情報源符号化の定式化・続

「情報源 alphabet と頻度との組」
が符号化の対象

S : 情報源 alphabet(有限集合)

$P : S \longrightarrow [0,1] \subset \boldsymbol{R}$: **生起確率** $\left(\sum_{s \in S} P(s) = 1\right)$

情報源 $\mathcal{S} := (S, P)$

: 文字 $s \in S$ を確率 $P(s)$ で次々と発生

$\longrightarrow w \in S^+$ を発生

(ここでの) 仮定 : 情報源の無記憶性

各 $s \in S$ の生起確率は、s のみで決まり、先立って発生した文字に依らない。

情報源符号化の定式化・続

T : **符号 alphabet(有限集合)**

(しばしば $T = \{0, 1\}$)

$\mathcal{C} : S \longrightarrow T^+$: **符号**

$\longrightarrow$ **文字列を並べて** $\mathcal{C}^* : S^* \longrightarrow T^*$ **に延長**

$$L(\mathcal{C}) := \sum_{s \in S} P(s)|\mathcal{C}(s)|$$: $\mathcal{C}$ **の平均符号長**

(1 文字の符号語長の期待値)

符号への要請

- **一意符号:** $\mathcal{C}^* : S^* \longrightarrow T^*$ **: 単射**
- **瞬時符号:** $\mathcal{C}(\boldsymbol{x}) = \mathcal{C}(s)\boldsymbol{w} \Longrightarrow \boldsymbol{x} = s\boldsymbol{y}$
  (最初に届いた符号語 $\mathcal{C}(s)$ で
  最初の文字 s 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** $\cdots$ **平均符号長 $L(\mathcal{C})$ が小さい**
  (これは生起確率 P に依る)

生起確率を考慮に入れて、

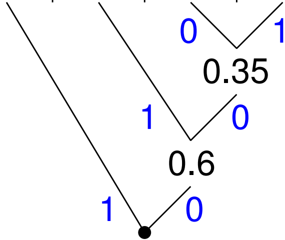
平均符号長の小さい符号の構成を考えよう

$\longrightarrow$ **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:** $1.95 (< 2)$

	1	01	000	001
S	a	b	c	d
P	0.4	0.25	0.2	0.15



Huffman 符号

- 瞬時符号の中での
  平均符号長 $L(\mathcal{C})$ の最小値を実現
  $\cdots$ **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

- 弱点: 予め生起確率が判らないと
  符号を構成できない

$\#S = 2$ **だったら、**

どうやっても

(生起確率に関わらず) $L(\mathcal{C}) = 1$ **か ?**

(何かうまい手はないか)

$\longrightarrow$ **"拡大情報源" を考える**

拡大情報源

情報源 $\mathcal{S} = (S, P)$ に対し、

“n 文字づつまとめた情報源” $\mathcal{S}^n$ を考える

$$\mathcal{S}^n := (S^n, P^{\otimes n})$$

$$S^n = S \times \cdots \times S = \left\{ s_{i_1} \ldots s_{i_n} \middle| s_{i_j} \in S \right\}$$

$$P^{\otimes n} : S^n \longrightarrow [0, 1] \subset \boldsymbol{R}$$

$$s_{i_1} \ldots s_{i_n} \longmapsto P(s_{i_1}) \cdots P(s_{i_n})$$

$\longrightarrow$ **この情報源 $\mathcal{S}^n$ を符号化せよ**

問題: 次の生起確率を持つ情報源
$\mathcal{S}=(S,P), S=\{a,b\}$ **について、**

S	a	b
P	0.8	0.2

(1) 2 **次の拡大情報源** $\mathcal{S}^2=(S^2,P^{\otimes 2})$
に対する Huffman 符号 $\mathcal{C}_2$ **を構成し、**
“1 **文字当たりの平均符号長”**
$L(\mathcal{C}_2)/2$ **を求めよ。**

(2) 3 **次の拡大情報源** $\mathcal{S}^3=(S^3,P^{\otimes 3})$
に対しても同様の計算をせよ。

一般に、

n **次の拡大情報源** $\mathcal{S}^n = (S^n, P^{\otimes n})$

の符号 $\mathcal{C}_n$ **で、**

n **を大きくしてゆくと、**

情報源 alphabet 1 **文字当たりの平均符号長**

$$\frac{L(\mathcal{C}_n)}{n}$$

を下げることが出来る

(符号は複雑になってゆくが)

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

情報源 alphabet 1 **文字当たりの平均符号長**

$$\frac{L(\mathcal{C}_n)}{n}$$

はどこまで下げられるか？

$\longrightarrow$ **情報源** $\mathcal{S}$ **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)"

n **次の拡大情報源** $\mathcal{S}^n = (S^n, P^{\otimes n})$

の符号 $\mathcal{C}_n$ **を利用すると、**

情報源 alphabet 1 **文字当たりの平均符号長**

$$\frac{L(\mathcal{C}_n)}{n}$$

はどこまで下げられるか？

$\longrightarrow$ **情報源** $\mathcal{S}$ **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)"

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

情報源 alphabet 1 文字当たりの平均符号長

$$\frac{L(\mathcal{C}_n)}{n}$$

はどこまで下げられるか？

$\longrightarrow$ 情報源 $\mathcal{S}$ 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？

“情報の量”

基本的なアイデア:

確率 $\frac{1}{4}$ で起きる出来事を

教えてもらうことの価値は、

確率 $\frac{1}{2}$ で起きる出来事を 2 つ

教えてもらうのと同じ

$\longrightarrow$ **“情報の量” が 2 倍**

「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) $I : (0, 1] \longrightarrow \boldsymbol{R}_{\geq 0}$ **: 連続関数**

(const. 0 ではない)

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

「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) $I : (0,1] \longrightarrow \boldsymbol{R}_{\geq 0}$ **: 連続関数**

(const. 0 ではない)

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

「事象 P が起こる」という "情報の価値" $I(P)$

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

定数 C の取り方
$\longleftrightarrow$ $\log$ **の底の取り方**
$\longleftrightarrow$ **"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)

「事象 P が起こる」という "情報の価値" $I(P)$

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

定数 C の取り方
$\longleftrightarrow$ $\log$ **の底の取り方**
$\longleftrightarrow$ **"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)

情報源のエントロピー

情報源 $\mathcal{S} = (S, P)$ **の**

1 **文字から得られる情報量の期待値**

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

：情報源 $\mathcal{S}$ **のエントロピー (entropy)**

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

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

情報源のエントロピー

特に $\#S=2$ **(**$S=\{$ **起きる**, **起きない** $\}$**) の時、**

- **起きる確率** p
- **起きない確率** $1-p=:\overline{p}$

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

: 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 ?)**

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

情報源 alphabet 1 **文字当たりの平均符号長**

$$\frac{L(\mathcal{C}_n)}{n}$$

はどこまで下げられるか？

$\longrightarrow$ **情報源** $\mathcal{S}$ **のエントロピー** $H(\mathcal{S})$ **より**
小さくはならないだろう

$\longrightarrow$ **まず一般に、符号** $\mathcal{C}$ **に対し、**
平均符号長 $L(\mathcal{C})$ **と** $H(\mathcal{S})$ **とを比べよう**

n **次の拡大情報源** $\mathcal{S}^n = (S^n, P^{\otimes n})$

の符号 $\mathcal{C}_n$ **を利用すると、**

情報源 alphabet 1 **文字当たりの平均符号長**

$$\frac{L(\mathcal{C}_n)}{n}$$

はどこまで下げられるか？

$\longrightarrow$ **情報源** $\mathcal{S}$ **のエントロピー** $H(\mathcal{S})$ **より**

小さくはならないだろう

$\longrightarrow$ **まず一般に、符号** $\mathcal{C}$ **に対し、**

平均符号長 $L(\mathcal{C})$ **と** $H(\mathcal{S})$ **とを比べよう**

定理

$\mathcal{S} = (S, P)$: **情報源**

$\mathcal{C}$: $\mathcal{S}$ **の一意符号 (瞬時符号)**

$$\Longrightarrow \qquad \boxed{L(\mathcal{C}) \geq H(\mathcal{S})}$$

(但し、$\log$ **の底は** $r := \#T$ **に取る)**

$\eta := \dfrac{H(\mathcal{S})}{L(\mathcal{C})}$: $\mathcal{C}$ **の効率 (efficiency)**

$\overline{\eta} = 1 - \eta$: $\mathcal{C}$ **の冗長度 (redundancy)**

Kraft の不等式(再掲)

$$S = \{s_1, \ldots, s_k\}, \qquad \#T = r$$

自然数列 $(\ell_1, \ldots, \ell_k)$ **に対し、**

各符号語長 $|\mathcal{C}(s_i)| = \ell_i$ **なる**

r **元瞬時符号が存在**

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

補題

$$x_i, y_i > 0 \ (i = 1, \ldots, k)$$

$$\sum_{i=1}^{k} x_i = \sum_{i=1}^{k} y_i = 1$$

$$\Longrightarrow \qquad \sum_{i=1}^{k} x_i \log \frac{1}{x_i} \leq \sum_{i=1}^{k} x_i \log \frac{1}{y_i}$$

(等号は $\forall i : x_i = y_i$ のとき)

- $L(\mathcal{C}) = H(\mathcal{S})$ **は実現できるのか？**

- $\inf_{\mathcal{C}} L(\mathcal{C}) = H(\mathcal{S})$ **であるか？**

Huffman 符号は確かに最適だが、
$L(\mathcal{C})$ の上からの評価は難しい

$\longrightarrow$ Shannon-Fano 符号
(Kraft-McMillan の不等式の利用)

- $L(\mathcal{C}) = H(\mathcal{S})$ **は実現できるのか？**

- $\inf_{\mathcal{C}} L(\mathcal{C}) = H(\mathcal{S})$ **であるか？**

Huffman 符号は確かに最適だが、
$L(\mathcal{C})$ の上からの評価は難しい

$\longrightarrow$ Shannon-Fano 符号
(Kraft-McMillan の不等式の利用)

- $L(\mathcal{C}) = H(\mathcal{S})$ **は実現できるのか？**

- $\inf_{\mathcal{C}} L(\mathcal{C}) = H(\mathcal{S})$ **であるか？**

Huffman 符号は確かに最適だが、
$L(\mathcal{C})$ の上からの評価は難しい

$\longrightarrow$ **Shannon-Fano 符号**
(Kraft-McMillan の不等式の利用)