符号 $\sigma : X \longrightarrow \bigcup_{n=1}^{\infty} \{0,1\}^n$ は.

$\tilde{\sigma} : \bigcup_{n=1}^{\infty} X^n \longrightarrow \bigcup_{n=1}^{\infty} \{0,1\}^n$ に自然に拡張できる.

命題 3.1.

符号が復号できる必要十分条件は. $\tilde{\sigma}$ が単射であることである.

定義 3.2.

$\sigma(x)$の長さ(ケタ数)

${}^{\forall}x \in X$ に対し. $|\sigma(x)|$ が一定であるとき. σ を等長符号という.

命題 3.3.

等長符号は復号可能である.

定義 3.4

どの符号語 $\sigma(x)$ も他の符号語 $\sigma(y)$ の接頭語になっていないとき.

σ をプレフィックス符号という.

命題 3.5.

プレフィックス符号は復号可能である.

∵ もし. $x = x_1 \cdots x_n \in X^n$, $y = y_1 \cdots y_m \in X^m$ に対し.

$\tilde{\sigma}(x) = \tilde{\sigma}(y)$ であったとすると.

$\sigma(x_1) = \sigma(y_1)$ がわかる

∵ もし $\sigma(x_1) \neq \sigma(y_1)$, $|\sigma(x_1)| < |\sigma(y_1)|$ なら.

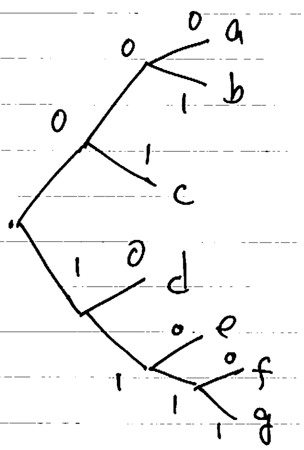
$\sigma(y_1)$ の接頭語が $\sigma(x_1)$ になっているので矛盾

以下. $\sigma(x_2) = \sigma(y_2), \cdots$ が帰納的にわかる.

例1. 樹形図から作る符号.

0,1で作られる樹形図に対し.符号を対応させるように作ると.

プレフィックス符号が作られる.



のように作る.

ある $\sigma(x)$ が $\sigma(y)$ の接頭語になるならば.

y は樹形図で x の下にないといけないが.

これは仮定に反する.

次に.符号の効率を考える.

$X=\{x_1,\cdots,x_n\}$ とし.文章に文字 x_i がでる確率を p_i とする.

よく使う文字は.符号が短い方がいいし.あまり使わない文字は長くてもよい.

命題3.6.

$X=\{x_1,\cdots,x_n\}$.出現確率 p_i とすると.符号語の平均語長 L を.

$S(p) \leq L \leq S(p)+1$ とするような符号 σ が存在する.ただし $p=(p_1,\cdots,p_n)$ である

(証).まず p_i に対して. $\frac{1}{2^{k_i-1}} > p_i \geq \frac{1}{2^{k_i}}$ をみたす $k_i \in \mathbb{N}$ をとる.

この k_i は. $k_1 \leq k_2 \leq \cdots \leq k_n$ をみたすとしてよい.

このとき. $\sum \frac{1}{2^{k_i}} \leq \sum p_i = 1$ をみたすことに注意

さて. X に対し.符号 σ を以下のように作る.

①. $|\sigma(x_i)| = k_i$ とする.

② $\sigma(x_1) = \underbrace{0\cdots0}_{k_1 コ}$ とする

③ $\sigma(x_2) = \underbrace{0\cdots0}_{k_1-1コ}1\underbrace{0\cdots0}_{k_2-k_1コ}$ とする.

以下. 樹形図を上から順に埋める形で符号を作れる.

このとき. 平均語長は

$L = \sum_{i=1}^{n} p_i k_i$ であるが $\frac{1}{2^{k_i-1}} > p_i \geq \frac{1}{2^{k_i}}$ より

$k_i \geq -\log p_i > k_i - 1$ をみたす. これより

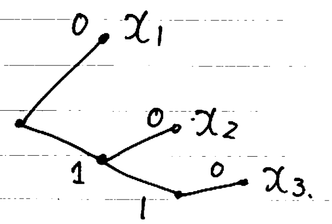
$-\log p_i \leq k_i < -\log p_i + 1$ である.

$\therefore\ -\sum p_i \log p_i \leq \sum p_i k_i < -\sum p_i \log p_i + \sum p_i$.

$\therefore\ S(p) \leq L < S(p) + 1$ となる.

例. $X = \{x_1, x_2, x_3\}$, $p_1 = \frac{1}{2}$, $p_2 = \frac{1}{3}$, $p_3 = \frac{1}{6}$ とすると.

$k_1 = 1$, $k_2 = 2$, $k_3 = 3$ となる. ここから符号σを.



と作れば. 命題をみたす符号になる.