

エントロピー (Entropy)

情報量

$P(a)$ を事象 a の発生確率とする. このとき事象 a の (自己) 情報量は以下のように定義される.

$$I(a) = \log \frac{1}{P(a)}$$

情報量の意味は, 珍しいほど情報の価値が高いという考え方をもとにして作られた関数. 発生確率 $P(a)$ について単調減少な関数となっている. 情報量の単位として, 対数の底が 2 のときはビット, e のときはナット, 10 のときはハートレイと呼ばれる. 今後の議論では, 対数の底が 2 であるとする.

定義域である発生確率は $0 \leq P(a) \leq 1$ なので, 値域である情報量は $0 \leq I(a) \leq \infty$ となる. よって情報量は $P(a) = 1$ のとき最小値 $I(a) = 0$ をとる.

エントロピー

事象の集合を事象系 $A = \{a_1, a_2, \dots, a_J\}$ と呼ぶ. 事象系は各事象の発生確率について以下を満たす.

$$\sum_{j=1}^J P(a_j) = 1$$

事象系 A の各事象が持っている情報量の平均を平均情報量 (エントロピー) と呼び, 以下のように定義する.

$$H(A) = \sum_{j=1}^J P(a_j) I(a_j) = \sum_{j=1}^J P(a_j) \log \frac{1}{P(a_j)}$$

エントロピーは不確かさを表す. または観測者が得る情報量の平均とも考えられる.

エントロピーの最小値

事象系 A をある事象 a_i の発生確率 $P(a_i)$ が 1 で, それ以外の事象の発生確率が 0 となるような事象系とする. このとき A のエントロピーが最小になることを示す. A のエントロピーは以下ようになる.

$$H(A) = -0 \log 0 - \dots - 0 \log 0 - 1 \log 1 - 0 \log 0 - \dots - 0 \log 0 = 0$$

よって A 以外の任意の事象系 A' について $H(A') \geq 0$ を示す. 事象系の性質より, A 以外の事象系では発生確率が 0, 1 以外の事象が 2 つ以上存在する. そしてその 2 つの事象の情報量は, 0 よりおおきい. また任意の a'_j について $P(a'_j)$ は非負なので, $H(A') \geq 0$ を満たす.

エントロピーの最大値

事象系 A をすべての事象が均等な発生確率であるような事象系とする。このとき A のエントロピーが最大になることを示す。 A のエントロピーは以下のようになる。

$$H(A) = \sum_{j=1}^J \frac{1}{J} \log J = \log J$$

よって A 以外の任意の事象系 A' について $H(A') - \log J \leq 0$ を示す。左辺から、

$$\begin{aligned} H(A') - \log J &= \sum_{j=1}^J P(a'_j) \log \frac{1}{P(a'_j)} - \log J \\ &= \sum_{j=1}^J \left(P(a'_j) \log \frac{1}{P(a'_j)} - \log J \right) \\ &= \sum_{j=1}^J P(a'_j) \log \frac{1}{P(a'_j) \log J} \\ &= \frac{1}{\ln 2} \sum_{j=1}^J P(a'_j) \ln \frac{1}{P(a'_j) \log J} \end{aligned}$$

任意の 0 より大きい x について $\ln x \leq x - 1$ を満たす。よって、

$$\begin{aligned} H(A') - \log J &\leq \frac{1}{\ln 2} \sum_{j=1}^J P(a'_j) \left(\frac{1}{P(a'_j) \log J} - 1 \right) \\ &= \frac{1}{\ln 2} \sum_{j=1}^J \left(\frac{1}{\log J} - P(a'_j) \right) \\ &= \frac{1}{\ln 2} (1 - 1) \\ &= 0 \end{aligned}$$

これらのことから $H(A') - \log J \leq 0$ を示せた。