エントロピー

(1). $H(A) \geq 0$ ($p(a_j) = 1$ となる a_j が存在する時、等号が成立)

(2). $H(A) \leq \log_2 m$ ($p(a_1) = p(a_2) = \cdots = p(a_m) = \frac{1}{m}$ の時、等号が成立)

....

$A = \{a_1, a_2, \ldots, a_m\}$: 事象の集合

Ω : 標本空間

$p : \Omega \to [0, 1] \subset \mathbb{R}$: 確率関数

- $0 \leq p(a_i) \leq 1$ $(i = 1, 2, \ldots, m)$: 各確率は 0 から 1 の間をとる
- $\log_2 p(a_i) \leq 0$ $(i = 1, 2, \ldots, m)$: 1 でない確率の対数は負の数
- $0 \leq \sum_{i=1}^{m} p(a_i) \leq 1$: 全ての確率の和は 1 を超えない

$H : 2^{\Omega} \to \mathbb{R}$: エントロピー関数

$H(A) = \sum_{i=1}^{m} p(a_i)\,(-\log_2 p(a_i))$: $-\log_2 p(a_i)$ の期待値

証明

$H(A) \geq 0$ ($p(a_j) = 1$ となる a_j が存在する時、等号が成立)

等号の場合

$p(a_j) = 1$ となる j が存在する場合、$i \neq j$ であるとき $p(a_i) = 0$ である。これを $H(A)$ に当てはめると次の式を得る。

$$
\begin{aligned}
H(A) &= \sum_{i=1}^{m} p(a_i)\,(-\log_2 p(a_i)) && (1)\\
&= 0(-\log_2 0) + \cdots + 1(-\log_2 1) + \cdots + 0(-\log_2 0) && (2)
\end{aligned}
$$

$\log_2 1 = 0$ より j 番目の項は 0 である。それ以外の項については次の式から 0 とする。

$$\lim_{x \to 0} x \log x = 0 \qquad (3)$$

これにより $H(A) = 0$ である。

等号でない場合

$0 < p(a_j) < 1$ の確率がある場合、$\log_2 p(a_j) < 0$ であるから、$p(a_j)(-\log_2 p(a_j)) > 0$ となる。

$p(a_j) = 0$ または $p(a_j) = 1$ のときは $p(a_j)(-\log_2 p(a_j)) = 0$ である。

よって $H(A) = \sum_{i=1}^{m} p(a_i)\left(-\log_2 p(a_i)\right)$ の全ての項は 0 以上となるので $H(A) > 0$

・・

$H(A) \leq \log_2 m$ ($p(a_1) = p(a_2) = \cdots = p(a_m) = \frac{1}{m}$ の時、等号が成立)

等号の場合

$p(a_i) = 1/m$ ($i = 1, \ldots, m$) を $H(A)$ に代入する

$$H(A) = \sum_{i=1}^{m} p(a_i)\left(-\log_2 p(a_i)\right) \tag{4}$$

$$= \sum_{i=1}^{m} \frac{1}{m}\left(-\log_2 \frac{1}{m}\right) = \log_2 m \tag{5}$$

等号でない場合

$H(A)$ の最大値を求め、この最大値が $\log_2 m$ であることを示す。

・・・・・・・・・・・・・・・・・・・・・・・[ラグランジュの未定係数法] ・・・・・・・・・・・・・・・・・・・・・・・

λ を定数、$\sum_{i=1}^{m} x_i = 1$ とし次の関数 F の最大値について考える。

$$F(x_1, x_2, \ldots, x_m) = -\sum_{i=1}^{m} x_i \log_2 x_i + \lambda\left(1 - \sum_{i=1}^{m} x_i\right) \tag{6}$$

$F(x_1, x_2, \ldots, x_m)$ を各 x_i について偏微分する。

$$\frac{\partial}{\partial x_i} F(x_1, x_2, \ldots, x_m) = -\log_2 x_i - \frac{1}{\log 2} - \lambda \quad (i = 1, \ldots, m) \tag{7}$$

最大値を考えるので偏導関数が 0 となる時を考える。

$$-\log_2 x_i - \frac{1}{\log 2} - \lambda = 0 \quad (i = 1, \ldots, m) \tag{8}$$

これを x_i について解くと次が得られる。

$$x_i = \exp(-1 - \lambda \log 2) \quad (i = 1, \ldots, m) \tag{9}$$

λ は定数であるので右辺は i によらずある値を指す。つまり各 $i = 1, \ldots, m$ において偏導関数が 0 となるときはいつも同じ値となる。

$$x_1 = \cdots = x_m = \exp(-1 - \lambda \log 2) \tag{10}$$

これを $\sum_{i=1}^{m} x_i = 1$ に代入し λ について解くと次のようになる。

$$\sum_{i=1}^{m} \exp(-1 - \lambda \log 2) = 1 \tag{11}$$

$$m \cdot \exp(-1 - \lambda \log 2) = 1 \tag{12}$$

$$m = \exp(1 + \lambda \log 2) \tag{13}$$

$$\log m = 1 + \lambda \log 2 \tag{14}$$

$$\lambda = \frac{1}{\log 2}(\log m - 1) \tag{15}$$

この式 (15) を式 (9) に代入すると $x_i = m^{-1}$ が得られる。

また、次の 2 式からこの極値の候補は最大値であることがわかる。

$$\frac{\partial^2}{\partial x_i \partial x_j} F(x_1, x_2, \ldots, x_m) = 0 \quad (i \neq j) \tag{16}$$

$$\frac{\partial^2}{\partial x_i^2} F(x_1, x_2, \ldots, x_m) = -\frac{1}{x_i \log 2} = -\frac{m}{\log 2} < 0 \tag{17}$$

よって、$x_1 = \cdots = x_m = m^{-1}$ の時最大値を取る。

....................[ラグランジュの未定係数法 ここまで] ....

上記の 未定係数法から $p(a_1) = \cdots = p(x_m) = 1/m$ である時、$H(A)$ が最大値を取ることがわかる。

これにより $H(A)$ は最大値 $\log_2 m$ より小さくなるため $H(A) < \log_2 m$ である。

等号の場合と等号でない場合を合わせ

$$H(A) \leq \log_2 m \tag{18}$$

がわかる。