1　一意復号化可能な符号

Definition 1. φ を Σ から A への符号とする. φ から自然に誘導される Σ^+ から A^+ への写像

$$\Phi^{(+)} : \Sigma^+ \ni (x_1, x_2, \ldots, x_k) \mapsto \varphi(x_1)\varphi(x_2)\cdots\varphi(x_k) \in A^+$$

が単射である時, φ は一意復号化可能であると言う.

Proposition 1. 非特異な等長符号は一意復号化可能である.

Example 1. 次の符号は等長符号であり一意復号化可能である;

情報源文字	符号語
a	00
b	01
c	10
d	11

Notetion 1. 特に $k = 1$ の時を考えることによって一意復号化可能であれば非特異であることが分かるが逆は真ではない. 実際, 非特異だが一意復号化可能でない符号が存在する.

Example 2. 次の符号は非特異であるが一意復号化可能でない;

情報源文字	符号語
a	0
b	1
c	01
d	11

関数 $\Sigma \to A^+$ とみて明らかに単射であるが

$$011$$

は cb と ad の二通りに解釈できる. 即ち $\Phi^{(*)}$ は単射ではない.

Notetion 2. 一意復号化可能であっても

$$\Sigma^\infty \ni (x_1, x_2, \ldots) \mapsto \varphi(x_1)\varphi(x_2)\cdots \in A^\infty$$

が単射とは言えない. つまり一意復号化が保証する復号は通信の終了時まで待つ必要がある.

Example 3. 次の符号は一意復号化可能であるが 2 つの意味に解釈できる無限列が存在する;

情報源文字	符号語
a	0
b	01
c	12
d	21

先ず一意復号化可能であることを見る.

$$w = \varphi(x_1)\cdot\cdots\cdot\varphi(x_k) = \varphi(x_1)\cdot\cdots\cdot\varphi(y_k)$$

とする, 符号文字 0 は常に符号語の先頭に来るので 0 が w の先頭の文字にしか現れないとして一般性を失わない. $|w|$ が偶数の時は 0 の数が 1 つのみなので $01, 12, 21$ の 3 符号語しか現れず, これらの符号語の長さが全て等しい事から $x_i = y_i$ を得る. $|w|$ が奇数の時は $0 = \varphi(x_1) = \varphi(x_1)$ より $a = x_1 = y_1$ で残りは上の議論によって等しい.

　一方で無限列

$$01212121212121212121212121\cdots$$

は $acccccccccc\cdots$ と $bdddddddddd\cdots$ の二通りに解釈できる.

1.1　一意復号化可能性の判定 (サーディナス-パターソン)

符号 $\varphi : \Sigma \to A^+$ について, 符号語全体の集合 $\mathcal{C} := \{\varphi(\sigma) \in A^+ \mid \sigma \in \Sigma\}$ によって $\mathcal{C}_0 := \mathcal{C}$ から帰納的に

$$\mathcal{C}_n := \{v \in A^+ \mid uv \in \mathcal{C},\ u \in \mathcal{C}_{n-1}\} \cup \{v \in A^+ \mid u \in \mathcal{C},\ uv \in \mathcal{C}_{n-1}\}$$

と置き, 更に

$$\mathcal{C}_\infty := \bigcup_{n=1}^{\infty} \mathcal{C}_n \quad (n = 0 \text{ を含まない事に注意})$$

とする.[1]

Proposition 2. φ が一意復号化可能である時, またそのときに限り $\mathcal{C} \cap \mathcal{C}_\infty = \phi$

Proof. 符号語の列 $u_i, v_i \in \mathcal{C}$ で

$$u_1 \dots u_n = v_1 \dots v_m w$$

なるなるものが存在するとき, またそのときのみ $w \in \mathcal{C}_\infty$ であることが $\mathcal{C}_n$ の定義から帰納的に分かる. □

Example 2 では

$$\mathcal{C} = \{0, 1, 01, 11\}, \quad \mathcal{C}_1 = \{1\}, \quad \mathcal{C}_2 = \{1\}, \quad \mathcal{C}_3 = \{1\}, \dots$$

であるから, $\mathcal{C} \cap \mathcal{C}_\infty = \{1\} \neq \phi$ であり, 一意復号化可能でない.

Example 3 では

$$\mathcal{C} = \{0, 01, 12, 21\}, \quad \mathcal{C}_1 = \{1\}, \quad \mathcal{C}_2 = \{2\}, \quad \mathcal{C}_3 = \{1\}, \quad \mathcal{C}_4 = \{2\}, \quad \mathcal{C}_5 = \{1\}, \dots$$

であるから, $\mathcal{C} \cap \mathcal{C}_\infty = \phi$ であり, 一意復号化可能である.

Example 4.

情報源文字	符号語
a	11
b	1101
c	0100
d	0001

2　符号語の下限

$\#\Sigma$ 個の源を含む情報源 Σ をできるだけ “短い符号” で符号化することを考える.

2.1　例：非特異等長符号

Definition 2. 符号 $\varphi : \Sigma \to A$ の符号アルファベット A の元の数が r であるとき, φ を r 元符号という.

r 個の種類のアルファベットのので構成された長さ l の語の数は r^l であるので, 符号長 l の r 元等長符号 φ が非特異であれば, 情報源アルファベット Σ の元の数との間に不等式

$$\#\Sigma \leq r^l \tag{1}$$

[1] $\mathcal{C}_n$ の最も長い語長を l_n とすると定義によって $l_n \leq l_{n-1} \vee l_0$ であるから帰納的に $l_n \leq l_0$. つまり $\mathcal{C}_\infty \subset \Sigma^{l_0}$ であり, $\mathcal{C}_\infty$ も有限集合である. 更に有理数が循環小数となるのと同じ論法により $\mathcal{C}_n$ は周期的であることが分かる.

が成り立つ.

この不等式は $\#\Sigma$ 個の元を持つ情報源についての, 非特異等長符号の符号長 l に対し, 下からの評価

$$\log_r \#\Sigma \leq l$$

を与えている.

全ての $\sigma \in \Sigma$ について $|\varphi(x)| = l$ であるから. 不等式 (1) は

$$\sum_{\sigma\in\Sigma} r^{-|\varphi(\sigma)|} = \sum_{\sigma\in\Sigma} r^{-l} = \#\Sigma r^{-l} \leq 1 \tag{2}$$

とも表現できる.

2.2 クラフト (マクミラン) の不等式

不等式 (2) の両端の項は等長符号以外でも意味を持つ事に注意せよ. 非特異な等長符号は瞬時復号化可能であり, よって一意復号化可能でもあるが, 一般の一意復号化可能な符号についてもこの不等式が成り立つ:

Theorem 1. *r 元符号 φ が一意復号化可能であれば,*

$$\sum_{\sigma\in\Sigma} r^{-|\varphi(\sigma)|} \leq 1$$

を満たす.

Proof. φ の符号語で最も長い語の長さを $\hat{l}$ とする, i.e., $\hat{l} := \max_{\sigma\in\Sigma} |\varphi(\sigma)|$. さらに, $\varphi^{(k)} : \Sigma^k \ni w = (\sigma_1, \ldots, \sigma_k) \mapsto \varphi(\sigma_1), \ldots, \varphi(\sigma_k) \in A^+$ とおけば任意の $w \in \Sigma^k$ について

$$|\varphi^{(k)}(w)| \leq k\hat{l}$$

であるから, Σ^k の語を長さで分類する事によって

$$\begin{aligned}
\left(\sum_{\sigma\in\Sigma} r^{-|\varphi(\sigma)|}\right)^k &= \sum_{\sigma_1,\ldots,\sigma_k\in\Sigma} r^{-|\varphi(\sigma_1)|} \cdots r^{-|\varphi(\sigma_k)|} \\
&= \sum_{\sigma_1,\ldots,\sigma_k\in\Sigma} r^{-(|\varphi(\sigma_1)|+\cdots+|\varphi(\sigma_k)|)} \\
&= \sum_{\sigma_1,\ldots,\sigma_k\in\Sigma} r^{-(|\varphi(\sigma_1)\cdots\varphi(\sigma_k)|)} \\
&= \sum_{(\sigma_1,\ldots,\sigma_k)\in\Sigma^k} r^{-(|\varphi^{(k)}(\sigma_1\cdots\sigma_k)|)} \\
&= \sum_{l=k}^{k\hat{l}} \sum_{w\in\Sigma^k:|\varphi^{(k)}(w)|=l} r^{-(|\varphi^{(k)}(w)|)} \\
&= \sum_{l=k}^{k\hat{l}} r^{-l} \sum_{w\in\Sigma^k:|\varphi^{(k)}(w)|=l} 1
\end{aligned}$$

が成り立つ.

一方で, 一意復号化可能性から $\varphi^{(k)}$ が単射であり, $\varphi^{(k)}$ の符号語を長さ l のものに限定しても一意的であることに注意すれば, $|\varphi^{(k)}(w)| = l$ なる語 $w \in \Sigma^k$ の数は A^l の元の数, 即ち r^l を超えない事;

$$\sum_{w\in\Sigma^k:|\varphi^{(k)}(w)|=l} 1 \leq r^l$$

が得られるので結局

$$\left(\sum_{\sigma\in\Sigma} r^{-|\varphi(\sigma)|}\right)^k \leq \sum_{l=k}^{k\hat{l}} r^{-l} r^l = k(\hat{l}-1) \leq k\hat{l}$$

を任意の k について得る．よって

$$\sum_{\sigma\in\Sigma} r^{-|\varphi(\sigma)|} \leq \lim_{k\to\infty}(k\hat{l})^{\frac{1}{k}} = \exp(\lim_{k\to\infty}\frac{\log k\hat{l}}{k}) = \exp(0) = 1.$$

□

2.3 平均符号長とエントロピー

Theorem 2. 憶情報源 $(\Sigma, (p_\sigma))$ の一意復号化可能な r 元符号 φ の平均符号長は r を底とした情報源のエントロピーを下回らない, *i.e.*,

$$\sum_{\sigma\in\Sigma} |\varphi(\sigma)| p_\sigma \geq H_r(\Sigma, (p_\sigma)) = \frac{1}{\log r} H(\Sigma, (p_\sigma))$$

Proof.

$$q_\sigma := r^{-|\varphi(\sigma)|}$$

とおけばクラフトの不等式から $\sum_\sigma q_\sigma \leq 1$ であるので下の lemma による． □

Lemma 1. $p_i > 0$, $q_i \geq 0$ が

$$\sum_{i=1}^m p_i = 1 \geq \sum_{i=1}^m q_i$$

を満たすならば

$$-\sum_{i=1}^m p_i \log p_i \leq -\sum_{i=1}^m p_i \log q_i$$

が成り立つ．

Proof. $-\log y$ の凸性と単調減少性から

$$(-\sum_{i=1}^m p_i \log q_i) - (-\sum_{i=1}^m p_i \log p_i) = -\sum (\log \frac{q_i}{p_i}) p_i \geq -\log(\sum \frac{q_i}{p_i} p_i) = -\log(\sum q_i) \geq -\log(1) = 0$$

□