B(2t+1, \varphi(\mathbf{a})) \supset \{0,1\}^n$$

を満たさないと仮定する。このとき $\hat{\mathbf{a}} \in \{0,1\}^n \setminus \bigcup_{\mathbf{a} \in \mathcal{C}} B(2t+1, \varphi(\mathbf{a}))$ が存在するが、 $\hat{\mathbf{a}}$ を加えた符号言語 $\hat{\mathcal{C}} = \mathcal{C} \cup \{\hat{\mathbf{a}}\}$ は $\hat{\mathbf{a}} \notin \bigcup_{\mathbf{a} \in \mathcal{C}} B(2t+1, \varphi(\mathbf{a}))$ であるから最小距離 $2t+1$ のままで、明らかに $\hat{\mathcal{C}} \supset \mathcal{C}$ であり、 $\mathcal{C}$ の極大性と矛盾する。 $\square$

1.2 ハミング符号

Proposition 2 の符号は一般に存在するが、ある c について、 $k := 2^c - 1 - c$, $n := 2^c - 1$ を満たす $[k, n]$ についてハミング距離 1 の誤りを訂正する $\# \mathcal{C} = 2^k$ なる符号が知られていている。特にこれは

$$2^k \left(1 + \binom{n}{1}\right) = 2^k (1 + n) = 2^{2^c - c - 1} (1 + 2^c - 1) = 2^{2^c - 1} = 2^n$$

であるから完全符合である。

$\{0,1\}^c$ のハミング距離に関する単位球、即ち全ての単位ベクトルと零ベクトルの集合

$$\mathcal{B}_c := \{(1, 0, \dots, 0)^t, (0, 1, 0, \dots, 0)^t, \dots, (0, \dots, 0, 1)^t\} \cup \{\mathbf{0}\} \subset \{0,1\}^c$$

を考える。明らかに $\#\mathcal{B}_c = c + 1$ 。

A を全ての $\{0,1\}^c \setminus \mathcal{B}_c$ の元を縦ベクトルとして横に並べた行列とする。 2^c から $c + 1$ 個を除いた $k = 2^c - c - 1$ 個の縦ベクトルを並べたので、 A は $c \times k$ 行列となる。

更に $c \times c$ 単位正方行列 I_c 又は $k \times k$ 単位正方行列 I_k を並べて

$$H := (A, I_c), \quad G := \begin{pmatrix} I_k \\ -A \end{pmatrix}$$

と置く。 $k + c = 2^c - 1 = n$ なることから H, G は各々 $c \times n, n \times k$ 行列となる。

今度は $\{0,1\}^n$ のハミング距離に関する単位球、

$$\mathcal{B}_n := \{(1, 0, \dots, 0)^t, (0, 1, 0, \dots, 0)^t, \dots, (0, \dots, 0, 1)^t\} \cup \{\mathbf{0}\} \subset \{0,1\}^n$$

とおけば

$$\mathcal{B} \ni \mathbf{b} \rightarrow H\mathbf{b} \in \{0,1\}^c$$

は H の定義より全単射となる。¹ その逆写像を f とする。 $(f$ は線形ではない)

$$HG = (A \ I_c) \begin{pmatrix} I_k \\ -A \end{pmatrix} = A - A = 0$$

であるので、符号と決定則を

$$\mathcal{C} = G\{0,1\}^k, \quad \Delta(\mathbf{y}) = \mathbf{y} - f(H\mathbf{y})$$

1

1. $\mathbf{b} \in \left\{ \begin{pmatrix} \tilde{\mathbf{b}} \\ \mathbf{0} \end{pmatrix} \in \mathcal{B}_n \mid \tilde{\mathbf{b}} \in \mathcal{B}_k \setminus \{\mathbf{0}\} \right\} =: \mathcal{B}_{n,k}$ について $H\mathbf{b} = A\tilde{\mathbf{b}}$ であるから A の定義より $\mathcal{B}_{n,k} \ni \mathbf{b} \mapsto A\tilde{\mathbf{b}} \in \{0,1\}^c \setminus \mathcal{B}_c$ は全単射。
2. $\mathbf{b} \in \left\{ \begin{pmatrix} \mathbf{0} \\ \tilde{\mathbf{b}} \end{pmatrix} \in \mathcal{B}_n \mid \tilde{\mathbf{b}} \in \mathcal{B}_c \right\} =: \mathcal{B}_{n,c}$ $H\mathbf{b} = I_c\tilde{\mathbf{b}} = \tilde{\mathbf{b}}$ であるから $\mathcal{B}_{n,k} \ni \mathbf{b} \mapsto \tilde{\mathbf{b}} \in \mathcal{B}_c$ は全単射。
3. $\mathcal{B}_n = \mathcal{B}_{n,k} \cup \mathcal{B}_{n,c}$ は素な分割である。

とすれば, $\mathbf{a} \in \{0, 1\}^k$ について送信符号 $\mathbf{x} = G\mathbf{a}$ がエラー $\mathbf{e}$ によって $\mathbf{y} := \mathbf{x} + \mathbf{e}$ に変化したとき

$$H\mathbf{y} = HG\mathbf{a} + H\mathbf{e} = H\mathbf{e}$$

を満たす.

ハミング距離 1 までの誤り $\mathbf{e} \in \mathcal{B}$ である限り

$$\Delta(\mathbf{y}) = \mathbf{y} - f(H\mathbf{y}) = \mathbf{y} - f(H\mathbf{e}) = \mathbf{y} - \mathbf{e} = \mathbf{x}$$

つまり訂正できる.

1.3 Example

任意に自然数 c を選ぶ毎に対応する $[k, n]$ を決定できるから, 無数のハミング符号が存在する.

c	2	3	4	5	6
$n = (2^c - 1)$	3	7	15	31	63
$k = n - c$	1	4	11	26	57
$2^n/2^k$	4	8	16	32	64
$n + 1$	4	8	16	32	64

特に $[4, 7]$ がオリジナルなハミング符号である. 具体的に

$$H := \begin{pmatrix} 1 & 0 & 1 & 1 & 1 & 0 & 0 \\ 1 & 1 & 0 & 1 & 0 & 1 & 0 \\ 0 & 1 & 1 & 1 & 0 & 0 & 1 \end{pmatrix}, \quad G := \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 1 & 0 & 1 & 1 \\ 1 & 1 & 0 & 1 \\ 0 & 1 & 1 & 1 \end{pmatrix},$$

$\mathbf{z}$	$f(\mathbf{z})$
$(000)^t$	$(0000000)^t$
$(001)^t$	$(0000001)^t$
$(010)^t$	$(0000010)^t$
$(011)^t$	$(0100000)^t$
$(100)^t$	$(0000100)^t$
$(101)^t$	$(0010000)^t$
$(110)^t$	$(1000000)^t$
$(111)^t$	$(0001000)^t$

ただし $\mathbb{Z}/2\mathbb{Z}$ 上の行列である (つまり $-1 = 1$).

例えば, 入力 $\mathbf{a} = (1, 0, 0, 0)^t$ は送信符号 $\mathbf{x} = G\mathbf{a} = (1, 0, 0, 0, 1, 1, 0)^t$ に変換される. ノイズ $\mathbf{e} = (1, 0, 0, 0, 0, 0, 0)^t$ がのって $\mathbf{y} = \mathbf{x} + \mathbf{e} = (0, 0, 0, 0, 1, 1, 0)^t$ として受信されたとき, 受信符号 $\mathbf{y}$ のみから

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

を計算して $\Delta(\mathbf{y}) = \mathbf{y} - f(H\mathbf{y}) = (1, 0, 0, 0, 1, 1, 0)^t - f((1, 1, 0)^t) = (1, 0, 0, 0, 1, 1, 0)^t$ により送信符号を復元できる. ここで出てきた $H\mathbf{y}$ はシンδροームと呼ばれることがある.