

## I. 情報源符号化

効率の良い情報源の符号化方法と復号方法について学習する。情報理論において、「効率がいい」符号化方法とは平均情報量が最も小さい符号化方法のことを指す。また「効率がいい」復号方法とは可逆で符号を前から見た時に瞬時に復号できる方法のことを指す。

効率の良い符号化方法と復号方法はどのような性質を持っていて欲しいかを考える。情報源アルファベットを  $A = \{A, B, C, D\}$  し確率分布が以下のような情報源を考える。

情報源記号	確率	$C_1$	$C_2$	$C_3$	$C_4$	$C_5$	$C_6$
A	0.6	00	0	0	0	0	0
B	0.25	01	10	10	01	10	10
C	0.1	10	110	110	011	11	11
D	0.05	11	1110	111	111	01	0

情報を 01 の符号で表現するとき、情報源記号 A は 00、B は 01、C は 10、D は 11 で表現され、これよりも効率的に情報を表現する方法はないかを考える。情報源それぞれの発生確率が異なっているため、情報源記号ごとに符号の長さを変えることで**平均符号長**をより小さくすることができる。

### i. さまざまな符号の特徴

上の表の符号化を例に符号化の特徴を考える。

#### 1. 特異符号 (singular code)

$C_6$  の符号化のように異なる情報源記号に同じ符号語が割り当てられる符号を特異符号という。特異符号は復号が困難であるため、情報源記号ごとに異なる符号語を割り当てることが望ましい。特異符号のように符号化が困難な符号は**一意復号不可能な符号**といい、他の  $C_1, C_2, C_3, C_4, C_5$  は**一意複合可能な符号**という。

#### ii. 瞬時符合 (instantaneous code)

情報源記号を前から見た時に、その時点でどの情報源記号であるかが瞬時にわかる符号を**瞬時符合**という。 $C_4$  は最終的には 0(A)111(D)111(D)0 と復元できるが、2 番目の 0 が来た時点で情報源記号が D であることがわかるため、瞬時符合ではない。0 がくるまでには 01(B)111(D)11 という可能性もはらんでいる。一方で、 $C_1, C_2, C_3, C_5$  は瞬時符合である。