第 7 章「統 計 学 に お け る 情 報 量」

1．Hartley 情報量

定義　N 個の要素からなる集合を $A[N]$ で表すとき，集合 $A[N]$ がもつ情報量

$$\Lambda(A[N]) = \log_2 N$$

を Hartley 情報量という．

・性質

(H1) $\Lambda(A[NM]) = \Lambda(A[N]) + \Lambda(A[M])$

(H2) $\Lambda(A[N]) \leq \Lambda(A[N+1])$

(H3) $\Lambda(A[2]) = 1$　　　　（単位集合は bit (binary unit)）

2．Shannon 情報量

定義 1　離散確率分布 $p = (p_1, \cdots, p_n)$ に対して,

$$H(p) = -\sum_{k=1}^{n} p_k \log_2 p_k$$

を，離散確率分布 p の Shannon 情報量という．熱力学では，**エントロピー**と呼ばれる．

•解釈　事象 A_k 上で $-\log_2 p_k$ という値を確率 p_k で取る確率変数 X を考えれば，$H(p) = E(X)$ となる．

・性質

(S1) $H(p) \geq 0$ であり，最小値 $H(p) = 0$ は一点分布（ある i に対して $p_i = 1$）のときに達成される．

(S2) $H(p)$ の最大値は $\log_2 n$ であり，離散一様分布（$p_i = 1/n$）の場合に達成される．

(S3) 不等式 $-p_i \log_2 p_i - p_j \log_2 p_j \geq -(p_i + p_j)\log_2(p_i + p_j)$ が成り立つ．すなわち，事象の合併によりエントロピーは減少する．

（注）　エントロピーは，加算無限の離散分布に対しても定義される．

（例）　非負整数上の確率分布の平均 $\mu\ (>0)$ が与えられているとき，エントロピーを最大にする分布は，幾何分布である．

定義 2　密度関数 $f(x)$ に対するエントロピーは,

$$H(f) = -\int_{-\infty}^{\infty} f(x) \log f(x)\, dx$$

により定義される．

(例)　密度関数の平均，分散が与えられたとき，エントロピーを最大にする分布は，正規分布である．

３．Kullback 情報量

定義 1　離散確率分布 $p = (p_1, \cdots, p_n)$ と $q = (q_1, \cdots, q_n)$ に対して，分布 p を q で置き換えたときの増加情報量

$$I(q||p) = \sum_{k=1}^{n} q_k \log_2 \frac{q_k}{p_k}$$

を離散分布に対する Kullback 情報量という．

- **統計学的解釈**　真の確率分布 q と，モデル p との分布の距離を測る．

・性質

1. $I(q||p) \geq 0$
2. $I(q||p) = 0 \quad \Leftrightarrow \quad p = q$

(証明)　値 p_k/q_k を確率 q_k で取る確率変数を Z とすると，

$$E(Z) = \sum_{k=1}^{n} \frac{p_k}{q_k} q_k = 1$$

また，

$$h(z) = -\log_2 z$$

を定義すると，$E(h(Z)) = I(q||p)$ である．ここで，$h(z)$ は，$h''(z) = 1/(z^2 \log 2) > 0$ であるから，狭義の凸関数であり，

$$h(z) \geq -z + 1$$

が成立する．これより，$E(h(Z)) \geq 0$ で，等号は $p = q$ のときに限る．

(問 1)　二項確率モデル $p = (0.7, 0.3)$ と真の分布 $q = (0.6, 0.4)$ があるとき，Kullback 情報量を求めよ．また，$r = (0.5, 0.5)$ と p では，どちらが q に近いか．

定義 2　真の密度関数を $g(x)$，モデルに基づく密度関数を $f(x)$ とするとき，

$$I(g||f) = \int_{-\infty}^{\infty} g(x) \log \frac{g(x)}{f(x)} \, dx$$

をモデルに関する真の分布の Kullback 情報量という．常に非負であり，0 になるのは $f(x) = g(x)$ のときに限る．

(問 2)　真の分布が N(0, 1) のとき，2 つの正規分布 N(0.5, 1) と N(0, 1.5) とでは，どちらが N(0, 1) に近いであろうか．

４．Fisher 情報量

定義　$L(\theta)$ をパラメータ θ をもつ対数尤度関数とするとき，

$$I(\theta) = E[(L'(\theta))^2]$$

を Fisher 情報量という．

正則性の条件のもとで，

$$I(\theta) = -E[L''(\theta)]$$

が成り立つ．

- **Kullback 情報量との関係**

$$\lim_{h\to 0} \frac{1}{h^2} I(\theta||\theta+h) = \frac{1}{2} I(\theta)$$

θ の近傍での Kullback 情報量 $I(\theta||\theta+h)$ が Fisher 情報量 $I(\theta)$ に比例していることから，Fisher 情報量は分布族の局所情報量を表している．

- **Fisher 情報量の有用性**

推定において，推定量のよさを測る量として，重要な役割をはたす．