

情報源符号化の定式化・続

(Formulation of source coding: continued)

S : source alphabet (a finite set)

$P : S \rightarrow [0, 1] \subset \mathbf{R}$: 生起確率
(occurrence probability, $\sum_{s \in S} P(s) = 1$)

情報源 (source) $\mathcal{S} := (S, P)$

文字 $s \in S$ を確率 $P(s)$ で次々と発生

(generates symbols $s \in S$

with probability $P(s)$ successively)

$\rightarrow w \in S^+$ を発生

(generates a sequence $w \in S^+$)

Here we assume the following property
for simplicity:

(ここでの) 仮定 : 情報源の定常・無記憶性
(Assumption : stationary, memoryless source)

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

**The occurrence probability $P(s)$ of $s \in S$
depends only on s ,
not on preceding symbols.**

情報源符号化の定式化・続

(Formulation of source coding: continued)

T : code alphabet (a finite set)

(typically $T = \{0, 1\}$)

$C : S \longrightarrow T^+$: a code

→ 文字列を並べて $C^* : S^* \longrightarrow T^*$ に延長
(extended by concatenation)

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

(average code-word-length)

符号への要請 (Requirement for good codes)

- 一意符号 (uniquely decodable code):

$$\mathcal{C}^* : S^* \longrightarrow T^* : \text{injective (単射)}$$

- 瞬時符号 (instantaneously decodable code):

$$\mathcal{C}^*(x) = \mathcal{C}(s)w \implies x = sy$$

... These do not depend on P .

- 効率が良い (efficient) : $L(\mathcal{C})$ is small.

... This depends on P .

Taking the occurrence probability P
into consideration,

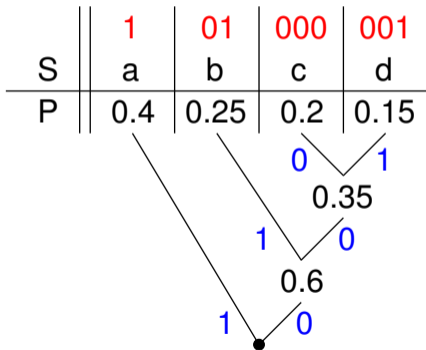
construct a code with small average length.

→ **Huffman code**

平均符号長の小さい符号の構成 (Huffman code)

(Construction of a code \mathcal{C} with small $L(\mathcal{C})$)

Ex. $\#S = 4 = 2^2$, average length: $1.95 (< 2)$



Huffman code

- 平均符号長 $L(\mathcal{C})$ の最小値を実現
(Attain the minimum average length $L(\mathcal{C})$)
... 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
(More effective
if $P(s)$'s are more “scattering”)
- 弱点: 予め生起確率が判らないと
符号を構成できない
(Weak point: must know $P(s)$'s in advance)

#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 ?
(If #S = 2, must one have $L(C) = 1$?)

何かうまい手はないか ?
(Is there a good way to improve ?)

→ Consider **extended sources** (拡大情報源)

拡大情報源 (Extended source)

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
(For the source $\mathcal{S} = (S, P)$, consider
the source \mathcal{S}^n “packed every n symbols”)

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

$$S^n = S \times \cdots \times S = \{s_{i_1} \cdots s_{i_n} \mid s_{i_j} \in S\}$$

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

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

→ Encode this source \mathcal{S}^n .

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) 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。

By considering the code \mathcal{C}_n
for the extended source $\mathcal{S}^n = (\mathcal{S}^n, P^{\otimes n})$
of degree n ,

the average length

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

per each symbol can be decreased generally

(though the code turns more complex).

For the code \mathcal{C}_n
for the extended source $\mathcal{S}^n = (S^n, P^{\otimes n})$
of degree n ,

what is the infimum of $\frac{L(\mathcal{C}_n)}{n}$?

→ It will not be smaller than
“the information content”
which the source \mathcal{S} has properly.

the notion of “entropy(エントロピー)”

For the code \mathcal{C}_n
for the extended source $\mathcal{S}^n = (S^n, P^{\otimes n})$
of degree n ,

what is the infimum of $\frac{L(\mathcal{C}_n)}{n}$?

→ It will not be smaller than
“the information content”
which the source \mathcal{S} has properly.

the notion of “entropy(エントロピー)”

For the code \mathcal{C}_n
for the extended source $\mathcal{S}^n = (\mathcal{S}^n, P^{\otimes n})$
of degree n ,

what is the infimum of $\frac{L(\mathcal{C}_n)}{n}$?

→ It will not be smaller than
“the information content”
which the source \mathcal{S} has properly.

the notion of “entropy(エントロピー)”

“the information content (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？

**How should one quantify
“the information content”
that an event P occurs ?**

“the information content (情報の量)”

Basic idea:

The value to know

that an event with probability $\frac{1}{4}$ occurs

is the same as the one to know

that two events with probability $\frac{1}{2}$ occur.

(確率 $\frac{1}{4}$ で起きる出来事を知ることの価値は、

確率 $\frac{1}{2}$ で起きる出来事を 2 つ知ると同じ)

→ “the content” is double (“情報の量” が 2 倍)

「事象 P が起こる」という“情報の量” $I(P)$

(“the information content” $I(P)$)

that an event P occurs)

要請 (Requirement):

(1) depends only on the occurrence probability p

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

(2) for two independent events P_1, P_2 ,

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

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

(3) $I : (0, 1] \longrightarrow \mathbf{R}_{\geq 0}$: continuous (not const. 0)

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

「事象 P が起こる」という“情報の量” $I(P)$

(“the information content” $I(P)$)

that an event P occurs)

要請 (Requirement):

(1) depends only on the occurrence probability p

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

(2) for two independent events P_1, P_2 ,

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

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

(3) $I : (0, 1] \longrightarrow \mathbf{R}_{\geq 0}$: continuous (not const. 0)

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

「事象 P が起こる」という“情報の量” $I(P)$

(“the information content” $I(P)$

that an event P occurs)

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

the choice of the constant C

↔ the choice of the base of log

↔ the choice of the unit of “information”

Usually we choose 2 as the base; $I\left(\frac{1}{2}\right) := 1$.

→ the unit of “information”: bit (binary digit)

「事象 P が起こる」という“情報の量” $I(P)$

(“the information content” $I(P)$

that an event P occurs)

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

the choice of the constant C

⟷ the choice of the base of log

⟷ the choice of the unit of “information”

Usually we choose 2 as the base; $I\left(\frac{1}{2}\right) := 1$.

→ the unit of “information”: **bit** (binary digit)

情報源のエントロピー (the entropy of a source)

the expected value of the information

from the source $\mathcal{S} = (S, P)$ per each symbol:

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

: the **entropy** of the source \mathcal{S}

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

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

情報源のエントロピー (the entropy of a source)

In particular when $\#S = 2$ ($S = \{\text{yes, no}\}$),

- the probability to occur: p
- the probability not to occur: $1 - p =: \bar{p}$

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

: **binary entropy function**

**For the code \mathcal{C}_n
for the extended source $\mathcal{S}^n = (\mathcal{S}^n, P^{\otimes n})$
of degree n ,**

what is the infimum of $\frac{L(\mathcal{C}_n)}{n}$?

**→ It will not be smaller than
the entropy $H(\mathcal{S})$ of \mathcal{S} .**

→ For a code \mathcal{C} ,
first compare the average length $L(\mathcal{C})$
with the entropy $H(\mathcal{S})$.

**For the code \mathcal{C}_n
for the extended source $\mathcal{S}^n = (\mathcal{S}^n, P^{\otimes n})$
of degree n ,**

what is the infimum of $\frac{L(\mathcal{C}_n)}{n}$?

**→ It will not be smaller than
the entropy $H(\mathcal{S})$ of \mathcal{S} .**

**→ For a code \mathcal{C} ,
first compare the average length $L(\mathcal{C})$
with the entropy $H(\mathcal{S})$.**

Theorem

$\mathcal{S} = (S, P)$: a source

\mathcal{C} : a uniq. (or instant.) decodable code for \mathcal{S}

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

(Here, the base of \log is chosen as $r := \#T$.)

$\eta := \frac{H(\mathcal{S})}{L(\mathcal{C})}$: the efficiency (効率) of \mathcal{C}

$\bar{\eta} = 1 - \eta$: the redundancy (冗長度) of \mathcal{C}

Theorem

$\mathcal{S} = (S, P)$: a source

\mathcal{C} : a uniq. (or instant.) decodable code for \mathcal{S}

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

(Here, the base of log is chosen as $r := \#T$.)

$\eta := \frac{H(\mathcal{S})}{L(\mathcal{C})}$: the **efficiency** (効率) of \mathcal{C}

$\bar{\eta} = 1 - \eta$: the **redundancy** (冗長度) of \mathcal{C}

Kraft の不等式 (Kraft's inequality)

$$S = \{s_1, \dots, s_k\}, \quad \#T = r \text{ (r-ary code)}$$

For a sequence (ℓ_1, \dots, ℓ_k) of natural numbers,

\exists an r-ary instant. decodable code \mathcal{C}
with $|\mathcal{C}(s_i)| = \ell_i \ (\forall i)$

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

Lemma

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

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

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

(Equality iff $\forall i : x_i = y_i$)