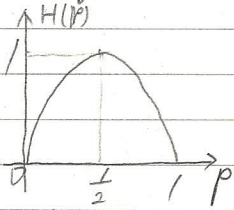


例. $P_i = \frac{1}{\Omega}$ $H(P) = -\sum_{i=1}^{\Omega} \frac{1}{\Omega} \log_2 \frac{1}{\Omega}$
 $= \log_2 \Omega$

$\Omega = 2$ $P_1 = p$ $P_2 = 1-p$ $0 \leq p \leq 1$
 $H(P) = -p \log_2 p - (1-p) \log_2 (1-p)$
 $0 \log 0 = 0$ とする



Shannon エントロピーの相加性

2つの独立な確率的な系 — system

- 系1 事象 $i=1, 2, \dots, \Omega_1$ 確率分布 $P^{(1)} = (P_1^{(1)}, \dots, P_{\Omega_1}^{(1)})$
- 系2 事象 $j=1, 2, \dots, \Omega_2$ 確率分布 $P^{(2)} = (P_1^{(2)}, \dots, P_{\Omega_2}^{(2)})$

両者を合わせた全系

事象 (i, j) $\Omega = \Omega_1, \Omega_2$ 通りの事象
 (i, j) の出現する確率 $P(i, j) = P_i^{(1)} P_j^{(2)}$
 $0 \leq P(i, j) \leq 1$ である

$$\sum_{(i,j)} P(i, j) = \sum_{i=1}^{\Omega_1} \sum_{j=1}^{\Omega_2} P_i^{(1)} P_j^{(2)} = \left(\sum_{i=1}^{\Omega_1} P_i^{(1)} \right) \left(\sum_{j=1}^{\Omega_2} P_j^{(2)} \right) = 1$$

全系の確率分布 $P = (P(i, j))_{\substack{i=1, \dots, \Omega_1 \\ j=1, \dots, \Omega_2}}$

$H(P) = H(P^{(1)}) + H(P^{(2)})$ 相加性

$$\begin{aligned} H(P) &= -\sum_{(i,j)} P(i, j) \log_2 P(i, j) \\ &= -\sum_{i=1}^{\Omega_1} \sum_{j=1}^{\Omega_2} P_i^{(1)} P_j^{(2)} \left[\log_2 P_i^{(1)} + \log_2 P_j^{(2)} \right] \\ &= -\sum_{i=1}^{\Omega_1} \sum_{j=1}^{\Omega_2} P_i^{(1)} P_j^{(2)} \log_2 P_i^{(1)} \\ &\quad - \sum_{i=1}^{\Omega_1} \sum_{j=1}^{\Omega_2} P_i^{(1)} P_j^{(2)} \log_2 P_j^{(2)} \\ &= -\sum_{i=1}^{\Omega_1} P_i^{(1)} \log_2 P_i^{(1)} - \sum_{j=1}^{\Omega_2} P_j^{(2)} \log_2 P_j^{(2)} \end{aligned}$$

情報源の符号化 (source coding)

例. 情報源 A, B, C, D, E の5種類の文字を
 N 個並べた任意の文字列

BACA DECA CABA $N=12$

↓ coding

0, 1 を使った文字列 → 長さ L 短くなる

↓ 展開

BACA DECA CABA 7-9 圧縮

[1] A → 000

B → 001

BACA

C → 010

00/0000/0000

D → 011

E → 100

長さ L だけ E を混ぜてみる

A 00

BACADE

[2] B 01

0100/000/110/111

C 10

D 110

左端から E を解読可能

接頭符号 (prefix code)

E の符号を他の符号の接頭部 n にはしない

A → 0

[3] B → 10

C → 110

D → 1110

E → 11110

A, B, C, D, E N 文字

↓ coding

2, 1, 0 L 文字

[1] のとき $L = 3N$

[2] のとき $L \leq 3N$