本講義の概要

- **情報通信の数理**
  - ⋆ 情報理論
  - ⋆ 符号理論
  - ⋆ 暗号理論
- **それを支える数学**
  - ⋆ 有限体とその上の線型代数・Galois 理論
  - ⋆ 情報量の理論・計算量の理論

情報通信を行なう際の要請

- **効率的に** $\longrightarrow$ **情報理論**
- **確実に** $\longrightarrow$ **符号理論**
- **安全に** $\longrightarrow$ **暗号理論**

情報理論 (情報源符号化・情報量の理論)

- 伝えるべき情報をより効率良く伝えるには

- 「効率の良さ」を計る
  - ⋆ 伝えるべき「情報の量」を計る
  - ⋆ 伝える為の「手間」を計る

$\longrightarrow$ Shannon
「情報の量は伝えるのに必要な手間と一致」

情報の符号化

アナログデータ (連続データ)

⇓ サンプリングなど

ディジタルデータ (離散・有限データ)
↑ ここで扱う “情報” (元データ・情報源)

⇓ ⟵ ここで扱う “符号化”

通信用の形式のディジタルデータ
(特定 (一般には少数) の種類の文字の列)

文字データ $\longrightarrow$ ASCII code

	0	1	2	3	4	5	6	7	8	9	a	b	c	d	e	f
0	(control characters)															
1	(control characters)															
2		!	"	#	$	%	&	’	(	)	*	+	,	-	.	/
3	0	1	2	3	4	5	6	7	8	9	:	;	<	=	>	?
4	@	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O
5	P	Q	R	S	T	U	V	W	X	Y	Z	[	\	]	^	_
6	‘	a	b	c	d	e	f	g	h	i	j	k	l	m	n	o
7	p	q	r	s	t	u	v	w	x	y	z	{	|	}	~	

7 bit で 1 文字を表す (実装上は 8 bit にする)

ASCII code

A	01000001
B	01000010
C	01000011
D	01000100
E	01000101
F	01000110
G	01000111
H	01001000
I	01001001

J	01001010
K	01001011
L	01001100
M	01001101
N	01001110
O	01001111
P	01010000
Q	01010001
R	01010010

S	01010011
T	01010100
U	01010101
V	01010110
W	01010111
X	01011000
Y	01011001
Z	01011010

モールス符号 (Morse code)

A	・ ー
B	ー ・・・
C	ー ・ ー ・
D	ー ・・
E	・
F	・・ ー ・
G	ー ー ・
H	・・・・
I	・・

J	・ ー ー ー
K	ー ・ ー
L	・ ー ・・
M	ー ー
N	ー ・
O	ー ー ー
P	・ ー ー ・
Q	ー ー ・ ー
R	・ ー ・

S	・・・
T	ー
U	・・ ー
V	・・・ ー
W	・ ー ー
X	ー ・・ ー
Y	ー ・ ー ー
Z	ー ー ・・

モールス符号 (Morse code)

1 **文字のための符号長が区々**

$\longleftarrow$ **頻度の高い文字は短く、低い文字は長く**

$\longrightarrow$ **頻度まで考慮して符号長の期待値を短く**

$\cdots$ **頻度 (出現確率) を考慮して**
符号効率の定式化を考えている

モールス符号 (Morse code)

1 **文字のための符号長が区々**

$\longleftarrow$ **頻度の高い文字は短く、低い文字は長く**

$\longrightarrow$ **頻度まで考慮して符号長の期待値を短く**

$\cdots$ **頻度 (出現確率) を考慮して**
符号効率の定式化を考えている

モールス符号 (Morse code)

1 **文字のための符号長が区々**

$\longleftarrow$ **頻度の高い文字は短く、低い文字は長く**

$\longrightarrow$ **頻度まで考慮して符号長の期待値を短く**

$\cdots$ **頻度 (出現確率) を考慮して**
符号効率の定式化を考えている

モールス符号 (Morse code)

1 **文字のための符号長が区々**

$\longleftarrow$ **頻度の高い文字は短く、低い文字は長く**

$\longrightarrow$ **頻度まで考慮して符号長の期待値を短く**

$\cdots$ **頻度 (出現確率) を考慮して**
符号効率の定式化を考えている

モールス符号 (Morse code)

1 **文字のための符号長が区々**

$\longrightarrow$ 1 **文字毎の区切りは判るのか ？**

$\longrightarrow$ **実はモールス符号では、**
文字間・単語間の送信間隔が
定められている

モールス符号 (Morse code)

1 文字のための符号長が区々

$\longrightarrow$ 1 文字毎の区切りは判るのか ？

$\longrightarrow$ 実はモールス符号では、
文字間・単語間の送信間隔が
定められている

A	・ー
B	ー ・・・
C	ー ・ ー ・
D	ー ・・
E	・
F	・・ ー ・
G	ー ー ・

H	・・・・
I	・・
J	・ ー ー ー
K	ー ・ ー
L	・ ー ・・
M	ー ー
N	ー ・

O	ー ー ー
P	・ ー ー ・
Q	ー ー ・ ー
R	・ ー ・
S	・・・
T	ー

U	・・ ー
V	・・・ ー
W	・ ー ー
X	ー ・・ ー
Y	ー ・ ー ー
Z	ー ー ・・

例: the cat and the dog

ー○ ・・・・ ○ ・ ○○○ー ・ ー ・ ○ ・ ー○ー○○
○・ー○ー・○ー ・・○○○ー○ ・・・・ ○ ・ ○○○ー ・・
○ー ー ー○ー ー ・ **(○は・1 個分の間隔)**

⟶ ーと・と○とを用いて符号化している

情報源符号化の定式化

情報源 alphabet S : 有限集合

$S^+ := \bigsqcup_{n \geq 1} S^n$: S **の元の** 1 **個以上の列**

$S^0 := \{\varepsilon\}$: **空語**

$S^* := \bigsqcup_{n \geq 0} S^n$: S **の元の** 0 **個以上の列**

$$= S^+ \sqcup \{\varepsilon\}$$

$w \in S^n$ **に対し、** $|w| := n$ **(文字列の長さ)**

情報源符号化の定式化

符号 (伝送)alphabet T : 有限集合
(しばしば $T = \{0, 1\}$)

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

$\mathcal{C}$ **の像の元: 符号語 (code word)**

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

符号への要請

- **一意復号可能 (uniquely decodable) か ？**
- **一意復号可能とした上で、瞬時復号可能 (instantaneously decodable) か ？**
- **その上で効率が良いか ？**

一意復号可能でない例

$$S = \{a, b, c\}, T = \{0, 1\}$$

$$a \longmapsto 0$$

$$b \longmapsto 01$$

$$c \longmapsto 001$$

「001」が ab か c か判らない

$\longrightarrow$ 一意復号可能でない!!

一意復号可能でない例

$S = \{a, b, c\}, T = \{0, 1\}$

$$a \longmapsto 0$$
$$b \longmapsto 01$$
$$c \longmapsto 001$$

「001」が ab か c か判らない

$\longrightarrow$ 一意復号可能でない!!

一意復号可能でない例

$$S = \{a, b, c\}, T = \{0, 1\}$$

$$a \longmapsto 0$$

$$b \longmapsto 01$$

$$c \longmapsto 001$$

「001」が ab か c か判らない

$\longrightarrow$ **一意復号可能でない!!**

瞬時復号可能でない**例**

$S = \{a, b, c\}, T = \{0, 1\}$

$$a \longmapsto 0$$

$$b \longmapsto 01$$

$$c \longmapsto 11$$

一意復号可能ではあるが、

「011$\cdots$」まで見ただけでは

ac$\cdots$ か bc$\cdots$ か判らない

(「0111」なら bc、「01111」なら acc)

$\longrightarrow$ 瞬時復号可能でない

瞬時復号可能でない例

$S = \{a, b, c\}, T = \{0, 1\}$

$$a \longmapsto 0$$

$$b \longmapsto 01$$

$$c \longmapsto 11$$

一意復号可能ではあるが、

「011$\cdots$」まで見ただけでは

ac$\cdots$ か bc$\cdots$ か判らない

(「0111」なら bc、「01111」なら acc)

$\longrightarrow$ **瞬時復号可能でない**

瞬時復号可能でない例

$S = \{a, b, c\}, T = \{0, 1\}$

$$a \longmapsto 0$$

$$b \longmapsto 01$$

$$c \longmapsto 11$$

一意復号可能ではあるが、

「011$\cdots$」まで見ただけでは

ac$\cdots$ か bc$\cdots$ か判らない

(「0111」なら bc、「01111」なら acc)

$\longrightarrow$ 瞬時復号可能でない

瞬時復号可能でない例

$S = \{a, b, c\}, T = \{0, 1\}$

$$a \longmapsto 0$$

$$b \longmapsto 01$$

$$c \longmapsto 11$$

一意復号可能ではあるが、

「011$\cdots$」まで見ただけでは

ac$\cdots$ か bc$\cdots$ か判らない

(「0111」なら bc、「01111」なら acc)

$\longrightarrow$ **瞬時復号可能でない!!**

瞬時復号可能な例

$S = \{a, b, c\}, T = \{0, 1\}$

$$a \longmapsto 0$$
$$b \longmapsto 10$$
$$c \longmapsto 11$$

$\longrightarrow \mathcal{C}(S) = \{0, 10, 11\} \subset T^+$

だけを見て判る特徴があるか ？

符号への要請

- **一意符号:**

$$\mathcal{C}^* : S^* \longrightarrow T^* : \textbf{単射}$$

- **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)

- **効率が良い** $\cdots$ **符号長** $|\mathcal{C}(s)|$ **が小さい**

瞬時符号の性質

- $\mathcal{C}$：瞬時符号 $\Longrightarrow$ $\mathcal{C}$：一意符号
- $\mathcal{C}$：瞬時符号 $\Longleftrightarrow$ $\mathcal{C}$：語頭符号

$$\left(\mathcal{C}(s') = \mathcal{C}(s)\boldsymbol{x} \Longrightarrow s' = s, \boldsymbol{x} = \varepsilon\right)$$

瞬時符号の作り方

「符号語木」を考えよう

瞬時符号の性質

- $\mathcal{C}$：瞬時符号 $\Longrightarrow$ $\mathcal{C}$：一意符号
- $\mathcal{C}$：瞬時符号 $\Longleftrightarrow$ $\mathcal{C}$：語頭符号

$$\left(\mathcal{C}(s') = \mathcal{C}(s)\boldsymbol{x} \Longrightarrow s' = s, \boldsymbol{x} = \varepsilon\right)$$

瞬時符号の作り方

「符号語木」を考えよう