

## “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 ?**

「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)

## 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  $\mathcal{S}$

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

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

**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  $\mathcal{C}$

## Kraft の不等式 (Kraft's inequality)

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

For a sequence  $(\ell_1, \dots, \ell_k)$  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  $\forall i : x_i = y_i$ )**



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

- **Can  $L(\mathcal{C}) = H(\mathcal{S})$  be attained for some  $\mathcal{C}$  ?**
- $\inf_{\mathcal{C}} L(\mathcal{C}) = H(\mathcal{S})$  ?

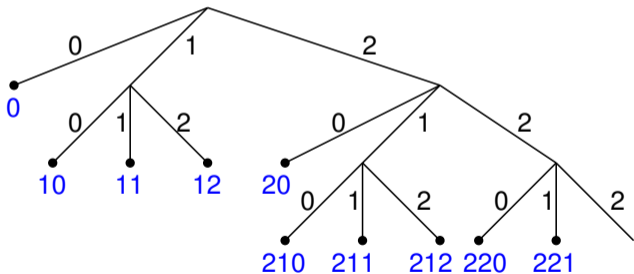
**Though Huffman code is optimal,  
the estimate of  $L(\mathcal{C})$  from above is difficult.**

→ **Shannon-Fano code**  
(use of Kraft-McMillan inequality)

## 符号木 (code tree)

Ex.  $T = \{0, 1, 2\}$

$\mathcal{C}(S) = \{0, 10, 11, 12, 20, 210, 211, 212, 220, 221\}$



## Shannon-Fano codes

For  $S = \{s_1, \dots, s_k\}$ ,  $P(s_i) = p_i$ ,  $\#T = r$ ,

put  $l_i := \left\lceil \log_r \left( \frac{1}{p_i} \right) \right\rceil$

$$\longrightarrow \sum_{i=1}^k \frac{1}{r^{l_i}} \leq 1$$

$\longrightarrow \exists$  an  $r$ -ary instant. decodable code  $\mathcal{C}$   
with  $|\mathcal{C}(s_i)| = l_i$  for all  $i$

For this  $\mathcal{C}$ , we have  $H(S) \leq L(\mathcal{C}) < 1 + H(S)$ .

## Theorem

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

$\mathcal{C}$  : an optimal code for  $\mathcal{S}$

$\implies$

$$\boxed{H(\mathcal{S}) \leq L(\mathcal{C}) < 1 + H(\mathcal{S})}$$

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

## Shannon's Noiseless Coding Theorem

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

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

: the extended source of  $\mathcal{S}$  of degree  $n$

$\mathcal{C}_n$  : an optimal code for  $\mathcal{S}^n$

$$\implies \boxed{\lim_{n \rightarrow \infty} \frac{L(\mathcal{C}_n)}{n} = H(\mathcal{S})}$$

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

情報源を効率良く符号化する話は一段落

**Here we stop saying about the source coding  
and go to the next topic.**

情報通信を行なう際の要請

**(What is required in communication ?)**

- 効率的に (efficiently)  
→ 情報理論 (Information Theory)
- 確実に (certainly)  
→ 符号理論 (Coding Theory)
- 安全に (safely)  
→ 暗号理論 (Cryptography)

情報通信にはノイズ(雑音)が妨げとなる  
**Noises are obstructive in communication.**

雑音の入る通信路を介して情報通信を行なう際、  
通信途中での誤りに如何に対処するか?

**When trying to communicate  
with noisy channel,  
how can one overcome the errors ?**

誤りに対処しつつ、如何に効率的に伝達するか?  
**Dealing with the errors,  
how efficiently can one communicate ?**

→ **Coding theory(符号理論)**  
• **Error-correcting codes(誤り訂正符号)**

## 誤り訂正符号 (お話)(Error-correcting codes)

### How to overcome errors occurring in channels:

- **Physical technology: Error suppression**  
誤りの発生を抑える (物理技術による対処)  
(導線の高品質化・ノイズの遮蔽・等々)
- **Social technology: The thought of “fail-safe”**  
誤りが起きても致命的にならないように  
(フェイルセーフの発想・社会技術)
- **Mathematical technology: Error correction**  
多少の誤りなら検出・訂正できる仕組み  
→ 数理技術により実現 (誤り訂正符号)



## 誤り訂正符号 (お話)(Error-correcting codes)

情報通信中に誤り (らしきこと) に出遭ったら？

**How should we do when we meet**

**a (suspicious) error in communication ?**

例: 「じゃあヨツバ駅で待ち合わせね」

**“We will meet at Yotsuba station.”**

- 聞き直す (より安全なプロトコルの採用)  
**Ask again (more secure protocol)**
- 見当を付ける (誤りの自動訂正)  
**Make a guess (error-correcting)**

## 誤り訂正符号 (お話)(Error-correcting codes)

例: 「じゃあヨツバ駅で待ち合わせね」

“We will meet at **Yotsuba** station.”

→ きっとヨツヤ駅だろう

It should be **Yotsuya** station.

- **Why can one be aware of the error?**  
→ **No station has the name “Yotsuba”.**
- **Why can one guess the correct name?**  
→ **No other station has a similar name.**

← **Only few strings are correct names.**

## 誤り訂正符号 (お話)(Error-correcting codes)

- **Why can one be aware of the error?**  
→ **No station has the name “Yotsuba”.**
- **Why can one guess the correct name?**  
→ **No other station has a similar name.**

← **Only few strings are correct names.**

**If all strings were names of some stations,  
we could not correct the error.**

→ **冗長性**を利用して**確実性**を確保した  
**Use of redundancy to keep certainty**

## 誤り訂正符号 (お話)(Error-correcting codes)

「誤りの自動訂正」が出来るように

予め適切に冗長性を持たせて通信する

**Give some controlled redundancy beforehand  
to do automatic error-correction.**

- 誤り訂正性能は高く  
(High error-correction ability)
- とは言えなるべく効率的に  
(As efficient as possible)

→ 有限体上の線型代数・代数幾何などの利用  
(Use of linear algebra and algebraic geometry  
over finite fields)