

符号への要請

- 一意符号: $C^* : S^* \longrightarrow T^* : \text{単射}$
- 瞬時符号: $C(x) = C(s)w \implies x = sy$
(最初に届いた符号語で最初の文字が復元できる)
(以上は生起確率 P には依らない)
- 効率が良い... 平均符号長 $L(C)$ が小さい
(これは生起確率 P に依る)

Kraft の不等式

$$S = \{s_1, \dots, s_k\}, \quad \#T = r$$

自然数列 (ℓ_1, \dots, ℓ_k) に対し、

各符号語長 $|C(s_i)| = \ell_i$ なる
 r 元 瞬時符号 が存在

$$\iff \sum_{i=1}^k \frac{1}{r^{\ell_i}} \leq 1$$

McMillan の不等式

$$S = \{s_1, \dots, s_k\}, \quad \#T = r$$

自然数列 (ℓ_1, \dots, ℓ_k) に対し、

各符号語長 $|C(s_i)| = \ell_i$ なる
 r 元 一意符号 が存在

$$\iff \sum_{i=1}^k \frac{1}{r^{\ell_i}} \leq 1$$

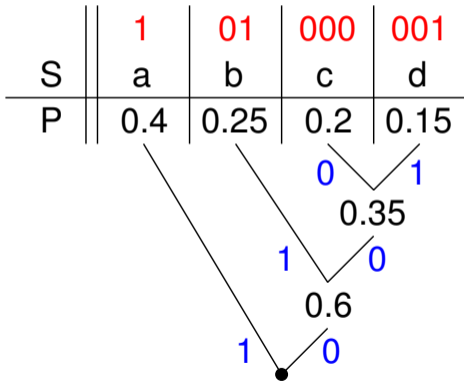
次に、生起確率を考慮に入れて、

平均符号長の小さい符号の構成を考えよう

→ **Huffman 符号**

平均符号長の小さい符号の構成 (Huffman 符号)

例: $\#S = 4 = 2^2$, 平均符号長: $1.95 (< 2)$



Huffman 符号

- 平均符号長 $L(C)$ の最小値を実現
… 「最適符号 (optimal code)」

- 各文字の生起確率 $P(s)$ の
“ばらつき” が大きいほど効果的

$S = 2$ だったら、

どうやっても

(生起確率に関わらず) $L(C) = 1$ か？

(何かうまい手はないか)

→ “拡大情報源” を考える

問題: 次の生起確率を持つ情報源

$S = (S, P), S = \{a, b\}$ について、

S	a	b
P	0.8	0.2

- (1) 2 次の拡大情報源 $S^2 = (S^2, P^{\otimes 2})$ に対する Huffman 符号 C_2 を構成し、“1 文字当たりの平均符号長” $L(C_2)/2$ を求めよ。
- (2) (時間に余裕があれば) 3 次の拡大情報源 $S^3 = (S^3, P^{\otimes 3})$ に対しても同様の計算をせよ。

n 次の拡大情報源 $S^n = (S^n, P^{\otimes n})$
の符号 C_n を利用すると、

情報源 alphabet 1 文字当たりの平均符号長

$$\frac{L(C_n)}{n}$$

はどこまで下げられるか？

→ 情報源 S が本来持っている“情報の量”
より小さくはならないだろう。

“エントロピー (entropy)”

“情報の量”

「或る事象 P が起こる」

という“情報の価値”は

どう評価したら良いか？

「事象 P が起こる」という“情報の価値” $I(P)$

要請:

(1) 生起確率 p のみに依る

$$\longrightarrow I(p) := I(P)$$

(2) 独立な事象 P_1, P_2 に対して、

$$I(P_1 \wedge P_2) = I(P_1) + I(P_2)$$

$$\longrightarrow I(p_1 p_2) = I(p_1) + I(p_2)$$

(3) $I : [0, 1] \longrightarrow \mathbf{R}_{\geq 0}$: 連続関数

$$\longrightarrow \boxed{I(p) = C \log \frac{1}{p} = -C \log p \quad (C \geq 0)}$$

$$I(p) = C \log \frac{1}{p} = -C \log p \quad (C \geq 0)$$

定数 C の取り方

↔ \log の底の取り方

↔ “情報量” の単位の取り方

通常 2 を底に取って ($I(\frac{1}{2}) := 1$)、

“情報量” の単位とする: **bit (binary digit)**

情報源 $\mathcal{S} = (S, P)$ の

1 文字から得られる平均情報量

$$H(\mathcal{S}) := \sum_{s \in S} P(s) I(P(s))$$

: 情報源 \mathcal{S} の エントロピー (entropy)

$S = \{s_1, \dots, s_k\}$, $P(s_i) = p_i$ の時は、

$$H(\mathcal{S}) := \sum_{i=1}^k p_i \log \frac{1}{p_i} = - \sum_{i=1}^k p_i \log p_i$$

特に $\#S = 2$ の時、

- 起きる確率 p
- 起きない確率 $1 - p =: \bar{p}$

$$H(p) := H(S) = p \log \frac{1}{p} + \bar{p} \log \frac{1}{\bar{p}}$$

: 2 値エントロピー関数

問: $y = I(p), y = H(p)$ のグラフの概形を描け。
(最大値をとる p とその時の値は ?)