参考資料： 前回と今回の定理の証明など

工学部 情報エレクトロニクス学科

情報科学研究院 情報理工学部門
大規模知識処理研究室

堀山 貴史

補助定理A.1[シャノンの補助定理]

補助定理A.1

$p_1, p_2, \cdots, p_M$ および $q_1, q_2, \cdots, q_M$ を

$$p_1 + p_2 + \cdots + p_M = 1,$$

$$q_1 + q_2 + \cdots + q_M \leq 1$$

を満たす任意の非負の数とする(ただし，$p_i \neq 0$ のときは $q_i \neq 0$ とする)．このとき，

$$-\sum_{i=1}^{M} p_i \log_2 q_i \geq -\sum_{i=1}^{M} p_i \log_2 p_i \qquad \text{(A.3)}$$

が成立する．等号は $q_i = p_i$ $(i = 1, 2, \cdots, M)$のとき，またそのときに限って成立する

．

証明は教科書を参照

つまり，確率分布$P = \{p_i\}_{i=1}^{M}$ とちょっと違う分布 q_i (ただし総和が1以下)を持ってきて，$\log_2$ の内側の p_i と置き換えると，元よりも少し大きくなる.

定理2.1の証明

Xのエントロピー$H(X)$は

$$H(X) = -\sum_{i=1}^{M} p_i \log_2 p_i \ .$$

$-\log_2 p_i \geq 0$ だから

$0 \leq p_i \leq 1$ なので，明らかに $0 \leq H(S)$ であり，$H(S) = 0$ が成立するのは，$p_1, p_2, \cdots, p_k$ のうち一つが 1 で他が 0 の場合である.

$\sum_{i=1}^{M} p_i = 1$ だから

補助定理A.1（シャノンの補助定理）を$q_i = 1/M$として適用すると，

$$\begin{aligned} H(X) &= -\sum_{i=1}^{M} p_i \log_2 p_i \\ &\leq -\sum_{i=1}^{M} p_i \log_2 \frac{1}{M} \\ &= \log_2 M \ . \end{aligned}$$

補助定理A.1より

等号が成立するのは $p_i = q_i = 1/M$ のときのみである. □

定理2.2の証明

［証明］ 結合エントロピーの定義より$0 \leq H(X,Y)$は明らかである.
よって, $H(X,Y) \leq H(X) + H(Y)$を証明する.

$$H(X) = -\sum_{i=1}^{M_X} P(x_i) \log_2 P(x_i) = -\sum_{i=1}^{M_X} \sum_{j=1}^{M_Y} P(x_i, y_j) \log_2 P(x_i),$$

$$H(Y) = -\sum_{j=1}^{M_Y} P(y_j) \log_2 P(y_j) = -\sum_{j=1}^{M_Y} \sum_{i=1}^{M_X} P(x_i, y_j) \log_2 P(y_j).$$

したがって,

$$H(X) + H(Y) = -\sum_{j=1}^{M_Y} \sum_{i=1}^{M_X} P(x_i, y_j) \log_2 P(x_i) P(y_j)$$

定理2.2の証明(つづき)

A.1節の補助定理A.1(シャノンの補助定理)を適用すると,

$$-\sum_{j=1}^{M_Y}\sum_{i=1}^{M_X} P(x_i, y_j) \log_2 P(x_i)P(y_j)$$

$$\geq -\sum_{j=1}^{M_Y}\sum_{i=1}^{M_X} P(x_i, y_j) \log_2 P(x_i, y_j)$$

が成り立つ. すなわち, $H(X) + H(Y) \geq H(X, Y)$ となる.

等号が成り立つのは, シャノンの補助定理の統合条件より, すべてのi, jに対して$P(x_i, y_j) = P(x_i)P(y_j)$が成立する場合である.
これは, X と Y が独立であるときに他ならない. □

定理2.3(2)の証明

［証明］ 結合エントロピーと条件付き確率の定義から,

$$H(X,Y) = -\sum_{i=1}^{M_X}\sum_{j=1}^{M_Y} P(x_i, y_j)\log_2 P(x_i, y_j)$$

$$= -\sum_{i=1}^{M_X}\sum_{j=1}^{M_Y} P(x_i, y_j)\log_2 \frac{P(x_i, y_j)P(x_i)}{P(x_i)}$$

ベイズの定理

$$= -\sum_{i=1}^{M_X}\sum_{j=1}^{M_Y} P(x_i, y_j)\left\{\log_2 P(x_i) + \log_2 P(y_j|x_i)\right\}$$

$$= H(X) + H(Y|X)$$

が成立する.

$H(X,Y) = H(Y) + H(X|Y)$ も同様にして証明できる. □

自己情報量が対数関数である理由(1/3)

まず，コーシー（Cauchy）の関数方程式

$$f(x+y)=f(x)+f(y)$$

を満たす連続関数が $f(x)=kx$（kは定数）であることを示す．

$x=y=0$ を代入すると，$f(0)=f(0)+f(0)$ より，$f(0)=0$.
次に，$y=-x$ を代入すると，

$$\begin{aligned} f(x-x)&=f(x)+f(-x),\\ 0&=f(x)+f(-x). \end{aligned}$$

より，$f(-x)=-f(x)$ が成り立つ（つまり，$f(x)$ は奇関数）．
n が自然数のとき，

$$\begin{aligned} f(n)=f\big(1+(n-1)\big)&=f(1)+f(n-1)\\ &=f(1)+f(1)+f(n-2)=\cdots=nf(1). \end{aligned}$$

$f(x)$が奇関数であることから，$f(-n)=-nf(1)$も成り立つ．すなわち，任意の整数について$f(n)=nf(1)$が成り立つ．

自己情報量が対数関数である理由(2/3)

同様の考えにより，任意の実数 x と自然数 m に対して，
$f(mx) = mf(x)$ が成り立つ．m が自然数，n が整数のとき，

$$f(n) = f\left(\frac{n}{m} \times m\right) = mf\left(\frac{n}{m}\right)$$

より，

$$f\left(\frac{n}{m}\right) = \frac{f(n)}{m} = \frac{n}{m} f(1).$$

したがって，任意の有理数 x に対して，次が成り立つ．

$$f(x) = xf(1).$$

$f(1)$は定数なので，これを k と置くと，$f(x) = kx$ と書ける．
有理数の稠密性から，連続関数に限定するとコーシーの関数方程式を満たす解は $f(x) = kx$ のみであることが言える．

どんなに微小な区間をとっても，その間に有理数が存在する

自己情報量が対数関数である理由(3/3)

コーシーの関数方程式の解を応用して，自己情報量の三つの性質を満たす関数 $I(p)$ が対数関数で表されることを示す.

ある定数 $a > 1$ に対して，$f(x) = I(a^x)$ とおく. このとき，

$$f(x+y) = I(a^{x+y}) = I(a^x \cdot a^y) = I(a^x) + I(a^y)$$
$$= f(x) + f(y)$$

$I(p)$の加法性から

が成り立つ. コーシーの関数方程式の解から，

$$f(x) = kx$$

と書ける（kは定数）. ここで，$p = a^x$ とおくと，最初の式から，

$$I(p) = f(x) = kx = k \log_a p\,.$$

$x = \log_a p$

$I(p)$ は $0 < p \leq 1$ で単調減少関数なので，$k < 0$ でなければならない. $k = -1$ ととれば，$I(p) = -\log_a p$ となる.