

1 コンパクト符号

瞬時復号化可能符号全体は無限個あるので最小元の存在は自明ではないが, 十分短い符号は有限個しかないので “最小元” が存在する:

Proposition 1. 瞬時復号化可能符号には平均符号長についての “最小元” が存在する.

Proof. シヤノン-ファノの方法により, 任意の情報源 $(\Sigma, (p_\sigma))$ において符号長 $H_r((\Sigma, (p_\sigma))) + 1$ 以下の瞬時復号化可能な r 元符号 φ_s が構成できる.

$$|\varphi| = \sum_{\sigma} |\varphi(\sigma)| p_{\sigma} \geq |\varphi(\hat{\sigma})| p_{\hat{\sigma}} \geq |\varphi(\hat{\sigma})| \min_{\sigma} p_{\sigma}$$

特に $|\varphi| \geq \max_{\hat{\sigma}} |\varphi(\hat{\sigma})| \min_{\sigma} p_{\sigma}$ を満たすので, $|\varphi| \leq |\varphi_s|$ ならば $|\varphi_s| / \min_{\sigma} p_{\sigma} \geq \max_{\hat{\sigma}} |\varphi(\hat{\sigma})|$.

よって $L = |\varphi_s| / \min_{\sigma} p_{\sigma}$ とおいて

$$\mathcal{A} := \left\{ \varphi \mid \varphi : \text{瞬時復号化可能}, \max_{\sigma} |\varphi(\sigma)| \leq L \right\}$$

なる符号の集まりを考えれば, $|\varphi| \leq |\varphi_s|$ なる任意の瞬時復号化可能な符号 φ は $\mathcal{A}$ に含まれる. 特に $\varphi_s \in \mathcal{A}$ より $\mathcal{A}$ は空でない. しかも $\mathcal{A}$ は $\sigma \rightarrow \bigcup_{l \leq L} A^l$ の関数全体の集合の部分集合であるから

$$\#\mathcal{A} \leq \#\{f : \sigma \rightarrow \bigcup_{l \leq L} A^l\} = (\#\bigcup_{l \leq L} A^l)^{\#\Sigma} = (r^{L+1} - r)^{\#\Sigma} < \infty.$$

□

Definition 1. 情報源 $(\Sigma, (p_\sigma))$ 上の瞬時復号化可能な r 元符号の中で平均符号長が最短であるものを $(\Sigma, (p_\sigma))$ の r 元コンパクト符号と呼ぶ.

上述の議論により, 任意の情報源においてコンパクト符号は少なくとも 1 つ存在する (複数存在することが多い). シヤノン-ファノ符号の存在から,

Proposition 2. $\hat{\varphi}$ が情報源 $(\Sigma, (p_\sigma))$ 上の r 元コンパクト符号ならば

$$H_r(\Sigma, (p_\sigma)) \leq |\hat{\varphi}| < H_r(\Sigma, (p_\sigma)) + 1$$

が成り立つ.

2 ハフマン符号

コンパクト符号の一つ, ハフマン符号と呼ばれる符号の構成法を与える.

Definition 2. 情報源 $(\Sigma, (p_\sigma)_{\sigma \in \Sigma})$ と $(\Sigma', (p'_{\sigma'})_{\sigma' \in \Sigma'})$ について全射 $\iota : \Sigma \rightarrow \Sigma'$ が任意の $\sigma' \in \Sigma'$ に対し,

$$p'_{\sigma'} = \sum_{\sigma : \iota(\sigma) = \sigma'} p_{\sigma}$$

を満たすとき ι を縮退化写像と言い, $(\Sigma', (p'_{\sigma'})_{\sigma' \in \Sigma'})$ は $(\Sigma, (p_\sigma)_{\sigma \in \Sigma})$ の縮退化情報源であると言う.

Lemma 1. $1 < R \leq r - 1$, 情報源 $(\Sigma, (p_\sigma))$ について $\#\Sigma = \#\Sigma' + R$ なる情報源 $(\Sigma', (p'_{\sigma'}))$ への縮退化写像 $\iota : \Sigma \rightarrow \Sigma'$ で次を満たすものが存在するとする; $\hat{\sigma}' \in \Sigma'$, $\{\sigma_1, \dots, \sigma_d\} \subset \Sigma$ (ただし $d = R + 1$) が存在して

1. $\iota(\sigma_i) = \hat{\sigma}'$ for $i = 1, 2, \dots, d$

2. $\iota|_{\Sigma \setminus \{\sigma_1, \dots, \sigma_d\}}$ は単射.

3. $\{\sigma_1, \dots, \sigma_d\}$ は確率最小のもの. i.e., $p_{\sigma_i} \leq p_\sigma$ for any i and $\sigma \in \Sigma \setminus \{\sigma_1, \dots, \sigma_d\}$

このとき情報源 $(\Sigma', (p'_{\sigma'}))$ の r 元コンパクト符号 φ' と $\{a_1, \dots, a_d\} \subset A$ から

$$\varphi(\sigma) = \begin{cases} \varphi'(\hat{\sigma}')a_i & \text{if } \sigma = \sigma_i \\ \varphi'(\iota(\sigma)) & \text{otherwise} \end{cases}$$

で与えられる符号 φ は情報源 $(\Sigma, (p_\sigma))$ の r 元コンパクト符号となる.

与えられた $m := \#\Sigma$ 個の元を持つ情報源 $(\Sigma, (p_\sigma))$ に対し, Lemma の仮定を満たすような縮退化の系列を次の様に構成できる.

縮退化系列の構成

1. $m - 1$ を $r - 1$ で割った商を切り上げたものを n とする;

$$m - 1 = (n - 1)(r - 1) + R, \quad (0 < R \leq r - 1).$$

2. $\Sigma^{(0)} := \Sigma$ とし, $d_0 := R + 1, d_1 = d_2 = \dots = d_{n-1} := r$ とする.

3. $k = 0, 1, \dots, n - 1$ について次の方法で $(\Sigma^{(k+1)}, (p_\sigma^{(k+1)}))$ を構成する.

(a) $\Sigma^{(k)}$ の元のうち確率の小さい順から d_k 個選んで $\{\sigma_1^{(k)}, \dots, \sigma_{d_k}^{(k)}\}$ とする.

(b) $\Sigma^{(k+1)} := \Sigma^{(k)} \setminus \{\sigma_1^{(k)}, \dots, \sigma_{d_k}^{(k)}\} \cup \{\sigma_1^{(k)} \dots \sigma_{d_k}^{(k)}\}$ とする.

(c) $p_\sigma^{(k+1)} := \begin{cases} p_\sigma^{(k)} & (\sigma \in \Sigma^{(k)} \setminus \{\sigma_1^{(k)}, \dots, \sigma_{d_k}^{(k)}\}) \\ \sum_{i=1}^{d_k} p_{\sigma_i^{(k)}}^{(k)} & (\sigma = \sigma_1^{(k)} \dots \sigma_{d_k}^{(k)}) \end{cases}$ とする.

(d) $\iota^{(k)} : \Sigma^{(k)} \rightarrow \Sigma^{(k+1)}$ を $\iota^{(k)}(\sigma) := \begin{cases} \sigma & (\sigma \in \Sigma^{(k)} \setminus \{\sigma_1^{(k)}, \dots, \sigma_{d_k}^{(k)}\}) \\ \sigma_1^{(k)} \dots \sigma_{d_k}^{(k)} & (\sigma \in \{\sigma_1^{(k)}, \dots, \sigma_{d_k}^{(k)}\}) \end{cases}$ で定める.

毎ステップ $d_k - 1$ 個元の数が減るので, $k > 0$ の時, $\#\Sigma^{(k)} = (n - k)(r - 1) + 1$ であり, 特に $\#\Sigma^{(n-1)} = (r - 1) + 1 = r$ である.

一方で $\#\Sigma \leq r$ の時は, 符号長 1 の非特異な等長符号が存在し, その平均符号長の最小性と瞬時復号化可能性は明らかである. この事実と lemma, 及び上述の縮退化系列は帰納的にコンパクト符号を構成する方法を与える. これを**ハフマン符号**と呼ぶ.

2.1 ハフマン符号の例

2 元符号化, 特に $A = \{0, 1\}$ とする.

(Σ, P)		$(\Sigma^{(1)}, P^{(1)})$		$(\Sigma^{(2)}, P^{(2)})$	
情報源文字	確率	情報源文字	確率	情報源文字	確率
a	0.55	a	0.55	a	0.55
b	0.25	b	0.25	bcd	0.45
c	0.15	cd	0.2		
d	0.05				

情報源	$(\Sigma^{(2)}, P^{(2)})$			$(\Sigma^{(1)}, P^{(1)})$			(Σ, P)		
	情報源文字	確率	符号語	情報源文字	確率	符号語	情報源文字	確率	符号語
	a	0.55	1	a	0.55	1	a	0.55	1
	bcd	0.45	0	b	0.25	01	b	0.25	01
				cd	0.2	00	c	0.15	001
							d	0.05	000
平均符号長		1			1.45			1.65	
エントロピー		0.99			1.44			1.60	