第 8 回　エントロピー（平均情報量）

（教科書 1.2.1 自己エントロピー）

自己情報量

事象 a の生起確率 $P(a)$ とする。

- a が起こったことにより得られる情報量 $I(a)$
  $I(a) \propto \dfrac{1}{P(a)}$ $\cdots$ 事前の確率 $P(a)$ が小さいほど、それが生じた場合の情報量が大きい
- $P(a)$ が小 $\to$ $I(a)$ が大 $\Longrightarrow$ $P(a) = 0 \to I(a) = \infty$
  $\Updownarrow$
  $P(a)$ が大 $\to$ $I(a)$ が小 $\Longrightarrow$ $P(a) = 1 \to I(a) = 0$
- 情報量の加法性
  2 つの独立した事象から得られる情報量は、各々から得られる情報量の和

  $$I(a \cap b) = I(a) + I(b)$$

情報量の満たすべき条件

(1) $I(a) = I(P(a))$

(2) $I(a)$ は $P(a)$ の減少関数

(3) $I(a \cap b) = I(a) + I(b)$

この 3 条件を満たす関数 $\Rightarrow$ 対数関数

$$\underbrace{I(a) = I(P(a)) = \log \frac{1}{P(a)} = -\log P(a)}_{\textbf{自己情報量}}$$

情報理論においては一般に底を 2 として

$I(a) = -\log_2 P(a)$ [bit]

エントロピー（平均情報量）

1 つの情報源から発生する通報 $x_i (i = 1, 2, \cdots, n)$,

確率変数 $X = \{x_1, x_2, \cdots, x_n\}$ ，各々の生起確率 $P(x_i)$

$$\sum_{i=1}^{n} P(x_i) = 1$$

自己情報量 $I(x_i)$ の期待値 $H(X)$

$$\underbrace{H(X) = \sum_{i=1}^{n} P(x_i) I(x_i) = -\sum_{i=1}^{n} P(x_i) \log_2 P(x_i) \text{ [bit]}}_{\textbf{平均情報量} \longrightarrow \textbf{情報源のエントロピー}}$$

エントロピー関数（entropy function）

$n = 2$ の場合（2 元情報源）

$$X = \left\{ \begin{array}{cc} x_1, & x_2 \\ P(x_1), & P(x_2) \end{array} \right\} = \left\{ \begin{array}{cc} x_1, & x_2 \\ p, & 1-p \end{array} \right\}$$

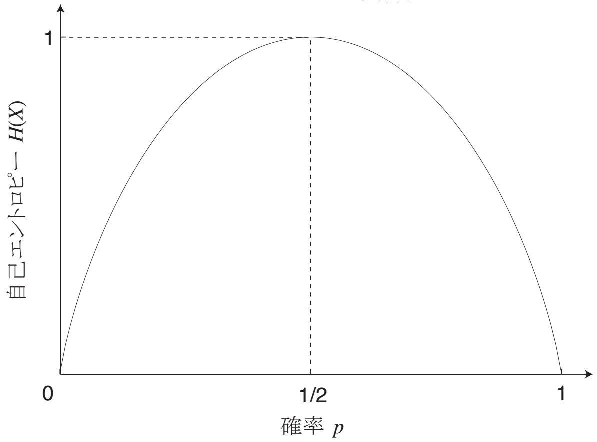
平均情報量（情報源のエントロピー）

$\underline{H(X) = -p\log_2 p - (1-p)\log_2(1-p) \ [\text{bit}]}$

$H(X)$ は p のみで与えられる関数

$\underline{H(p) = -p\log_2 p - (1-p)\log_2(1-p) \ [\text{bit}]}$

エントロピー関数



n 元事象系

平均情報量（情報源のエントロピー）

$$H(X) = -\sum_{i=1}^{n} P(x_i)\log_2 P(x_i) \ [\text{bit}]$$

$0 \le H(X) \le \log_2 n$ を満足

$P(x_i) = 1$，$P(x_j) = 0 \ (i \ne j)$（生起する事象が自明）のとき，$H(X) = 0$（最小値）

$P(x_i) = 1/n$ （等確率）のとき，$H(X) = \log_2 n$（最大値）

↓

全く予想がつかない

参考文献

[1] 久保田一, 大石邦夫, 福本昌弘, ”1.2.1 自己エントロピー”, C 言語による情報理論入門, pp.15–18, コロナ社, 2007 (ISBN978–4–339–02521–0).