〈情報学共同研究〉

新しい情報の測度とパターン情報処理*

鈴　木　昇　一

New Information Measures and Pattern-Information Processing*

Shoichi Suzuki

Abstract

The Shannon's amount of information transmitted by a discrete channel represents the amount of all the available uncertainty (,i. e. which is missing in the received symbols or which is removed from the transmitted symbols) about the joint probability of occurrence of a pair 〈transmitted symbols, received symbols〉. This information measure based on a probabilistic interpretation, that is to say, the amount of information contained about the transmitted symbols contained in the received symbols does not involve any attempt to measure similarities between patterns to be recognized. The two mutual information-measures between patterns presented here are different from the various types of information measure heretofore explained in this paper in that the idea of probabilistic interpretation concerning the information transmitted will not stand up to an examination. The measure may be used to compare relative goodnesses of measurements produced by different recognition-models.

要　　約

Shannon 相互情報量は二つの確率変数間の結合確率を用いて定義されており，この結合確率は統計的な仮定を導入しないと，実際の情報処理の場面においては計算しにくい。そのため，Shannon 相互情報量は応用が限られ，変革が求められていた。本論文では，従来の各種情報量を様々な観点から考察している。その結果，結合確率を導入しないで，パターン間に，一種の相互情報量が定義されている。提案されたこの新しいパターン間相互情報量は従来の相互情報量では表現しにくい内容を，パターン間類似度を用い計量化したものであり，パターン認識の働き同志を比較検討する上で有用な量になることが期待される。

*本研究の一部は株式会社インターナショナルソフトウエア（〒112東京都文京区）によって支持された。

1．まえがき

情報処理とは情報を簡約・整理して不要なものを捨て去ることであり，要素の多い集合を要素の少ない集合に対応させることである[6]。処理対象のあいまいさ[3] (fuzziness)，不確かさ (uncertainty)，平均情報量 (average amount of information)，エントロピー[1] (entropy) を減少させる操作が情報処理という訳である[7]。

事実，確率分布

$\{p(x) \mid x \in X\}$

を持つ確率変数 X (random variable) のエントロピー

$H(X) = -\sum_{x \in X} p(x)\, \ell og\, p(x)$

については，情報処理の働き f によって

$Y = f(X)$

と変換されたとしよう。ここに ℓog は自然対数 (natural logarithm) の意である。写像

$f : X \to Y$

が

任意の $x \in X$ に対し，唯一つの $y \in Y$ を対応させる一価関数 (a completely specified single-valued function)

とすると，

$H(f(X)) \leqq H(X)$

が成立し[1]，確かに情報処理の操作 f によって不確かさが減少し，注目している属性がどの値をとるかを予測するときのむずかしさ (unpredictability) は減少しているのである。

要素 $x \in X$ と，その x がファジイ集合 (fuzzy set) A に帰属する程度 (degree) $\mu_A(x)$ が与えられたときの，ファジイ集合[3]

$A = \{(x, \mu_A(x) \mid x \in X\}$

ここに　$O \leqq \mu_A(x) \leqq 1$

についても，一種のエントロピー

$$-\sum_{x \in X} \frac{\mu_A(x)}{\sum_{x \in X} \mu_A(x)}\, \ell og \frac{\mu_A(x)}{\sum_{x \in X} \mu_A(x)} - \sum_{x \in X} \frac{1-\mu_A(x)}{\sum_{x \in X}[1-\mu_A(x)]}\, \ell og \frac{1-\mu_A(x)}{\sum_{x \in X}[1-\mu_A(x)]}$$

を減少させる働きが fuzzy 情報処理であることは，もう既に半順序原理 (a principle of partial ordering) に基づき，説明されている[7]。

計算機プログラムとは

計算機プログラム＝〔データ構造＋アルゴリズム〕(データ構造を操作するアルゴリズム)

と解釈されるようになって久しい。対応して，
情報処理とは

情報処理＝〔情報構造＋知能〕（情報構造を操作する知能）

と解釈して，本論文では，この種の情報処理の働きに関連する

新しい情報の測度，尺度（an information measure）

を提案する。

2．情　報　量

ある系[2]の秩序はこの系を表現するのに必要な情報量に等しい

が Shannon 流情報理論の採用している基本的な立場である。このように，情報量は表現の多様性と秩序（圧縮：compression [8]）の双方に関係している。

この二つの観点から情報量を説明してみよう。

〔表現の多様性〕

ある性質のもつあいまいさ（表現の多様性）が，p_j を第 $j(=1\sim N)$ 番目の表現の生起確率として

$$H=\sum_{j=1} p_j \ell og_2 p_j$$

ここに，$0\leqq p_j\leqq 1, \sum_{j=1}^{N} p_j=1$

の形の平均情報量で測られるとすれば，表現の総数を 2^N とし，各表現が等確率で生じるという条件の下で，

$$\max H=-\sum_{j=1}^{2N} 2^{-N} \ell og_2 2^{-N}=N$$

を得，この性質は高々 $N[bit]$ で，つまり，N 桁の 2 進数（符号）で表現できる。

〔表現の圧縮（秩序；order)〕

入力の情報量が $k[bit]$ であるとは：

各入力ベクトル x_i（第 i 番の入力ベクトル）は，あらかじめ求められている典型的なベクトル（代表ベクトル）の集合

$\{y_1, y_2 \cdots, y_N\}$（コードブック；codebook）

の一つである $y_{\hat{n}}$（コードベクトル）に置換されるとしよう。ここに，$\hat{n}$ は，

$$\hat{n}=arg \min d(x_i, y_n)$$

（n を変えて得る x_i と y_n との間の距離 $d(x_i, y_n)$ の最小値を与える最小の n)。

この置換の働きにより，もしコードブックサイズ N が

$$2^{k-1}<N\leqq 2^k$$

であれば，各ベクトル x_i の秩序（構造）は $k〔bit〕$ に圧縮される。

□

生起確率 p_j をもつ第 $j(=1\sim N)$ 番目の確率事象（random event）e_j の集まり

$$E=\{e_j \mid j=1\sim N\}$$

を考える。ここに，

$0 \leqq p_j, \sum_{j=1}^{N} p_j = 1$。

a を1より大なる整数とする。生起確率 p_j がある非負整数 n_j を用い，

$p_j = a^{-nj}$

と表現される場合を考えよう。事象 e_j を
a 者択一（互いに同等に確からしい a 個の内から一つを選択する操作）
を繰り返して確定させるためには，
a 者択一の操作が $1/a$ の確率で実施される
と想定すると，

$p_j = a^{-1}.\ a^{-1}.\cdots a^{-1}$（$n_j$ 回の積）

であるから，n_j 回の a 者択一操作が必要であることが知られる。
よって，起り得る可能性のある事象が多数回生起して，この内の一つの事象の生起を知る場合，

$M = \sum_{j=1}^{N} p_j \cdot n_j$

は，一つの任意の事象を選択し確定させるために必要な“a 者択一の操作の平均回数”を表現している。この平均回数 M を
事象 e_j の確率 p_j が a^{-nj} である場合の事象系 E の，一事象当りのエントロピー，あるいは平均情報量
という。
ところで，

$p_j = a^{-nj} \quad \therefore \quad -\ell og_a p_j = n_j$

であるから，a 者択一操作の平均回数 M は

$M = -\sum_{j=1}^{N} p_j \ell og\, p_j$

と再表現される。
一般に，各事象 e_j の生起確率 p_j が a^{-nj} と表現され得ない場合でも，a 者択一操作の平均回数 M に関する上述の解釈を拡張して，

$H(E) \equiv -\sum_{j=1}^{N} p_j \ell og_a p_j$〔$a$ 元単位〕

という量を考え，これを事象系 E の，一事象あたりのエントロピーという。
次の補助定理２．１を適用すれば，

$$H(E) \equiv -\sum_{j=1}^{N} p_j \ell og_a p_j$$
$$\leqq -\sum_{j=1}^{N} p_j \ell og_a a^{-nj} = \sum_{j=1}^{N} p_j \cdot n_j = M$$

ここに，$\sum_{j=1}^{N} a^{-nj} = 1$，　$M \equiv \sum_{j=1}^{N} p_j \cdot n_j$
が成立し，

E のエントロピー $\mathrm{H}(E)$ は平均記述長（平均符号長）M の最小値である

との結論が得られる。これが **Shannon の第 1 符号化定理**[10]（Shannon's first coding theorem）である。

$0\,\ell og\,0 = \lim_{x\to+0} x \lim x = 0.$

に注意して，

〔補助定理２．１〕

$x_j > 0 \wedge \sum_{j\in J^+} y_j \leqq \sum_{j\in J^+} x_j$

を満たす添字 $j \in J$ の集合 J^+

を導入すると，二つの数列 $\{x_j\}_{j\in J}$, $\{y_j\}_{j\in J}$ に関し，不等式

$$-\sum_{j\in J^+} x_j\,\ell og\,x_j \leqq \sum_{j\in J+} x_j\,\ell og\,y_j$$

が成り立つ。ここで，等号は

$$[\forall\, j \in J^+, y_j/x_j = 1] \wedge \sum_{j\in J^+} y_j = \sum_{j\in J^+} x_j$$

が成立するとき，かつそのときのみに限る。

（証明）$f(x) = x - 1 - \ell og\,x$

とおくと，

$$\lim_{x\to+0} f(x) = +\infty, f(1) = 0$$

$$\lim_{x\to+\infty} f(x) = +\infty \quad \because \lim_{x\to+0} \frac{1}{x}\cdot \ell og\,x = 0$$

であり，

$$(d/dx)f(x) = 1 - x^{-1}$$
$$< 0 \text{ if } x < 1, \quad = 0 \text{ if } x = 1, > 0 \text{ if } x > 1$$

が成り立つ。よって，不等式

$$\ell og\ \ x \leqq x - 1$$

が成立し，$x = 1$ のときに限り，等号は成立する。この不等式を適用すれば，

$$\sum_{j\in J^+} x_j\,\ell og\,\frac{y_i}{x_j} \leqq \sum_{j\in J^+} x_j \cdot \left[\frac{y_i}{x_j} - 1\right]$$
$$= \sum_{j\in J^+} y_j - \sum_{j\in J^+} x_j \leqq 0$$
$$\therefore\ -\sum_{j\in J^+} x_j\,\ell og\,x_j \leqq -\sum_{j\in J^+} x_j\,\ell og\,y_j$$

を得て，等号は，

$$[\forall j \in J^+, y_j/x_j = 1] \wedge \sum_{j\in J^+} y_j = \sum_{j\in J+} x_j$$

のときに限り成立することがわかる。 □

〔備考2．1〕　対数の底の変換公式を使えば，

$\ell og_a x_j = \ell og\, x_j / \ell og\, a \ (a > 1)$

などが成立するから，上述の補定2．1は実は，任意の対数の底 a につき成立する。　□

a 者択一の平均所要回厳 M が大であることは，確率事象系 E が記述しているモデルが複雑であることを意味する。ある任意の一つの事象を確定させる操作の困難さの程度（難度，degree of difficulty）を表していることになる。

我々は a 者択一操作を平均的に最小 $H(E)$ 回（計算量，computational complexity）反復することにより，実際に生起された事象が確定し，その生起した事象のモデル構造の推定に伴う不確かさが消滅すると考える。その消滅する不確かさが a 者択一操作を繰り返した主体の得た知識の程度を反映したものであり，従って主体の得た“情報の量”（an amount of information ; the contents of an information）と考えていることになる。

次の命題2．2のiiはもともと，不確かさの存在しない事象系 E を処理しても得られるものはないことを意味し，命題2．2のiiiは等確率分布を備えた事象系 E は最大の複雑さをもったモデルに対応していることを指摘している。

〔命題2．2〕

(i)　$0 \leqq H(E) \leqq \ell og_a N$

(ii)　$H(E) = 0$

$\Leftrightarrow \exists j, [p_j = 1] \wedge [\forall\ k \in \{1, 2, \cdots, N\} - \{j\}, p_k = 0]$

(iii)　$H(E) = \ell og_a N$

$\Leftrightarrow \forall\ j (= 1 \sim N), p_j = 1/N.$　□

3．一般化された情報量

与えられたデータをモデル自身の記述も含めて最も短く語頭符号化できる様な確率モデル（数学的制約の入ったデータの確率分布のこと，例えば回帰モデル（regression model），マルコフモデル（Markov model）など）が最良のモデルである，と主張する

MDL（Minimum Description Length）原理[10]

は，データ X の確率を $p(x)$ とすると，

自己情報量（amount of self information）$I(X) = -\ell og\, p(X)$ は X を語頭符号化するための最適な記述長であり，一データ当りの平均自己情報量（average amount of self information）即ちエントロピー

$$H(X) = \Sigma_X p(X) \cdot I(X)$$

は平均符号長の下限である

という **Shannon の第1符号化定理**（Shannon's noiseless coding theorem）に支えられている。なお，一つの符号が他の任意の符号のある先頭部分に一致することはないという条件を語頭条件（prefix condition）といい，一意的に復号化できるための条件である。語頭条件を満たす符号（の組）を語頭符号（prefix　code）といっている。

条件

$x_i \geqq 0\ (i = 1 \sim N)$，$\Sigma_{i=1}^{N} x_i = 1$

$c_i \geqq 0\ (i = 1 \sim N)$

を満たす量

$$F\ (x_1, x_2, \cdots, x_N) \equiv -\textstyle\sum_{j=1}^{N} [c_i / \sum_{k=1}^{N} c_k] \cdot \ell og_a\, x_i \quad (a > 1)$$

を一般化された情報量（generalized entropy）と称することがある。一般化された情報量 $F\,(x_1, \cdots, x_N)$ が最小となるのは，補助定理２．１から分るように，

$$x_i = c_i / \textstyle\sum_{k=1}^{N} c_k,\ i = 1 \sim N$$

のときであり，かつこのときのみである。

もし，クラス t の成員が $n[t]$ 個あるとすると，ある成員がクラス t に属するということを指摘するために必要な記述はその全記述において $n[t]$ 回起る。それで，クラス t の確率を $p[t]$ とすると，Shannon の意味では[12]，クラス t に属する成員の記述をなしとげるのに必要とされる情報量（２進符号の長さ）は

$$-n[t] \cdot \ell og\, p[t]$$

であり，よって，すべての成員の記述をなしとげるのに必要とされる情報量は

$$I = -\textstyle\sum_t n[t] \cdot \ell og\, p[t]$$

ということになる。この量 I の最小値は，
$p[t]$ が

$$p[t] = n[t] / \textstyle\sum_t n[t]$$

であるときに生じる[11]。これはまさに，すべてのクラス t についても一成員当りの，一般化された情報量の概念

$$\begin{aligned} &- \textstyle\sum_t \{n[t] / \sum_s n[t]\} \cdot \ell og\, p[t] \\ &= [\textstyle\sum_t n[t]]^{-1} \cdot I \end{aligned}$$

に通じるものである。

４．これ迄の情報量と，S. Suzuki の情報量研究

Maximum entropy has been used to estimate the power spectrum corresponding to a measured autocorrelation function.The princicle of maximum entropy therefore seems to be useful for attacking a wide class of estimation problems[13].

上記の最大エントロピー原理に関連した話題を簡単に紹介しておこう。

S. Suzuki の提唱した測度的不変量検出の理論[14],[15]では，パターン φ は内積，ノルムを各々

$$(\phi, \eta), \|\phi\| = \sqrt{(\phi, \phi)}$$

とする可分な一般抽象 Hilbert 空間 $\mathfrak{H}$ の元であるとして，正値自己共役作用素 G と φ との定める測度的不変量（metrical invariants）

$$
\begin{aligned}
&(G\varphi, \varphi)/(\varphi, \varphi)\\
&= |(G^{\frac{1}{2}}\varphi\|\varphi\|^{-1}, G^{\frac{1}{2}}\varphi\|\varphi\|^{-1})|\\
&= \|G^{\frac{1}{2}}\varphi\|^2/\|\varphi\|^2 \geqq 0
\end{aligned}
$$

がパターン φ から抽出される基本的な特徴量（primitive features）であるとされている。これは，量子力学的観測理論では，力学的状態 φ における“物理量”G の（スペクトル）の期待値（expectation）

$$
\begin{aligned}
&\langle G\rangle_{\varphi} \equiv (G\varphi, \varphi)/(\varphi, \varphi)\\
&= \|G^{\frac{1}{2}}\varphi\|\varphi\|^{-1}\|^2
\end{aligned}
$$

であり，three-dimensional model-based vision system が二つのパターン φ, η から抽出する projectively invariant measurement[16]

$$
\begin{aligned}
&|(G\varphi, \eta)|^2/[(G\varphi, \varphi)\cdot(G\eta, \eta)]\\
&= |(G^{\frac{1}{2}}\varphi, G^{\frac{1}{2}}\eta)|^2/[\|G^{\frac{1}{2}}\varphi\|^2\cdot\|G^{\frac{1}{2}}\eta\|^2]\\
&= |(G^{\frac{1}{2}}\varphi\|G^{\frac{1}{2}}\varphi\|^{-1}, G^{\frac{1}{2}}\eta\|G^{\frac{1}{2}}\eta\|^{-1})|^2
\end{aligned}
$$

と比較してみることは興味深い。

S. Suzuki は，パターン φ から抽出され，自己共役作用素 H の関数 $f(H)$（正値自己共役作用素）に関連した測度的不変量

$$\mathfrak{F}(\varphi) \equiv (f(H)\varphi, \varphi)/(\varphi, \varphi)$$

を，射影作用素 $\theta_{\ell}(H)$ の和に直交分解した形式

I（恒等作用素）$= \theta(H) = \sum_{\ell\in L}\theta_{\ell}(H)$

を導入し，

$$\mathfrak{F}(\varphi) = \sum_{\ell\in L}\mathfrak{F}_{\ell}(\varphi)$$

ここに，$\mathfrak{F}_{\ell}(\varphi) \equiv (f(H)\cdot\theta_{\ell}(H)\varphi, \varphi)/(\varphi, \varphi)$

という具合に，$\mathfrak{F}_{\ell}(\varphi), \ell \in L$ の和に直交直和分解して，パターン φ のエントロピー

$$
\begin{aligned}
&-\sum_{\ell\in L} q_{\ell}\,\ell og\, q_{\ell}\\
&= \ell og_2\mathfrak{F}(\varphi) - [\sum_{\ell\in L}\mathfrak{F}_{\ell}(\varphi)\,\ell og_2\mathfrak{F}_{\ell}(\varphi)]/\mathfrak{F}(\varphi)
\end{aligned}
$$

ここに，$q_{\ell} = \mathfrak{F}_{\ell}(\varphi)/\mathfrak{F}(\varphi)$

を提案し[14],[15]，このエントロピーが手書き漢字 φ の形状的複雑さに対応する量であることを計算機シミュレーションで確かめた[17]。

また，鈴木などは，純粋連続スペクトル型自己共役作用素 H のスペクトル表現

$$H = \int_{-\infty}^{+\infty}\lambda\, dE(\lambda)$$

の規格化スペクトル密度（power spectrum density）

$$p_\varphi(\lambda) = (d/d\lambda)(E(\lambda)\varphi, \varphi)/\int_{-\infty}^{+\infty} d(E(\mu)\varphi, \varphi)$$

に注目して，測度的不変量

$$\langle f(H)\rangle_\varphi = (f(H)\varphi, \varphi)/(\varphi, \varphi)$$

$$= \int_{-\infty}^{+\infty} f(\lambda)\, d(E(\lambda)\varphi, \varphi)/\int_{-\infty}^{+\infty} d(E(\mu)\varphi, \varphi)$$

の特別なものとして，パターン φ の微分スペクトルエントロピー

$$\langle -\ell og_e p_\varphi(H)\rangle_\varphi$$
$$= -\int_{-\infty}^{+\infty} d\lambda\, p_\varphi(\lambda)\, \ell og_e p_\varphi(\lambda)$$

を，測度的不変量検出理論[(14),(15)]，認識の量子論[(5)]，情報の量子論[(5)]の立場から提案し，スペクトル分散

$$\langle (H-m)^2\rangle_\varphi = \sigma^2$$
ここに，$m = \langle H\rangle_\varphi$

が一定の条件の下で，

$$\langle -\ell og_e P_\varphi(H)\rangle_\varphi$$

を最大にするスペクトル密度 $p_\varphi(\lambda)$ は，変分法（the variational principle）を適用し，

$$P_\varphi(\lambda) = [2\pi\sigma^2]^{-1/2}\cdot exp[-(\lambda-m)^2/(2\sigma^2)]$$

であり，このときの最大微分エントロピーが

$$\mathrm{MAX}\langle -\ell og_e p_\varphi(H)\rangle_\varphi = \ell og_e\sqrt{2\pi e\sigma^2}$$

であることを示した。そして，

$\mathfrak{H} = L_2(R^2)$ (an infinite-dimensional separable Hilbert space)

$(\varphi, \eta) = \int_{-\infty}^{+\infty} dx_1 \int_{-\infty}^{+\infty} dx_2\, \varphi(x_1, x_2)\cdot\bar{\eta}(x_1, x_2)$, $\bar{\eta}$ はその複素共役

$H = i^{-1}\partial/\partial x_1,\ i = \sqrt{-1}$

と選んだ場合，

$$(e^{-itH}\varphi, \varphi)$$

$= \int_{-\infty}^{+\infty} dx_1 \int_{-\infty}^{+\infty} dx_2\, \varphi(x_1 - t, x_2)\cdot\bar{\varphi}(x_1, x_2)$，ここに $\bar{\varphi}$ は φ の複素共役

が自己相関関数（autocorrelation function）になるが，最大微分エントロピー

$$\mathrm{MAX}\langle -\ell og_e p_\varphi(H)\rangle_\varphi$$

をもたらすパターン $\varphi=\varphi(x_1,x_2)$ はそのノルム $\|\varphi\|=\sqrt{(\varphi,\varphi)}$ が1の下では

$$\int_{-\infty}^{+\infty} dx_2 |\eta(x_2)|^2 = 1$$

を満たす任意の η を導入すると，ガウス型関数

$$\varphi(x_1,x_2)=\sqrt[4]{2\sigma^2/\pi}\cdot e^{+imx_1}\cdot e^{-\sigma^2 x_1^2}\cdot\eta(x_2)$$

で与えられることを示した[18]。

更に，S. Suzuki は，

the Kullback information distance from the similarity measures SM $(\varphi,\omega_j), j\in J$ to the probalities of occurrence of each category

の形での認識情報量（amount of recognition information）

$$\mathrm{REIN}\ \{\mathfrak{C}/\varphi\}$$
$$=\sum_{j\in J}\mathrm{SM}(\varphi,\omega_j)\,\ell og\,[\mathrm{SM}(\varphi,\omega_j)/p(\mathfrak{C}_j)]$$

ここに，$\mathrm{SM}(\varphi,\omega_j)$ はパターン φ と，生起確率として $p(\mathfrak{C}_j)$ をもつ第 $j\in J$ 番目のカテゴリ $\mathfrak{C}_j$ の代表パターン ω_j との間の類似度

を提案し[19]，計算機シミュレーションでこの量を計算した[20]。REIN $\{\mathfrak{C}/\varphi\}$ は，

$\exists\, j\in J, \mathrm{SM}(\varphi,\omega_j)=1$（パターン φ は代表パターン ω_j と確定的な類似関係にある）

という具合に，

パターン φ が第 $j\in J$ 番目のカテゴリ $\mathfrak{C}_j$ に帰属する

と認識推断されたとき，

$$\mathrm{REIN}\ \{\mathfrak{C}/\varphi\}=-\ell og\, p(\mathfrak{C}_j)$$

という，$\mathfrak{C}_j$ の Shannon 形自己情報量が得られる性質を持っている。 □

元来，Shannon 形エントロピーは次のように定義される[21]。

有限個の可測集合の系

$$\xi=\{A_i \mid 1\leqq i\leqq n\}$$

で，条件

$A_i\cap A_j=\phi$（空集合）$(i\neq j)$

$\cup_{i=1}^{n} A_i=\Omega$

を満たすものを，Ω の有限分割という。例えば，n 次元ユークリッド空間 R^n の部分集合 A の体積（ルベーグ測度）を $\mu(A)$ と書く。一般には，可算加法的測度を μ とする。$\mathfrak{B}$ は Ω の適当な部分集合族のなす $\sigma-$ 代数，つまり，

$\Omega\in\mathfrak{B}$

$A_j\in\mathfrak{B}\ (1\leqq i<\infty)$ ならば $\Omega\setminus A_i$ (A_j の補集合) $\in\mathfrak{B}$, $\cup_{i=1}^{n}A_i\in\mathfrak{B}$

であるようなものであるとする。このとき，測度空間 $(\Omega, \mathfrak{B}, \mu)$ を考えると，

$A_i \in \mathfrak{B}$
$i \neq j$ に対して　$A_i \cap A_j = \phi$ ならば $\mu(\cup_{i=1}^{n} A_j) = \sum_{i=1}^{n} \mu(A_i)$

が成り立つ。ここに，

$$\mu(\Omega) = 1$$

と規格化して，ξ のエントロピー $H(\xi)$ は

$$H(\xi) = -\sum_{k=1}^{n} \mu(A_k)\, \ell og\ \ \mu(A_k)$$

で定義される。

ちなみに，画像関数

$$f(i) \quad (0 \leqq f(i) \leqq 1)$$

のエントロピーとして，

$$H(f) = -\sum f(i)\, \ell og\, f(i)$$

を定義し，相関エントロピー（Correlation Entropy）

$$CH(f) = H(f * f)$$

ここに，

$$(f * f)(t) = \sum_i f(i) \cdot f(i + t)$$

を定義している研究[22]もある。

また，時点 τ_k に着眼し，局所エントロピー・フロー（local entropy flow）と呼ばれる 1 次元信号 $f(t)$ のエントロピー

$$H(\tau_k) = -\sum_{j=1}^{+N} p(x_j\,;\tau_k) \cdot \ell og\, p(x_j\,;\tau_k),\ k = 0, \pm 1, \pm 2, \cdots$$

を定義している研究[23]もある。ここに，時点 τ_k に着眼したときに，信号レベルが x_j となる確率 $p(x_j\,;\tau_k)$ を

$$p(x_j\,;\tau_k) = [\sum_{(1)} w(t_i\,;\tau_k)]^{-1} \cdot \sum_{(2)} w(t_i\,;\tau_k)$$

と表している。$\sum_{(1)}$ は

$|t_i - \tau_k| \leqq B/2$ を満たす i についての総和

であり，$\sum_{(2)}$ は

$f(t_i) = x_j$ かつ $|t_i - \tau_k| \leqq B/2$ を満たす i についての総和

である。

上記に登場した正定数 B，関数 w につき説明しておこう。

信号 $f(t)$ が $\Delta = (2\mathrm{W})^{-1}$[sec] 間隔でサンプリングし量子化して得られる1次元離散信号を

$$x_j = f(t_i), i = 0, \pm 1, \pm 2, \cdots; j = 0, \pm 1, \pm 2, \cdots, \pm N$$

とし，着眼点を

$$\tau_k, k = 0, \pm 1, \pm 2, \cdots$$

としており，また，正定数 B を

α : 局所性を定める1パラメータ（ある有限の正整数）

として，

$$B = 2\alpha \cdot \Delta = \alpha / \mathrm{W}$$

とおく。重み関数（weighting function）

$$W(t_i ; \tau_k)$$

は次のように三角形に選ぶ：

底辺の中心を原点にもつ，符号を含めた高さが $a(>0, <0)$ であり，底辺の長さが $c(>0)$ である三角形を表す関数

$$y = g(x)$$

は

$$g(x) = \begin{cases} a\left[1 - \dfrac{|x|}{2^{-1} \cdot c}\right] & if \ \ |x| \leqq 2^{-1} \cdot c \\ 0 \quad if \ \ |x| > 2^{-1} \cdot c \end{cases}$$

と表現されるから，

$$W(t_i ; \tau_k) = \begin{cases} 1 - \dfrac{|t_i - \tau_k|}{2^{-1} \cdot B} & if \ \ |t_i - \tau_k| \leqq 2^{-1} \cdot B \\ 0 \quad if \ \ |t_i - \tau_k| > 2^{-1} \cdot B \end{cases}$$

□

情報量の概念は上記のごとく信号に対してだけではなく，言語に対しても有益なものとなる。W. Kuich によって導入された言語 L のエントロピー H を紹介しておこう。

Shannon[12] defines the entropy or the channel capacity to be the quantity

$$\lim_{T \to \infty} \frac{1}{T} \cdot \ell og\, N(T)$$

where $N(T)$ is the number of allowed signals of duration T transmitted by a discrete channel.

上記の Shannon の定義に基づき，$u(n)$ を

$u(n)$: 言語 L に含まれる長さ n の互いに異なる語の数

と定義し，言語 L のエントロピー H を

$$H = \overline{\lim_{n\to\infty}} \frac{1}{n} . \ell og\, u(n)$$

と定義すると[24]，言語 L を特徴づける指標が得られるという。ここで，数列 $\{a_n\}_{n=1,2,\cdots}$ の上極限（the limit superior）$\overline{\lim_{n-\infty}} a_n$ は

$$\overline{\lim}_{n\to\infty} a_n = inf_n \sup_k \{a_k \mid k \geqq n\}$$

と定義されている。

5．しきい値の決定（最大エントロピー原理の適用(1)）

パターン φ から抽出された第 $\ell \in L$ 番目の特徴量が

$$v(\varphi, \ell)$$

であるとき，これから，

$$u(\varphi, \ell) \in \{0, 1\}$$

なる 2 値特徴量（binarized feature）を

$$u(\varphi, \ell) = \begin{cases} 1 & \text{if} \quad v(\varphi, \ell) \geqq b_\ell \\ 0 & \text{if} \quad v(\varphi, \ell) < b_\ell \end{cases}$$

の形で決定する際のしきい値（threshold value）b_ℓ の組

$$b_\ell, \ell \in L$$

を，最大エントロピー原理で決定してみよう。

パターンが M 個あるものとし，

$$\varphi_1, \varphi_2, \cdots, \varphi_M$$

とする。

$p(\ell, y) = (1/M) \cdot$〔$u(\varphi_k, \ell) = y$ となる k の個数〕

をすべての $\ell \in L$ につき，量子化（quantization）された y について求めておく。

二つの出現度数分布（frequency-distribution of occurrence）

$$g(0, d_\ell, y) = p(\ell, y) / \textstyle\sum_{y<d_\ell} p(\ell, y)$$
$$g(1, d_\ell, y) = p(\ell, y) / \textstyle\sum_{y \geqq d_\ell} p(\ell, y)$$

を導入し，二つのエントロピーの和

$$E(d_\ell) = -\textstyle\sum_{y<d_\ell} g(0, d_\ell, y)\, \ell og_2\, g(0, d_\ell, y)$$
$$-\textstyle\sum_{y \geqq d_\ell} g(1, d_\ell, y)\, \ell og_2\, g(1, d_\ell, y)$$

を求め，

$E(d_\ell)$ が最大になる d_ℓ，つまり

$$\arg \max_{d_\ell} E(d_\ell) = b_\ell$$

なる b_ℓ が求めるしきい値である。これは,

$$y < d_\ell,\quad y \geqq d_\ell$$

なる双方の y 領域において，各 $u(\varphi_k, \ell)$ の変化が共に大きいような d_ℓ をしきい値 b_ℓ として採用していることになる。

もともとの Kapur et al. の方法[25]はパターン φ の振幅を 2 値化（binarization）する際のしきい値 t_q に適用されたものである。

それは

$$\varphi(x_q) = \begin{cases} 1 \text{ if } \varphi(x_q) \geqq t_q \\ 0 \text{ if } \varphi(x_q) < t_q \end{cases}$$

でのしきい値 t_q を次の様にして求めるものである。

$$p(q, y) = [\varphi_k(x_q) = y \text{ となる } k \text{ の個数}] / M$$

から定義される 2 種類の量

$$f(0, d_q, y) = p(q, y) / \textstyle\sum_{y<d_q} p(q, y)$$
$$f(1, d_q, y) = p(q, y) / \textstyle\sum_{y \geqq d_q} p(q, y)$$

を導入する。

$$F(d_q) = -\textstyle\sum_{y<d_q} f(0, d_q, y)\, \ell og_2 f(0, d_q, y)$$
$$\quad - \textstyle\sum_{y \geqq d_q} f(1, d_q, y)\, \ell og_2 f(1, d_q, y)$$

を求め,

$$\arg_{d_q} \max F(d_q) = t_g$$

が求めるものである。

6．特徴量の復元（最大エントロピー原理の適用(2)）

Frieden は

$$I_m = \textstyle\sum_{j=1}^{J} S(y_m, x_j) \cdot \hat{O}_j + \hat{N}_m - B,\ m = 1 \sim M$$

の形のシステム方程式に

the principle of maximum entropy

を適用し,

$$\hat{O}_j, j = 1 \sim J$$

を推定している。すなわち，additive noise

$$N_m, m = 1 \sim M$$

によって画像が低下させられた（degraded）画像データ集合

$$I_m, m = 1 \sim M$$

から，point-spread function

$$S(y_m, x_j), m = 1 \sim M, j = 1 \sim J$$

を用いて，元の an incoherent object scene

$$\hat{O}_j \geqq 0, j = 1 \sim J$$

を復元する機構（restoring scheme）について，

$$\begin{aligned} K \equiv & - \textstyle\sum_{j=1}^{J} \hat{O}_j \, \ell og \, \hat{O}_j - \rho \cdot \sum_{m=1}^{M} \hat{N}_m \, \ell og \, \hat{N}_m \\ & - \textstyle\sum_{m=1}^{M} \lambda_m \cdot [\sum_{j=1}^{J} S(y_m, x_j) \cdot \hat{O}_j + \hat{N}_m - B \\ & - I_m] - \mu(\textstyle\sum_{j=1}^{J} \hat{O}_j - p_0) \rightarrow MAX \end{aligned}$$

とすることを考えている。解として，

$$\frac{\partial K}{\partial \hat{O}_j} = O \quad \therefore \hat{O}_j = \exp[-1 - \mu - M_{M=1}^{M} \lambda_m \cdot S(y_m, x_j)]$$

$$\hat{N}_m = \exp[-1 - \lambda_m / \rho]$$

を得ている(13)。

上記の研究から，hint を得て，

scheme for restoring the extracted features from knowledge of its degraded features and the point-impulse response characteristic, based on the principle of maximum entropy

が以下のように，提案される。

パターン φ から抽出される第 $\ell \in L$ 番目の特徴量が $u(\varphi, \ell)$ と観測されたとき，point-impulse response S によって

$$\begin{aligned} u(\varphi, m) &= \textstyle\sum_{\ell \in L} S(m, \ell) \cdot v(\varphi, \ell) \\ u(\varphi) &= \{u(\varphi, \ell) \mid \ell \in L\} \end{aligned}$$

と表現されているシステムを考えよう。元の特徴量の組

$$\begin{aligned} & v(\varphi) = \{v(\varphi, \ell) \mid \ell \in L\} \\ & \quad \forall \ \ell \in L, v(\varphi, \ell) \geqq 0 \end{aligned}$$

を，パターン φ がミクロカノニカル集団（第11章を参照）に属するパターンとみなし，条件

$$\textstyle\sum_{\ell \in L} v((\varphi, \ell) \equiv V(\varphi) = \text{constant}$$

が満たされているとし，the entropy criterion

$$\begin{aligned} & H(v(\varphi)) \\ & = - \textstyle\sum_{\ell \in L} v(\varphi, \ell) \, \ell og \, v(\varphi, \ell) \rightarrow MAX \end{aligned}$$

とする形で推定することを考えよう。

$u(\varphi)$ は真の特徴量の組 $v(\varphi)$ の a set of linearly distorted features である
と考えていることになる。この推定方法は

The most likely $v(\varphi)$ has a maximum entropy

と想定していることになる。

解 $v(\varphi)$ を

$u(\varphi)$

$\{S(m, \ell) \mid m, \ell \in L\}$

$V(\varphi)$

が与えられているものとして，ラグランジュの未定乗数法（the Lagrange multiplier method）を
適用し，求めよう。

$$
\begin{aligned}
K \equiv & - \textstyle\sum_{\ell \in L} v(\varphi, \ell) \log v(\varphi, \ell) \\
& + \textstyle\sum_{m \in L} \lambda_m \cdot [\sum_{\ell \in L} S(m, \ell) \cdot v(\varphi, \ell) - u(\varphi, m)] \\
& + \mu \cdot [\textstyle\sum_{\ell \in L} v(\varphi, \ell) - V(\varphi)]
\end{aligned}
$$

とおき，

$$\frac{\partial K}{\partial \lambda_m} = 0, \quad m \in L \tag{6. 1}$$

$$\frac{\partial K}{\partial \mu} = 0 \tag{6. 2}$$

$$\frac{\partial K}{\partial v(\varphi, \ell)} = 0, \quad \ell \in L \tag{6. 3}$$

を解けばよい。式（6．1）から

$$
\begin{aligned}
& \textstyle\sum_{\ell \in L} S(m, \ell) \cdot v(\varphi, \ell) \\
& = u(\varphi, m)
\end{aligned} \tag{6. 4}
$$

を得，式（6．2）から

$$\textstyle\sum_{\ell \in L} v(\varphi, \ell) = V(\varphi) \tag{6. 5}$$

が得られる。式（6．3）を計算すれば，

$$
\begin{aligned}
& v(\varphi, \ell) \\
& = \exp[\mu - 1] \cdot \exp[\textstyle\sum_{m \in L} \lambda_m \cdot S(m, \ell)]
\end{aligned} \tag{6. 6}
$$

となるが，式（6．5）から

$$
\begin{aligned}
& \exp[\mu - 1] \\
& = V(\varphi) / \textstyle\sum_{\ell \in L} \exp[\sum_{m \in L} \lambda_m \cdot S(m, \ell)]
\end{aligned}
$$

を得て，結局，

$v(\varphi, \ell)$
$= q_\ell \cdot V(\varphi)$, ここに

$$q_\ell = \frac{\exp[\sum_{m \in L} \lambda_m \cdot S(m, \ell)]}{\sum_{\ell \in L} \exp[\sum_{m \in L} \lambda_m \cdot S(m, \ell)]} \tag{6. 7}$$

と求められる。未定乗数 $\lambda_m, m \in L$ を決定するには，式（6．4）に式（6．7）を代入して得られる方程式

$$[\sum_{\ell \in L} S(m, \ell) \cdot q_\ell] \cdot V(\varphi)$$
$$= u(\varphi, m), m \in L$$

つまり，

$$[\sum_{\ell \in L} S(m, \ell) \cdot \exp[\sum_{m \in L} \lambda_m \cdot S(m, \ell)]$$
$$= u(\varphi, m) \cdot V(\varphi)^{-1} \cdot \sum_{\ell \in L} \exp[\sum_{m \in L} \lambda_m \cdot S(m, \ell)] \tag{6. 8}$$

を解けばよい。

7．制約付き推定問題(1)

確率モデルの推定問題へ，
定エネルギーの下での最大エントロピー原理（maximum entropy principle under constant energy）
を適用しよう。これは
constrained estimation problen（制約付き推定問題）
への，
maximum entropy estimation
の適用であり，下記の3条件 i ～iiiを制約条件（constraints）として，エントロピー H を最大にする確率分布

$$p(x_\alpha), \alpha = 1, 2, \cdots, n$$

を求める問題として定式化される：
maximize

$$H \equiv -\sum_\alpha p(x_\alpha)\, \ell og\, p(x_\alpha)$$

Subject to

(i) $\sum_\alpha p(x_\alpha) \cdot E(x_\alpha) = \langle E(x) \rangle =$ constant m
(ii) $\sum_\alpha p(x_\alpha) = 1$
(iii) $0 \leqq p(x_\alpha) \leqq 1$ for all α. □

上述の問題は変分原理（variational principle）におけるラグランジュの未定乗数（Lagrange multiplier）法によって解くことができ，
probability measure $p(x_\alpha)$

$$= [Z(T)]^{-1} \cdot \exp\left[-\frac{1}{T} \cdot E\left(x_\alpha\right)\right] \quad (7.\ 1)$$

partition function

$$Z(T)] = \Sigma_\alpha \exp\left[-\frac{1}{T} \cdot E(x_\alpha)\right] \quad (7.\ 2)$$

となる。温度（temperature）と呼ばれるパラメータ T を決定する方程式は

expectation $m = \langle E(x) \rangle$

$$= -\left[\partial / \partial\left(\frac{1}{T}\right)\right] \ell og\, Z(T) \quad (7.\ 3)$$

である。また，二つの等式

variance $\sigma^2 = \langle \{E(x) - m\}^2 \rangle$

$$= \Sigma_\alpha p(x_\alpha) \cdot [E(x_\alpha) - m]^2$$

$$= \langle E^2(x) \rangle - \langle E(x) \rangle^2$$

$$= \left[\partial^2 / \partial\left(\frac{1}{T}\right)^2\right] \ell og\, Z(T) \quad (7.\ 4)$$

$$\max H = \ell og\, Z(T) + \frac{1}{T} \cdot \langle E(x) \rangle \quad (7.\ 5)$$

も成立している。

$Z(T)$ は統計物理学モデル（statistical physical model）では分配関数と呼ばれている[26]。式（7．1）の確率測度 $p(x_\alpha), \alpha = 1, 2, \cdots, n$ はギブス分布（Gibbs distribution）と呼ばれており，特に，エネルギー関数（an energy function）$E(x_\alpha)$ が x_α の 2 次形式のときは

ボルツマン分布（Boltzmann distribution）と呼ばれている[27]。

5式（7．1）～（7．5）を導いておこう。

ε は十分小さなパラメータ

$\eta(x_\alpha)$ は任意の関数

$$q(x_\alpha) = p(x_\alpha) + \varepsilon \cdot \eta(x_\alpha)$$

とおく。ここで，

$$q(x_\alpha)\,|_{\varepsilon=0} = p(x_\alpha)$$

に注意しておく。さらに，

$$p_\alpha = p(x_\alpha),\ q_\alpha = q(x_\alpha),$$
$$\eta_\alpha = \eta(x_\alpha),\ e_\alpha = E(x_\alpha)$$

とおく。

$$F(p_\alpha) \equiv -\Sigma_\alpha p_\alpha\, \ell og\, p_\alpha$$

$$+\lambda\cdot[\Sigma_\alpha p_\alpha\cdot e_\alpha-m]+\mu\cdot[\Sigma_\alpha p_\alpha-1]$$

を導入し，p_a の代りに q_a を代入して得られる $F(q_\alpha)$ は $\varepsilon=0$ で極値をとるものとすれば，少なくとも

$$(\partial/\partial\lambda)\,F\,(q_\alpha)\,|_{\varepsilon=0}=0 \qquad (7.\ 6)$$

$$(\partial/\partial\mu)\,F\,(q_\alpha)\,|_{\varepsilon=0}=0 \qquad (7.\ 7)$$

$$(\partial/\partial\varepsilon)\,F\,(q_\alpha)\,|_{\varepsilon=0}=0 \qquad (7.\ 8)$$

が成り立たねばならない。

式（7．6）から

$$0=\Sigma_\alpha p_\alpha\cdot e_\alpha-m \qquad (7.\ 9)$$

を得，式（7．7）から

$$0=\Sigma_\alpha p_\alpha-1 \qquad (7.\ 10)$$

を得る。また，

$$(\partial/\partial\varepsilon)\,F\,(q_\alpha)$$
$$=-\Sigma_\alpha\eta_\alpha\cdot[\ell og\,q_\alpha+1-\lambda e_\alpha-\mu]$$

であるから，

$$0=-\Sigma_\alpha\eta_\alpha\cdot[\ell og\,q_\alpha+1-\lambda e_\alpha-\mu]$$

を得，η_α は任意関数であるから，

$$\forall\alpha,\ell og\,p_\alpha+1-\lambda e_\alpha-\mu=0$$
$$\therefore\quad p_\alpha=\exp[\mu-1]\cdot\exp[\lambda\,e_\alpha] \qquad (7.\ 11)$$

が得られる。式（7．11）を式（7．10）に代入して，

$$\exp[\mu-1]=[\Sigma_\alpha\exp[\lambda\,e_\alpha]]^{-1}$$

を得て，よって，式（7．11）から

$$\forall\alpha,p_a=[\Sigma_\alpha\exp[\lambda\,e_\alpha]]^{-1}\cdot\exp[\lambda\,e_\alpha] \qquad (7.\ 12)$$

が得られる。

$$\lambda=-\frac{1}{T}=-\beta,\beta=\frac{1}{T} \qquad (7.\ 13)$$

とおけば，2式（7．1），（7．2）が得られる。

式（7．9）に式（7．12）を代入すれば，T を決定する方程式

$$\langle E(x)\rangle=\Sigma_\alpha\frac{1}{Z(T)}\cdot\exp\left[-\frac{e_\alpha}{T}\right]\cdot e_\alpha=m \qquad (7.\ 14)$$

が得られる。ところで,

$$-\left[\partial/\partial\left(\frac{1}{T}\right)\right]\ell og\,Z(T)$$

$$=-\frac{1}{Z(T)}\cdot\left[\partial/\partial\left(\frac{1}{T}\right)\right]Z(T)$$

$$=-\frac{1}{Z(T)}\Sigma_\alpha e_\alpha\cdot\exp\left[-\frac{1}{T}\cdot e_\alpha\right]$$

$$=\Sigma_\alpha e_\alpha\cdot p_\alpha=\langle E(x)\rangle$$

であるから,式(7.14)より

$$\langle E(x)\rangle=-\left[\partial/\partial\left(\frac{1}{T}\right)\right]\ell og\,Z(T)=m$$

を得て,式(7.3)が示された。

式(7.4)を示そう。

$$\left[\partial^2/\partial\left(\frac{1}{T}\right)^2\right]\ell og\,Z(T)$$

$$=\left[\partial/\partial\left(\frac{1}{T}\right)\right]\left[\frac{1}{Z(T)}\cdot\frac{\partial Z(T)}{\partial(1/T)}\right]$$

$$=\frac{1}{Z(T)}\cdot\left[\partial^2/\partial\left(\frac{1}{T}\right)^2\right]Z(T)$$

$$-Z(T)^{-2}\cdot\left[[\partial/\partial\left(\frac{1}{T}\right)\right]Z(T)\Big]^2$$

$$=Z(T)^{-1}\cdot\Sigma_\alpha e_\alpha^2\cdot\exp\left[-\frac{1}{T}e_\alpha\right]$$

$$-\left[Z(T)^{-1}\cdot\Sigma_\alpha(-e_\alpha)\cdot\exp\left[-\frac{1}{T}e_\alpha\right]\right]^2$$

$$=\Sigma_\alpha e_\alpha^2\cdot p_\alpha-[\Sigma_\alpha e_\alpha p_\alpha]^2$$

$$=\langle E^2(x)\rangle-\langle E(x)\rangle^2$$

$$=\Sigma_\alpha[e_\alpha-m)^2\cdot p_\alpha$$

$$=\langle\{E(x)-m\}^2\rangle$$

を得て,式(7.4)が示された。

最後に,式(7.5)を示そう。

式(7.1)の $p_\alpha=p(x_\alpha)$ を代入したものが $\max H$ であるから,

$$\max\quad H=\Sigma_\alpha p_\alpha\,\ell og\,p_\alpha$$

$$=-\Sigma_\alpha p_\alpha\left[-\ell og\,Z(T)-\frac{e_\alpha}{T}\right]$$

$$=\ell og\,Z(T)+\frac{1}{T}\langle E(x)\rangle$$

を得て，示された。

8．制約付き推定問題(2)

前章とは異なり，

定分散の下での最大エントロピー原理（maximum entropy principle under constant variance）を示そう。

Maximize

$$H \equiv -\sum_\alpha p(x_\alpha)\, \ell og\, p(x_\alpha)$$

Subject to

(i) $\sum_\alpha p(x_\alpha) \cdot [E(x_\alpha) - m]^2 = \text{constant } \sigma^2$

(ii) $\sum_\alpha p(x_\alpha) = 1$

(iii) $0 \leqq p(x_\alpha) \leqq 1$ for all x_α. □

解の確率測度 $p(x_\alpha), \alpha = 1, 2, \cdots, n$ は

$$p(x_\alpha) = \exp[\lambda(E(x_\alpha) - m)^2] / \sum_\alpha \exp[\lambda(E(x_\alpha) - m)^2]$$

であり，未知パラメータ λ を決定する方程式は，

$$\sum_\alpha (E(x_\alpha) - m)^2 \cdot \exp[\lambda(E(x_\alpha) - m)^2] = \sigma^2 \sum_\alpha \exp[\lambda(E(x_\alpha) - m)^2] \qquad (8.\ 1)$$

である。

定数 m は例えば，期待値（expected value）$\sum_\alpha p(x_\alpha) \cdot E(x_\alpha) = m$

と考えられる。したがって，

$$\sum_\alpha E(x_\alpha) \cdot \exp[\lambda(E(x_\alpha) - m)^2] = m \cdot \sum_\alpha \exp[\lambda(E(x_\alpha) - m)^2] \qquad (8.\ 2)$$

を得て，2式（8．1），（8．2）が二つの未知パラメータ λ，m を決める方程式である。

上述の　$p(x_\alpha)$　を導こう。

$$p_\alpha = p(x_\alpha), \eta_\alpha = \eta(x_\alpha)$$
$$q_\alpha = q(x_\alpha) = p(x_\alpha) + \varepsilon \cdot \eta(x_\alpha)$$
$$= p_\alpha + \varepsilon \cdot \eta_\alpha$$
$$e_\alpha = E(x_\alpha)$$

とおく。ここに，ε は十分小さいパラメータであり，$\eta(x_\alpha)$ は任意の関数である。

関数　$F(q_\alpha)$　を

$$F(q_\alpha)$$
$$= -\sum_\alpha q_\alpha\, \ell og\, q_\alpha + \lambda(\sum_\alpha q_\alpha (e_\alpha - m)^2$$
$$- \sigma^2) + \mu(\sum_\alpha q_\alpha - 1)$$

とおく。$F(q_\alpha)$ は $\varepsilon=0$ で極値をとるから，

$$(\partial/\partial\lambda)\, F\ (q_\alpha)\,|_{\varepsilon=0}=0 \qquad (8.\ 3)$$

$$(\partial/\partial\mu)\, F\ (q_\alpha)\,|_{\varepsilon=0}=0 \qquad (8.\ 4)$$

$$(\partial/\partial\varepsilon)\, F\ (q_\alpha)\,|_{\varepsilon=0}=0 \qquad (8.\ 5)$$

が成立しなければならない。

式（8．3）から

$$\Sigma_\alpha p_\alpha (e_\alpha - m)^2 - \sigma^2 = 0 \qquad (8.\ 6)$$

を得，式（8．4）から

$$\Sigma_\alpha p_\alpha - 1 = 0 \qquad (8.\ 7)$$

を得る。また，式（8．5）から，

$$\begin{aligned}&(\partial/\partial\varepsilon)\, F\ (q_\alpha)\,|_{\varepsilon=0}\\ &= -\Sigma_\alpha \eta_\alpha \cdot [\ell og\, p_\alpha + 1 - \lambda(e_\alpha - m)^2 - \mu] = 0\end{aligned}$$

を得て，η_α は任意であるから，

$$\begin{aligned}&\forall\alpha, \ell og\, p_\alpha + 1 - \lambda(e_\alpha - m)^2 - \mu = 0\\ &\therefore\ \ p_\alpha = \exp[\mu - 1]\cdot\exp[\lambda(e_\alpha - m)^2]\end{aligned} \qquad (8.\ 8)$$

が得がれる。

式（8．8）を式（8．7）に代入すれば，

$$\exp[\mu - 1] = [\Sigma_\alpha \exp[\lambda(\mathrm{e}_\alpha - m)^2]]^{-1} \qquad (8.\ 9)$$

を得て，この式（8．9）を式（8．8）に代入すれば，所要の

$$\forall\alpha, p_\alpha = \frac{\exp[\lambda(\mathrm{e}_\alpha - m)^2]}{\Sigma_\alpha \exp[\lambda(\mathrm{e}_\alpha - m)^2]} \qquad (8.\ 10)$$

が得られ，この式（8．10）を式（8．6）に代入すれば，式（8．1）が得られ，証明が終った。

9．制約付最適化問題(1)

確率モデルの最適化問題へ，

定エントロピーの下での最小平均エネルギー原理（the principle of minimum average energy under constant entropy）

を適用しよう。これは，

constrained optimization problem（制約付き最適化問題）

への，

minimum energy optimization

の適用であり，下記の３条件ｉ～iiiを制約条件として，平均エネルギー $\langle E(x)\rangle$ を最小にする確率分布

$$p(x_\alpha),\ \alpha=1,2,\cdots,n$$

を求める問題として定式化される：

Minimize

$$\langle E(x)\rangle \equiv \sum_\alpha p(x_\alpha)\cdot E(x_\alpha)$$

Subject to

(i) $-\sum_\alpha p(x_\alpha)\,\ell og\, p(x_\alpha) = \text{constant } H$

(ii) $\sum_\alpha p(x_\alpha)=1$

(iii) $0\leqq p(x_\alpha)\leqq 1$ for all x_α. □

上述の問題は変分原理におけるラグランジュの未定乗数法によって解くことができ，
確率測度

$$p(x_\alpha)=\frac{1}{Z(T)}\cdot\exp\left[-\frac{E(x_\alpha)}{T}\right] \tag{9.1}$$

分配関数

$$Z(T)=\sum_\alpha \exp\left[-\frac{E(x_\alpha)}{T}\right] \tag{9.2}$$

となる。ここに，T は定エントロピー条件の下での温度と呼ばれている。温度 T を決定するのに必要とされる方程式は

$$\ell og\, Z(T)+\frac{1}{T}\cdot\langle E(x)\rangle = H \tag{9.3}$$

である。

３式（９．１）～（９．３）を導こう。
ε は十分小さいパラメータ
$\eta(x_\alpha)$ は任意の関数
として，

$$q(x_\alpha)=p(x_\alpha)+\varepsilon\cdot\eta(x_\alpha)$$

とおく。ここで，

$$q(x_\alpha)\,|_{\varepsilon=0}=p(x_\alpha)$$

の成立に注意しておく。

$$p_\alpha=p(x_\alpha),\ q_\alpha=q=(x_\alpha),\ \eta_\alpha=\eta(x_\alpha),$$

$e_\alpha = E(x_\alpha)$

とおく。ラグランジュの未定乗数法によって解く。

$$F(q_\alpha)$$
$$\equiv \Sigma_\alpha q_\alpha \cdot e_\alpha + \lambda \cdot [-\Sigma_\alpha q_\alpha \ell og\, q_\alpha - H]$$
$$+ \mu \cdot [\Sigma_\alpha q_\alpha - 1]$$

は　$\varepsilon = 0$　で極値をとるものとすれば，少なくとも

$$(\partial / \partial\lambda)\, F\, (q_\alpha)\,|_{\varepsilon=0} = 0 \qquad (9.\ 4)$$

$$(\partial / \partial\mu)\, F\, (q_\alpha)\,|_{\varepsilon=0} = 0 \qquad (9.\ 5)$$

$$(\partial / \partial\varepsilon)\, F\, (q_\alpha)\,|_{\varepsilon=0} = 0 \qquad (9.\ 6)$$

が成立しなければならない。
式（9．4）から

$$-\Sigma_\alpha p_\alpha p_\alpha - H = 0 \qquad (9.\ 7)$$

を得，式（9．5）から

$$\Sigma_\alpha p_\alpha - 1 = 0 \qquad (9.\ 8)$$

を得る。また，

$$(\partial / \partial\varepsilon)\, F\, (q_\alpha)$$
$$= \Sigma_\alpha \eta_\alpha [e_\alpha - \lambda\, \ell og\, q_\alpha - \lambda + \mu]$$

であるから，式（9．6）から

$$0 = \Sigma_\alpha \eta_\alpha [e_\alpha - \lambda\, \ell og\, p_\alpha - \lambda + \mu]$$

を得るが，η_α は任意であるから，

$$\forall \alpha, e_\alpha - \lambda\, \ell og\, p_\alpha - \lambda + \mu = 0$$
$$\therefore \quad p_\alpha = \exp[-1 + \lambda^{-1}\mu] \cdot \exp[e_\alpha / \lambda] \qquad (9.\ 9)$$

が成立しなければならない。式（9．9）を式（9．8）に代入すれば，

$$\exp[-1 + \mu / \lambda] = \frac{1}{\Sigma_\alpha \exp[e_\alpha / \lambda]} \qquad (9.\ 10)$$

を得，改めて，式（9．10）に式（9．9）を代入すれば，

$$\forall \alpha, p_\alpha = \frac{\exp[e_\alpha / \lambda]}{\Sigma_\alpha \exp[e_\alpha / \lambda]}$$

が成立する。ここで，

$$\lambda^{-1} = -T^{-1}$$

とおけば，解として，2式（9．1），（9．2）が得られる。

温度 T を決定する方程式については，未使用の式（9．7）に，式（9．1）を代入して，

$$-\sum_{\alpha} p_{\alpha} \cdot [-\ell og\, Z(T) - T^{-1} \cdot e_{\alpha}] - H = 0$$
$$\therefore \quad \ell og\, Z(T) + T^{-1} \cdot \langle E(x) \rangle = H$$

が得られ，証明が終った。

上述は次の事実を指摘している：定エントロピーの下での最小平均エネルギー原理は，目的と制約とを交換し，最小化を最大化に置き換えて得る“定エネルギーの下での最大エントロピー原理であり，互いに双対的である。

□

上記の定エントロピー下の最小平均エネルギー原理の一つの適用例を示しておこう。

M 個のパターンの集合

$$\Phi = \{\varphi_1, \varphi_\alpha, \cdots, \varphi_M\}$$

がミクロカノニカル集団（第11章を参照）である場合，

$\forall k (=1 \sim M), \sum_{\ell \in L} n_k(\ell) = N_k$（パターン φ_k に含まれている粒子の総数）＝一定

であるが，ここで，この制限をはずそう。

第 k 番目のパターン φ_k から抽出される第 $\ell \in L$ 番目の特徴量（連続変量；continuous variable）を

$$u(\varphi_k, \ell) \geqq 0$$

とすれば，$n_k(\ell)$ は不等式

$$n(\ell) \cdot \Delta u \leqq u(\varphi_k, \ell) < [n_k(\ell) + 1] \cdot \Delta u$$

を満たす非負整数（離散変量，discrete variable）であり，$u(\varphi_k, \ell)$ に含まれる粒子（量子 Δu を単位とする特徴量子）の総数と解釈されてよい。

パターン集合 Φ 内の粒子の総数 N は

$\sum_{k=1}^{M} n_k(\ell) = N(\ell)$（第 ℓ 特徴に関する Φ 内の粒子の総数）

を導入して，

$$N = \sum_{k=1}^{M} \sum_{\ell \in L} n_k(\ell) = \sum_{k=1}^{M} N_k$$
$$= \sum_{\ell \in L} N(\ell)$$

と表現される。さて，

$$E(x_\ell) \leftrightarrow N(\ell)$$

という対応を考えると，Φ 内に含まれている

平均粒子数 $= \langle N(\ell) \rangle = \sum_{\ell \in L} p_\ell \cdot N(\ell)$

を最小にする，つまり，

粒子数 $n_k(\ell) \in \{0, 1, 2, \cdots,\}, k = 1 \sim M, \ell \in L$

に関して，<u>Φ の形状を全体的に簡素化する粒子数の確率分布</u> $p_\ell, \ell \in L$ は，

定エントロピー条件

$$-\sum_{\ell\in L} p_\ell \log p_\ell = H$$

の下では,

$$p_\ell = \frac{1}{Z(T)}\cdot\exp\left[-\frac{N(\ell)}{T}\right],\ ここに,$$

$$Z(T) = \sum_{\ell\in L}\exp\left[-\frac{N(\ell)}{T}\right]$$

であり,温度 T は

$$\log Z(T) + T^{-1}\cdot\sum_{\ell\in L} p_\ell\cdot N(\ell) = H$$

から決められる。 □

熱容量の大なる物体と熱平衡[35] (thermal equilibrium) を保つ任意の力学系の統計的エネルギー分布はカノニカル分布 (canonical distribution) であり,この力学系(粒子集団)を

カノニカル集団 (canonical ensemble)

ということは統計力学 (statistical mechanics) 上知られている[26]。カノニカル分布の場合,エネルギー E_ℓ をもつ量子状態 ℓ の規格化された実現確率 $p(E_\ell)$ は

$$p(E_\ell) = \exp[-E_\ell/T]/\sum_k \exp[-E_k/T]$$

であることが示されているから,上記の定エントロピー下での最小平均エネルギー原理を適用して得られた"粒子数の確率分布" p_ℓ, $\ell\in L$ は一種のカノニカル分布であり,式(9.4)のすぐ上の関数値 $F(q_\alpha)$ は実はカノニカル粒子集団の自由エネルギーに相当する。

また,Hopfield neural net binary model[7] の状態遷移動作を確率化して得られる Boltzmann machine[35] でも,その遷移確率 (transition probability),定常確率は一種のカノニカル分布である(ただし,その温度 T は simulated annealing algorithm に基づき,熱平衡に到達する様,状態変化させながら徐々に下げさせられる)ことに留意すれば,定エネルギー下の最大エントロピー原理あるいは定エントロピー下での最小平均エネルギー原理の基本的重要性が了解できる。

統計力学では,カノニカル粒子集団は,エネルギー E_ℓ をもつ量子状態 ℓ について,ε_ℓ を一粒子としての量子状態 ℓ におけるエネルギーとして,

(イ) $E_\ell\in\{0,\varepsilon_\ell\}$ の場合

フェルミ・デイラック (Fermi-Dirac) の統計に,

(ロ) $E_\ell\in\{0,\varepsilon_\ell,2\varepsilon_\ell,3\varepsilon_\ell,\cdots\}$ の場合

ボーズ・アインシュタイン (Bose-Einstein) の統計に

従うと称せられるが,これに対して,

分配関数 $Z(T)$,確率 p_ℓ

を計算しておこう。上記の特徴抽出の場面における Δu を ε_ℓ と考えれば,

(i)(フェルミ・デイラック統計)

$$Z(T) = \sum_{k=0}^{1}\exp\left[-\frac{k}{T}\right] = 1+\exp\left[-\frac{1}{T}\right]$$

であるから,

$$p_\ell = \frac{1}{Z(T)} \cdot \exp\left[-\frac{N(\ell)}{T}\right] = \frac{\exp\left[-\frac{N(\ell)}{T}\right]}{1+\exp\left[-\frac{1}{T}\right]}$$

$$= \left[1+\exp\left[-\frac{1}{T}\right]\right]^{-1} \text{ if } N(\ell)=0$$

$$= \left[1+\exp\left[+\frac{1}{T}\right]\right]^{-1} \text{ if } N(\ell)=1$$

(ii) (ボーズ・アインシュタイン統計)

$N(\ell) \in \{0, 1, 2, \cdots\}$ の場合

$Z(T)$

$$= \Sigma_{k=0}^{\infty} \exp\left[-\frac{k}{T}\right] = \frac{1}{1-\exp\left[\frac{1}{T}\right]}$$

を得,

$$p_\ell = \frac{1}{Z(T)} \cdot \exp\left[-\frac{N(\ell)}{T}\right] = \frac{\exp\left[-\frac{N(\ell)}{T}\right]}{1+\exp\left[-\frac{1}{T}\right]}.$$

□

10. 制約付き最適化問題(2)

前章とは異なり,

定エントロピーの下での最小分散原理 (minimum variance principle under constant entropy) を示そう。これは8章での定分散下での最大エントロピー原理の双対にあたる。

Minimize

$$\langle \sigma^2(x)] \equiv \Sigma_\alpha p(x_\alpha) \cdot [E(x_\alpha) - m]^2$$

Subject to

(i) $-\Sigma_\alpha p(x_\alpha)\, \ell og\, p(x_\alpha) = \text{constant } H$

(ii) $\Sigma_\alpha p(x_\alpha) = 1$

(iii) $0 \leqq p(x_\alpha) \leqq 1$ for all x_α. □

解の確率測度 $p(x_\alpha),_\alpha = 1, 2, \cdots, n$ は

$$p(x_\alpha) = \exp[\lambda^{-1}(E(x_\alpha) - m)^2] \\ / \Sigma_\alpha \exp[\lambda^{-1}(E(x_\alpha) - m)^2]$$

であり, 未知パラメータ λ, m を決める方程式は

$$\ell og\, \Sigma_\alpha \exp[\lambda^{-1}(E(x_\alpha) - m)^2]$$

$$= \lambda^{-1} \sum_\alpha p(x_\alpha) \cdot (E(x_\alpha) - m)^2 + H \tag{10. 1}$$

$$\sum_\alpha E(x_\alpha) \cdot \exp[\lambda^{-1}(E(x_\alpha) - m)^2]$$
$$= m \cdot \sum_\alpha \exp[\lambda^{-1}(E(x_\alpha) - m)^2] \tag{10. 2}$$

である。

上述の $p(x_\alpha)$ を導こう。

$$p_\alpha = p(x_\alpha), q_\alpha = q(x_\alpha), \eta_\alpha = \eta(x_\alpha), e_\alpha = E(x_\alpha)$$
$$q(x_\alpha) = p(x_\alpha) + \varepsilon \cdot \eta(x_\alpha)$$

とし，関数

$$F(q_\alpha) = \sum_\alpha q_\alpha (e_\alpha - m)^2$$
$$+ \lambda \cdot [- \sum_\alpha q_\alpha \log q_\alpha - H]$$
$$+ \mu(\sum_\alpha q_\alpha - 1)$$

を考える。ここに，

ε：十分小さなパラメータ

$\eta(x_\alpha)$：任意関数

である。

$F(q_\alpha)$ は $\varepsilon = 0$ で極値をとるものとすれば，

$$(\partial / \partial \lambda) F(q_\alpha) |_{\varepsilon=0} = 0 \tag{10. 3}$$

$$(\partial / \partial \mu) F(q_\alpha) |_{\varepsilon=0} = 0 \tag{10. 4}$$

$$(\partial / \partial \varepsilon) F(q_\alpha) |_{\varepsilon=0} = 0 \tag{10. 5}$$

が成り立つ。

式（10. 3）から，

$$- \sum_\alpha p_\alpha \log p_\alpha - H = 0 \tag{10. 6}$$

を得，式（10. 4）から

$$\sum_\alpha p_\alpha - 1 = 0 \tag{10. 7}$$

を得る。また，

$$(\partial / \partial \varepsilon) F(q_\alpha)$$
$$= \sum_\alpha \eta_\alpha [(e_\alpha - m)^2 - \lambda \log q_\alpha - \lambda + \mu]$$

であるから，式（10. 5）より

$$0 = \sum_\alpha \eta_\alpha [e_\alpha - m)^2 - \lambda \log p_\alpha - \lambda + \mu]$$

が成り立ち，η_α は任意であるから，結局，

$$\forall \alpha, (e_\alpha - m)^2 - \lambda \log p_\alpha - \lambda + \mu = 0$$
$$\therefore \quad p_\alpha = \exp[-1 + \lambda^{-1} \mu] \cdot \exp[\lambda^{-1}(e_\alpha - m)^2] \tag{10. 8}$$

が得られる。式（10. 8）を式（10. 7）に代入すれば，

$$\exp[-1+\lambda^{-1}\mu]=[\Sigma_\alpha \exp[\lambda^{-1}(e_\alpha - m)^2]]^{-1} \qquad (10.\ 9)$$

を得て，この式（10. 9）を式（10. 8）に代入すれば，所要の

$$p_a=\exp[\lambda^{-1}(e_\alpha - m)^2]/\Sigma_\alpha \exp[\lambda^{-1}(e_\alpha - m)^2] \qquad (10.\ 10)$$

が得られる。未定乗数 λ を決める方程式については，式（10. 10）を式（10. 6）に代入して

$$\ell og\, \Sigma_\alpha \exp[\lambda^{-1}(e_\alpha - m)^2]$$
$$-\Sigma_\alpha p_\alpha \ell og \exp[\lambda^{-1}(e_\alpha - m)^2]-H=0$$

を得，これを整理したものが式（10. 1）である。

最後に，式（10. 2）は，m を平均値と考え，

$$\Sigma_\alpha E(x_\alpha)\cdot p(x_\alpha)=m$$

に式（10. 10）の $p(x_\alpha)$ を代入すれば得られる。

11. 頻度分布，相互情報量とパターンのミクロカノニカル粒子集団

一般に，Shannon の発見した二つの変数 X，Yの相互情報量（mutual information），$I(X, Y)$ はX，Y間の**相互依存性**（mutual dependence）の計量化であり，

(a) X，Yが確率変数であれば，統計的独立性（statistial independence）

(b) X，Yが確率変数でなければ，相互排除性（mutual exclusion）

の程度が激しい程，小さな値をとる非負量である。

For any event A whose probability of occurrence is $P(A)$, the amount of information we receive as a result of being told that A has occurred is defined by

$$i(A)=-\ell og\, p(A).$$

This information is zero when $P(A)=1$, since we knew already that A will occur, and approaches infinity when $P(A)$ approaches zero. In the same manner, the conditional information that we receive if we already know that B has occurred and told that A has occurred is

$$i(A/B)=-\ell og\, p(A/B).$$

The contribution of B to the information about A is expressed by the mutual information

$$i(A;B)=i(A)-i(A/B)$$
$$=\ell og[p(A/B)/p(A)].$$

Note that if A is highly correlated with B, $p(A/B)$ should be close to one, so that $i(A/B)$ is close to zero, making $i(A;B)$ high (close to $i(A)$); while if A is negatively correlated with B, $P(A/B)$ will be close to zero, so that $I(A;B)$ will be very small[28].

上記英文説述内容に関連し，上述 b での相互排除性の計量化については第18章で研究すること

にし，本章では，a での統計的確立性の計量化について説明する。

$p(x, y)$: x と y との結合確率
(the joint probability of pair $\langle x, y\rangle$)
$p_1(x)$: x の確率
$p_2(y)$: y の確率

を導入すると，

$$\textstyle\sum_y p(x, y) = p_1(x)$$
$$\textstyle\sum_x p(x, y) = p_2(y)$$

が成り立ち，
$p_1(x/y)$: y の値が知られているときの，x の条件付確率（the conditional probability）
$p_2(y/x)$: x の値が知られているときの，y の条件付確率
について，

$$p_1(x/y) = p(x, y)/p_2(y)$$
$$p_2(y/x) = p(x, y)/p_1(x)$$

が成立する。

$$\begin{aligned} H(X) &= -\textstyle\sum_x p_1(x)\,\ell og\, p_1(x) \\ &= -\textstyle\sum_x \sum_y p(x, y)\,\ell og\, p_1(x) \\ &= -\textstyle\sum_x \sum_y p_2(y)\cdot p_1(x/y)\,\ell og \\ &\qquad \textstyle\sum_y p_2(y)\cdot p_1(x/y) \geqq 0 \end{aligned}$$

は，確率変数 X のもつ不確さであるから，確率変数 Y が特定の y という値をとったとき，X の不確かさは

$$\begin{aligned} &H(X/y) \\ &= -\textstyle\sum_x p_1(x/y)\,\ell og\, p_1(x/y) \geqq 0 \end{aligned}$$

であるといえる。多数の y が生起した場合の平均の不確かさは

$$\begin{aligned} &H(X/y) \\ &= \textstyle\sum_y p_2(y)\cdot H(X/y) \\ &= -\textstyle\sum_y p_2(y)\, p_1(x/y)\,\ell og\, p_1(x/y) \\ &= -\textstyle\sum_y p(x, y)\,\ell og\, p_1(x/y) \geqq 0 \end{aligned}$$

と表現されることになる。そうすると，

$$\begin{aligned} &I(X, Y) \\ &= H(X) - H(X/Y) \end{aligned}$$

＝〔もともと，X の持っていた不確かさ〕－〔Y が知らされた後でも残存している X の不確

かさ〕
は
多数の y が生起した場合に Y によって X から取り出された不確かさ
と解釈され，システムが受け取った情報量であるといっていい。この $I(X,Y)$ はシャノンによれば，
X，Y の相互情報量（mutual information content），あるいは伝達情報量（報知高）
と呼ばれている。$I(X, Y)$ は

$$
\begin{aligned}
&I(X,Y)\\
&= -\sum_x \sum_y p(x,y)\,\ell og\, p_1(x)\\
&\quad + \sum_x \sum_y p(x,y)\,\ell og\, p_1(x/y)\\
&= \sum_x \sum_y p(x,y)\,\ell og \frac{p_1(x/y)}{p_1(x)}\\
&= \sum_x \sum_y p(x,y)\,\ell og \frac{p(x,y)}{p_1(x)\cdot p_2(y)}
\end{aligned}
$$

と表現され，X，Y につき対称である：

$$I(X,Y)=I(Y,X).$$

$I(X,Y)$ が非負量であることを示すために，次の補助定理11．1を用意する。
〔補助定理11．1〕　関数

$$f(x) = -x\,\ell og\, x \quad (0 \leqq x)$$

については，

$$
\begin{aligned}
&f(0)=f(1)=0\\
&(d/dx)f(x) = -\ell og\, x - 1\\
&(d^2/dx^2)f(x) = -1/x \leqq 0
\end{aligned}
$$

が成り立ち，$f(x)$ は上に凸であり，
$x=e^{-1}$ のとき，最大値 $f(e^{-1})=e^{-1}$
をとり，
$0 \leqq x \leqq 1$ に対し $f(x) \geqq 0$
$1 < x$ に対し，$f(x) < 0$
である。さらに，

$$
\begin{aligned}
&\lambda_i \geqq 0 \quad (i=1\sim n)\\
&\textstyle\sum_{i=1}^{n} \lambda_i = 1
\end{aligned}
$$

を満たす任意の $\lambda=\{\lambda_i \mid i=1\sim n\}$ について，不等式

$$\textstyle\sum_{i=1}^{n} \lambda_i \cdot f(x_i) \leqq f(\sum_{i=1}^{n} \lambda_i x_i)$$

が成り立つ。等号の成立は

$$x_1 = x_2 = \cdots = x_n$$

のときに限る。 □

上述の補定11. 1を $f(x) = -x\,\ell og\,x$ として，

$$\lambda_i \leftrightarrow p_2(y)$$

と対応させて適用すれば，

$$\begin{aligned}
&\textstyle\sum_y p_2(y)\,[-p_1(x/y)\,\ell og\,p_1(x/y)] \\
&= \textstyle\sum_y p_2(y)\cdot f(p_1(x/y)) \\
&\leqq f(\textstyle\sum_y p_2(y)\cdot p_1(x/y)) \\
&= f(p_1(x)) = -p_1(x)\,\ell og\,p_1(x)
\end{aligned} \tag{11.1}$$

を得て，変数 x につき総和をとれば，

$$\begin{aligned}
H(X/Y) &= \textstyle\sum_x \sum_y p_2(y)\,[-p_1(x/y)\,\ell og\,p_1(x/y)] \\
&\leqq -\textstyle\sum_x p_1(x)\,\ell og\,p_1(x) = H(X)
\end{aligned}$$

が成立し，

$$I(X, Y) \geqq 0$$

が知れ，等号の成り立つのは
変数 y を含まない変数 x のみの関数 $q(x)$ が存在して，$p_1(x/y) = q(x)$ が成立する，
つまり

$$\forall x, \forall y, p(x, y) = p_1(x)\cdot p_2(y)$$

（X，Y が統計的に独立（statistical independence）
の場合に限ることが判明した。 □

さて，　$x = x_k$　なる事態が　n_k 回　生起したとすれば，規格化頻度分布（normalized frequency-distribution of occurrence）

$\{n_k/N \mid k \in K\}$，ここに $N = \sum_{k\in K} n_k$
に注目し，

$$p_1(x_k) = n_k/N, k \in K$$

とおけるから，

$$H(X) = -\textstyle\sum_k p_1(x_k)\,\ell og\,p_1(x_k)$$

については

$$N\cdot H(X) = H' + N\,\ell og\,N$$

ここに，$H' \equiv -\sum_k n_k\,\ell og\,n_k$
が得られる。

Nが一定であるとき，$H(X)$ の代わりに，
$N \cdot H(X)$ を用いる
ことを考えると，

$$N \ell og\, N + H' = (\Sigma_k n_k)\, \ell og\, (\Sigma_k n_k) \\ - \Sigma_k n_k \ell og\, n_k$$

を情報量とみなせる。H' は頻度分布

$$\{n_k \mid k \in K\}$$

から計算される情報測度と呼ばれている(29)。

パターン情報処理への一つの応用を述べよう。

m 個のパターン　$\varphi_1, \varphi_2, \cdots, \varphi_M$

を考え，第 k 番目のパターン φ_k から抽出される第 $\ell \in L$ 番目の非負特徴量を

$$u(\varphi_k, \ell) \geqq 0$$

と表す。不等式

$$n_k(\ell) \cdot \Delta u \leqq u(\varphi_k, \ell) \\ < [n_k(\ell) + 1] \cdot \Delta u$$

を満たす非負整数　$n_k(\ell)$　を考えよう。ここに，Δu は特徴量子（feature-quantum）と呼ばれ，固定化した正の量が採用されねばならない。

$n_k(\ell)$ は φ_k の第 ℓ 特徴量 $u(\varphi_k, \ell)$ に対応する粒子数と称えられる。

統計物理学においては，エネルギーが一定である力学系の集団を
ミクロカノニカル集団（microcanonical ensemble）
というが(26)，パターン情報処理学においては**エネルギーに対応する概念は特徴量の総和**と考え，
特徴量の総和　$\Sigma_{\ell \in L} u(\varphi_k, \ell)$　が一定であるパターンはミクロカノニカル集団に属する粒子から成っている
ということにしよう。

全粒子数（the number of particles contained in the pattern φ_k）は

$$N_k = \Sigma_{\ell \in L} n_k(\ell)$$

と表現されるが，特徴量の総和は不等式

$$N_k \cdot \Delta_u \leqq \Sigma_{\ell \in L} u(\varphi_k, \ell) < [N_k + \Sigma_{\ell \in L} 1] \cdot \Delta u$$

を満たす。この様に“粒子の概念”を導入すれば，パターン φ_k からの特徴抽出（feature-extraction）の働きとは
N_k 個の粒子の内，第 $\ell \in L$ 番目の特徴軸に $n_k(\ell)$ 個の粒子を割り当てる
ことであるといいかえられる。

この場合，各 k にわたって　$N_k =$ 一定　であれば，

$[\sum_{\ell\in L} n_k(\ell)]\ \ell og\,[\sum_{\ell\in L} n_k(\ell)]$
$-\sum_{\ell\in L} n_k(\ell)\ \ell og\, n_k(\ell)$

がミクロカノニカル集団に属する粒子から成っているパターン φ_k の情報量である，といえる。

頻度分布から定まる伝達情報量について論じよう。

クラス i に関して性質 j に属する要素の個数を　a_{ij}　と表す。このとき，

$$r_i = \sum_j a_{ij} \quad , \quad c_j = \sum_i a_{ij}$$

として，

$$N \equiv \sum_i \sum_j a_{ij} = \sum_i r_i = \sum_j c_j$$

であるから，

$X \leftrightarrow$ クラス i の集合
$Y \leftrightarrow$ 性質 j の集合

という対応の下で，伝達情報量

$$I(X, Y)$$
$$\equiv \sum_i \sum_j \frac{a_{ij}}{N}\ell og\left[[a_{ij}/N]/\left[\frac{r_i}{N}\cdot\frac{c_j}{N}\right]\right]$$

を計算してみよう。

$$I_r \equiv -\sum_i r_i\, \ell og\, r_i$$
$$I_c \equiv -\sum_j c_j\, \ell og c_j$$
$$I_a = -\sum_i \sum_j a_{ij}\, \ell og\, a_{ij}$$

として，

$$N\cdot H(X) = N\,\ell og\,N - \sum_i r_i\,\ell og\,r_i$$
$$N\cdot H(X/Y)$$
$$= -N\cdot\sum_j \frac{c_j}{N}\cdot[\sum_i\cdot\{[a_{ij}/N]/[c_j/N]\}\ \ell og\left[\frac{a_{ij}}{N}/\frac{c_j}{N}\right]$$
$$= -N\cdot\sum_i\cdot\sum_j\cdot\frac{a_{ij}}{N}\ell og\left[\frac{a_{ij}}{N}/\frac{c_j}{N}\right]$$
$$= \sum_j [c_j\,\ell og\,c_j - \sum_i a_{ij}\,\ell og\,a_{ij}]$$
$$N\cdot H(Y) = N\,\ell og\,N - \sum_j c_j\,\ell og\,c_j$$

が成立するから，

$$N\cdot I(X, Y) = N\cdot H(X) - N\cdot H(X/Y)$$

に代入すれば，

$$N\cdot I(X, Y)$$

$$
\begin{aligned}
&= -I_a + I_r + I_c + N \ell og\, N \\
&= (\textstyle\sum_i \sum_j a_{ij})\, \ell og\, (\sum_i \sum_j a_{ij}) - \sum_i r_i\, \ell og\, r_i \\
&\quad - \textstyle\sum_j c_j\, \ell og\, c_j + \sum_i \sum_j a_{ij}\, \ell og\, a_{ij}
\end{aligned}
$$

と計算される。よって，Nが一定であれば，$I(X, Y)$ の代わりに，

$$N \cdot I(X, Y)$$

を伝達情報量といって良い(29)。

ミクロカノニカル集団（に属する粒子から成るパターン φ_k の集合）

$$\Phi = \{\varphi_1, \varphi_2, \cdots, \varphi_M\}$$

に関し，

第 k 番目のパターン φ_k の第 ℓ 特徴粒子数 $n_k(\ell)$

に注目し，

$$n_i(j) \leftrightarrow a_{ij}$$

という対応を考えれば，対応

$$\textstyle\sum_i \sum_j a_{ij} \leftrightarrow \sum_{k=1}^{M} \sum_{\ell \in L} n_k(\ell) = \sum_{k=1}^{M} N_k = N$$

の下での，Φ 内の粒子数の総和　$\sum_{k=1}^{M} N_k$ は一定であるから，伝達情報量

$$
\begin{aligned}
& N \cdot I(X, Y) \\
&= N \cdot H(X) - N \cdot H(X/Y) \\
&= [\textstyle\sum_{k=1}^{M} \sum_{\ell \in L} n_k(\ell)]\, \ell og\, [\sum_{k=1}^{M} \sum_{\ell \in L} n_k(\ell)] \\
&\quad - \textstyle\sum_{k=1}^{M} [\sum_{\ell \in L} n_k(\ell)]\, \ell og\, [\sum_{\ell \in L} n_k(\ell)] \\
&\quad - \textstyle\sum_{\ell \in L} [\sum_{k=1}^{M} n_k(\ell)]\, \ell og\, [\sum_{k=1}^{M} n_k(\ell)] \\
&\quad + \textstyle\sum_{k=1}^{M} \sum_{\ell \in L} n_k(\ell)\, \ell og\, n_k(\ell) \\
&= N \ell og\, N - \textstyle\sum_{k=1}^{M} N_k\, \ell og\, N_k - \sum_{\ell \in L} N(\ell)\, \ell og\, N(\ell) \\
&\quad + \textstyle\sum_{k=1}^{M} \sum_{\ell \in L} n_k(\ell)\, \ell og\, n_k(\ell)
\end{aligned}
$$

ここに，$N_k \equiv \sum_{\ell \in L} n_k(\ell)$

$N(\ell) \equiv \sum_{k=1}^{M} n_k(\ell)$

は，

$X \leftrightarrow$ パターン $\varphi_k (k = 1 \sim M)$

$Y \leftrightarrow$ 第 ℓ 特徴粒子数 $n_k(\ell)$ $(\ell \in \ L)$

という対応の下で，

ミクロカノニカル集団 Φ 内の各パターン φ_k から粒子数 $n_k(\cdot)$ に関し特徴抽出したために，任意の１個のパターンの各種粒子数 $n.(\ell)$ の不確かさに関し取り去られた量（粒子数に関する処理量，an information measure for the number of particles）

と解釈される。

12. 確率変数間の疎遠量，親近量

二つの確率変数X，Yの確率分布間の違いの量，つまり

X，Y間の**親近量**（amount of intimacy）

を情報の量として，第11章での諸記号を使用し表現すれば，次のⅠ～Ⅲの３種類があげられる。

Ⅰ．Shannon の距離（Shannon's distance）$d(X, Y)$

$H(X/Y)$ と同様に，Xが知らされたときのYの条件付エントロピー（conditional entropy）

$$H(Y/X)$$
$$=\sum_x p_1(x)\cdot H(Y/x) \geqq 0$$

ここに，

$$H(Y/x) = -\sum_y p_2(y/x)\ell og\, p_2(y/x) \geqq 0$$

を導入して，定義される非負量

$$d(X, Y) \equiv H(X/Y) + H(Y/X)$$
$$= -\sum_x \sum_y p(x, y)\, \ell og\left|\frac{p(x, y)}{\sqrt{p_1(x)}\cdot\sqrt{p_2(y)}}\right|^2 \geqq 0$$

がシャノンの距離と称せられているものである。

$$H(X/Y) = 0$$

$\Leftrightarrow$ $p_2(y) > 0$ となる y について，ある x に対し，$p_1(x/y) = 1$

が成立しているから，

$$d(X, Y) = 0$$

となるのは，次のⅰ，ⅱが共に成立する場合に限る：

(ⅰ) $p_2(y) > 0$ となる y について，
ある x に対し，$p_1(x/y) = 1$

(ⅱ) $p_1(x) > 0$となる x について，
ある y に対し，$p_2(y/x) = 1$。 □

なお，自己情報量（amount of self-information）$-\ell og\, p_1(x)$，$-\ell og\, p_2(x)$ に対し，

$$-\ell og\, p_1(x/y), -\ell og\, p_2(y/x)$$

は条件付自己情報量（amount of conditional self-information）と呼ばれている。

Ⅱ．shannon の相互情報量（Shannon's average mutual information）$I(X, Y)$

これは第11章で説明済のものであり，

$$I(X, Y)$$
$$\equiv H(X) - H(X/Y)$$
$$\equiv H(Y) - H(Y/X)$$

$$= + \sum_x \sum_y p(x,y)\ \ell og \frac{p(x,y)}{p_1(x)\cdot p_2(y)} \geqq 0$$

と表現せられているものである。

$I(X, Y) = 0$

$\Leftrightarrow \quad \forall x, \forall y, p(x,y) = p_1(x)\cdot p_2(y)$ （X, Y は統計的に独立）

が成立している。

Ⅲ．S. Kullback の逸脱量の対称化（symmetrized Kullback's divergence）$D(X, Y)$

二つの確率変数 X, Y のとる値が同一の集合に属し，確率分布 $p_1(z), p_2(z)$ が異なる場合

$D_1(X, Y)$

$\equiv + \sum_Z p_1(z)\ \ell og \frac{p_1(z)}{p_2(z)}$

$= - \sum_Z p_1(z)\ \ell og\, p_2(z) - [- \sum_Z p_1(z)\ \ell og\, p_1(z)] \geqq 0$

と定義される非負量（補助定理２．１を参照）は

S. Kullback の**逸脱量**（divergence）あるいは判別関数（discriminant fumction）として知られめており[30]，

p_2 の，p_1 からのへだたり

として解釈される。

$D_1(X,) = 0 \Leftrightarrow \forall z, p_1(z) = p_2(z)$

が成立している。X, Y を入れ換えた

$D_2(Y, X)$

$= + \sum_Z p_2(z)\ \ell og \frac{p_2(z)}{p_1(z)}$

をも導入し，X, Y に関し対称化された量[31]

$D(X, Y)$

$= D_1(X, Y) + D_2(Y, X)$

$= + \sum_Z [p_1(z) - p_2(z)] \cdot \ell og\, [p_1(z)/p_2(z)]$

に関し，

$D(X, Y) = 0$

$\Leftrightarrow \quad \forall z, p_1(z) = p_2(z)$

が成立し，X, Y 間の親近量として採用できることがわかる。 □

さて，著者は，上述のⅡでの $I(X, Y)$ とは異なり，

X, Y が統計的に独立であればある程，大きい値をとる**疎遠量**（amount of estrangement）として，次の $SI(X, Y)$ を提案する。

Ⅳ. S. Suzukiの提案する情報疎遠量 $SI(X, Y)$

一変数 u の threshold function

$$psn(n) = 1 \quad \text{if} \quad u \geqq 0, = 0 \quad \text{if} \quad u < 0$$

を導入して,

$$SI(X, Y)$$
$$= -\frac{1}{2} \Sigma_x \Sigma_y psn(p_1(x) \cdot p_2(y) - p(x, y)) \cdot p(x, y)$$
$$\cdot \ell og \left| 1 - \frac{p(x, y)}{p_1(x) \cdot p_2(y)} \right|^2 \geqq 0.$$

一般に,

$$0 \leqq p(x, y)/[p_1(x) \cdot p_2(y)]$$
$$= p_1(x/y)/p_1(x)$$

であるが, X, Yが統計的に独立であればある程,

$$|1 - p(x, y)/[p_1(x) \cdot p_2(y)|^2$$

は小となるから, Ⅱでの $I(X, Y)$ 内の主要項

$$\ell og \frac{p(x, y)}{p_1(x) \cdot p_2(x)} = \ell og \left[1 - \left\{1 - \frac{p(x, y)}{p_1(x) \cdot p_2(y)}\right\}\right]$$

に注目し,

$$\ell og \left[1 - \frac{p(x, y)}{p_1(x) \cdot p_2(y)}\right]$$

を取り出し,

$$p_1(x) \cdot p_2(y) \geqq p(x, y) > 0$$
$$\Leftrightarrow \quad 1 > 1 - \frac{p(x, y)}{p_1(x) \cdot p_2(y)} \geqq 0$$

を考慮し, この取り出した量を

$$p_1(x) \cdot p_2(y) \geqq p(x, x)$$

を満たす x, y について結合確率分布 $p(x, y)$ で平均化したものが $SI(X, Y)$ である。

Shannon 相互情報量 $I(X, Y)$ は

X, Yが統計的に独立であればあるほど小さい値をとる量, あるいは X が Yに関する情報をどの程度含んでいるかを示す量（**Yに含まれる X の情報の量**; the amount of information about the random variable X contained in the variable Y)

であるのに対し, $SI(X, Y)$ は

X, Yが統計的に独立であればある程大きい値をとる量（XがYに関する情報をどの程度含んでいないかを示す量）

である。なお，X, Yが確率密度$p_1(x)$, $p_2(y)$をもち，その結合確率密度が $p(x,y)$である場合は，その微分量（differential amount）として，

$$SI(X,Y)$$
$$= -\frac{1}{2}\int dx \int dy\, p(x,y)\cdot psn(p_1(x)\cdot p_2(y)-p(x,y))$$
$$\ell og\left|1-\frac{p(x,y)}{p_1(x)\cdot p_2(x)}\right|^2$$

なるごとく定義されねばならない。 □

$SI(X,Y)$の応用として，X, Yが確率変数でなく，二つのパターンφ, ϕの場合には例えば，次の様に考えれば良い。

二つのパターンφ, ϕは共にHilbert空間$\mathfrak{H}=L_2(M;dm)$　の元とする。内積(φ, ϕ)は，Mをn次元ユークリッド空間R^nの可測部分集合として，

$$(\varphi,\phi)=\int_M dm(z)\,\varphi(z)\cdot\bar{\phi}(z),\quad {}^{-}\text{は複素共役の意}$$

と定義され，φのノルム$\|\varphi\|$は

$$\|\varphi\|=\sqrt{(\varphi,\varphi)}$$

である。$dm(z)$はM上でのLebesgue-Stieltjes式測度である。

$$\eta_1(z)\equiv\frac{\varphi(z)\cdot\phi(z)}{\sup_z|\varphi(z)\cdot\phi(z)|}$$

とおくと，不等式

$$|\eta_1(z)|\leqq 1$$

が成立する。点$z\in M$において

$$\varphi(z)\cdot\phi(z)=\sup_z|\varphi(z)\cdot\phi(z)|$$

が満たされる確率は，つまり

$\eta_1(z)=+1$が満たされる確率

は

$$\eta_2(z)\equiv\frac{\eta_1(z)-(-1)}{+1-(-1)}=\frac{1}{2}\cdot[\eta_1(z)+1]$$

である。よって，

$$(T\varphi)(z)=\varphi(z)/\sup_z|\varphi(z)|$$

$$(F\varphi)(z) = \frac{1}{2}\cdot[(T\varphi)(z) + 1]$$

として定義される非負量

$$pI(\varphi\,;\phi)$$

$$= -\frac{1}{2}\cdot\int_M dm(z)$$

$$\cdot\,\mathrm{psn}((F\varphi)(z)\cdot(F\phi)(z) - (F\varphi\phi)(z))$$

$$\cdot\,\ell og\left|1 - \frac{(F\varphi\phi)(z)}{(F\varphi)(z)\cdot(F\phi)(z)}\right|^2 \geqq 0$$

については次の解釈が可能である：

$pI(\varphi\,;\phi)$ は，二つのパターン φ，ϕ の振幅のピーク（peak）を与える二つの座標点が一致する座標点 z の集合

$$\{z\in M\,|\,(T\varphi)(z) = (T\phi)(z) = +1 \wedge (F\varphi)(z)\cdot(F\phi)(z) \geqq (F\varphi\phi)(z)\}$$

の測度に対応する情報量であり，

φ，ϕ の**ピーク一致情報量**（amount of information about the degree of that two patterns φ and ϕ coincide in the peak of amplitudes）

と称せられてよい。

13. 情報密度とモデル構成作用素

Gelomb によれば，Shannon entropy function

$$h(x,1-x)$$
$$\equiv -x\,\ell og\,x - (1-x)\,\ell og\,(1-x),\ 0\leqq x\leqq 1$$

ついて，

$$\int_x^z du\,\ell og\frac{u}{1-u} = h(x,1-x) - h(z,1-z)$$

が成立することが示されており[32]，尤度比の対数（likelihood ratio）

$$\ell og\,[u/(1-u)]$$

は<u>情報量の密度</u>（Information density）と解釈されてよい。更に，

$u_i = p_1(a_i)$：第 i 番目の送信シンボル a_i の生起確率 $(i=1,2)$

$v_j = p_1(a_i/b_j)$：受け取ったシンボル y が　$y=b_j$（第 j 番目の出力シンボル）であるとき，送信したシンボル x が　$x=a_i$　であることの条件付確率 $(i,j=1,2)$

とすると，

入力シンボル a_1，a_2 の集合　$A=\{a_1,a_2\}$ 自身が事前に持っている不確定さ　$H(A)$

A 内のいずれか一つを送信し，第 j 番目のシンボル b_j を受け取った後でも A 自身について残存している不確定さ　$H(A/b_j)$
について

$$H(A) = h(u_1, 1-u_1)$$
$$H(A/b_j) = h(v_1, 1-v_1)$$

という表現が成り立っているが，

$$\int_{u_1}^{v_j} du\, \ell og \frac{u}{1-u} = H(A) - H(A/b_j)$$

を得，$A, B = \{b_1, b_2\}$ 間の相互情報量

$$I(A;B)$$
$$= \Sigma_{i=1}^{2} \Sigma_{j=1}^{2} p(a_i, b_j)\, \ell og \frac{p(a_i, b_j)}{p_1(a_i) \cdot p_2(b_j)}$$

は

$$\Sigma_{i=1}^{2} p_2(b_j) \cdot \int_{u_1}^{v_j} du\, \ell og \frac{u}{1-u} = I(A;B)$$

と表現され，まさに，関数　$\ell og[u/(1-u)]$ は情報密度である感を深くさせられる。ここに，
$p_2(b_j)$：第 j 番目の受信シンボル b_j の生起確率 $(j = 1, 2)$
$p(a_i, b_j)$：a_i と b_j との結合確率
であり，
B を受け取った後でも残存している A 自身の不確定さ　$H(A/B)$
を導入し，

$$I(A;B) = H(A) - H(A/B)$$ □

上記の情報密度に hint を得て，Hilbert 空間 $\mathfrak{H} = L_2(M; dm)$（第12章を参照）の元としての実数値パターン　$\varphi = \varphi(x)$ について
振幅の peak，bottom に関する情報密度について検討してみよう。

$$s = \sup_x \varphi(x),\ i = inf_x \varphi(x)$$

とおくと，
点 $x \in M$ において $\varphi(x) = s$ が満たされる確率 $p(\varphi(x) = s) \equiv p_s(x)$
点 $x \in M$ において，$\varphi(x) = i$ が満たされる確率 $p(\varphi(x) = i) \equiv p_i(x)$
は各々，

$$p_s(x) = [\varphi(x) - i]/[s - i] \geqq 0$$
$$p_i(x) = [\varphi - i(x)]/[s - i] \geqq 0$$

と表現されることがわかる[33]。この設定が好都合であることは，

$$\forall x, p_s(x) + p_i(x) = 1$$

が成り立っており，然も点 $x \in M$ での振幅の期待値

$$\begin{aligned}&\text{Expec}\,[\varphi(x)]\\ &\equiv s \cdot p_s(x) + i \cdot p_i(x)\end{aligned}$$

を計算してみると，

$$\text{Expec}\,[\varphi(x)] = \varphi(x)$$

であることから分る。

点 $x \in M$ での，パターン φ の振幅の peak，bottom に関する不確かさに関するエントロピー密度 $I(\varphi\,;x)$ は

$$\begin{aligned}I(\varphi\,;x) &= h(p_s(x), 1 - p_s(x))\\ &= -\,p_s(x)\,\ell og\, p_s(x) - p_i(x)\,\ell og\, p_i(x)\end{aligned}$$

と定義されてよく，これは具体的に

$$\begin{aligned}&I(\varphi\,;x)\\ &= \ell og\,[s - i] - \frac{1}{[s-i]} \cdot [\{\varphi(x) - i\}\,\ell og\{\varphi(x) - i\} + \{s - \varphi(x)\}\,\ell og\{s - \varphi(x)\}]\end{aligned}$$

と計算される。よって，パターン φ の振幅のエントロピー $I(\varphi)$ は規格化積分の形で

$$I(\varphi) = \int_M dm(x) \cdot I(\varphi\,;x) \Big/ \int_M dm(z)$$

と定義でき，

$$\begin{aligned}I(\varphi) = \ell og[s - i] - &\left[\int_M dm(z)\right]^{-1}\\ &[s - i]^{-1} \cdot \left[\int_M dm(x) \cdot \{\varphi(x) - i\}\,\ell og\{\varphi(x) - i\}\right.\\ &\left. + \int_M dm(x) \cdot \{s - \varphi(x)\}\,\ell og\{s - \varphi(x)\}\right]\end{aligned}$$

と再表現される。

さて，S. Suzuki のパターン認識の数学的理論[34]によれば，下記の定理13. 1の3条件(i)，(ii)，(iv)と今一つの巾等性

$$(\text{iii})_2 \quad \forall\, \varphi \in \Phi, TT\varphi = T\varphi$$

とを満たす写像 $T : \Phi \to \Phi$ を適用して得られるパターン $\varphi \in \Phi$の像 $T\varphi \in \Phi$ は φ の構造モデル（structural model）と呼ばれる。写像 $T : \Phi \to \Phi$ はモデル構成作用素（model construction operator）といわれるが，パターン情報システムは，入力パターン φ を受け入れた後，モデル $T\varphi$ に変換するものとされている。

下の定理13．１は，

$$(T\varphi)\,(x) = \frac{\varphi(x)}{\sup_x|\varphi(x)|}\cdot p_s(x) \qquad (13.\ 1)$$

と定義される写像　$T:\Phi\to\Phi$ について

$\varphi(x)\in\{0,1\}$　であれば
(a) $\forall x, \varphi(x)=0$　のとき，
$(T\varphi)\,(x)=0=\varphi(x)$
(b) $\exists x, \varphi(x)\neq 0$　のとき
$(T\varphi)\,(x)=[\varphi(x)]^2=\varphi(x)$

がいえ，結局

$\forall x\in M, (T\varphi)\,(x)=\varphi(x)$

が成立し，(iii)$_2$ が自明的に成り立っている。
なお，$c>0$ なる定数 c の下で

$(T_1\varphi)\,(x)=c\cdot p_s(x)$
$(T_2\varphi)\,(x)=c\cdot\varphi(x)/\sup_x|\varphi(x)|$

と定義される写像 T_1，T_2 は共に，４条件(i)，(ii)，(iii)$_2$，(iv)を満たし，モデル構成作用素の一種であることが証明されている（文献（34）第25部の定理２．１，定理３．１を参照)。
［定理13．１］（モデル構成可能定理）
式（13．１）の様に定義された写像　$T:\Phi\to\Phi$　は，次の(i)，(ii)，(iii)$_1$，(vii) を満たす。ただし，$T\varphi$ 内の分数計算において　$0/0=0$ と約束する：

(i)　$\varphi=0\in\Phi$ に対し　$T\varphi=\varphi$
(ii)　任意の正定数 a に対し，

$\forall\varphi\in\Phi, T(a\varphi)=T\varphi$

(iii)$_1$　$T\varphi=0$　に対しては

$TT\varphi=T\varphi$

であり，$\sup_x\varphi(x)>\mathit{inf}_x\,\varphi(x)\geqq 0$ であれば

$(TT\varphi)\,(x)=[T\varphi(x)]^2$
(iv)　$T\varphi\neq 0$　を満たす $\varphi\in\Phi$ が存在する。
(証明)　<u>ｉの成立</u>：　$\varphi=0$ に対しては，$\sup_x\varphi(x)=\mathit{inf}_x\,\varphi(x)=0$　を得，
$\varphi(x)/\sup_x|\varphi(x)|=0$，　$p_s(x)=0$
$\therefore$　$(T\varphi)\,x)=0$

<u>iiの成立</u>：$\sup_x[a\,\varphi(x)]=a\cdot\sup_x\varphi(x), \mathit{inf}_x[a\,\varphi(x)]=a\cdot\mathit{inf}_x\,\varphi(x)$　であるから，

$a\varphi(x)/\sup_x|a\,\varphi(x)| = \varphi(x)/\sup_x|\varphi(x)|$
$p((a\varphi)(x) = \sup_x[a\varphi)(x)])$
$= p(\varphi(x) = \sup_x[\varphi(x)])$

を得て,

$(Ta\varphi)(x) = (T\varphi)(x).$
ivの成立:
(イ) $c_1 > 0 > c_2 \wedge c_2 \geqq |c_2|$ であり,
$\varphi(x_1) = c_1, \varphi(x_2) = c_2$
(ロ) $\varphi(x) = 0$ if $x \neq c_1 \wedge x \neq c_2$
とすれば,

$\sup_x \varphi(x) = c_1$, $inf_x\,\varphi(x) = c_2$
であるから,

$$(T\varphi)(x) = \frac{\varphi(x)}{c_1}\cdot\frac{\varphi(x)-c_2}{c_1-c_2}$$

と表現され,

$(T\varphi)(x_1) = 1 \wedge (T\varphi)(x_2) = 0$

(iii)$_1$ の成立:

$$\eta(x) = (T\varphi)(x)$$
$$= \frac{\varphi(x)}{\sup|\varphi(x)|}\cdot\frac{\varphi(x)-inf\,\varphi(x)}{\sup\varphi(x)-inf\,\varphi(x)}$$

とおく。
iii—1 $\eta = 0$ の場合
i より $T\eta = 0$ を得て,

$TT\varphi = 0 = 0\cdot 0 = \eta\cdot\eta = T\varphi\cdot T\varphi$

iii—2 $\eta \neq 0$ の場合
$\varphi = 0$ とすると, $\eta = T\varphi = 0$ となって矛盾するから, $\varphi \neq 0$ である。

$\therefore$ $\sup|\varphi(x)| \neq 0.$

このとき,

$$inf\,\eta(x) = \frac{inf\varphi(x)}{\sup|\varphi(x)|}\cdot\frac{inf\,\varphi(x)-inf\,\varphi(x)}{\sup\varphi(x)-inf\,\varphi(x)} = 0$$

であるが,
$\sup\varphi(x) = 0$, $\neq 0$ のいずれの場合でも,

$$\sup|\eta(x)| = \frac{\sup|\varphi(x)|}{\sup|\varphi(x)|}\cdot\frac{\sup\varphi(x) - \mathit{inf}\,\varphi(x)}{\sup\varphi(x) - \mathit{inf}\,\varphi(x)} = 1$$

が成り立つ。

$$\sup\eta(x) = \frac{\sup\varphi(x)}{\sup|\varphi(x)|}\cdot\frac{\sup\varphi(x) - \mathit{inf}\,\varphi(x)}{\sup\varphi(x) - \mathit{inf}\,\varphi(x)}$$

$$= \sup\varphi(x)/\sup|\varphi(x)| \in \{+1, -1\}$$

であるが，$\mathit{inf}\,\eta(x) = 0$　より，結局

$$\sup\eta(x) = 1$$

を得て，

$$(T\eta)\ (x)\ \frac{\eta(x)}{\sup|\eta(x)|}\cdot\frac{\eta(x) - \mathit{inf}\,\eta(x)}{\sup\eta(x) - \mathit{inf}\,\eta(x)}$$

$$= \frac{n(x)}{1}\cdot\frac{\eta(x) - 0}{1-0} = \eta(x)\cdot\eta(x)$$

$$= [(T\varphi)\ (x)]\cdot[(T\varphi)\ (x)]$$ □

14. 特徴量の変動エントロピー，edge-entropy

M 個のパターンから成る場合

$$\Phi = \{\varphi_1, \varphi_2, \cdots, \varphi_M\}$$

を考え，第 k 番目のパターン φ_k から抽出された第 $\ell \in L$ 番目の非負特徴量

$$u(\varphi_k, \ell) \quad \geqq 0$$

に関し，固定した　$\Delta u > 0$　の下での不等式

$$n_k(\ell)\cdot\Delta u \leqq u(\varphi_k, \ell)$$
$$< [n_k(\ell) + 1]\cdot\Delta u$$

の解

$$n_k(\ell) \in \{0, 1, 2, \cdots\}$$

を導入する。ここに，

$$u : \Phi\times L \to \{r\,|\,0 \leqq r\}$$

は特徴量抽出写像である。

$$N_k \equiv \textstyle\sum_{\ell\in L} n_k(\ell),\ k = 1 \sim M$$
$$N(\ell) \equiv \textstyle\sum_{k=1}^{M} n_k(\ell),\ \ell \in L$$

とすると，

$$N \equiv \sum_{k=1}^{M} \sum_{\ell \in L} n_k(\ell) = \sum_{k=1}^{M} N_k = \sum_{\ell \in L} N(\ell)$$

が成立している（第6，9，11章を参照）。

$N_k, N(\ell), N$ は各々

パターン φ_k に含まれている粒子の総数

第 ℓ 特徴に関する Φ 内の粒子の総数

パターン集合 Φ 内の粒子の総数

であり，Φ がミクロカノニカル集団であれば

$$\forall_k (=1 \sim M), N_k = 一定, \quad \therefore \quad N = M \cdot N_k$$

が成立している。

I．　変動エントロピー $VH(\Phi)$

本節での変動エントロピー　$VH(\Phi)$　は　$1 \cdot \Delta u \leqq u(\varphi_k, \ell) < 2 \cdot \Delta u$　のときパターン φ_k は第 $\ell \in L$ 番目の特徴量を持っているといい，$0 \cdot \Delta u \leqq u(\varphi_k, \ell) < 1 \cdot \Delta u$　のとき持っていないという解釈の下でのみ，つまり，

$$n_k(\varphi, \ell) \in \{0, 1\}$$

の場合のみ定義されると考えてよい。

統計力学では，

一つの量子状態にある粒子の個数が0または1であるような粒子集団はフェルミ・デイラックの統計に従う[26]

ということに対応して（9章を参照）

$n_k(\ell) \in \{0, 1\}$ であるようなパターン集合 Φ

をフェルミ・ディラック集団ということにする。$n_k(\ell) \in \{0, 1\}$ であるから，

$N(\ell)$ は第 $\ell \in L$ 特徴をもつパターン φ_k の総数

という直観的解釈が許される。このとき，

$$b \equiv N/M = \sum_{\ell \in L} N(\ell)/M$$

は，1パターン当たりの粒子の平均存在個数である。また，各 φ_k 内に少なくとも1個の粒子が存在するものとすれば，つまり，

$$\forall_k (= 1 \sim M), 1 \leqq N_k = \sum_{\ell \in L} n_k(\ell)$$

と仮定すれば，

$$q_\ell \equiv M/N = 1/b$$

について

$$0 \leqq q_\ell = M / \sum_{\ell \in L} N(\ell)$$
$$\leqq \sum_{k=1}^{M} 1 / \sum_{k=1}^{M} N_k \quad \leqq 1 \quad \therefore \quad 1 \leqq \mathrm{b} \wedge \sum_{\ell \in L} \frac{N(\ell)}{M} \geqq 1$$

が得られる。

パターン総数 M は一定であるが，特徴抽出写像 u も固定して，粒子総数 N も一定と考えよう。よって，b は一定である。なお，$\# B$ は集合 B 内の要素の総数という記法を採用すれば，$\# L$ は特徴軸の総数であり，

$$L^{+} \equiv \{\ell \in L \mid N(\ell) > 0\}$$

とすると，$\# L^{+}$ は少なくとも一個の粒子が存在する様な特徴軸の総数である。

さて，パターン集合 Φ の全エントロピー（total entropy）

$$H(\Phi) = -\sum_{\ell \in L} \frac{N(\ell)}{N} \log \frac{N(\ell)}{N}$$

に対し定義される

$$VH(\Phi) = -\sum_{\ell \in L} \frac{N(\ell)}{N} \log \frac{N(\ell)}{N}$$

$$= -\frac{1}{b} \sum_{\ell \in L} \frac{N(\ell)}{M} \log \frac{N(\ell)}{M}$$

を，Φ の**変動エントロピー**（variation entropy）ということにする。なお，文献（36），（37）では，この種の同様な変動エントロピーは特徴抽出写像 u を導入せずに，

$$\varphi_k(x) \in \{0, 1\}$$

なる各パターン φ_k に対し直接定義されている。

両エントロピーの間に，次の命題14．1のような関係がある。

〔命題14．1〕

$$H(\Phi) = \log b + VH(\Phi) \qquad \square$$

それは，$N = b \cdot M$ であるから，

$$H(\Phi) = -\sum_{\ell \in L} [N(\ell)/N] \log [N(\ell)/\{bM\}]$$

と書き直され，これから明らかである。

〔命題14．2〕　不等式

$$0 \leqq H(\Phi) \leqq \log \# L^{+}$$

が成立し，

(i)　$H(\Phi) = 0 \Leftrightarrow \exists\, \ell \in L^{+}, N(\ell) = N.$

(ii)　$H(\Phi) = \log \# L^{+}$

$\Leftrightarrow \forall\, \ell \in L^{+}, N(\ell) = N / \# L^{+}.$ $\square$

〔命題14.　3〕(変動エントロピー定理)

M, Nは共に一定，よってbが一定とならば，

(i)　$VH(\Phi) = 0 \Leftarrow$

$$\forall\, \ell \in L, N(\ell) = 0 \vee N(\ell) = M$$

(ii)　分布　$\{N(\ell) \mid \ell \in L\}$　を変動させて得られる$VH(\Phi)$の最大値$\max VH(\Phi)$，最小値$VH(\Phi)$については，

(イ)　$\forall\, \ell \in L^+, N(\ell)/N = 1/\#L^+$

$\Leftrightarrow$　$\max VH(\Phi) = -\ell og\, b + \ell og\, \#L^+$,

(ロ)　$\exists\, \ell \in L^+, N(\ell)/N = 1$

$\Leftrightarrow$　$\min VH(\Phi) = -\ell og\, b$

(証明)

iの証明：補助定理11.　1より関数　$f(x) = -x \ell og\, x$　は　$x \in \{0, 1\}$　の場合のみ0となることから明らかである。

iiの証明：

$$VH(\Phi) = -\Sigma_{\ell \in L} \frac{N(\ell)}{N} \ell og \frac{N(\ell)}{M}$$

$$= -\Sigma_{\ell \in L^+} \frac{N(\ell)}{N} \ell og \left[\frac{N(\ell)}{N} \cdot \frac{N}{M}\right]$$

$$= -\ell og \frac{N}{M} - \Sigma_{\ell \in L^+} \frac{N(\ell)}{N} \ell og \frac{N(\ell)}{N}$$

と変形されることに注意する。よって，

$M =$ 一定　かつ　$N =$ 一定

の下では，分布　$\{N(\ell) \mid \ell \in \in L\}$　を変動させて得られる$VH(\Phi)$の最大値は，

$\forall\, \ell \in L^+, N(\ell)/N = 1/\#L^+$

のとき生じ，

$$\max VH(\Phi = -\ell og \frac{N}{M} - \Sigma_{\ell \in L^+} \frac{1}{\#L^+} \ell og \frac{1}{\#L^+}$$

$$= -\ell og \frac{N}{M \cdot \#L^+} = -\ell og\, b / \#L^+.$$

であり，最小値は

$\exists\, \ell \in L^+, N(\ell)/N = 1$

のとき生じ，

$$\min VH(\Phi) = -\ell og\frac{N}{M} = -\ell og\, b. \qquad \square$$

上述の命題14. 3のｉについて検討しよう。

$N(\ell) = \sum_{k=1}^{M} n_k(\ell)$ であるから，

$$H(\Phi) = \ell og\, b$$
$$\Leftrightarrow \quad VH(\Phi) = 0$$
$$\Leftrightarrow \quad \forall \ell \in L, N(\ell) = 0 \vee N(\ell) = M$$
$$\left(\Leftrightarrow \quad \forall \ell \in L, N(\ell) = 0 \vee \frac{N(\ell)}{N} = \frac{M}{b \cdot M} = \frac{1}{b}\right)$$
$$\Leftrightarrow \quad \forall \ell \in L, [\forall k(=1 \sim M), n_k(\ell) = 0] \vee [\forall k(=1 \sim M), n_k(\ell) = 1]$$
$$\Leftrightarrow \quad \forall \ell \in L^+, \forall k(=1 \sim M), n_k(\ell) = 1$$
$$\Leftrightarrow \quad \forall \ell \in L, \forall k(=1 \sim M), u(\varphi_k, \ell) = 0 \vee [\Delta u \leqq u(\varphi, k) < 2 \cdot \Delta u]$$

を得て，

各パターン φ_k からどんな $\ell \in L$ についても同一特徴量が抽出された場合，

変動エントロピー $VH(\Phi) = 0$ かつ全エントロピー $H(\Phi) = \ell og\, b$

となる。この事実から主張できることは次の通りである：

$\ell og\, b$ は Φ 内の各パターン φ_k に共通な特徴に関する情報量

と解釈され，更に重要なことは，

$VH(\Phi)$ は，各パターン φ_k から同一特徴軸について相異なる特徴量が抽出される程度，つまり抽出される特徴量の変動の程度

を表わしており，Φ 内の各パターン φ_k が共に同一カテゴリに属するならば，この変動エントロピー $VH(\Phi)$ が小となる様な特徴抽出写像

$$u : \Phi \times L \to \{r \mid 0 \leqq r\}$$

を選ぶことが望ましいといえる。

なお，上述の論は，$n_k(\ell) \in \{0, 1, 2,, \cdots\}$ に対しても３命題14. 1～14. 3はそのまま成立するが，この場合は，$VH(\Phi)$ のもっている変動性の計量化機能はうまく生かされなくなる場合も生じることに注意しておこう。

Ⅱ. edge-entropy $EDH(\Phi)$

パターンのエッジとは，振幅の険しい変化をもたらす座標点の集合である。文献 (38)，(39) によれば，the Laplacian and the gradient operations などは振幅の変化の程度（the rate of change of amplitude）のみに対応する検出法であるが，entropy operator は局所座標領域の平均的な振幅（the average amplitude in a local region）にも反応する演算であるという。本節では，この entropy of erator に hint を得て，特徴抽出写像

$$u : \Phi \times L \to \{j \mid 0 \leqq r\}$$

を導入し，Φ から抽出される特徴量の集合上の edge を検出する方法を研究しよう。

統計力学では

一つの量子状態に粒子の個数が任意の非負整数であるような粒子集団はボーズ・アインシュタインの統計に従う[26]

ということに対応して（9章も参照)

$n_k(\ell) \in \{0, 1, 2, \cdots\}$ となる如きパターン集合 Φ をボーズ・アインシュタイン集団ということにし，本節ではこの集団につき論じよう。

Φ の粒子数分布　$\{N(\ell) \mid \ell \in L\}$　について

$$v_\ell = N(\ell) \quad \text{if} \quad \ell \in L, \ = 0 \quad \text{if} \quad \ell \notin L$$

として，

$$H(v_\ell) \equiv - \textstyle\sum_{m=-w}^{+w} p[\ell + m]\, \ell og\, p[\ell + m]$$

ここに，

$$P[\ell + m] = v_{\ell+m} / \textstyle\sum_{n=-w}^{+m} v_{\ell+m}$$

を求める。ここに，$w = 3$ 位が適当である。

$\ell \in L$ を動かして得る集合

$$\{H(v_\ell)\ \ell \in L\}$$

の極小値を与える特徴軸番号 $\ell \in L$，つまり

$$\arg\min_{\ell \in L} H(v_\ell)$$

が Φ から定まる粒子数分布　$\{N(\ell) \mid \ell \in L\}$　の edge に相当する。この様な edge は w を $w = 3$ と選んでいる場合

(i)　$v_{\ell-1} > v_\ell \fallingdotseq v_{\ell+1} = v_{\ell+2} = v_{\ell+3}$

(ii)　$v_{\ell-3} \fallingdotseq v_{\ell-2} = v_{\ell-1} = v_\ell < v_{\ell+1}$

を満たす添字 ℓ の集合であると期待される。このとき，Φ の edge-entropy $EDH(\Phi)$　を

$$EDH(\Phi) = \textstyle\sum_{\ell \in L} H(v_\ell)$$

と定義できよう。

15. 相互情報量の線形近似としての相関係数

二つの実数値パターン φ，η 間の内積 (φ, η) と，φ，η のノルム $\|\varphi\| = [(\varphi, \varphi)]^{1/2}$, $\|\eta\|$ とを導入すると，

$$-1 \leqq cor(\varphi, \eta) \equiv \frac{(\varphi, \eta)}{\|\varphi\| \cdot \|\eta\|} \leqq +1$$

が Schwarz の不等式から成立し，

$\varphi \neq 0 \wedge \eta \neq 0$　なる条件の下で，

(i)〔ある負の定数 a が存在して，$\varphi = a\eta$〕$\Leftrightarrow$ $cor(\varphi, \eta) = -1$

(ii)〔φ が η と直交する，つまり $(\varphi, \eta) = 0$〕

$\Leftrightarrow$ $cor(\varphi, \eta) = 0$

(iii)〔ある正の定数 b が存在して，$\varphi = b\eta$〕$\Leftrightarrow$ $cor(\varphi, \eta) = +1$

が成り立つ。φ と η との間の線形性の程度を計量化する機能をもった，この様な $cor(\varphi, \eta)$ は

相関係数（correlation coefficient）

と呼ばれるものの典型である。

本章では，第12章で説明されたShannonの相互情報量を近似し，確率変数間の新しい相関係数を提案する。勿論，二つの確率変数 x, y の結合確率 $p(x, y)$ を導入して，内積 (φ, η) を

$$(\varphi, \eta) = \sum_x \sum_y p(x, y) \cdot \varphi(x, y) \cdot \eta(x, y)$$

と設定すれば，上述の $cor(\varphi, \eta)$ も相関係数の一種となるが，本章で提案されるのはこれとは別のものである。

x, y の各々の生起確率 $p_1(x), p_2(y)$

x, y の同時生起確率 $p(x, y)$

x, y の条件付生起確率 $p_1(x/y) = p(x, y)/p_2(y)$,

$$p_2(y/x) = p(x, y)/p_1(x)$$

を導入し，H. Yamamoto は probabilistic labeling process（確率的ラベル付け過程）[41] での stochastic relaxation operator の構成に必要な compatibility coefficient[28]（両立係数，適合係数）として，次の命題15. 1と同様な相関係数を提案した[40]。

相互情報量は $\ell og[p(x, y)/\{p_1(x) \cdot p_2(y)\}]$ を $p(x, y)$ で平均化したものであるが，不等式

$$\ell og\, x \leqq x - 1 \quad \text{if} \quad x > 0$$

を適用すれば，

$$\ell og \frac{p(x, y)}{p_1(x) \cdot p_2(y)} \leqq \frac{p(x, y)}{p_1(x) \cdot p_2(y)} - 1 \tag{15. 1}$$

と線形近似評価可能な事実が次の命題15. 1を導き出すhintとなる。なお，

$$\frac{p(x, y)}{p_1(x) \cdot p_2(y)} = \frac{p_1(x/y)}{p_1(x)} = \frac{p_2(y/x)}{p_2(y)}$$

であるから，次の約束を計算上しておく：

$$\frac{p(x, y)}{p_1(x) \cdot p_2(y)} = 0 \quad \text{if} \quad p_1(x) = 0 \vee p_2(y) = 0. \qquad \square$$

〔命題15. 1〕$I(x \leftarrow y)$ を

(a) $p_1(x) \cdot p_2(y) < p(x, y) \wedge p_1(x) \neq 1$ の場合

$$I(x \leftarrow y) = \frac{1}{[1 - p_1(x)]} \cdot \left[1 - \frac{p_1(x) \cdot p_2(y)}{p(x, y)}\right] > 0$$

(b)　$p_1(x)\cdot p_2(y) \geqq p(x,y) \vee p_1(x)=1$ の場合

$$I(x \leftarrow y) = \frac{p_1(x)}{p_1(x)\cdot p_2(y)} - 1$$

と定義すると，不等式

$$-1 \leqq I(x \leftarrow y) \leqq +1$$

が成立し，

(i)　$I(x \leftarrow y) = +1 \Leftrightarrow p_1(x/y) = 1$

(ii)　$I(x \leftarrow y) = 0 \Leftrightarrow p(x,y) = p_1(x)\cdot p_2(y) \vee p_1(x) = 1$

(iii)　$I(x \leftarrow y) = -1$

$\Leftrightarrow\ p(x,y) = 0 \wedge [p_1(x) > 0 \vee p_2(y) > 0]$

が成立する。なお，iiiから，次のivの成立は明らかである。

(iv)　$p_1(x) = 0 \vee p_2(y) = 0$

$\Rightarrow p(x,y) = 0 \Leftrightarrow I(x \leftarrow y) = -1.$

(証明)

i の証明：$p_1(x) \neq 1$ とすれば

$$0 < p_1(x/y) \leqq 1 \Leftrightarrow -\frac{p_1(x)}{p_1(x/y)} \leqq -p_1(x)$$

$$\Leftrightarrow \frac{1}{1-p_1(x)}\cdot\left[1-\frac{p_1(x)}{p_1(x/y)}\right] \leqq 1$$

より明然。

ii，iiiの証明：

$$\begin{aligned} & p_1(x) = 1 \Rightarrow p_2(y/x) = p_2(y) \\ & \Leftrightarrow p(x,y) = p_1(x)\cdot p_2(y) \\ & \Leftrightarrow I(x \rightarrow y) = 0 \end{aligned} \tag{15. 2}$$

に注意すれば，

$$\begin{aligned} & p(x,y) \leqq p_1(x)\cdot p_2(y) \\ & \Leftrightarrow p(x,y)/[p_1(x)\cdot p_2(y)] - 1 \leqq 0 \end{aligned}$$

より明然。 □

ここで，$I(x \leftarrow y)$ を結合確率 $p(x,y)$ で平均化して得られる

$$\begin{aligned} & I(X \leftarrow Y) \\ & \equiv \textstyle\sum_x \sum_y p(x,y)\cdot I(x \leftarrow y) \\ & = \textstyle\sum_y p_2(y) \sum_x p_1(x/y)\cdot I(x \leftarrow y) \end{aligned}$$

$$= \sum_x p_1(x) \sum_y p_2(y/x) \cdot I(x \leftarrow y) \tag{15. 3}$$

については，上記の命題15. 1から，

〔命題15. 2〕　不等式

$$-1 \leqq I(X \leftarrow Y) \leqq +1$$

が成り立ち，

(i)　$I(X \leftarrow Y) = +1 \Leftarrow \exists x, \forall y, p_1(x/y) = 1$

(ii)　$I(X \leftarrow Y) = 0$

$\Leftarrow [\forall x, \forall y, p(x, y) = p_1(x) \cdot p_2(y)]$

$\vee [\exists x, p_1(x) = 1]$

も成立する。　□

さて，一変数 u の関数

$$\mathrm{psn}(u) = 1 \quad \text{if} \quad u \geqq 0, = 0 \quad \text{if} \quad u < 0$$

を導入すると，$I(x \leftarrow y)$ は

$$\begin{aligned} &I(x \leftarrow y) \\ &= [1 - \mathrm{psn}(p_1(x) \cdot p_2(y) - p(x,y))] \\ &\quad [1 - p_1(x)]^{-1} \cdot \left[1 - \frac{p_1(x) \cdot p_2(y)}{p(x,y)}\right] \\ &\quad + psn(p_1(x) \cdot p_2(y) - p(x,y)) \cdot \left[\frac{p(x,y)}{p_1(x) \cdot p_2(y)} - 1\right] \end{aligned} \tag{15. 4}$$

と表現されることがわかる。式（15. 2）より $0/0 = 0$ と約束すると，$p_1(x) = 1$ のときもこの表式は正確であることに注意しておく。

〔命題15. 3〕($I(X \leftarrow Y)$ の表現)

$$\begin{aligned} &I(X \leftarrow Y) \\ &= \sum_x \sum_y \frac{p(x,y) - p_1(x) \cdot p_2(y)}{1 - p_1(x)} \\ &+ \sum_x \sum_{\substack{y \\ p_1(x) \cdot p_2(y) \geqq p(x,y)}} [p_1(x) \cdot p_2(y) - p(x,y) \cdot \left[\frac{1}{1 - p_1(x)} - \frac{p(x,y)}{p_1(x) \cdot p_2(y)}\right] \end{aligned}$$

(証明)　式（15. 3）に式（15. 4）を代入すれば，

$$\begin{aligned} &I(X \leftarrow Y) \\ &= \sum_x \sum_y p(x,y) \cdot [1 - \mathrm{psn}(p_1(x) \cdot p_2(y) - p(x,y))] \\ &\quad \cdot [1 - p_1(x)]^{-1} \cdot [1 - p_1(x) \cdot p_2(y)/p(x,y)] \\ &\quad + \sum_x \sum_y p(x,y) \cdot \mathrm{psn}(p_1(x) \cdot p_2(y) - p(x,y)) \\ &\quad \cdot [p(x,y)/[p_1(x) \cdot p_2(y)] - 1] \\ &= \sum_x \sum_y p(x,y) \cdot [1 - p_1(x)]^{-1} \cdot [1 - p_1(x) \cdot p_2(y)/p(x,y)] \end{aligned}$$

$$+\sum_x \sum_y p(x,y)\cdot \mathrm{psn}(p_1(x)\cdot p_2(y)-p(x,y))$$
$$\cdot\left[\frac{(-1)}{-p_1(x)}\cdot\left\{1-\frac{p_1(x)\cdot p_2(y)}{p(x,y)}\right\}-\frac{p_1(x)\cdot p_2(y)-p(x,y)}{p_1(x)\cdot p_2(y)}\right]$$
$$=\sum_x \sum_y [1-p_1(x)]^{-1}\cdot[p(x,y)-p_1(x)\cdot p_2(y)]$$
$$+\sum_x \sum_y \mathrm{psn}(p_1(x)\cdot p_2(y)-p(x,y))$$
$$\cdot[p_1(x)\cdot p_2(y)-p(x,y)]\cdot\left[\frac{1}{1-p_1(x)}-\frac{p(x,y)}{p_1(x)\cdot p_2(y)}\right]$$

を得て，示された。 □

$I(x\leftarrow y)$ を変数 x, y に関し，対称化しよう。

$$I(x,y)=2^{-1}\cdot[I(x\leftarrow y)+I(y\leftarrow x)]$$

と定義すると，$I(x,y)$ は変数 x, y に関し対称であり，

(イ) $p_1(x)\cdot p_2(y)<p(x,y)$
$\wedge\ [p_1(x)\neq 1\wedge p_2(y)\neq 1]$ の場合

$$I(x,y)=\frac{\left[1-\dfrac{p_1(x)+p_2(y)}{2}\right]}{[1-p_1(x)\cdot[1-p_2(y)]}\cdot\left[1-\frac{p_1(x)\cdot p_2(y)}{p(x,y)}\right]>0$$

(ロ) $p_1(x)\cdot p_2(y)\geqq p(x,y)$
$\wedge\ [p_1(x)=1\vee p_2(y)=1]$ の場合

$$I(x,y)=\frac{p(x,y)}{p_1(x)\cdot p_2(y)}-1 \qquad (15.\ 5)$$

と表現される。

命題15. 1から次の命題15. 4の成立がいえ，命題15. 1の $I(x\leftarrow y)$ と同様，$I(x,y)$ は確率分布 $p(x,y)$ に関連した相関係数としての意味を備えていることがわかる。

〔命題15. 4〕　不等式

$$-1\leqq I(x,y)\leqq +1$$

が成り立ち，

(i) $I(x,y)=+1$
$\Leftarrow p_1(x/y)=p_2(y/x)=1$

(ii) $I(x,y)=0\Leftrightarrow p(x,y)=p_1(x)\cdot p_2(y)\vee[p_1(x)=1\wedge p_2(y)=1]$

(iii) $I(x,y)=-1$
$\Leftarrow p(x,y)=0\wedge[p_1(x)>0\vee p_2(y)>0]$ □

さて，$I(x,y)$ を $p(x,y)$ で平均化して，

$$I(X,Y)\equiv\sum_x \sum_y p(x,y)\cdot I(x,y)$$

と定義すれば，

$$I(X, Y) = 2^{-1} \cdot [I(X \leftarrow Y) + I(Y \leftarrow X)]$$

が成り立つ。ここに，

$$I(Y \leftarrow X) = \sum_x \sum_y p(x, y) \cdot I(y \leftarrow x).$$

そうすれば，命題15．2から次の命題15．5が成立し，$I(X, Y)$の相関係数としての性質がいえ，また，命題15．3から命題15．6の成立が示され，$I(X, Y)$の具体的表現が得られる。

〔命題15．5〕　不等式

$$-1 \leqq I(X, Y) \leqq +1.$$

が成り立ち，

(i) $I(X, Y) = +1$
$\Leftarrow [\exists\, x, \forall\, y, p_1(x/y) = 1]$
$\wedge [\exists\, y, \forall\, x, p_2(y/x) = 1]$

(ii) $I(X, Y) = 0$
$\Leftarrow [\forall\, x, \forall\, y, p(x, y) = p_1(x) \cdot p_2(y)]$
$\vee [[\exists\, x, p_1(x) = 1] \wedge [\exists\, y, p_2(y) = 1]$ □

〔命題15．6〕　（$I(X, Y)$ の表現）

$$I(X, Y)$$

$$= \sum_x \sum_y [p(x, y) - p_1(x) \cdot p_2(y)] \cdot \frac{\left[1 - \dfrac{p_1(x) + p_2(y)}{2}\right]}{[1 - p_1(x)] \cdot [1 - p_2(y)]}$$

$$+ \sum_x \sum_{\substack{y \\ p_1(x) \cdot p_2(y) \geqq p(x,y)}} [p_1(x) \cdot p_2(y) - p(x, y)] \cdot \left[\frac{1 - \dfrac{p_1(x) + p_2(y)}{2}}{[1 - p_1(x)] \cdot [1 - p_2(y)]} - \frac{p(x,y)}{p_1(x) \cdot p_2(y)}\right] \quad \square$$

S. Suzuki の“パターン認識の数学的理論”[34]への応用を簡単に述べておこう。

〈応用１〉

文献（７）の第６章，あるいは文献（34）で登場した二つの記号

$SM : \Phi \times \Omega \rightarrow \{s \mid 0 \leqq s \leqq 1\}$
$\hat{\gamma}(\varphi) \in 2^J$

を使う。ここに，

$\langle \varphi, \gamma \rangle$: パターン $\varphi \in \Phi$ はカテゴリ $\mathfrak{C}_j, j \in J$ のいずれかに帰属するという知識

$SM(\varphi, \omega_j)$: パターン φ が第 $j \in J$ 番目のカテゴリ $\mathfrak{C}_j$ の代表パターン ω_j に似ている程度で，

$$0 \leqq SM(\varphi, \omega_j) \leqq 1 \wedge \sum_{j \in J} SM(\varphi, \omega_j) = 1$$

$\hat{\gamma}(\varphi) : \langle \varphi, \gamma \rangle \in \langle \Phi, 2^J \rangle$ の最大実効カテゴリ番号のリスト

である。

$I(\langle\phi,\tau\rangle \leftarrow \langle\varphi,\gamma\rangle)$
$I(\langle\phi,\tau\rangle, \langle\varphi,\gamma\rangle)$

は次の対応の下で定義できることがわかる：

$p_1(x) \leftrightarrow p_1(\langle\phi,\tau\rangle) = \sum_{k\in\hat{\tau}(\phi)} SM(\phi,\omega_k)$
　$: \langle\phi,\tau\rangle \in \langle\Phi, 2^J\rangle$ の存在確率（probability of existence）
$p_2(y) \leftrightarrow p_2(\langle\varphi,\gamma\rangle)$
$p(x,y) = p_1(x/y)\cdot p_2(y)$
$\leftrightarrow\ p(\langle\phi,\tau\rangle, \langle\varphi,\gamma\rangle) = p_1(\langle\phi,\tau\rangle/\langle\varphi,\gamma\rangle)\cdot p_2(\langle\varphi,\gamma\rangle)$

ここに，
推移確率（transition probability）といわれる　$p_1(\langle\phi,\tau\rangle/\langle\varphi,\gamma\rangle)$ は

$$p_1(\langle\phi,\tau\rangle/\langle\varphi,\gamma\rangle) = \begin{cases} 0 \quad \text{if} \quad p_2(\langle\varphi,\gamma\rangle) = 0 \\ \dfrac{\sum_{j\in\hat{\tau}(\phi)\cap\hat{\gamma}(\varphi)} SM(\varphi,\omega_j)}{\sum_{k\in\hat{\gamma}(\varphi)} SM(\varphi,\omega_k)} \\ \quad \text{if} \quad p_2(\langle\varphi,\gamma\rangle) > 0 \ . \end{cases}$$

□

16. stochastic relaxation-labeling への相関係数の応用

本章では，文献（28），（40）での２研究に hint を得て，命題15．４（前章）の応用を，
digitized pattern $\varphi(x,y)$ の符号化
に関し説明する。なお，文献（35）では，neural net functioning と本章で論じる relaxation-labeling[41] との間の関係につき，
(i) Neural learning amounts to discovering the constraints between the states of connected neurons.
(ii) Neural learning itself is interpreted as an extended form of relaxation-labeling.
が指摘されている。
さて，n 個の対象物

$A = \{a_1, a_2, \cdots, a_n\}$

内の任意の a_i に，m 個のラベル

$\Lambda = \{\lambda_1, \lambda_2, \cdots, \lambda_m\}$

内の λ_j を割り当てる確率

$p_i(\lambda_j)$

を決めることを考えよう。ここに，各 $p_i(\lambda_j)$ は，

$0 \leqq p_i(\lambda_j) \leqq 1,$
$\sum_{j=1}^{m} p_i(\lambda_j) = 1$ for all $a_i \in A$

を満たしていなければならない。

Relaxation Labeling Process

では，$p_i(\lambda_j)$ の初期値

$$p_i(\lambda_j\,;k)\,|_{k=0}$$

を与え，第 k 段階での

$$p_i(\lambda_j\,;k)$$

を，第 $(k+1)$ 段階での確率

$$p_i(\lambda_j\,;k+1)$$

へと，次の形式で更新する：

$$p_i(\lambda_j\,;k+1)=\frac{p_i(\lambda_j\,;k)\cdot[1+q_i(\lambda_j\,;k)]}{\sum_{\ell=1}^{m}\ p_i(\lambda_j\,;k)\cdot[1+q_i(\lambda_\ell\,;k)]}.$$

ここで，$q_i(\lambda_j\,;k)$ は，不等式

$$-1\leqq\gamma_{i_1,i_2}(\lambda_j,\lambda_{j2})\leqq1 \qquad (16.\ 1)$$

を満たす両立係数，適合係数（compatibility coeffficient）$\gamma_{i_1,i_2}(\lambda_{j_1},\lambda_{j_2})$ を使って，

$$q_{i_1}(\lambda_{j_1}\,;k)$$
$$=\frac{1}{n}\cdot\sum_{i_2=1}^{n}\sum_{j_2=1}^{m}\gamma_{i_1,\,i_2}(\lambda_{j_1},\lambda_{j_2})\cdot p_{i_2}(\lambda_{j_2}\,;k)$$

と与えられる。

適合係数 $\gamma_{i_1,\,i_2}(\lambda_{j_1},\lambda_{j_2})$は次の 4 性質 a，b，c，d を満たさなければならない[28]：

(a) ラベル λ_{j_1} と λ_{j_2} とが各々，対象物 a_{i_1} と a_{i_2} とに両立できる（compatible）ならば，$\gamma_{i_1,\,i_2}(\lambda_{j_1},\lambda_{j_2})>0$。

(b) ラベル λ_{j_1} と λ_{j_2} とが各々，対象物 a_{i_1} と a_{i_2} とに両立しないならば，

$$\gamma_{i_1,\,i_2}(\lambda_{j_1},\lambda_{j_2})<0.$$

(c) 対象物 a_{i_1} に λ_{j1} をラベル付けることと，かつ，対象物 a_{i_2} に λ_{j_2} をラベル付けることとが制約になっていないのなら，

$$\gamma_{i_1,\,i_2}(\lambda_{j_1},\lambda_{j_2})=0$$

(d) $\gamma_{i_1,\,i_2}$ は両立性の強さを表わす。 □

2 次元平面上のディジタル化画像関数 $\varphi=\varphi(x_1,x_2)$ の符号化

$$\varphi(x_1,x_2)\in\Lambda \text{ for all points } (x_1,x_2)$$

を行なうことを考えよう。

この符号化は，以下の両立係数 $\gamma_{(x_1,x_2),\,(x_1+k_1,\,x_2+k_2)}(\lambda_{j_1},\lambda_{j_2})$ を採用し，ある終了条件（some termination criterion）を満たすまで，k を増加させていけば，

$p_{(x_1,x_2)}(\lambda_j)$ for all points (x_1, x_2)

が得られるから，

$\varphi(x_1, x_2) = \lambda_j$
, where $j = \arg\max_k p_{(x_1,x_2)}(\lambda_k)$

とすれば得られる。

適合係数 γ を近傍でのラベル間相互情報量（the mutual information of the labels at neighboring points）の形で，式（16. 2）のごとく与えるのが文献（28）の手法である。

$p_{(x_1,x_2)}(\lambda)$: the initial estimate of the probability of labeling point (x_1, x_2) with the label λ
を導入して，

$$p(\lambda) = \frac{1}{n} \cdot \sum_{(x_1,x_2)} p_{(x_1,x_2)}(\lambda)$$

: the estimate of the probability of any point having the label λ

$$p_{k_1, k_2}(\lambda_{j_1}, \lambda_{j_2}) = \frac{1}{n} \cdot \sum_{(x_1,x_2)} p_{(x_1,x_2)}(\lambda_{j_1}) \cdot p_{(x_1+k_1,x_2+k_2)}(\lambda_{j_2})$$

: the estimate of the joint probability of a pair $\langle (x_1, x_2), (x_1+k_1, x_2+k_2) \rangle$ of points having labels λ_{j_1} and λ_{j_2}

$$p_{k_1,k_2}(\lambda_{j_1}/\lambda_{j_2}) = \frac{P_{k_1,k_2}(\lambda_{j_1}, \lambda_{j_2})}{p(\lambda_{j_2})}$$

: the estimate of the conditional probability that (x_1, x_2) is labeled λ_{j_1} given that (x_1+k_1, x_2+k_2) is labeled λ_{j_2}

を計算し，これから算出される相互情報量

$$I_{k_1, k_2}(\lambda_{j_1}; \lambda_{j_2}) = \ell og \frac{p_{k_1,k_2}(\lambda_{j_1}/\lambda_{j_2})}{p(\lambda_{j_1})} = -\ell og\, p(\lambda_{j_1}) - [-\ell og\, p_{k_1,k_2}(\lambda_{j_1}/\lambda_{j_2})]$$

を，

$$\gamma_{(x_1,x_2)(x_1+k_1,x_2+k_2)}(\lambda_{j_1}, \lambda_{j_2})$$

に用いればよい。点 (x_1, x_2) にラベル λ_{j_1} を付与した事態に対し，点 (x_1+k_1, x_2+k_2) にラベル λ_{j_2} を付与する事態が寄与する程度を表現している $I_{k_1,k_2}(\lambda_{j_1}, \lambda_{j_2})$ は

$$I_{k_1,k_2}(\lambda_{j_1}, \lambda_{j_2}) = \ell og \frac{n \cdot \sum_{(x_1,x_2)} p_{(x_1,x_2)}(\lambda_{j_1}) \cdot p_{(x_1+k_1,x_2+k_2}(\lambda_{j_2})}{[\sum_{(a,b)} p_{(a,b)}(\lambda_{j_1})] \cdot [\sum_{(c,d)} p_{(c,d)}(\lambda_{j_2})]} \qquad (16.\ 2)$$

と具体的に表現される。これが文献（28）の考えである。然しながら不等式（16. 1）を満たすようにするため，実は，

$$\gamma_{(x_1,x2)(x_1+k_1,x_2+k_2)}(\lambda_{j_1},\lambda_{j_2})$$
$$=\begin{cases}+1 \text{ if } I_{k_1,k_2}(\lambda_{j_1},\lambda_{j_2}) \geqq +1\\ I_{k_1,k_2}(\lambda_{j_1},\lambda_{j_2}) \text{ if } -1 < I_{k_1,k_2}(\lambda_{j_1},\lambda_{j_2}) < +1\\ -1 \text{ if } I_{k_1,k_2}(\lambda_{j_1},\lambda_{j_2}) \leqq -1\end{cases} \tag{16. 3}$$

とする訳である。

文献（40）は，上述の相互情報量 $I_{k_1,k_2}(\lambda_{j_1},\lambda_{j_2})$ を採用せずに，改良して，適合係数

$\gamma_{(x_1,x2)(x_1+k_1,x_2+k_2)}$ の推定値

として，命題15. 1の非対称性の線形近似

$$I((x_1,x_2)\leftarrow(x_1+k_1,x_2+k_2))$$

を用いれば良いことを指摘しているが，本論文では，S. Suzuki の提案する命題15. 4でいう対称性の線形近似

$$I((x_1,x_2),(x_1+k_1,x_2+k_2))$$

を使用する方が良いことを指摘しておこう。

17. 命題のエントロピーと充真率

岡本・中島・大澤[43]によれば，ある命題の<u>確信度</u>（confidence factor）とは，

その命題の確かさを表す指標であり，命題の全領域の内のいかほどの部分から情報が得られているかを示す指標

であるという。また，ある命題の<u>主観確率</u>（subjective probability）とは，

その命題の全領域内にいかほどの割合で真の要素が含まれているかの指標

であるという。S. Suzuki によれば，主観確率とは真が満たされている程度であり，その命題の

<u>充真率</u>＝［命題を真にする入力の総数／全入力の総数］ (17. 1)

といって良いと思う。

命題 φ の主観確率が y であることを，認識主体 i が確信度 x で信じている，つまり

Agent i believes with a confidence facter x that the proposition φ has a subjective probability y

という命題を，岡本らは

$$B_i[x,y]\,\varphi$$

と表した。確信度 x と主観確率（uncertainity）y とは互いに独立であるとして，$B_i[x,y]$ を

確率信念オペレータ（probability belief operator）

と称している。積合成則

$$B_i[x,y]\cdot B_j[z,w]\,\varphi$$
$$\rightarrow\ B_i[x\,y\,z,w]\,\varphi$$

を証明し,

$B[z, w][\phi \leftarrow \varphi] \wedge B[x, y]\,\varphi$

ならば, $B[x\,y\,z, w]\,\phi$

という三段論法 (modus ponens) を提唱している。更に, 確信度 x のもつ不定さ $1-x$ を解消するのに必要な観測回数 (a 者沢一の回数) を表す関数として,

$$h(x) = -\,\ell og_a(1-x), \quad \text{where} \quad a > 1$$

を発見している。この関数 $h(x)$ の意味は S. Suzuki によれば, 次の通りである:

確信度 x が

$$x = [a^n - a^k]/a^n = 1 - a^{-(n-k)}$$

と表現されるとすれば,

$$1 - x = a^{-(n-k)} = \frac{1}{a} \cdot \frac{1}{a} \cdots \cdot \frac{1}{a} \quad (\frac{1}{a} \text{を} (n-k) \text{回かけたもの})$$

であり, 不定さ $1-x$ は a 者沢一を $(n-k)$ 回行えば解消することがわかる。このとき

$$h(x) = -\,\ell og_a(1-x) = n - k$$

である。 □

また, $B[x_1, y_1]\,\varphi \wedge B[x_2, y_2]\,\varphi$ を合成して, $B[x, y]\,\varphi$ になるとすれば,

$$x = x_1 + x_2 - x_1 \cdot x_2$$
$$y = \textstyle\sum_{i=1}^{2} y_i \cdot [\ell og_a(1-x_i)/\sum_{j=1}^{2} \ell og_a(1-x_j)]$$

とすれば良いという“独立な命題の合成則”も提唱している。更に, 確率信念命題 $B[x, y]$ について命題 φ が真である確率 p は, $x \cdot y$ は最小値の確実度と考え, 最大の確実度はそれに, 現在不定の部分が全部真となった場合の項 $1-x$ が付加されるとして, 不等式

$$x \cdot y \leqq p \leqq x \cdot y + (1-x)$$

を満たさなければならないとしている。よって, 命題 φ の Shannon 自己情報量 $-\,\ell og\,p$ は不等式

$$-\,\ell og\,xy \geqq -\,\ell og\,p \geqq -\,\ell og\,xy - \ell og\left[1 + \frac{1-x}{x} \cdot \frac{1}{y}\right]$$

を満たし, その上限は確信度 x, 主観確率 y のもたらす自己情報量の和

$$-\,\ell og\,xy = -\,\ell og\,x - \ell og\,y$$

である。

本章では, 月本の研究[(45), (46)]での論理エントロピーを S. Suzuki の提案する情報量の立場から意味づけ, 命題 φ の上述の主観確率 y を S. Suzuki のいう充真率とみて, y の決め方を論じよう。

S. Suzuki の提案する情報量 ISS とは次の様なものである。

M : 処理を想定している対象の総数

N : M 個の対象の内，処理可能な対象の総数 $(N \leqq M)$

N^+ : N 個の対象の内，有意味な対象の総数 $(N^+ \leqq N)$

N^- : N 個の対象の内，無意味な対象の総数 $(N^- \leqq N)$

とすれば，四つの関係

$$M = N + (M - N), N = N^+ + N^-$$
$$N^+ \leqq N \leqq M, N^- \leqq N \leqq M$$

が成立する。このとき，S. Suzuki の提唱する情報量 ISS とは

$$ISS = \ell og_2[M/N^+]$$

であると定義される。上述の関係式から

$$ISS = \ell og_2[1 + (N^-/N^+) + (M - N)/N^+]$$

と表現される。

一般に，n 個の論理変数 $X_2, X_2, \cdots, X_n$ の命題　$F(X_1, X_2, \cdots, X_n)$　は

$$\varphi(x_1, x_2, \cdots, x_n) \in \{0, 1\}$$
$$\text{, where} \quad x_i \in \{0, 1\}\ (i = 1 \sim n)$$

という 2 値関数と一対一対応をもつ。

何故ならば，$F(X_1, X_2, \cdots, X_n)$ において，$X_i \in \{\text{false, true}\}$ に対応して，$x_i \in \{0, 1\}$ なる通常の変数 x_i を考え，

F 内の X_i の否定$\sim X_i$ を $1 - x_i$ に

F 内の連言 $X_i \wedge X_j$ を $x_i \cdot x_j$ に

F 内の選言 $X_i \vee X_j$ を $x_i + x_j - x_i \cdot x_j$ に

F 内の含意 $X_i \rightarrow X_j$ を $1 - x_i + x_i x_j$ に

F 内の同値 $X_i \leftrightarrow X_j$ を $1 - x_i - x_j + 2x_i x_j$ に

置き換えて得られる関数を

$$\varphi(x_1, x_2, \cdots, x_n)$$

とすれば，任意の $i(= 1 \sim n)$ につき，

$$X_i = \text{false} \quad \leftrightarrow \quad x_i = 0$$
$$X_i = \text{true} \quad \leftrightarrow \quad x_i = 1$$

という対応の下で，

$$F(X_1, \cdots, X_n) = \text{false} \leftrightarrow \varphi(x_1, \cdots, x_n) = 0$$
$$F(X_1, \cdots, X_n) = \text{true} \leftrightarrow \varphi(x_1, \cdots, x_n) = 1$$

が成立するからである。

さて，2値変数 $x_i \in \{0, 1\}$ を連続値形に，$0 \leqq x_i \leqq 1$ と拡張し，月本[(45), (46)]は

$$-\ell og_2 \int_0^1 dx_1 \int_0^1 dx_2 \cdots \int_0^1 dx_n \mathfrak{T}(\varphi^2)(x_1, x_2, \cdots, x_n)$$

をこの命題 φ の論理エントロピーといったが，本研究では，φ を複素数値に拡張し，命題 φ の，岡本ら[(43)]のいう確信度 y は充真率と考え直すと，実は

$$y = \int_0^1 dx_1 \cdots \int_0^1 dx_n \mathfrak{T}(\varphi \cdot \bar{\varphi})(x_1, \cdots, x_n)$$

ここに，$\bar{\varphi}$ は φ の複素共役

と与えられ，

$M \equiv N = 2^n$ の場合　$ISS = -\ell og\, y$

であることが，定理17．1で示される。ここに

$\mathfrak{T}(\psi)(x_1, x_2, \cdots, x_n)$ は
$\psi(x_1, x_2, \cdots, x_n)$ を　$(x_1^2 - x_1) \cdot (x_2^2 - x_2) \cdots\cdots (x_n^2 - x_n)$ で割って得られる余り

である。

実は，n 変数論理変数　$X_1, X_2, \cdots, X_n \in$　{false, true}} の命題　$F(X_1, X_2, \cdots, X_n)$ は，false, true の表現として各々0，1を用いると，選言標準形（disjunctive normal from）では

$$\begin{aligned} &F(X_1, X_2, \cdots, X_n) \\ &= \vee_{e_1=0}^1 \vee_{e_2=0}^1 \cdots \vee_{e_n=0}^1 F(e_1, e_2, \cdots e_n) \\ &\quad \wedge X_1^{e_1} \wedge X_2^{e_2} \wedge \cdots \wedge X_n^{e_n} \end{aligned}$$

ここに，$X_k^{e_k} = \sim X_k$　if　$e_k = 0$，$= X_k$　if　$e_k = 1$

と表現される。このとき，

$M = N = 2^n$ の場合 F を真にする入力の総数 N^+ は

$$N^+ = \sum_{e_1=0}^{T} \sum_{e_2=0}^{1} \cdots \sum_{e_n=0}^{1} F(e_1, e_2, \cdots, e_n), \text{ where } F(e_1, e_2, \cdots, e_n) \in \{0, 1\}$$

である。

各変数 x_i が　$0 \leqq x_i \leqq 1$　であるような複素数値関数　$\varphi = (x_1, \cdots, x_n)$ は，各 x_i を　$x_i \in \{0,1\}$ と制限し，

$$\varphi = (x_1, \cdots, x_n) \in \{0,1\}$$

が満たされるとすれば，

ある命題 $F(X_1, \cdots, X_n)$ の拡張表現

とみなされてよい。満たされていないと，

無限多値論理（パターン）の表現

である。

φ が一実変数 $x\ (0 \leqq x \leqq 1)$ のみの関数で，$\varphi = \varphi(x)$ のとき

$(\mathfrak{T}\varphi)\ (x) = \varphi(0)\cdot(1-x) + \varphi(1)\cdot x$
$= \sum_{e=0}^{1} \varphi(e)\cdot x^{e}$

であり，この $\mathfrak{T}\varphi$ は

$$(\mathfrak{T}\varphi)\ (x) = \varphi(0) + \frac{\varphi(1)-\varphi(0)}{1-0}\cdot(x-0)$$

と再表現される。単位閉区間〔0, 1〕での関数 $\varphi(x)$ のグラフを直線で近似したときの補間公式を与えている。$(\mathfrak{T}\varphi)\ (x_1, x_2, \cdots, x_n)$ は，精確には

$(\mathfrak{T}\varphi)\ (x_1, x_2, \cdots, x_n)$
$= \sum_{e_1=0}^{1} \sum_{e2=0}^{1} \cdots \sum_{e_n=0}^{1} \varphi(e_1, e_2, \cdots, \mathrm{e}_n)\ x_1^{e_1} x_2^{e_2} \cdots x_n^{e_n}$

ここに，

$x_k^{e_k} = 1 - x_k \text{ if } e_k = 0, = x_k \text{ if } e_k = 1$

と表現され，$\mathfrak{T}\varphi$ は，もとの φ の，n 次元超立方体の頂点での関数値の集合

$\{\varphi(e_1, e_2, \cdots, e_n)\,|\, e_k \in \{0, 1\}, 1 \leqq k \leqq n\}$

によって一意的に決まってしまう事実に注目しよう。この様にして，パターンの表現とみなされた　$\varphi(x_1, x_2, \cdots, x_n)\ (0 \leqq x_i \leqq 1, i = 1 \sim n)$ については，上述の

$(\mathfrak{T}\varphi)\ (x_1, x_2, \cdots, x_n)$

が φ の持つ構造を簡約化したものであり，その正規化パターン（normalized　pattern）である，と考えても良いことは次の二事実，イ，ロから正当化されよう：

$x \in \{0, 1\}$ に対し，
　$x^{e} = 1 \text{ if } x = e, = 0 \text{ if } x \neq e$

の成立が容易に確かめられるから

$(\mathfrak{T}\varphi)\ \{e_1, e_2, \cdots, e_n) = \varphi(e_1, e_2, \cdots, e_n) \text{ for all } e_k \in \{0, 1\}$

が得られ，

(イ) 連続値変数 $x_i (0 \leqq x_i \leqq 1, i = 1 \sim n)$ に対し

$(\mathfrak{T}\mathfrak{T}\varphi)\ \{x_1, x_2, \cdots, x_n) = (\mathfrak{T}\varphi)\ (x_1, x_2, \cdots, x_n)$

(ロ) 特に，$x_i \in (0, 1\}\ (1 \leqq i \leqq n)$ と制限すると，

$(\mathfrak{T}\varphi)\ \{x_1, x_2, \cdots, x_n) = \varphi(x_1, x_2, \cdots, x_n).$ □

特に，上のロは次の事実を指摘している：

命題 $\varphi \in \{0, 1\}$ はパターン表現の正規化表現 $\mathfrak{T}\varphi$ と一致するごときものであり，命題論理とはパターン表現の特別な場合である。

〔定理17．1〕(充真率定理)

各実変数 x_i は　$0 \leqq x_i \leqq 1$　として，

$\varphi\{x_1, x_2, \cdots, x_n) \in \{0, 1\}$

とする。このとき，

$$q \equiv \int_0^1 dx_1 \int_0^1 dx_2 \cdots \int_0^1 \mathfrak{T}(\varphi \cdot \bar{\varphi})\ \{x_1, x_2, \cdots, x_n)$$

$$= \frac{[\varphi \text{ が } 1 \text{ になる入力の総数}]}{\text{入力の総数}}$$

が成立し，

$M \equiv \varphi = 2^n$ として，

$N^+ : \varphi$ が1になる入力 $\{x_1, x_2, \cdots, x_n)$ の総数

とすると，q は

$q = N^+ / N =$〔命題 φ の全領域内にどれだけの割合で真の要素が含まれているかの指標（充真率)〕

と解釈され，この場合の Suzuki 情報量 ISS は

$$ISS \equiv \ell og_2 (N / N^+) = -\ell og_2 q$$

(証明) $\varphi \cdot \bar{\eta}$ の正規化表現 $\mathfrak{T}(\varphi \cdot \bar{\eta})$ は

$$\mathfrak{T}(\varphi \cdot \bar{\eta})\ (x_1, x_2, \cdots, x_n)$$
$$= \Sigma_{e_1} \Sigma_{e_2} \cdots \Sigma_{e_n} \varphi(e_1, e_2, \cdots, e_n) \cdot \bar{\eta}(e_1, e_2, \cdots, e_n)$$
$$\cdot x_1^{e_1} \cdot x_2^{e_2}, \cdots, x_n^{e_n}$$

であることが容易にわかる。ここで，

$$e = 0 \text{ のとき，} \int_0^1 ax \cdot x^e = \int_0^1 dx \cdot (1 - x) = \frac{1}{2}$$

$$e = 1 \text{ のとき，} \int_0^1 ax \cdot x^e = \int_0^1 dx \cdot x = \frac{1}{2}$$

であり，

$$\int_0^1 dx_1 \int_0^1 dx_2 \cdots \int_0^1 dx_n\, x_1^{e_1} x_2^{e_2} \cdots x_n^{e_n} = \frac{1}{2^n}$$

であることに注意すると，

$$q \equiv \int_0^1 dx_1 \int_0^1 dx_2 \cdots \int_0^1 dx_n \mathfrak{T}(\varphi \cdot \bar{\eta})\ (x_1, x_2, \cdots, x_n)$$
$$= \Sigma_{e}, \Sigma_{e_2} \cdots \Sigma_{e_n} \varphi(e_1, e_2, \cdots, e_n) \cdot \bar{\eta}(e_1, e_2, \cdots, e_n)$$
$$\int_0^1 ax_1 \int_0^1 ax_2 \cdots \int_0^1 dx_n\, x_1^{e_1} x_2^{e_2} \cdot \cdots \cdot x_n^{e_n}$$

$$=\frac{1}{2^n}\Sigma_{e_1}\Sigma_{e_2}\cdots\Sigma_{e_n}\varphi(e_1,e_2,\cdots,e_n)\cdot\bar{\eta}(e_1,e_2,\cdots,e_n)$$

を得て，$\varphi=\eta$ とおき，仮定

$$\varphi(e_1,e_2,\cdots,e_n)\in\{0,1\}$$

より得られる

$$\forall e_1,\cdots,\forall e_n\in\{0,1\},|\varphi(e_1,\cdots,e_n)|^2=\varphi(e_1,\cdots,e_n)^2=\varphi(e_1,\cdots,e_n)$$

を代入すると，

$$q=\frac{1}{2^n}\Sigma_{e_n}\cdots\Sigma_{e_n}|\varphi(e_1,\cdots,e_n)|^2$$

$$=\frac{1}{2^n}\Sigma_{e_1}\cdots\Sigma_{e_n}\varphi(e_1,\cdots,e_n)$$

＝〔{$\varphi(x_1,\cdots,x_n)=1$ になる入力 $(x_1,\cdots,x_n)$ の総数〕／〔入力の総数2^n〕

となって証明された。

脳内に外界の何んらかの表象を形成させるという“認識のメカニズム”が，情報処理による“表現の変換”を介して“計算モデル”（computation model）の形で組立てられると考えよう。

この想定は，

「認識に必要な情報を得るためにはどの様な計算を行い，それからいかなる情報を抽出すればよいか」，

「表象が外界の単なるコピーではなくして，特徴抽出の過程を経て，データ構造をもつ新しい意味ある表象が生成されるためには（物体が見え，言語音声が聞こえるためには），何が計算されねばならないか」

を主題とする

Marr の計算論(44)

と同様な立場である。単純化していえば，認識システムを計算機プログラム理論でいう“計算モデル”とみなす立場である。

上述の“表現の変換”を

実数各入力の重みつき総和についてのしきい動作を行なう計算素子（ニューロン）同志の結合に基づく局所的計算によるネットワーク（神経回路網）状態（全体状態）の変換

とみなすのが，情報の表現物はパターンであることを主張する

コネクショニズム（結合主義：connectionism）

の立場である。*AI* における“記号主義”（symbolism）と対立する立場である。記号主義は，情報の表現物はシンボルであることに固執している古典的人工知能学の基盤をなす思想である。*AI* における記号主義が

置換（replacement），変換（conversion），還元（reduction），代入（substitution），簡約（simplification）

という5操作（記号操作）によってあからさまに表現の変換を達成しようとするのに対し，コネ

クショニズムはその内の“パターン変換”という操作のみを用い，表現の変換を行なうのである。

本章で登場した2表現F，φを使用して説明すれば，論理命題　$F(X_1, X_2, \cdots, X_n)$　を扱い，記号による推論を可能にするのが命題論理の範囲での記号主義であり，$F(X_1, X_2, \cdots, X_n)$ の拡張表現であるパターン表現 $\varphi(x_1, x_2, \cdots, x_n)$ を扱い，パターン同志の結合による推論を実現しようというのがコネクショニズムである。

記号主義の利点は豊かに複雑なデータ構造をあからさまに容易に設定でき，高次推論（深い推論）の implementation が可能であると知られていることである。一方，結合主義ではこのようなことが容易に可能かなどにつき，いまだ試行錯誤的研究の対象となっていることである。結合主義の利点は，学習に伴う知識の獲得が容易であることであろう。記号主義では，知識のこの獲得の場面においては，組合わせの爆発（the combinatorial explosion）が起こり，学習の implementation が計算量の観点から困難であることである。

記号主義，結合主義のいずれの立場をとるにしろ，情報の概念を確立しなければならない。情報の表現は，記号主義においては記号（symbol）によって，結合主義ではパターン（pattern）によって得られるとしているが，情報の“状況”依存性を考慮する限り，いずれの立場も情報の概念を確立することは不可能に近い。このような情報から知識を学習によって獲得し，構造表現し，推論によって新しい知識を引き出す“知能の働き”も状況依存的にならざるを得なくて，知能の概念を確立とすることも不可能といわざるを得ない。

低次推論（浅い推論）ではパターン表現が用いられているとする有力な証拠があるといわれている。低次（感覚・知覚・学習・記憶），高次（学習・記憶・推論・思考）双方の認識活動に必要な知識（表象）の表現とはパターンか？　シンボルか？　状況によって使いわけるべきか？前述したごとく，(ロ)の意味するところからすると，どうやら，論理とはパターンの正規化された形式であることを暗示しているが。

パターンとは構造をもち，それ自身意味をもつ小さな単位（構成単位；primitive component）に分解できるものである。構成単位が一つだけでは通常意味をもたず単位間の結合関係が意味をもつような情報である。一方，シンボルとは構造をもたなくて分解すべき単位をもたない。強制的に分解すれば意味をもたなくなり，（大きな）構成単位（記号それ自身）が一つのみあり，その一つだけで意味（記号に割り当てられた意味；記号意味内容）をもつような情報である。シンボルとは構造をもたないパターンであり，またパターンとは量子化してみれば，シンボルの集合体である[(7)]。

S. Suzuki の構築によるパターン認識の数学的理論[(34)]は，記号主義での有用な推論規則としての導出原理における単一化置換（unifier）の役割と同様な役割を持つ操作として，構造受精変換

$$\langle \varphi, \gamma \rangle \rightarrow TA(\gamma)\, T \cdot \langle \varphi, \gamma \rangle$$

を基本的に用い，あたかも記号主義での記号処理を行なっているかのごとく，パターン情報処理における推論を計算論的に構築しようとする一つの試みである。精密化の方向に進むのが学問である。ならば，認識の働きを計算論的に定式化しようという試みが，パターン表現による認識の計算モデルが思考メカニズムの解明に対しても通用するかどうかの検証のためにも，これからなされるであろうことは疑いがない。

18. 一般化情報量の拡張としての新しい相互情報量

いよいよ，第11章で投げかけていた問題の解決に入る。本論文での high light である。これ迄の他の研究者による提案とは異なる新しい相互情報量の提案に入ろう。

第11，12章で説明された相互情報量の性質，つまり

情報処理によって不確かさが減少し，その減少した量が情報処理量（相互情報量）である

という性質が備えている量が，第３章での一般化された情報量を使ってパターン情報処理[(5)]につき提案される。

以下，S. Suzuki の数学的理論[(19),(34)]での類似度関数

$$SM : \Phi \times \Omega \rightarrow \{s \mid 0 \leqq s \leqq 1\}$$

と，パターン φ が帰属している有効な候補カテゴリ番号リスト

$$\hat{\gamma}(\varphi)$$

とが登場するが，この定義，意味についての詳細は著者の研究[(7),(19),(34)]を見られたい。

ちなみに，$SM(\varphi, \omega_j)$ は

$$\textstyle\sum_{j \in J} SM(\varphi, \omega_j) = 1 \wedge 0 \leqq SM(\varphi, \omega_j) \leqq 1$$

を満たし，

パターン φ が第 $j \in J$ 番目のカテゴリ $\mathfrak{C}_j$ の代表パターン $\omega_j \in \Omega$ に似ている程度

を表している。

確率変数 X，Y 間で定義されたシャノン相互情報量 $I(X, Y)$ に相当するものを，二つのパターン φ，$\phi \in \Phi$ 間で定義するには，

$I(X, Y)$ は一方の確率変数のとる値を知って，他方の確率変数の不確かさがどれほど減少するかを示す量（一方の確率変数に含まれる他方の確率変数のもつ情報の程度）である

ことに注意して，

一種の自己エントロピーが一種の条件付エントロピーより大きな値を持つようにしなければならない。その差を一種の相互情報量とすればよい。情報処理によってエントロピーが減少しなければならないからである[(7)]

: 一種の相互情報量

＝〔一種の自己エントロピー〕

－〔一種の条件付自己エントロピー〕。 □

今少し詳しく説明しよう。二つの自己情報量

$-\ell og\, p(x_j)$: 自己情報量

$-\ell og\, p_1(x_j / y_k)$: 条件付自己情報量

の差は

$$-\ell og\, p_1(x_j) - \left[-\ell og \frac{P(x_j, y_k)}{p_2(y_k)}\right]$$

$$= -\ell og\, p_1(x_j) - [-\ell og\, p_1(x_j / y_k)]$$

$$= \ell og\,[p_1(x_j/y_k)/p_1(x_j)]$$
$$= \ell og\left[\frac{p(x_j, y_k)}{p_1(x_j)\cdot p_2(y_k)}\right]$$

と表現されることに留意しよう。相互情報量 $I(X, Y)$ は自己情報量の差

$\ell og\,[p(x_j, y_k)/\{p_1(x_j)\cdot p_2(y_k)\}]$

を結合確率 $p(x_j, y_k)$ で平均化した量であるから,

$$I(X, Y) = \textstyle\sum_j \sum_k p(x_j, y_k)\,\ell og \dfrac{p(x_j, y_k)}{p_1(x_j)\cdot p_2(y_k)}$$

と表わされ，関数 $-x\,\ell og\,x$ の凸性（第11章の式（11. 1））

$$\begin{aligned}
&-\textstyle\sum_k p(x_j, y_k)\,\ell og\,[p(x_j, y_k)/p_2(y_k)] \\
&= \textstyle\sum_k p_2(y_k)\,[-p_1(x_j/y_k)\,\ell og\,p_1(x_j/y_k)] \\
&\leqq -p_1(x_j)\,\ell og\,p_1(x_j),\ \text{ここに},\ p_1(x_j) = \textstyle\sum_k p_2(y_k)\cdot p_1(x_j/y_k) \\
&= -\textstyle\sum_k p(x_j, y_k)\,\ell og\,p_1(x_j)
\end{aligned}$$

から，相互情報量 $I(X, Y)$ の非負性

$$\begin{aligned}
&I(X, Y) \\
&= \textstyle\sum_j \Big[-\sum_k p(x_j, y_k)\,\ell og\,p_1(x_j) \\
&\quad -\Big\{-\textstyle\sum_k p(x_j, y_k)\,\ell og \dfrac{p(x_j, y_k)}{p_2(y_k)}\Big\}\Big] \geqq 0.
\end{aligned}$$

が証明された事実に思い起こし，**一般化された相互情報量** (generalized mutual information content)

$MI(\varphi, \psi; F)$,
$MI(\langle\varphi, \gamma\rangle, \langle\psi, \tau\rangle; F)$

が各々，I—i，I—iiで提案される。

I．第１種相互情報量

<u>I．i</u>　パターン間の非対称性相互情報量 $MI(\varphi, \psi)$

Φ を対象とするパターン φ の集合として,

$$F: \Phi \times \Phi \to \Phi$$

という写像 F を導入する。$\eta = F(\varphi, \psi) \in \Phi$ は，二つのパターン φ, $\psi \in \Phi$ にある情報処理を施こし，一つのパターン $\eta \in \Phi$ を得る操作を表わしている。例えば，写像 B, T を

$B: \Phi \to \Phi$
$T: \Phi \to \Phi$（収縮写像[34]）

とすると，

$$
\begin{aligned}
&F(\varphi,\phi)(x)\\
&=\varphi(x)\cdot\phi(x)\\
&\vee\ \varphi(x)\cdot\phi(x)-\varphi(x)\cdot\phi(x)\\
&\vee\ \sup_x\varphi(x)-\varphi(x)+\phi(x)\\
&\vee\ \sup_x[\varphi(x)\cdot\phi(x)]-\varphi(x)+\varphi(x)\cdot\phi(x)\\
&\vee\ \sup_x[\varphi(x)\cdot\phi(x)]-\varphi(x)-\phi(x)+2\cdot\varphi(x)\cdot\phi(x)\\
&\vee\ (B\varphi)(x)-\phi(x)\\
&\vee\ (T\varphi)(x)-\phi(x)
\end{aligned}
$$

などが考えられる。（付録 1 を参照）

さて，3つの量

$$
\begin{aligned}
&x_j=SM(F(\varphi,\phi),\omega_j)\\
&a_j=SM(\varphi,\omega_j)\\
&\mathrm{b}_j=SM(\phi,\omega_j)
\end{aligned}
$$

を導入し（付録 2 を参照），更に，

$$
\begin{aligned}
&x_j>0\ \wedge\\
&\textstyle\sum_{j\in J^+}a_j\leqq\sum_{j\in J^+}x_j\leqq 1\ \wedge\ \sum_{j\in J^+}b_j\leqq\sum_{j\in J^+}x_j\leqq 1
\end{aligned}
$$

を満たす添字 j の集合 $J^+\equiv J^+(\varphi,\phi)\equiv J^+(\varphi,\phi\,;F)\ (\subseteqq J)$
も導入し，情報処理機能 F についての，パターン φ，$\phi\in\Phi$ 間の相互情報量（<u>情報処理量</u>）

$$
\begin{aligned}
MI(\varphi,\phi)&\equiv MI(\varphi,\phi\,;F)\\
&\equiv\textstyle\sum_{j\in J^+}x_j\,\ell og\dfrac{x_j}{a_j\cdot b_j}
\end{aligned}
$$

を定義する。更に，次の諸量をも定義しておく：

$$
\begin{aligned}
&I(\varphi)_\phi\equiv I(\varphi\,;F)_\phi\\
&\quad\equiv-\textstyle\sum_{j\in J^+}x_j\,\ell og\,a_j\geqq 0\\
&I(\varphi)_\varphi\equiv I(\phi\,;F)_\varphi\\
&\quad\equiv-\textstyle\sum_{j\in J^+}x_j\,\ell og\,b_j\geqq 0\\
&I(\varphi,\phi)\equiv I(F(\varphi,\phi))\\
&\quad\equiv-\textstyle\sum_{j\in J^+}x_j\,\ell og\,x_j\geqq 0\\
&I(\varphi/\phi)\equiv I(\varphi/\phi\,;F)\\
&\quad\equiv-\textstyle\sum_{j\in J^+}x_j\,\ell og\,[x_j/b_j]\\
&\quad=I(\varphi,\phi)-I(\phi)\\
&I(\phi/\varphi)\equiv I(\phi/\varphi\,;F)\\
&\quad\equiv-\textstyle\sum_{j\in J}x_j\,\ell og\,[x_j/a_j]\\
&\quad=I(\varphi,\phi)-I(\varphi)
\end{aligned}
$$

□

そうすれば，次の命題18. 1のｉ，iiiは
二つのパターン φ, ϕ に対し，F という情報処理を行なうと，不等式

$$I(\varphi\,;F)_\phi \geqq I(\varphi/\phi\,;F)$$

が成立し，情報量が減少する
事実を指摘している。
〔命題18. 1〕(パターン間情報処理量定理)

(i) $MI(\varphi,\phi)$
$= I(\varphi)_\phi + I(\phi)_\varphi - I(F(\varphi,\phi))$
$= I(\varphi)_\phi - I(\varphi/\phi)$
$= I(\phi)_\varphi - I(\phi/\varphi)$
(ii) $F(\varphi,\phi)\,F(\phi,\varphi)$ であれば，
$MI(\varphi,\phi) = MI(\phi,\varphi)$
(iii) $MI(\varphi,\phi) \geqq 0$

が成り立ち，

$$I(\varphi)_\phi = I(F(\varphi,\phi))$$

となるのは，

$$\forall\, j \in J^+, SM(F(\varphi,\phi),\omega_j) = SM(\varphi,\omega_j)$$

のときに限る。

(iv) $I(\varphi)_\phi \geqq I(\varphi/\phi)$

が成り立ち，

$$\forall\, j \in J^+, SM(\varphi,\omega_j)\cdot SM(\phi,\omega_j) = SM(F(\varphi,\phi),\omega_j) > 0$$
$$\Rightarrow I(\varphi/\phi) = I(\varphi)_\phi$$

(証明)

ｉの証明：$\ell og\dfrac{x_j}{a_j\cdot b_j} = -\,\ell og\, a_j - \ell og\, b_j - [-\,\ell og\, x_j]$
$= -\,\ell og\, a_j - [-\,\ell og\, x_j/b_j]$
$= -\,\ell og\, b_j - [-\,\ell og\, x_j/a_j]$

の両辺に操作　$\sum_{j\in J^+} x_j\cdot$ を作用させればよい。
iiの証明：x_j' を

$$x_j' \equiv SM(F(\phi,\varphi),\omega_j)$$

と定義すると，仮定より $x_j = x_j', j\in J^+$ であるから

$$MI(\phi,\varphi) = \sum_{j\in J^+} x_j'\,\ell og\,[x_j'/\{b_j\cdot a_j\}]$$

$= \sum_{j \in J^+} x_j \,\ell og\,[x_j / \{b_j \cdot a_j\}]$
$= MI(\varphi, \phi).$

<u>iiiの証明</u>：ｉより

$MI(\varphi, \phi)$
$= I(\varphi)_\phi - I(F(\varphi, \phi)) + I(\phi)_\varphi$

である。ここで，

(a)　$I(\phi)_\varphi \geqq 0$

である。ところで，

(b)　$I(\varphi)_\phi - I(F(\varphi, \phi)) \geqq 0$

の成立は補定２．１より知れ，

(c)　$I(\varphi)_\phi = I(F(\varphi, \phi))$

となるのは

$\forall\, j \in J^+, SM(F(\varphi, \phi)) = SM(\varphi, \omega_j)$

のときに限る。(a)，(b)より，

$MI(\varphi, \phi) \geqq 0$

が知れた。(c)より，このiiiの後半が知れる。
なお，$f(x) = -x\,\ell og\,x\ (0 \leqq x)$
に補定11．１を適用しても，$J^+ = J$の下で

$\sum_{j \in J^+} b_j \cdot \left[-\frac{x_i}{b_j} \ell og \frac{x_i}{b_j} \right].$
$= \sum_{j^+} b_j \cdot f(x_j / b_j)$
$\leqq f(\sum_{j \in J^+} b_j \cdot x_j / b_j) \quad \because \sum_{j \in J^+} b_j = 1$
$= f(\sum_{j \in J^+} x_j) = f(1) = 0$
$\therefore\ -\sum_{j \in J^+} x_j \,\ell og\, x_j = I(F(\varphi, \phi))$
$\leqq -\sum_{j \in J^+} x_j \,\ell og\, b_j = I(\phi)_\varphi$

が知れるだけである。
<u>ivの証明</u>：ｉ，iiiより得られている

$MI(\varphi, \phi) = I(\varphi)_\phi - I(\varphi / \phi) \geqq 0$

を移項すれば，

$I(\varphi)_\phi \geqq I\varphi / \phi)$

が得られる。また，

$$\forall j \in J^+, a_j \cdot b_j = x_j > 0$$
$$\Rightarrow \forall j \in J^+, a_j = x_j / b_j$$
$$\Rightarrow I(\varphi)_\phi = I(\varphi / \phi) \qquad \square$$

なお，$0 \ell og 0$ に注意し，$J^+ = J^+(\varphi, \phi) = J^+(\varphi, \phi; F)$ なる定義からわかるように，

$a_j = 0 \wedge b_j = 0$ ならば，$x_j = 0$

を満たす $j \in J$ が存在すること

が成立する様な写像

$$F : \Phi \times \Phi \rightarrow \Phi$$

を想定すれば，この様な写像 F がパターン変換機能として有効であろうと予想される。

I. ii　パターンの帰属知識間の非対称性相互情報量　$MI(\langle\varphi, \gamma\rangle, \langle\phi, \tau\rangle)$

パターン $\varphi \in \Phi$，カテゴリ番号リスト $\gamma \in 2^J$ の対

$$\langle\varphi, \gamma\rangle \in \langle\Phi, 2^J\rangle$$

はパターン $\varphi \in \Phi$ がカテゴリ $\mathfrak{C}_j, j \in \gamma$ のいずれかに帰属しているという知識を表している。

$\langle\varphi, \gamma\rangle$，の台（support）$\hat{\gamma}(\varphi)$ のindicator

$$IND(k; \langle\varphi, \gamma\rangle)$$
$$= 1 \quad \text{if} \quad k \in \hat{\gamma}(\varphi), \quad = 0 \quad \text{if} \quad k \in J - \hat{\gamma}(\varphi)$$

を導入し，これと，$\langle\varphi, \gamma\rangle$ の重複度（the number of repeated times : multiplicity）とも称される $SM(\varphi, \omega_k)$　との積

$$SM(\langle\varphi, \gamma\rangle, \omega_k)$$
$$\equiv SM(\varphi, \omega_k; \gamma)$$
$$\equiv IND(k; \langle\varphi, \gamma\rangle) \cdot SM(\varphi, \omega_k)$$

を定義する。そして，

$$\mathfrak{S}\mathfrak{M}(\langle\varphi, \gamma\rangle, \omega_k)$$
$$\equiv \widehat{SM}(\langle\varphi, \gamma\rangle, \omega_k)$$
$$\equiv \widetilde{SM} \quad (\varphi, \omega_j; \gamma)$$
$$= \frac{SM(\varphi, \omega_j; \gamma)}{\sum_{k \in J} SM(\varphi, \omega_j; \gamma)} = \frac{SM(\varphi, \omega_j; \gamma)}{\sum_{k \in \hat{\gamma}(\varphi)} SM(\varphi, \omega_j)}$$
$$= \begin{cases} SM(\varphi, \omega_j) / \sum_{k \in \hat{\gamma}(\varphi)} SM(\varphi, \omega_k) & \text{if} \quad j \in \hat{\gamma}(\varphi) \\ 0 \quad \text{if} \quad j \in J - \hat{\gamma}(\varphi) & \end{cases}$$

も定義する。

$\hat{\gamma}(\varphi) = \phi$（空集合）のとき

$$\forall\, j \in J,\, SM\,(\langle\varphi, \gamma\rangle, \omega_k) = 0$$

に注意し，以後

$$\hat{\gamma}(\varphi) \neq \phi$$

としよう。

$\hat{\gamma}(\varphi) \neq \phi$　ならば，
$0 \leqq \widehat{SM}(\langle\varphi,\ \gamma\rangle, \omega_j) \leqq 1$
$\sum_{j\in J} \widehat{SM}(\langle\varphi,\ \gamma\rangle, \omega_j) = 1$

が成立している。

さて，情報処理の操作として，写像

$$F : \langle\Phi, 2^J\rangle \times \langle\Phi, 2^J\rangle \to \langle\Phi, 2^J\rangle$$

を導入し，また，

$$x_j \equiv SM\,(F(\langle\varphi, \gamma\rangle, \langle\phi, \tau\rangle), \omega_j)$$
$$a_j \equiv SM\,(\langle\varphi, \gamma\rangle, \omega_j)$$
$$b_j \equiv SM\,(\langle\varphi, \tau\rangle, \omega_j)$$

も導入し，更に，

$$x_j > 0 \wedge$$
$$\textstyle\sum_{j\in J^+} a_j \leqq \sum_{j\in J^+} x_j \wedge \sum_{j\in J^+} b_j = \sum_{j\in J^+} x_j$$

を満たす添字 j の集合

$$J^+ \equiv J^+(\langle\varphi, \gamma\rangle, \langle\phi, \tau\rangle)$$

を定義し，情報処理機能 F についての，二つのカテゴリ帰属知識

$$\langle\varphi, \gamma\rangle, \langle\phi, \tau\rangle \in \langle\Phi, 2^J\rangle$$

間の相互情報量

$$MI(\langle\varphi, \gamma\rangle, \langle\phi, \tau\rangle)$$
$$\equiv MI(\langle\varphi, \gamma\rangle, \langle\phi, \tau\rangle\,; F)$$
$$\equiv \textstyle\sum_{j\in J^+} x_j\, \ell og \dfrac{x_j}{a_j \cdot b_j}$$

を定義する。ここに，

$$I(\langle\varphi, \gamma\rangle) \equiv -\textstyle\sum_{j\in J^+} x_j\, \ell og\, a_j \geqq 0$$
$$I(\langle\varphi, \tau\rangle) \equiv -\textstyle\sum_{j\in J^+} x_j\, \ell og\, b_j \geqq 0$$
$$I(\langle\varphi, \gamma\rangle, \langle\phi, \tau\rangle)$$
$$\equiv -\textstyle\sum_{j\in J^+} x_j\, \ell og\, x_j$$

$I(\langle\varphi,\gamma\rangle/\langle\phi,\tau\rangle)$

$\equiv -\sum_{j\in J^+} x_j\, \ell og\,[x_j/b_j] = I(\langle\varphi,\gamma\rangle), \langle\phi,\tau\rangle) - I(\langle\phi,\tau\rangle)$

$I(\langle\phi,\tau\rangle/\langle\varphi,\gamma\rangle)$

$\equiv -\sum_{j\in J^+} x_j\, \ell og\,[x_j/a_j].$ □

命題18．1と同様に，次の命題18．2が証明される。

〔命題18．2〕（帰属知識間情報処理量定理）

(i) $MI(\langle\varphi,\gamma\rangle,\langle\phi,\tau\rangle)$

$= I(\langle\varphi,\gamma\rangle) + I(\langle\phi,\tau\rangle)$

$\quad - I(\langle\varphi,\gamma\rangle,\langle\phi,\tau\rangle)$

$= I(\langle\varphi,\gamma\rangle) - I(\langle\varphi,\gamma\rangle/\langle\phi,\tau\rangle)$

$= I(\langle\phi,\tau\rangle) - I(\langle\phi,\tau\rangle/\langle\varphi,\gamma\rangle)$

(ii) $MI\,(\langle\varphi,\gamma\rangle,\langle\phi,\tau\rangle) = F(\langle\phi,\tau\rangle\,\langle\varphi,\gamma\rangle)$ であれば，

$MI(\langle\varphi,\gamma\rangle,\langle\phi,\tau\rangle)$

$= MI(\langle\phi,\tau\rangle,\langle\varphi,\gamma>).$

(iii) $MI(\langle\varphi,\gamma\rangle,\langle\phi,\tau\rangle) \geqq 0$

が成立し，

$$I(\langle\varphi,\gamma\rangle) = I(\langle\varphi,\gamma\rangle,\langle\phi,\tau\rangle)$$

となるのは，

$$\forall\, j\in J^+, \widehat{SM}\,(F(\langle\varphi,\gamma\rangle,\langle\phi,\tau\rangle),\omega_j) = \widehat{SM}\,(\langle\varphi,\gamma\rangle,\omega_j)$$

のときに限る。

(iv) $I(\langle\varphi,\gamma\rangle) \geqq I\,(\langle\varphi,\gamma\rangle/\langle\phi,\tau\rangle)$

が成り立ち，

$$\forall\, j\in J^+, \widehat{SM}\,(\langle\varphi,\gamma\rangle,\omega_j)\cdot\widehat{SM}\,(\langle\phi,\tau\rangle,\omega_j) = \widehat{SM}\,(F(\langle\varphi,\gamma\rangle,\langle\phi,\tau\rangle),\omega_j)$$

$\Rightarrow\quad I(\langle\varphi,\gamma\rangle) = I(<\varphi,\gamma\rangle/\langle\phi,\tau\rangle).$ □

Ⅱ．第２種相互情報量

命題18．1のii，命題18．2のiiからわかるように，Ⅰでの相互情報量は必らずしも対称性を満たしていない。この点を改良しよう。

Ⅱ．i　パターン間対称性相互情報量

情報処理の働きとしての写像

$$F:\Phi\times\Phi\rightarrow\Phi$$

を，節Ⅰ．iにおけると同様に，導入する。（付録1を参照）

$x_j \equiv SM\,(F\,(\varphi,\phi),\omega_j)$

$y_k \equiv SM\,(F\,(\phi,\varphi),\omega_k)$

$a_i \equiv SM(\varphi, \omega_j)$
$b_k \equiv SM(\phi, \omega_k)$

とし（付録2を参照），また，

$x_j > 0 \wedge y_j > 0 \wedge$
$1 = \sum_{j\in J^+} a_j \leqq \sum_{j\in J^+} x_j \wedge$
$1 = \sum_{j\in J^+} b_j \leqq \sum_{j\in J^+} y_j$

を満たす添字 j の集合

$J^+ = J^+(\varphi, \phi) \subseteqq J$

を導入し，F についての，φ，$\phi \in \Phi$ 間の相互情報量　$MI(\varphi, \phi)$ を

$$MI(\varphi, \phi) \equiv MI(\varphi, \phi; F)$$
$$\equiv \sum_{j\in J^+} \sum_{k\in J^+} x_j \cdot y_k \,\ell og \frac{x_j \cdot y_k}{a_j \cdot b_k}$$

と定義する。次の諸量をも同時に定義する：

$I(\varphi) \equiv I(\varphi; F)_\phi$
$\equiv - \sum_{j\in J^+} x_j \,\ell og\, a_j$
$\geqq I(F(\varphi, \phi))$ $\because$ 補定2．1
ここに，
$I(F(\varphi, \phi)) \equiv - \sum_{j\in J^+} x_j \,\ell og\, x_j \geqq 0$
$I(\phi) \equiv I(\phi; F)_\varphi$
$\equiv - \sum_{k\in J^+} y_k \,\ell og\, b_k$
$\leqq I(F(\phi, \varphi))$ $\because$ 補定2．1
ここに，
$I(F(\phi, \varphi)) \equiv - \sum_{k\in J^+} y_k \,\ell og\, y_k \geqq 0$
$I(\varphi, \phi) \equiv I(F(\varphi, \phi), F(\phi, \varphi))$
$\equiv - \sum_{j\in J^+} \sum_{k\in J^+} x_j y_k \,\ell og\, [x_j y_k]$
$= I(F(\varphi, \phi)) + I(F(\phi, \varphi)) \geqq 0$
$I(\varphi / \phi) \equiv I(\varphi / \phi; F)$
$$\equiv - \sum_{j\in J^+} \sum_{k\in J^+} x_j y_k \,\ell og \frac{x_j \cdot y_k}{b_k}$$
$= I(F(\varphi, \phi), F(\phi, \varphi)) - I(\phi; F)_\varphi$
$I(\phi / \varphi) \equiv I(\phi / \varphi; F)$
$$\equiv - \sum_{j\in J^+} \sum_{k\in J^+} x_j y_k \,\ell og \frac{x_j \cdot y_k}{a_j}$$
$= I(\varphi, \phi) - I(\varphi)$ □

備考18．1：上述の $I(\phi / \varphi)$ は

$$I(\phi/\varphi) = -\textstyle\sum_{k\in J^+}\sum_{j\in J^+} x_k y_j \ell og \dfrac{x_k \cdot y_j}{a_k}$$

$$\equiv -\textstyle\sum_{j\in J^+}\sum_{k\in J^+} y_j x_k \ell og \dfrac{y_j \cdot x_k}{a_k}$$

とも表現できることに注意しておく。

備考18. 2：$J^+ = J^+(\varphi,\phi)$ なる定義からわかるように，

$a_j = 0 \wedge b_j = 0$　ならば　$x_j = 0 \vee y_j = 0$　が存在するような $j \in J$ が存在する様な写像 F がパターン変換機能として望ましいといえよう。

さて，命題18．1に対応して，次の命題18．3が成り立つ。

〔命題18．3〕（パターン間情報処理対称定理）

(i) $MI(\varphi,\phi)$
$= I(\varphi) + I(\phi) - I(\varphi,\phi)$
$= I(\varphi) - I(\varphi/\phi)$
$= I(\phi) - I(\phi/\varphi)$

(ii) $MI(\varphi,\phi) = MI(\phi,\varphi)$

(iii) $I(\varphi) \geqq I(F(\varphi,\phi)) \geqq I(\varphi/\phi)$

が成り立ち，

$$I(\varphi) = I(\phi/\phi)$$

が成り立つのは

添字 k を含まなくて，添字 j のみの関数 c_j が存在して，任意の $j \in J^+$ につき，

$$\forall\ k \in J^+, SM(\varphi,\omega_j)\cdot SM(F(\phi,\varphi),\omega_k) = c_j \cdot SM(\phi,\omega_k) \quad \text{for } SM(\phi,\omega_k) > 0$$

が成立するときのみに限る。

(iv) $MI(\varphi,\phi) \geqq 0$

が成立し，

任意の $k \in J^+$ について，

$$\forall\ j \in J^+, SM(F(\varphi,\phi),\omega_j)\cdot SM(F(\phi,\varphi),\omega_k) = SM(\varphi,\omega_j)$$
$$\Rightarrow I(\varphi,\phi) = I(\varphi).$$

（証明）

i の証明：

$$-\textstyle\sum_{j\in J^+}\sum_{k\in J^+} x_j \cdot y_k \ell og\, a_j$$
$$= -\textstyle\sum_{j\in J^+} x_j \ell og\, a_j [\sum_{k\in J^+} y_k]$$
$$= -\textstyle\sum_{j\in J^+} x_j \ell og\, a_j$$
$$= I(\varphi) \qquad (18.\ 1)$$

であり，

$$- \sum_{j\in J^+} \sum_{k\in J^+} x_j \cdot y_k \ell og\, b_k$$
$$= - \sum_{k\in J^+} y_k \ell og\, b_k [\sum_{j\in J^+} x_j]$$
$$= - \sum_{k\in J^+} y_j \ell og\, b_k$$
$$= I(\phi) \qquad (18.\ 2)$$

である。よって，

$$\ell og \frac{x_j \cdot y_k}{a_j \cdot b_k}$$

の両辺に，$\sum_{j\in J^+} \sum_{k\in J^+} x_j \cdot y_k \cdot$ を作用させると，

$$MI(\varphi, \phi)$$
$$= I(\varphi) + I(\phi) - I(\varphi, \phi)$$

が得られる。また，

$$\ell og \frac{x_j \cdot y_k}{a_j \cdot b_k} = - \ell og\, a_j - \left[- \ell og \frac{x_j \cdot y_k}{b_k} \right]$$
$$= - \ell og\, b_k - \left[- \ell og \frac{x_j \cdot y_k}{a_j} \right]$$

より，式（18. 1），式（18. 2）を考慮すると，

$$MI(\varphi, \phi)$$
$$I(\varphi) - I(\varphi / \phi)$$
$$= I(\phi) - I(\phi / \varphi)$$

が得られる。

ii の証明：

$$MI(\phi, \varphi) \equiv \sum_{j\in J^+} \sum_{k\in J^+} y_j \cdot x_k \ell og \frac{y_j \cdot x_k}{b_j \cdot a_k}$$
$$= \sum_{j\in J^+} \sum_{k\in J^+} x_k \cdot y_j \ell og \frac{x_k \cdot y_j}{a_k \cdot b_j}$$
$$= \sum_{k\in J^+} \sum_{j\in J^+} x_k \cdot y_j \ell og \frac{x_k \cdot y_j}{a_k \cdot b_j}$$
$$= MI(\varphi, \phi).$$

iiiの証明：$f(x) = - x \ell og\, x$

として，補定11. 1を適用すれば，

$$\sum_{k\in J^+} b_k \cdot f\left(\frac{x_j \cdot y_k}{b_k} \right)$$
$$\leqq f\left(\sum_{k\in J^+} b_k \cdot \frac{x_j \cdot y_k}{b_k} \right) \quad \because \quad \sum_{k\in J^+} b_k = 1 \qquad (18.\ 3)$$
$$= f(\sum_{k\in J^+} x_j \cdot y_k)$$

$= f(x_j) \quad \because \quad \sum_{k \in J^+} y_k = 1$

を得て，この両辺を　$j \in J^+$　につき総和をとれば，

$$\sum_{j \in J^+} \sum_{k \in J^+} b_k \cdot f\left(\frac{x_j \cdot y_k}{b_k}\right)$$
$$\leqq \sum_{j \in J^+} f(x_j)$$
$$\therefore \quad I(\varphi / \phi) \leqq I(F(\varphi, \phi))$$

が成立する。この不等式に，補定２．１を適用して，

$$I(F(\varphi, \phi))$$
$$\leqq -\sum_{j \in J^+} x_j \ell og\, a_j \qquad (18.\ 4)$$
$$= I(\varphi)$$

が得られる。つまり，不等式

$$I(\varphi / \phi) \leqq I(F(\varphi, \phi)) \leqq I(\varphi)$$

が成立する。

$$I(\varphi / \phi) = I(\varphi)$$

が成立するのは，

$I(\varphi / \phi) = I(F(\varphi, \phi)$，つまり

補定11．１から，式（18．３）において任意の$j \in J^+$につき，あるc_j（kに無関係なjのみの関数）が存在して，

$$\forall\ k \in J^+,\ \frac{x_j \cdot y_k}{b_k} = c_j \quad \text{for } b_k > 0$$

がいえ，しかも，

補定２．１から，式（18．４）において，

$$\forall\ j \in J^+,\ x_j = a_j$$

が成立する場合に限る。結局，$I(\varphi / \phi) = I(\varphi)$ が成立するのは，

$\forall\ j \in J^+$，c_jが存在し，

$$\forall\ k \in J^+, a_j \cdot y_k = c_j \cdot b_k \quad \text{for } b_k > 0$$

のときに限ることがわかる。

ivの証明 ::前半の $MI(\varphi, \phi)$ の非負性は，iiiを i の公式

$$MI(\varphi, \phi) = I(\varphi) - I(\varphi / \phi)$$

に適用すればよい。後半は，

$$I(\phi / \varphi)$$

において，式（18. 1）を考慮すれば明らかである。 □

Ⅱ．ii　帰属知識間の対称性相互情報量

節Ⅰ．iiでの相互情報量とは異なり，対称性を有する相互情報量を，同様な諸記号を用い提案しよう。

帰属知識間の変換操作

$$F : \langle \Phi, 2^J \rangle \times \langle \Phi, 2^J \rangle \to \langle \Phi, 2^J \rangle$$

を導入する。

$$x_j \equiv \widehat{SM}\,(F(\langle \varphi, \gamma \rangle, \langle \phi, \tau \rangle), \omega_j)$$
$$y_k \equiv \widehat{SM}\,(F(\langle \phi, \tau \rangle, \langle \varphi, \gamma \rangle), \omega_j)$$
$$a_j \equiv \widehat{SM}\,(\langle \varphi, \gamma \rangle, \omega_j)$$
$$b_k \equiv \widehat{SM}\,(\langle \varphi, \gamma \rangle, \omega_k)$$

とし，また，

$$x_j > 0 \wedge y_j > 0 \wedge$$
$$1 = \textstyle\sum_{j \in J^+} a_i \leqq \sum_{j \in J^+} x_j \wedge$$
$$1 = \textstyle\sum_{j \in J^+} b_i \leqq \sum_{j \in J^+} y_j$$

を満たす添字 j の集合

$$J^+ = J^+(\langle \varphi, \gamma \rangle, \langle \phi, \tau \rangle) \subseteqq J$$

を導入し，F についての，二つのカテゴリ帰属知識

$$\langle \varphi, \gamma \rangle, \langle \phi, \tau \rangle \in \langle \Phi, 2^J \rangle$$

間の相互情報量

$$MI(\langle \varphi, \gamma \rangle, \langle \phi, \tau \rangle)$$
$$\equiv MI(\langle \varphi, \gamma \rangle, \langle \phi, \tau \rangle\,; F)$$
$$\equiv \textstyle\sum_{j \in J^+} \sum_{k \in J^+} x_j \cdot y_k \,\ell og \dfrac{x_j \cdot y_k}{a_j \cdot b_k}$$

を定義する。ここに，

$$I(\langle \varphi, \gamma \rangle)$$
$$\equiv - \textstyle\sum_{j \in J^+} x_j \,\ell og\, a_j \geqq I(F(\langle \varphi, \gamma \rangle, \langle \phi, \tau \rangle))$$
$\geqq 0$ $\because$ 補定２．１

ここに，

$$I(F(\langle \varphi, \gamma \rangle, \langle \phi, \tau \rangle))$$
$$\equiv - \textstyle\sum_{j \in J^+} x_j \,\ell og\, x_j$$
$$I(\langle \phi, \tau \rangle)$$
$$\equiv - \textstyle\sum_{j \in J^+} y_k \,\ell og\, b_k \geqq I(F(\langle \phi, \tau \rangle; \langle \varphi, \gamma \rangle))$$

$\geqq 0$ ∵ 補定2．1

ここに，

$I(F(\langle\phi, \tau\rangle, \langle\varphi, \gamma\rangle))$

$\equiv - \sum_{k\in J^+} y_k \,\ell og\, y_k$

$I(\langle\varphi, \gamma\rangle, \langle\phi, \tau\rangle)$

$\equiv - \sum_{j\in J^+} \sum_{k\in J^+} x_j y_k \,\ell og\, [x_j \cdot y_k]$

$= I(F(\langle\varphi, \gamma\rangle, \langle\phi, \tau\rangle))$

$+ I(F(\langle\phi, \tau\rangle, \langle\varphi, \gamma\rangle))$

$I(\langle\varphi, \gamma\rangle / \langle\phi, \tau\rangle)$

$$\equiv - \sum_{j\in J^+} \sum_{k\in J^+} x_j y_k \,\ell og\, \frac{x_j \cdot y_k}{b_k}$$

$= I(\langle\varphi, \gamma\rangle, \langle\phi, \tau\rangle) - I(\langle\phi, \tau\rangle)$

$I(\langle\phi, \tau\rangle / \langle\varphi, \gamma\rangle))$

$$\equiv - \sum_{j\in J^+} \sum_{k\in J^+} x_j y_k \,\ell og\, \frac{x_j \cdot y_k}{a_j}$$

$= I(\langle\varphi, \gamma\rangle, \langle\phi, \tau\rangle) - I(\langle\varphi, \gamma\rangle).$ □

このとき，命題18．2に対応して，次の命題18．4が成り立つ。

〔命題18．4〕(帰属知識間情報処理量対称定理)

(i) $MI(\langle\varphi, \gamma\rangle, \langle\phi, \tau\rangle))$

$= I(\langle\varphi, \gamma\rangle) + I(\langle\phi, \tau\rangle) - I(\langle\varphi, \gamma\rangle, \langle\phi, \tau\rangle)$

$= I(\langle\varphi, \gamma\rangle) - I(\langle\varphi, \gamma\rangle / \langle\phi, \tau\rangle)$

$= I(\langle\phi, \tau\rangle) - I(\langle\phi, \tau\rangle / \langle\varphi, \gamma\rangle)$

(ii) $MI(\langle\varphi, \gamma\rangle), \langle\phi, \tau\rangle) = MI(\langle\varphi, \tau\rangle, \langle\varphi, \gamma\rangle)$

(iii) 不等式

$I(\langle\varphi, \gamma\rangle) \geqq I(F\langle\varphi\, \gamma\rangle, \langle\phi,\tau\rangle)$

$\geqq I(\langle\varphi\, \gamma\rangle / \langle\phi,\tau\rangle)$

が成り立ち，等式

$I(\langle\varphi, \gamma\rangle) = I((\varphi, \gamma\rangle / \langle\phi, \tau\rangle)$

が成立するのは，

添字 k を含まない添字 j の関数が c_j 存在して，任意の $j \in J^+$ につき

$\forall\ k \in J^+, SM(\langle\varphi, \gamma\rangle, \omega_j) \cdot SM(F(\langle\phi, \tau\rangle, \langle\varphi, \gamma\rangle), \omega_k)$

$= c_j \cdot SM(\langle\phi, \tau\rangle, \omega_k)$ for $SM(\langle\phi, \tau\rangle, \omega_k) > 0$

が成立するときのみに限る。

(iv) $MI(\langle\varphi, \gamma\rangle, \langle\phi, \tau\rangle) \leqq 0$

が成り立ち，

任意の $k \in J^+$ について

$$\forall j \in J^+, SM(F\langle\varphi,\gamma\rangle, \langle\psi,\tau\rangle), \omega_j) \cdot SM(F\langle\psi,\tau\rangle, \langle\varphi,\gamma\rangle), \omega_k) = SM(\langle\varphi,\gamma\rangle, \omega_j)$$
$$\Rightarrow \quad I(\langle\varphi,\gamma\rangle, \langle\psi,\tau\rangle) = I(\langle\varphi,\gamma\rangle). \qquad \square$$

19. むすび

情報量の，Shannon による解釈（第 2 章）に始まり，一般化された情報量（第 3 章），これ迄の情報量（第 4 章）を説明し，最大エントロピー原理（第 7， 8 章）の応用（第 5 章， 6 章），最小平均エネルギー原理（第 9 章），最小分散原理（第10章），頻度分布による相互情報量の表現（第11章）を，主として，

認識の量子論[(5),(14)] (quantum theory of recognition)

での特徴抽出に的を絞り，論じた。

更に，情報疎遠量 $SI(X,Y)$ の提案（第12章），二つのパターン間のピーク一致情報量（第12章），パターンの振幅エントロピーの提案（第13章）の後，長谷など[(36),(37)]の研究にヒントを得て，

認識の量子論[(5)]でのフェルミ・ディラック集団的特徴粒子の集合が抽出されたパターン集合に対し，特徴量の変動エントロピーを提案し（第14章），edge-entropy（第14章）もボーズ・アインシュタイン集団的パターン集合に対し，同時に提案した。

また，相互情報量を線形近似し，絶対値が 1 より大きくない新しい相関係数をも提案し（第15章），その応用を

relaxation labeling process[(41)]

に関し論じた。システムへの有意味．無意味な入力を区別することで得られた情報量 ISS を提案し（第17章），続いて，岡本ら[(43)]の命題確信度を，月本[(45),(46)]の論理エントロピーと関連し，ISS で表現した（第17章）。

1948年に公表された Shannon 情報理論[(12)]においては，自己情報量 $-\ell og\, p_1(x)$ はデータ x を 2 進系列で符号化するときの最適な記述長であり，平均情報量（エントロピー）

$$H(X) = \sum_x p_1(x) \cdot [-\ell og_2\, p_1(x)]$$

は平均符号化長（ 2 進系列の平均桁数）の下限である（第 2， 3 章）。これが情報量の解釈についての基本である。条件付自己情報量 $-\ell og\, p_1(x/y)$ についても同様な解釈が可能であり，両者各々を平均化し，その差として

$$\begin{aligned} I(X,Y) &\equiv \sum_x p_1(x) \cdot [-\ell og\, p_1(x)] \\ &\quad - [\sum_x \sum_y p(x,y)\,[-\ell og\, p_1(x/y)\}] \\ &= \sum_x \sum_y p(x,y)\, \ell og\,[p(x,)/\{p_1(x) \cdot p_2(y)\}] \end{aligned}$$

と定義された従来の相互情報量は二つの確率変数 X， Y 間に導入されたものである。

相互情報量 $I(X,Y)$ が確率変数間で定義されていることはその応用が限られ，このため，変革が求められていた。本研究では，パターン φ， ψ（可分な一般抽象 Hilbert 空間）間に，相互

情報量　$MI(\varphi, \phi)$ を，また，二つのカテゴリ帰属知識 $\langle\varphi, \gamma\rangle$，$\langle\phi, \tau\rangle$ 間にも相互情報量 $MI(\langle\varphi, \gamma\rangle, \langle\phi, \tau\rangle)$ を定義し，従来の相互情報量 $I(X, Y)$ とほぼ同様な諸性質が備わっていることを確認した（第18章）。相互情報量 MI は構造性情報（パターン）の間に導入されたものであり，確率変数間に導入された従来の相互情報量では表現しにくい内容を計量化したものである。パターン情報処理機能を解明するのに役立つ一種の情報処理量であることが期待される。この際，第２章の一般化情報量に関し，**補助定理２．１**，**補助定理11．１**が有効に適用されたが，この２つの補助定理は従来の説述内容と異なり，各々

$$x_j > 0 \wedge \textstyle\sum_{j \in J^+} y_j \leqq \sum_{j \in J^+} x_j \quad ; \quad 0 \leqq x$$

である場合に拡張されていることに注意しておこう。このため，本研究で S. Suzuki が提案した相互情報量（パターンに関する情報処理量）MI は確かに，Shannon 相互情報量，Kullback 情報量（第12章のⅡ，Ⅲ）の双方を折衷し，止揚した形式を備えている。

本研究は著者の博士論文[14]の続編といえなくはない。このためには，用意しておいたにも拘わらず，紙面の都合上割愛された２つの内容につき，論及しておかなければならないだろう。

①パターン φ にパターン ϕ が含まれている程度（ϕ に含まれている φ の程度）のもたらす情報量（the amount of information about the pattern φ contained in the pattern ϕ）$I(\varphi, \phi)$ は，内積（　，　）を用いて定義されたノルム $\|\cdot\| = [(\cdot, \cdot)]^{1/2}$ に関し，

ある η が存在し，$\|\varphi\|^2 = \|\phi\|^2 + \|\eta\|^2$

ここに，$(\phi, \eta) = 0$（直交性）

と直交分解されたとき，不等式

$$0 \leqq [\|\eta\|^2/\|\varphi\|^2] = 1 - [\|\phi\|^2/\|\varphi\|^2] \leqq 1$$

が成立するから，

$$I(\varphi, \phi) = -\frac{1}{2} \cdot \ell og\left(1 - \frac{\|\phi\|^2}{\|\varphi\|^2}\right)$$

と表わされても良いことに留意し，

①．１　２つのパターン φ，ϕ に関し，

非零定数 c が存在して，$\varphi = c \cdot \phi$

が成立すればするほど大きい値をとる情報量

$$I(\varphi, \phi) = -\frac{1}{2} \cdot \ell og\left[1 - \left|\frac{(\varphi, \phi)}{\|\varphi\| \cdot \|\phi\|}\right|^2\right]$$

①．２　パターン φ が，１次独立なパターンの集合 $\phi_k, k \in K$ によって，

$$\varphi = \textstyle\sum_{k \in K} a_k \cdot c_k \cdot \phi_k \|\phi_k\|^{-1} \wedge [\exists\ k \in K, a_k \neq 0]$$

と表現されればされる程大きい値をとる情報量

$$I(\varphi\,;\, c_k \cdot \phi_k \|\phi_k\|^{-1}, k \in K)$$
$$= -\frac{1}{2} \ell og\left[1 - \frac{\|\sum_{k \in K} a_k \cdot c_k \cdot \phi_k \|\phi_k\|^{-1}\|^2}{\|\varphi\|^2}\right].$$

この情報量 $I(\varphi; c_k \cdot \phi_k \| \phi_k \|^{-1}, k \in K)$ については，従来の回帰分析（regression analysis）への応用が期待される。

②．ファジィ集合論への応用について

②．1　二つのファジィ情報 x，y に関し，

$$0 \leqq y \leqq x \leqq h \quad \vee \quad h \leqq x \leqq y \leqq 1$$

ここに，h はしきい値

であれば，不等式

$$I(x) \leqq I(y)$$

を満たす情報量としてのファジィ情報測度（fuzzy information measure）

$$I(x) = -\frac{1}{2}\left[psn(h-x) \cdot \ell og \frac{x}{h} + psn(x-h) \cdot \ell og \frac{1-x}{1-h}\right]$$

ここに，$psn(u) = 1 \quad \text{if} \quad u \geqq 0, = 0 \quad \text{if} \quad u < 0.$

②．2　ファジィ集合

$$A = \{\langle x_j, \mu_A(x_j)\rangle \mid j = 1 \sim n\}$$

ここに，$\mu_A(x_j)$: x_j が集合 A に帰属する程度

内の各標識 x_j の規格化ファジイ情報測度

$$IZ(j, A) = I(j, A) / \textstyle\sum_{k=1}^{n} I(k, A)$$

ここに，

$$y_j \equiv \mu_A(x_j)$$

$$I(j, A) = -\frac{1}{2} \cdot \left[\mathrm{psn}(h_j - y_j)\, \ell og \frac{y_j}{h_j} + \mathrm{psn}(y_j - h_j) \cdot \ell og \frac{1-y_j}{1-h_j}\right]$$

□

なお，更に進んだ相互情報量の取り扱いについては，著者が現在研究続行中のパターン認識の数学的理論(34)での **fuzzy 集合論**でなされることを最後に論及しておこう。

文　献

(1)　堀部安一：情報エントロピー論，森北出版，1989—10

(2)　翁長健治他：情報システムの基礎，朝倉書店，1983—03

(3)　水本雅晴：ファジィ理論とその応用，サイエンス社，1989—06

(4)　松本　元・大津展之共編：ニューロコンピューティングの周辺，培風館，1991—07

(5)　鈴木昇一：認識工学（上），柏書房，1975—02

(6)　沢田康次：自然法則と情報処理，電子情報通信学会誌，Vol. 75，No. 4，pp. 344—349，1992—04

(7)　鈴木昇一：半順序と情処理，情報研究（文教大学・情報学部），Vol. 12，pp. 121—174，1991—12

(8) 有本　卓：情報理論（共立数学講座22），共立出版，1976—02
(9) 中川聖一：確率モデルによる音声認識，電子情報通信学会（コロナ社販取次販売），1988—07
(10) 山西健司，韓太舜：MDL 入門（情報理論の立場から，人工知能学会誌，Vol. 7，No. 3，pp. 427—434，1992—05
(11) C. S. Wallace, D. M. Boulton : An information measure of classification, Computer Journal, Vol. 11, No. 2, pp. 185—194, 1968
(12) Shannon, C. E. : A mathematical theory of communications, Bell system Technical Journal, Vol. 27, No. 3, pp. 379—423, 1948—07
(13) B. Roy Friedn : Restoring with Maxinum Likelihood and Maximum Entropy, Journal of the Optical Society of America, Vol. 62, No. 4, pp. 511—518, 1972—04
(14) 鈴木昇一：測度的不変量検出形認識系に関する研究，博士論文（工学院大学・博乙第1号），1975年3月22日
(15) 鈴木昇一：測度的不変量検出形認識系の構成理論，電子通信学会論文誌 D，Vol. 55—D，No. 8，pp. 531—538，1972—08
(16) David Forsyth et al. : Invariant Descriptors for 3—D Object Recognition and Pose, IEEE TRANS. on PAMI, Vol. 13, No. 10, pp. 971—991, 1991—10
(17) 鈴木昇一：画像情報量とその手書き漢字への応用，画像電子学会誌，Vol. 4，No. 1，pp. 4—12，1975—04
(18) 鈴木昇一，柴山秀雄，福永一保，古田普吾：認識の量子論と画像の微分エントロピー，芝浦工業大学研究報告理工系編，Vol. 23，No. 1，pp. 117—125，1979—03
(19) 鈴木昇一：パターン認識の数学的理論，第Ⅸ部（帰属関係あいまい度と認識情報量），電子（情報）通信学会技術研究報告［パターン認識と理解］，PRU87—28，pp. 1—10，1987—07
(20) 鈴木昇一：帰属係数法に基づく類似度，帰属関係あいまい度，認識情報量の計算機シミュレーション，情報研究（文教大学情報学部），Vol. 11，pp. 51—68，1990—12
(21) 丸山儀四郎：エルゴード理論とくに力学系のエントロピー理論，数理科学，Vol. 7，No. 4，pp. 36—40，1969—04
(22) 福井郁生：相関エントロピーと分散による画像の巨視的性質の評価法，電子通信学会論文誌，Vol. J68—D，No. 4，pp. 515—522，1985—04
(23) 田中初一，半田志郎：局所エントロピー・フローと情報源の符号化（定エントロピーの原理），電子通信学会論文誌，Vol. J65—A，No. 7，pp. 679—684，1982—07
(24) Werner Kuich : On the Entropy of Context-Free Languages, Information and Control, Vol. 16, pp. 173—200, 1970
(25) Ling-Hwei Chen : A two-phase area-level line/edge detector, Pattern Recognition, Vol. 25, No. 1, pp. 55—63, 1992
(26) 久保亮吾：統計力学（共立全書11），共立出版，1965—10
(27) 深尾　毅：大規模システムと統計物理モデル，電子情報通信学会誌，Vol. 74，No. 8，pp. 821—830，1991—08
(28) Shmuel Peleg, Azriel Rosenfeld : Determining Compatibility Cofficients for Curve Enhancement Relaxation Processes, IEEE Transactions on Systems, Man and Cybernetics, Vol. SMC—8, No. 7, pp. 548—555, 1978—07
(29) 中谷和夫：多変量解析，新曜社，1978—10，pp. 94—98
(30) 石井直宏，方野健次，岩田　彰，鈴村宣夫：距離尺度による時系列パターンの分類，電子通信学会論文誌，

Vol. J64—D, No. 2, pp. 93—100, 1981—02

(31) 大想俊一郎，篠原靖典，土井幹生：2次元場自己回帰モデルとピラミッドリンキングを用いたテクスチャ画像の領域分割法，電子情報通信学会論文誌，D—Ⅱ，Vol. J75—D—Ⅱ，No. 7，pp. 1132—1142，1992—07

(32) Solomon W. Gelomb : A New Derivation of the Entropy Expressions, IRE transactions on information theory, Vol. IT—7, No. 3, pp. 166—167, 1961—07

(33) Rosenfeld, A. and Smith, R. C. : Thresholding Using Relaxation, IEEE Trans. on PAMI—3, Vol. PAMI—3, No. 5, pp. 598—606, 1981—09

(34) 鈴木昇一：パターン認識の数学的理論

第Ⅰ部（考え方，PRL84—6，pp. 1—10，1984—05）

第Ⅱ部（認識抽象と公理系，定理系，PRL84—30，pp. 65—74，1984—09）……

第XⅦ部（再帰領域方程式と標本化，PRU92—1，pp. 1—8，1992—05）

第25部（画像前処理，PRU92—18，pp. 1—8，1992—06）

第26部（線形歪を持った多次元パターンの，モーメントによる正規化，PRU92—25，PP. 1—8，1992—09）

電子（情報）通信学会技術研究報告〔パターン認識と学習，パターン認識と理解〕

(35) Genis, C. T. : Relaxation and Neural Learning (Points of Convergence and Divergence), Journal of Parallel and Distributed Computing, Vol. 6, pp. 217—,244, 1989

(36) 長谷博行，米田政明，酒井　充，吉田順作：変動エントロピーによる文字変動の評価，電子情報通信学会論文誌D，Vol. J71—D，No. 6，pp. 1048—1056，1988—06

(37) 米田政明，長谷博行，酒井　充：文字の変動評価に関する一考察，電子情報通信学会論文誌D—Ⅱ，Vol. J75—D—Ⅱ，No. 1，pp. 103—110，1992—01

(38) A. Siozaki : Edge Extraction Using Entropy Operator, Computer Vision, Graphics, and Image Processing, Vol. 36, pp. 1—9, 1986

(39) 柳原圭雄，菅原徹雄，宇山親雄：X線映画撮影法を用いた冠動脈造影図における血管径計測のためのフィルタの基礎的検討，電子情報通信学会論文誌D，Vol. J71—D，No. 6，pp. 1141—1148，1988—06

(40) Hiromichi Yamamoto : A Method of Deriving Compatibility Coefficients for Relaxation Operators, Computer Graphics and Image Processing, Vol. 10, pp. 256—271, 1979

(41) 鈴木昇一：Rosenfeld型の確率的弛緩ラベリング法の基本的諸性質，情報研究（文教大学情報学部），Vol. 11，pp. 163—181，1990—02

(42) Robert M. Haralick. Stanley R. Sternberg and Xinhua Zhuang : Image Analysis Using Mathematical Morphology, IEEE TRANS. ON PAMI–9, NO.4, pp.532–550, 1987–07

(43) 岡本義則，中島秀之，大澤一郎：確信度と主観確率を持つ信念推論システム，人工知能学会誌，Vol. 7，No. 2，pp. 263—270

(44) 安西祐一郎：認識の情報科学への計算論的アプローチ，人工知能学会誌，Vol. 3，No. 3，pp.248—256，1988—05

(45) 月本　洋：連続値論理数学について，情報処理学会研究報告〔知識工学と人工知能〕，Vol. 88，No. 66，60—5，pp. 1—6，1988—09

(46) 月本　洋：命題論理の幾何モデル，情報処理学会論文誌，Vol. 31，No. 6，pp. 783—791，1990—06

付録1（18章，Ⅰ—iの $F(\varphi,\ \phi)$ の別例）

(イ) $\varphi(x,\ \phi(x)) \in \{0,\ 1\}$ の場合

$M(\varphi) \equiv \{x \in Z^n \mid \varphi(x) = 1\}$.

ここに，$Z = \{0,\ \pm 1,\ \pm 2, \cdots\}$ ((a set of integers)

Z^n : the set of n-tuples of integers

を定義し，

$$F(\varphi,\ \phi) = inf_{b \in M(\phi)} \varphi(x+b)$$
$$F(\varphi,\ \phi) = sup_{b \in M(\phi)} \varphi(x-b)$$
$$F(\varphi,\ \phi) = sup_{c \in M(\phi)} \eta(x-c)$$

ここに，$\eta(x) \equiv inf_{b \in M(\phi)} \varphi(x+b)$

(ロ) $\varphi(x)$，$\phi(x) \in E^1$ の場合(42)

E^N : N-dimensional Enclidean space

$A, B \subseteqq E^N$ に対し

$$A \ominus B = \{x \in E^N \mid x + b \in A \quad \text{for every } b \in B\}$$
$$A \oplus B = \{x \in E^N \mid x = a+b \quad \text{for every } a \in A \text{ and } b \in B\}$$

を用意し，

$\varphi : A \rightarrow E^1$

$\phi : B \rightarrow E^1$，ここに $A,\ B \subseteqq E^N$

として，

$$F(\varphi,\ \phi) = inf_{z \in B} \{\varphi(x+z) - \phi(z)\},\ x \in A \ominus B$$
$$F(\varphi,\ \phi) = sup_{z \in B,\ x-z \in A} \{\varphi(x-z) + \phi(z)\},\ x \in A \oplus B$$
$$F(\varphi,\ \phi) = sup_{z \in B,\ x-z \in A \ominus B} \{\eta(x-z) + \phi(z)\},\ x \in (A \ominus B) \oplus B$$

ここに，$\eta(x) = inf_{z' \in B} \{\varphi(x+z') - \phi(z')\}$ for $x \in A \ominus B$

付録2 (18章 I－i での x_j, a_j, b_j と，II－i での x_j, y_k, a_j, b_k との設定方法)

II－i と同様であるから，I－i の場合を説明しよう。

$x_j = SM(F(\phi,\ \phi),\ \omega_j)$ を次の様に考え直せばよい。

$$x_j = \begin{cases} SM(F(\varphi,\ \phi),\ \omega_j) & \text{if} \quad \sum_{k \in J} SM(F(\phi,\ \phi),\ \omega_k) \cdot BSC(F(\varphi,\ \phi),\ k) \\ & \qquad = 0 \\ \dfrac{SM(F(\varphi,\ \phi),\ \omega_j) \cdot BSC(F(\varphi,\ \phi),\ j)}{\sum_{k \in J} SM(F(\varphi(\varphi,\ \phi),\ \omega_k) \cdot BSC(F((\varphi,\ \phi),\ k)} & \text{otherwise} \end{cases}$$

a_j, b_j の変更についてはこれから推察がつくだろう。

(付録1，2了)

[〈情報学共同研究，新しい情報の測度とパターン情報処理，情報研究（第13号）投稿論文，鈴木昇一，1992年9月24日〕