

3. 情報理論 (情報源符号化)

問 3-1. 一意符号ではあるが瞬時符号ではないような符号の例を構成し、それと同じ符号語長列を持つ瞬時符号を (なるべく systematic に) 構成せよ。

問 3-2. 瞬時符号であることと語頭符号であることが同値であることを示せ。

問 3-3. 情報源 $S = (S, P)$ に対し、2 元 Huffman 符号 $C : S \rightarrow \{0, 1\}^+$ が最適符号であることを、情報源 alphabet の個数 $\#S$ に関する帰納法を用いて、以下の手順で示せ。

- (1) 2 元瞬時符号では、最長の符号語は $x0, x1$ ($x \in \{0, 1\}^*$) の形で組になって現れる。
- (2) 2 元最適符号 D で、出現確率が最小とその次 (等しいこともある) の 2 つの S の元 s_i, s_j に対して、 $D(s_i) = x0, D(s_j) = x1$ となるものが存在する。
- (3) s_i, s_j をまとめて s' とした情報源 $S' = (S', P')$ を定式化せよ。 (s_i, s_j を縮退させた情報源と呼ぶ。)
- (4) S' の符号 D' を $D'(s') = x, D'|_{S \setminus \{s_i, s_j\}} = D|_{S \setminus \{s_i, s_j\}}$ で定めるとき、 $L(D) - L(D') = P(s_i) + P(s_j)$ となる。
- (5) S の Huffman 符号 C について、 $C(s_i) = y0, C(s_j) = y1$ ($y \in \{0, 1\}^*$) となっていて一般性を失わない。これより同様に S' の符号 C' を構成するとき、 C' は S' の Huffman 符号である。従って、帰納法の仮定により S' の最適符号である。
- (6) $L(C) - L(C') = P(s_i) + P(s_j)$ となる。このことから $L(C) = L(D)$ となり、 C は S の最適符号である。

問 3-4. 英文の alphabet の出現頻度の統計を何処かで調べ、それに従って 2 元 Huffman 符号化の符号化関数 $C : \{a, b, \dots, z\} \rightarrow \{0, 1\}^+$ を構成せよ。

問 3-5. 情報源 $S = (S, P)$ について、 $\#S \geq 4$ であるとする。生起確率 $P(s)$ が最大な情報源 alphabet がただ一つ (単独最大) あって、それを $s_1 \in S$ とする時、 S の 2 元 Huffman 符号化関数 $C : S \rightarrow \{0, 1\}^+$ について、

- (1) $P(s_1) < \frac{1}{3} \implies |\mathcal{C}(s_1)| \geq 2$
- (2) $P(s_1) > \frac{2}{5} \implies |\mathcal{C}(s_1)| = 1$
- (3) 上 2 つの $P(s_1)$ に関する条件は best possible (これより緩めたら成り立たない最良の限界) であるか。

(参考: 東大 2001 年前期入試問 5)

問 3-6. 関数 $y = p \log \frac{1}{p}$ ($0 < p < 1$)、及び 2 値エントロピー関数 $y = H(p) = p \log \frac{1}{p} + \bar{p} \log \frac{1}{\bar{p}}$ (ここに $\bar{p} := 1 - p$) のグラフの概形を描き、最大値をとる p とその時の値を求めよ。

問 3-7. 2 つの無記憶情報源 $S^{(1)} = (S^{(1)}, P^{(1)}), S^{(2)} = (S^{(2)}, P^{(2)})$ に対し、その直積情報源 $(S^{(1)} \times S^{(2)}, P)$ を考える。 $S^{(1)}, S^{(2)}$ が独立な情報源であるとは、任意の $(s^{(1)}, s^{(2)}) \in S^{(1)} \times S^{(2)}$ に対し、 $P((s^{(1)}, s^{(2)})) = P^{(1)}(s^{(1)}) \cdot P^{(2)}(s^{(2)})$ であることを言い、この時、直積情報源を $S^{(1)} \times S^{(2)}$ と書く。

- (1) 独立な無記憶情報源 $S^{(1)}, S^{(2)}$ に対し、 $H(S^{(1)} \times S^{(2)}) = H(S^{(1)}) + H(S^{(2)})$.
- (2) 無記憶情報源 S の n 次の拡大情報源 S^n について、 $H(S^n) = nH(S)$.

問 3-8. $\frac{1}{2} < p < 1$ である適当な p と幾つかの小さい n (または適当な系列に属する n) について、 $S = \{a, b\}, P(a) = p$ である無記憶情報源 $S = (S, P)$ に対し、 S の n 次の拡大情報源 S^n の Huffman 符号 C^n を構成し、その平均符号長 $L(C^n)$ と情報源 S のエントロピー $H(S) = H(p)$ とを比較してみよ。