「情報」に関する１３章

– 私家版・情報学入門 –

伊庭幸人

掲載元：　物性研究 78-2 (2002) p.172-193

この原稿は２００１年１０月に統計数理研究所で行われた研究会「動的システムの情報論」における研究発表と討論をもとに，とくに後半部分に大幅に加筆したものである．限定配布の報告集「動的システムの情報論」に収録されたあと，上記に掲載された．

目 次

1 いわゆる情報理論の初歩 4

2 しかし … 4

3 辞書 5

4 辞書はどうやって符号化するの? 7

5 汎化 9

6 予測 10

7 乱数の謎 11

8 危ない人たちと幻の規則 – 暗号幻想 – 13

9 共通の構造 16

10 統計学・情報理論・情報学 17

11 どういうカテゴリーの学問なのか 19

12 意図と差別 20

13 表現と近代とわたしたち 21

この小論の目的

情報理論や統計学にどういう問題意識が内在しているのか，なんとなく「感じ」がわかるようにしたいです．とくに，よくいわれる

- (シャノン流の) 情報理論は「意味」を扱えない
- 情報理論は情報量を求めるのが目的
- 新しい情報理論 (意味や時間を扱える) が必要

というような考えは，ある意味で誤解だと思うのですが，なぜそう思うかについても伝えたいです．

そもそも，「情報」ということばはよく意味がわかりません．うちの父もいつもそうこぼしていましたが，一般の人だけでなく，学問をした人でも同じじゃないでしょうか．「情報量」は「驚き度」である，っていうようなことをいいますが，「情報」ってのはそうすると，「驚きの源」でしょうか．当たらずとも遠からず，かもしれませんが，やっぱりいまいちよくわからない．さらに，「情報の学問」と「情報理論」の関係はどうなのでしょうか．このへんのことについても考えてみたいです．

とりあげる話題

以下のような問題について，ゆるく漫然と話をします．

- 情報量のパラドックス
- 汎化のパラドックス
- 乱数の検定をめぐる問題
- 危ない人たちの問題，あるいは，一般化フレーム問題

さらに，

- これらをより一般の文脈の中に位置づけるとどうなるか
- 情報学・統計学は科学か数学かそれとも?
- 物理科学，あるいは，「複雑系」との質的な違い

についても触れたく思います．

1　いわゆる情報理論の初歩

いわゆる情報理論の入門講義で教わることを復習します．シャノン情報量の定義は，

$$-N\sum_i p_i \log p_i$$

です．事象 i が起こる確率が p_i であるときに，長さ $-\log p_i$ の記号列を事象 i に割り当てるのが符号長の期待値を短くする上でベストで，そのときの符号長の期待値がシャノン情報量です．もとの記号列の長さを N としてます．もし，i の確率が q_i だと思ってる人がいたら，符号長の期待値は

$$-N\sum_i p_i \log q_i$$

となり，符号長の期待値は

$$N\sum_i p_i\{\,(-\log q_i)-(-\log p_i)\,\}=N\sum_i p_i \log\frac{p_i}{q_i}\ >0$$

だけ損になる．損の程度を **KL-divergence** といいます．これが正または零であるというのは $0<p_i<1, 0<q_i<1, \sum_i p_i=\sum_i q_i=1$ のもとで簡単に証明できる数学の定理で，これがすぐ上に述べたシャノン情報量の最良性を保証します．逆の方向，$-\sum_i p_i\log p_i$ に近い長さでの符号化が実際に可能だというのは別にやらないといけませんが，大体そのへんまでをやって学部の講義[1]はおしまいになるのではないでしょうか．もっといろいろやるかもしれませんが，大体の人が情報理論について覚えているのはこのくらいでしょう．

2　しかし …

さて，それで，実際の場面になると，

```
0100100010100111001110101001101110011
autreasjkloiuxdecollkfdryuhhgreuiolpjhgfreasjoijnbvcxzswert
```

のような記号列の情報量を計算してみろ，ということになるのですが …
そこで，あらためて

$$-N\sum_i p_i \log p_i$$

をみると，疑問がおきます．

確率 p_i をどうやって知るのか?

情報理論の初歩の本では，なにか謎の方法で p_i をあらかじめ正確に知っていることになっているのですが，わからないほうが普通ですよね，これは．とりあえず，n_i を事象 i の出現頻度，$N=\sum_i n_i$ を総数として，

$$p_i = n_i/N$$

とすることが考えられますが，よく考えると，

事象 i ってそもそも何 ?

[1]私は物理学科なのでぜんぜん教わらなかった．そんなんでいいのか物理学科 (笑)．

ということも自明じゃないです．

各文字，1 文字ずつの出現をそれぞれ「事象」とするのが一番簡単です．これは，

0 1 0 0 1 0 0 0 1 0 1 0 0 1 1 1 0 0 1 1 1 0 1 0 1 0 0 1 1 0 1 1 1 0 0 1 1

と区切ることに相当しますが，2 文字ずつ，さらに 3 文字，4 文字，5 文字ずつに区切ったものを「事象 i」とすることも考えられます．

01 00 10 00 10 10 01 11 00 11 10 10 10 01 10 11 10 01 (1)
010 010 001 010 011 100 111 010 100 110 111 001 (1)
0100 1000 1010 0111 0011 1010 1001 1011 1001 (1)
01001 00010 10011 10011 10101 00110 11100 (11)

1 文字区切りでは無視されていた相関の情報が 2 文字区切りには含まれ，さらに多くの情報が 3 文字区切りには取り込まれているわけです [2]．

多いほうがいいのであれば，考える文字数をどんどん増やしていった極限を考えればよい，ということになります．でも，問題は，データとなる文字列が有限であること，右側に無限を示す「…」が無いということです．それで，最後までいくと，区切りの長さが文字列の長さと同じになってしまいます [3]．

0100100010100111001110101001101110011

そうすると，「いま目の前にある文字列」が i のとき $n_i = 1$，それ以外の i については $n_i = 0$ ですから，情報量は，$0 \log 0 = 0$ として，

$$0 \log 0 + 0 \log 0 + \ldots + 1 \log 1 = 0$$

これはもとの列の長さがどんなに長くても，どんな列でも同じですから，われわれは次の「大定理」を得たことになります．

「どんな列の情報量もゼロ」

考えてみるとこれは当然かもしれません．無理矢理に，

「私は autreasjkloiuxdecollkfdryuhhgreuiolpjhgfreasjoijnbvcxzswert を 1 と呼ぶのだ！」

という人にとっては，この文字列はとっても簡単なわけで，呼ぶにも 1 ビット (0 ビット?) で済んじゃうわけです．これは「私的言語」と呼んでもいいでしょう．

3 辞書

「情報理論の意味がわかってないからそんな変なことを考えるんだ」．たしかにそうです．情報というのはアンサンブルに対して相対的に定義されるものですから，「私的言語」で表現されていても意味がありません．逆にいえば，そういう「情報」の性質をあらわすために確率というものを導入したともいえます．しかし，「情報量」の裏にあるそういう構造は，講義を一通り聞いただけ

[2] 嘘です．これがいえるには，区切りをずらした区切り方を全部考えなきゃいけない．いまのような確率モデルでこれをやるのは面倒で，むしろ「0010 のあとが 1 になる確率」のような見方– 高次のマルコフ連鎖– で考えたほうが便利ですが，それでやっても，このあと本文で述べるのと本質的に同じことが起きます．

[3] マルコフ連鎖であれば，長さ N の列に $N-1$ 次のマルコフ連鎖を考えた場合に相当．

では十分意識されないことが多いのではないでしょうか．また，「情報理論の意味がわかっている」人にとっても，実際の応用局面で，何を「要素となる事象」とみて，その確率をどう推定するかは決して自明ではないでしょう．

逆に，ことの深さに気づいている読者は「こういう問題に一般的に答える『理論』は無いに違いない」と考えるかもしれません．それは正しい直観かもしれませんが，あることに決定的な答えが無いということは，必ずしもそのことについて考えるのが無駄だということにはならないと思います．それで，もう少し話を続けましょう．興味ある考え方のひとつは次のようなものです．

解読のための「辞書」(対応表) を送る手間も含めて考える

折角，情報を送ってもそれを解読する「辞書」がなければ，相手にとって意味が無いはずです．それで，「辞書」を送る手間も考えなさい，というわけです．この考え方を MDL 原理 (Minimum Description Length principle) といいます．

MDL 原理 (最小記述原理)
「辞書の長さ」+「その辞書での最短記述長」を最小化

後半の「その辞書での最短記述長」がシャノン情報量 $-\sum_i p_i \log p_i$ に相当します．そうすると，

「私は autreasjkloiuxdecollkfdryuhhgreuiolpjhgfreasjoijnbvcxzswert
を 1 と呼ぶのだああ」

というためには，autreasjkloiuxdecollkfdryuhhgreuiolpjhgfreasjoijnbvcxzswert という列を辞書の一部として送らなければならないことになります．明らかに，これでは圧縮しないのと同じというか，かえって長くなるというか，駄目ですね．第 2 項がゼロになったぶん，第 1 項の辞書が延びてしまってます．一般に，第 1 項は長さ $O(1)$，第 2 項は $O(N)$ ですが，N が有限なので，オーダーの差で無視というわけにいかないということがポイントです．

より一般に考えると，「辞書」として，いちいち符号化法のリストを送らなくても，一般的な符号化法を決めておいて，「i に対して p_i の近似値 q_i を与えるプログラム」を送るという方法も考えられます．もうこうなると，q_i を計算するプログラムのサイズさえ小さくできれば事象 i はうんと複雑なものを取っていいことになります．いちいち，q_i の表を送る必要はないですから，「i に対して p_i の近似値 q_i を与えるプログラム」というのは，ある意味，確率モデルともいえますが，この意味で

「辞書」 $\Leftrightarrow$ 「確率モデル」

です．たとえば，ある量 y が値 $y_i < y \leq y_{i+1}$ を取ることを事象 i，事象 i の確率の推定値を q_i とすると，(離散化された) 正規分布

$$q_i \sim \exp\left(-\frac{(y_i-\theta)^2}{2\sigma^2}\right)\triangle_i, \qquad \triangle_i = y_{i+1} - y_i$$

のようなものも「辞書」と思っていいわけです．

MDL 原理という名前を考えた人は，一般に「統計モデル」というもの，さらに，世の中の「法則」「ルール」というものをこの延長でとらえようとしました．「モデルの当てはめとは 2 段階符号化である」というのがそのスローガンです．そこでは，「できるだけ簡単なモデルで近似的にデータを表現しようとする試み」は「最小記述を得ようとする試み」の一種として理解されます．たとえば，一般に，いくつかパラメータ θ を持つモデル $q_i = P(i|\theta)$ での当てはめでは，考えられるモデル P とパラメータ θ のうち

$$-\sum_i n_i \log P(i|\theta) + (\theta \text{ の符号長})$$

を最小化するような P, θ を選べばよい，ということになります．これだと，最小 2 乗法もテキストの圧縮も同じ枠組みで理解できる上に，モデルの当てはめだけでなく，モデルの選び方まで理論化できてしまう．すごい．

4 辞書はどうやって符号化するの?

誤魔化されてます．たぶん．辞書はどうやって送るのでしょうか．辞書も符号化するのでしょうか．辞書のパラメータ，というか，分布のパラメータ θ はどうやって送るのでしょうか．たとえば，θ が離散的な値をとるとして，$\theta = \theta_i$ である確率 π_i がわかれば，θ_i に長さ $-\log \pi_i$ の長さの符号を与えてやれば，パラメータをベストの方法で符号化したことになります．でも，これって，実はもとの問題に戻っちゃってるんじゃないか．いや，辞書の辞書の辞書 .. と階層を上げていくことで，どこかで，「誰もが納得する符号化」に達すればいいのですが ...

これは，ベイズ統計学における「事前分布」の問題と対応するものです [4]．そしてここで問題にしているのは，

「標準的な事前分布」はあるか?

というこの分野でおなじみの問題です．これにはいろいろな答があって，答え方によって

399 通りのベイズ統計学がある

といわれています．MDL 原理派の人はもちろんそれなりの答えを用意していて，「自分たちは問題を数学的に解決したから，単なるベイズとはひと味違う境地に達した」と考えています．それを詳しく紹介することはしませんが．別の話題にうつる前に，ちょっとだけ覗いてみましょう．

まず，「自然数」に関して，誰でも納得するような情報表現の方法があるか．という問題からはじめます．そんなのあるわけないと思うかもしれませんが，しかし，みんな 10 進法や 2 進法を「標準的」に使ってますね．こういう「普通の記数法」では，

n の符号長 $\sim$ 大体 $\log n$

となります．これは，

$$\pi(n) \sim 1/n$$

に相当します．ん，$\sum_n \pi(n) = \infty$ じゃないですか? これは，普通の記数法が「数字の区切り」の部分を考慮してないからで，実はそれを考えると，

n の符号長 $\sim \log^* n$

となります．ここで，$\log^* n$ は $\log n + \log\log n + \log\log\log n \ldots$ ととっていって，log の中が負になる手前で止めたものです [5]．これを使うと，対応する $\sum_n \pi(n)$ はちゃんと有限になってくれます．

[4] パラメータ θ に想定する分布 $\pi(\theta)$ のことを事前分布 (prior distribution) といいます．

[5] この長さまで伸ばせばよいというのを実際に構成してみせるのに，巧妙な再帰的議論 – 区切りの位置をあらわす表の区切りを第 2 の表で示し，第 2 の表の区切りを ... とやる – があって結構楽しいです．情報理論の人はこういうのが好きで研究やってるのかな，と思ったりします．

さて，自然数についてこの記数法がなぜ「特権的に良い」のかという議論ですが，MDL 原理の人の考え方は，ある条件のもとでの「ミニマックス性」をいうことです．実際 $\pi(n)$ が「平均 1000・分散 1000^2 の正規分布」なんていう仮定で記数法を構成していたら，10 万円貰ったら即アウト，紙に金額の数字が書けなくなり，「そんな金はわしには無限大じゃー」になってしまいます．で，「真の $\pi(n)$ 」が確率分布 $\sum_n \pi(n) = 1$ である限り，何であってもそれほど破滅しない無難な記数法だからいいのだ，というのがミニマックスの考え．とても数理的な方向ですね．逆方向にいちばん極端なのは「みんな 2 進法とか 10 進法と使ってるじゃん．だからいいの．」という考え方でしょう．経験的真実．この経験は，$\pi(n)$ のほうに直すと

$$\pi(n) \sim \text{大体 } 1/n$$

というのに相当します．これは，Zipf 則とか $1/f$ 揺らぎとかいうのに関係ありそうです．そういう「自然法則」，n が金額だったら「社会法則」，が経験的にあって，それが記数法を決めている，という議論もできそうです [6]．一方で，こうした「べき法則」を「導こう」という議論もあります．たとえば，2 つの確率変数の間のかけ算で決まる確率変数の分布，というのを再帰的にどんどん考えていって行き着く先を考えるとか．そうすると，$\pi(n)$ はまた数理的に「導かれる」．うーん，実はどの考えもみんな同じようなものかもしれませんが ... わかりません．

それで，MDL 原理派の戦略は，自然数 n についてのこの符号化から出発して，一般のパラメータ，プログラムの符号化に到達しよう，というものです．議論は省略しますが，ひとつの本質的な困難は，パラメータの範囲が有界でないときには，「考えている世界の大きさ」のような「不定」の定数が最適符号化に入ってくることです．次元のある量を自然数に直して符号化するのですから，一般には「基本的な長さ」L みたいなものがないといけない．で，これが最終の符号長に $\log L$ に比例するような不定の項をもたらします．符号化同士，モデル同士を比較するとき，この項がキャンセルすればいいのですが，パラメータの種類や個数の違う場合にはそれが陽に出てくるので，折角の数理的議論がいまいちしり抜けのようになってしまう．直観的にいって，3 次元のものと 2 次元のものを比較しようとしたら，2 次元というのは本当は「1 つの方向が薄い 3 次元」だ，と思わないと無理っぽいですよね．この薄い方向の厚さが $1/L$ だと思ってもよいです．これは，実は，「improper prior の困難」あるいは「Lindley paradox」としてベイズ統計ではよく知られた問題 [7] です．そういうわけで，

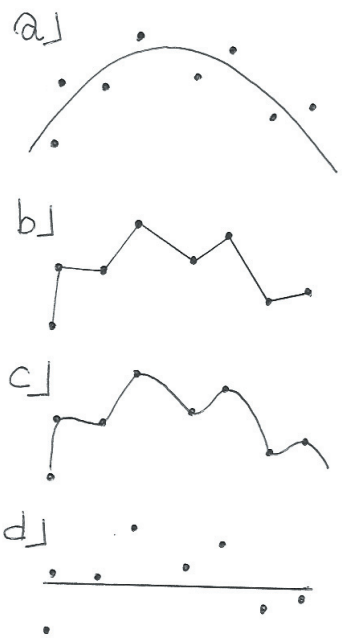
やっぱり **MDL**ってベイズじゃん

と思ってしまういけない私です．

[6] これは私のオリジナルかと思ったら，サイエンティフィック・アメリカンのコラムに出てたとか誰かいってました．

[7] ある統計学者によれば，「それって哲学みたいな問題でしょ，あんまり重要じゃないよ．」だそうです．先生，新しい特殊関数の研究頑張ってください ...

5 汎化



学生実験でグラフを書くとき，データを (b) のように繋ぐと怒られます．(a) のように「滑らかに」つなぐのが正しいというのです．(b) はまだしも，(c) となると，確かに変ですよね．点の上を線が通っていれば，これ以上正しい繋ぎ方はないはずなのに．といって，(d) のように線を引く人は少し変人かもしれません．ここで，曲線の方程式があらかじめ予想されていない場合，手で線を引くか，「最小二乗法」で多項式などをあてはめるかの違いは本質的ではありません．20 個の点に 19 次式，1000 個の点に 999 次式をあてはめれば，やはり (c) のようになってしまいます．そして，残差の最小二乗和でいえば，(b) や (c) がやっぱり一番いいのです．なにしろゼロなわけですから．

人間の代わりにロボットを考えてみます．遺伝的アルゴリズムか何かでロボットに歩き方を学習させることを考えます．うまく歩けなかった奴はボツにして，ましなやつを選抜してランダムに改良し，またテストをする．すると，確実に歩き方が上手になる [8]．すばらしい．しかし，この方法で「何が」学習されているのでしょうか．ロボット達がしていることが，

- 歩き方一般を学ぶ
- 訓練に使った廊下の構造を学ぶ
- その廊下のでこぼこやごみまで全部学ぶ

のどれであるかはわかりません．「起こりうる悪いことは全部起こる」という原則を信じれば，最後の項目である可能性が高いでしょう．特定の廊下ではなく，もっといろんな場所で訓練して総合点で選抜したら良いのでしょうか．しかし，「頭のいい」ロボットなら，それも全部，まるまる暗記してしまうかもしれません．

あきらかに，これは狭義の「学習能力」の大小の問題ではありません．実際，学習能力の大きいほど何でも丸暗記してしまい，ますます悪くなる可能性があるのです．さっきの曲線の例でいえば，曲線を表現する多項式の次数が高いほど (適応できるパラメータの数が多いほど)(c) に近くなる，というのにこれは対応しています．ニューラルネットの応用分野ではこれを「汎化」の困難の問題といいます．統計学の用語では overfit の困難ということになります．

では，人間はどうして学習できるのでしょう．「学習能力」が低いからでしょうか．ある意味ではそのとおりです．しかし，ただ単に低いだけでは (d) のようになってしまうかもしれません．長い歴史の中で環境に適応して，いや，環境と共進化してきた結果，「うまく学習できるような素質」「何を学習したらいいのかわかる素質」を獲得しているわけです．じゃあその適応はどうして獲得されたかというと，人間が人間になる以前からの素質が ... 本当でしょうか．少なくとも，部分

[8] 嘘ですよ．はじめに全員こけたらどうするんだ?

的には本当だと思うのですが．それから，学校の試験などでは丸暗記する人は必ずいますから，いつも何を学習したらいいのかわかる素質があるわけではありません．これは「試験」が人間にとって環境として異常なものだからとも考えられますね．

6 予測

今のような問題，たとえば，

なぜ，3 次式をあてはめるより，2 次式をあてはめたほうがいい場合があるのか?

というような問題を数理的に定式化することができるでしょうか．ひとつの方向は，

「予測の最良性」

に着眼することです．ここで「予測」というのは「未来」のことを論じることではありません．そうではなくて，過去でも未来でも現在でもいいから，「知らないこと」について論じることです．そもそも，過去とか未来というのは物理学の概念です．「情報」の世界で意味を持つのは「知っているか」「知らないか」の違いなのです．それで，

3 次式はいまあるデータに対しては良いけれど，予測の能力については 2 次式に劣る場合がある．このような場合には 2 次式のほうがよい．

という風なことがいいたいのですが，「予測能力」について語るためには，「今ここにないけど，あるかもしれないもの」を考えること，いいかえれば，データの背後にアンサンブル (「可能性の世界」) を想定することが必要です．

それを考えるには，

1. データにあてはめた確率分布から模擬データを発生させ，それに同じ方法であてはめをやり，その結果に基づいてモデルを自己評価する．
2. データの一部で学習して，残りで「実力テスト」をする．
3. 「事前分布」からモデルやパラメータが発生したとする．

というような方法があります．最後にあげた事前分布を考える方法はベイズ統計と呼ばれていますが，これは，さきに情報圧縮 [9] について述べたときのアプローチと深い関係があります．一方，1. の考えに基づくのが，AIC (赤池情報量規準) です．1. の考えはちょっと面倒ですが，データにあてはめた確率分布それ自体が「ありうるデータ」を仮想的に発生させる能力があるということをフルに使えば，事前分布，確率分布の確率分布というようなことを考えなくてもいいだろうという考えです．こういう考えを統計学では「正統派」とか「頻度主義」とかいいます [10]．

AIC の論理の詳しい説明や導出 [11] や赤池先生がどんな顔か [12] ということはここでは省いて，結果だけちょっと見ておきます．

[9]「情報圧縮」が「予測」に関係あることは，たとえばある状態で人間が右に動くだろうと予測できれば，左に動いたときだけ通報すれば良いという例からも明らかでしょう．これはさきにシャノン理論について論じたことの言い換えです．i を与えて p_i の近似値 q_i を計算するプログラム，とは，確率的予測のプログラムに他なりません．

[10] いうまでもなく，「頻度主義」というのはベイズの人の好む用語，「正統派」ってのはベイズじゃない人の好む用語ですね．

[11] AIC の導出の数理は尤度比検定のそれとかなり重なります．

[12] 私が撮影したポートレートが http://tswww.ism.ac.jp/kitagawa/HomePage/index.html にあります．ちなみに，研究のお話を直接伺ったことはほとんどありません．

「 $\log P$(データ $|\theta)$ - (θ の次元) 」を最大化するモデル・パラメータを選べ

ここで P は任意の確率モデルで θ はそのパラメータです．θ をベクトルと思っているので「θ の次元」と書きましたが，「個数」でもいいです．分散 σ^2 のガウス雑音下での曲線 $y = f_\theta(x)$ のあてはめなら，

$$P(\{y_i\}|\theta) \propto \exp\left(-\frac{1}{2\sigma^2}\sum_i (y_i - f_\theta(x_i))^2\right)$$

ですから，

$$\frac{1}{2\sigma^2}\sum_i (y_i - f_\theta(x_i))^2 + (\theta\text{の次元})$$

を最小化することに帰着します [13]．2 次式なら θ の次元は 3，3 次式なら 4 です．最小 2 乗誤差 $\sum_i (y_i - f(x_i))^2$ は 3 次式のほうがつねに小さいのですが，おまけの項の差が 1 だけあるので，3 次式にしてもそれだけ稼げなければ 2 次式が勝つ場合もあるわけです．大事なパラメータも，どうでもいいパラメータも，おまけの項に対して同じ寄与で，個数だけが効く，というのはちょっと面白いですね．どうでもいいパラメータはデータの変動に対して敏感で大きく揺らぎますが，その揺らぎは (もともと大事じゃないパラメータなので) あまり予測に効かない，というので打ち消しあって同じになります．統計力学の「等分配の法則」みたいですが，直接の数理的対応はつかないようです．

2. の考え (cross validation) でもや 3. の系統の考え (MDL,BIC, ベイズ) でも同じ式が出てくれば，めでたしめでたし，ベイズでも頻度主義も同じこと，予測も情報圧縮も統一的に扱えて良かったです，ということになるのですが，実は．1. と 2. は近い結果になるのに 3. はちょっと違う，ということがわかっています．なんじゃこりゃ，というわけで，ずっと揉めてたりして，それがまた面白いのですが，ここではそこまで深入りしません．

こうした数理的結果は，実際にパターン認識やいろいろなデータ解析，システム制御などで使われていますが，ロボットにものを教える問題などにどれだけ有効かは難しいところです．いわゆる人工ニューラルネットなどでは，非線形効果が強いので，AIC のような「線形揺らぎ」レベルの議論がうまくいくかどうか怪しいという問題もありますが，おもな困難は，「人間が使っているモデルの候補」を明示的に見出すのが難しいことにあります．

7 乱数の謎

少し違う話題として「乱数」について考えてみます．以前，「人間に乱数を書かせるとどうなるか」という研究をしていた時に，共同研究者に「乱数の定義は何ですか」と聞かれて参ったことがありました．まったく，朝から乱数の定義だなんて！ もちろん，「乱数の定義」はうっかりと口にしてはいけない話題です．うっかり聞くと情報の鬼もはだしで逃げ出すというくらいのものです．

次の疑問について考えてみましょう．

なんで乱数の検定法ってたくさんあるの?

さっきの質問をしてきた人は，「乱数は頻度で定義すれば良い」という意見のようでした．何の頻度かが問題ですが，たとえば，十分大きい m について，「m 個づつブロックに区切ったときに 2^m 種

[13] ここでは既知とした分散 σ^2 も，最尤法などでデータから推定できます．じつは，これは，実用面でも思想面でも重要なポイントで，推定法についての異論もあります．

類がほぼ均等な頻度に出現」という条件ではどうでしょう [14]．とりあえず，普通考え付くような規則性は大体これで捉えられるように思えます．ところが，実際には，乱数の検定というのは沢山の種類があって，それをいちいち試してみるのが普通のようなのです．その中には，ブロックに区切って頻度の一様性を見るという方法もあるのですが，衝突検定，連の検定，ポーカー検定，ランダムウォークを使った検定など，それ以外のいろんなやり方が本には出ています．何故でしょう．みんなが因習にとらわれていて，合理的な思考ができないからでしょうか．

確かに，無限列の場合，あるいは，考えに入れる規則性のパターンの長さの上限 m にくらべてずっとずっと列が長い場合には，「m 個づつブロックに区切ったときの 2^m 種類の頻度」というのは，考えられる規則的パターンのすべてを捉えているように思われます [15]．ところが，実際の列は有限列です．たとえば，254 個おきに確率 0.99 で 1 が出るような数列があったとします．この規則性を検出するには，少なくとも m が 256 のブロックの頻度を考える必要があります．$m = 256$ のブロックのありうる個数は 2^{256} 種類もあります．これは途方もない数です．これに対応する検定は自由度 $2^{256} - 1$ の χ^2 検定になりますが，これで問題の規則性を検出するには，ぼう大な長さの列が必要です．一方，「254 個おきの数字を調べる」あるいは一般に「いろいろな m について m 個おきの数字を調べる」という方法を思いつけば，ずっと短い列についても，規則性を捉えることができます．実際に，こんなに明確な規則性ではありませんが，ある疑似乱数について，250 個とかその位離れた位置での相関が原因でシミュレーションがおかしくなった，という例が報告されているので，この例は非現実的ではありません．一般に，検定に関しては，

大は小を兼ねない

のです．それでいろんな検定が必要になります．これは，AIC でパラメータの多いモデルが選択されるとは限らないということや，MDL で辞書が長いほど圧縮度が良いとは限らないことと，パラレルです．データの有限性の故に，揺らぎが無視できず，一般の仮説 (「大」) は特殊な仮説 (「小」) よりも揺らぎに弱いのです．χ^2 検定の数理はそれを表現しています．

少しだけ，無限列での「乱数の定義」の話にも触れておきます．「乱数の定義」については，万能チューリング機械を与えたときに，「ある列を表現するために必要な最短のプログラムの長さを乱数度とし，同じ長さの列で乱数度が最大のものが乱数」というのが，よく知られています (圧縮不能性による定義)．そして，与えられた列を表現する最短のプログラムの長さの計算が「決定不能問題」，つまり万能チューリング機械の上のプログラムによって原理的に計算不可能なことが，「鬼がはだしで逃げる」理由だと普通は考えられています．これと，無限列では妥当なように見える，ブロックの頻度の漸近的一様性による乱数の定義との関係はどうなのでしょう? 後者ではチューリング機械はいらず，鬼も逃げないように見えますが ... 本稿の主題からはちょっとはずれますが，考えてみます．ポイントは，無限列の規則性として，「自然数 $n = 1, 2, \ldots$ について，n^2 番目の要素はつねに 1」のようなものも可能だということですね．このような規則は，任意の m について m 文字のブロックの頻度の漸近的な一様性が保証されていても，それと両立します．規則的に存在する 1 の密度が先のほうでだんだん薄くなっていくからです．こういう隠れた規則性を一網打尽にするためには，「任意の部分列，任意の m について，m 文字のブロックの頻度が一様」としたいところです．ところが，この「任意の部分列」を普通の数学の意味にとると，「列をみてから部分列を取り出す」という「あと出し」が阻止できません．列をみてから，1 のところだけを部分列として取り出してもいいなら，1 が無限に含まれるどんな無限列からも，$11111111\ldots$ という全然ランダ

[14] 以前と同じように，m 通りある区切り方の全部を考えるか，マルコフ連鎖的に条件付き確率で考えるかしないといけません．ここでは，問題は「均等」ということだけですから，マルコフ連鎖を考えなくても，どういう区切り方をしても均等，ということでいいでしょう．

[15] 実はこれには落とし穴があるのですが，それは後回しにします．

ムでない列が取り出せます．そこで，万能チューリングマシンとか，その上でのプログラム (アルゴリズム) とかを定義して，「あと出し」の防止をしなきゃならなくなる．そういうものを定義しておいて，「何らかのアルゴリズムによって指定できるすべての部分列」とやるわけです．それで，頻度から出発しても，結局はチューリングマシンのようなものが必要になって，非決定性とかも導入してしまうことになります [16] ．これは「チューリングマシン」や「アルゴリズム」が必要とされる典型的な状況です．介護保険の一次判定をコンピュータプログラムでやるというのも同じ発想ですね．なお，「頻度から出発する方法」は統計学における「頻度主義」，「正統派」に立場が近いのに対し，万能チューリング機械に基づく方法は「ベイズ統計」に対応 (規準となるチューリング機械の選び方が「事前分布」の選び方に対応) するのですが，上の議論だけみるとここではベイズ有利の感じもします [17].

無限列の話は学問ぽい感じだけど，はじめの方でした有限列についての話 [18]は，なんだか頼りないという人もいるでしょう．しかし，本当に重要なのはそちらの方ではないかと思います．「大は小を兼ねない」をいいかえれば，

「複雑さ」は無限に多様である．

怪しい英語では，

Complexity is complicated.

もっとちゃんというと，Complexity is a complicated matter. でしょうか?，ということになります．

いわゆる「前期」のウィトゲンシュタインは，「そして，語りえないものについては語ってはならない」といいました．「語り得るもの」と「語り得ないもの」の間に線を引く考え方は，「無限列」の世界で「決定可能」と「決定不能」の間に線を引くことに対応していると考えることができます [19] ． これに対して，「後期」には，「説明」や「教育」の立場から，ことばの世界の多様性が語られます．こちらは，「有限列」の世界，統計学的な世界に対応すると考えられないでしょうか．柄谷行人が「探求 I，II」で語っている「他者とは超越的他者ではなく，そのへんに普通にいる他者である」というテーゼも，私には同じような意味に思えます．

8 危ない人たちと幻の規則 — 暗号幻想 —

ここでは「暗号幻想」の話をします．話といってもいろいろ例を並べるだけですが，人工知能の人のいう「一般化フレーム問題」にも通じる例だと思います．ただし，みんな人間の話です．

- シェークスピアの暗号
  シェークスピアの作品が，実は別の人の作で，しかもそのなかには「暗号」でメッセージが組

[16] これは序の口で，詳細はもっと難しいようですが，ここでは，わかった (ような気がした) ことだけを説明しました．Knuth の本の「準数値算法 乱数」の巻 (和訳，第 2 版) に詳しい議論がありますが，私はサトリをひらいてしまったので，ちゃんと勉強していません．なお，Knuth の新版 (英文) の該当部分はだいぶ変わっているらしいので，本気で興味のある人は新旧両方をみたほうがいいでしょう．

[17] じつは有限列の場合も，いろいろな「仮説」の間の関係や「仮説の出所」まで考えると頻度主義は苦しい面があります．しかし，「仮説は人間が作るものだ，その根拠は問うてはならないのじゃ」という風に主張するのもありかもしれません．これは科学哲学における「反証主義」を想起させます．反証主義の始祖であるポパーが，若いころ，頻度に基づく乱数の構成に凝ったことがあるというのは，「果てしなき探求—知的自伝」に出てきますね．

[18] $n \to \infty$, $m \to \infty$ での検定の漸近的な検出力を論じているとみれば，「無限列」の話なわけですが，有限性からくる効果を数学的に定式化するために，そうした議論が必要になるのだと考えたいです．一般に，数理的な結果を得るためには，何らかの極限操作が必要なことが多いですが，その場合も，どういう極限でどういう量を論じることが現実的かを常に反省しなくてはならないという点に，「有限性」が反映されます．

[19] 別の解釈として，これを，「反証主義」における「仮説の検証」と「仮説の創造」に対応させることもできますが．

み込まれている，というのは，ずいぶん沢山の人を熱中させたテーマのようです．何ページおきに何行目のアルファベットを見て，というのを複雑な数式で決めるやつとか，もっとすごいのは，活字の形が 2 タイプあって，それでメッセージが書かれているとか．後者は古い活字の形が一定しないで字ごとに少しづつ全部違うということに関係しています．熟練した「解読者」が見れば，それを 2 種類とかにグループ分けできて，その記号で秘密のメッセージが書かれているというわけです．

- **DNA** は暗号であるか
  似たような話で，DNA の同義コドンに関するものがあります．コドンというのは，アミノ酸を指定するための塩基の 3 つ組のことですが，同じアミノ酸を指定するのに複数の仕方がある場合があります．たとえば，アルギニンは TTG，CTT，CTC，など，手元の表によれば 11 通りのコドンで表現できます．日本の人ですけど，こういう同義性が，実は暗号に用いられている，という学説を唱えた人がいます．確か，ちゃんとした大学の先生だったように思うのですが，一般向き雑誌の記事には「解読した記号をたてよこ 2 次元にならべた絵」というのがあって，「手を上げて挨拶している人の姿が見えないか？」って．見えないってば．その絵，統計学の説明に使えると思ってコピーしてとってあったのですが，なくしてしまったようです．統計的有意性とは別に，進化の過程やウイルスの増殖の過程で，同義コドンは他の同義コドンで速く置換されるので，折角の宇宙人のメッセージも急激に消えてしまうと思うのですが，どうでしょう．

- オサマ・ビンラディンの手
  世界貿易センタービルのテロ事件のあと，関連を疑われているビンラディン氏の登場する映像が流されました．そのとき，話の内容とは別に，「手の合図」で世界の同志に秘密のメッセージを送っているのではないか，という説が出ました．この説は，人工知能の論文 [20] に，一般化フレーム問題の例として出てくる「子供の遊びの話」を想起させます．この遊びは，「いろいろなことを言っては相手に○か× かを答えさせる」というものです．「正解」かどうかを毎回教えてくれて，それによって何が○，何が× と分類されているのか速く学習しろ，というわけですが，まじめに考えてもできません．何故なら，実は○か× は言っていることに関係なくて，そのとき右手で上を指しているか，下を指しているかによるからです [21]．この遊びは 2 人でやっても意味がありません．大勢でやって，気がつかない人，最後までコードを共有できない人を仲間はずれにして笑う，という楽しい遊びです．ビンラディンの手のことを言い出したのは，もしかすると，これでいじめられた経験のある人かもしれませんね．

  こういう問題では，普通の意味で「データ」から規則を見つける，もしくは見つけたと信じ込む，というのと違って，どこまでが「データ」であるかの不定性が問題になっています．こういうのを，人工知能やロボット系の人は「アテンションの問題」というようです．もちろん，「データ」の範囲をはじめに十分広くとれば，同じことかもしれないので，きっちり区別できるわけではないですが「シェークスピアの暗号」も，活字の字体を問題にするとなると，これに近いでしょう．

- レイ・ライン
  古代遺跡などの地図上の配置をみると，それが一直線に並んでいる，あるいは，いくつかの

[20] 松原仁，橋田浩一「情報の部分性とフレーム問題の解決不可能性」，人工知能学会誌，Vol.4 No.6 pp.695-703(1989).
[21] 上の注の論文をちゃんと確認したところ，遊び「いいでんでん虫」及び「この置き方は誰」はもっとずっと洗練されたものであることが判明しました．お子様方にお詫びするとともに，興味を持たれた読者には原論文を参照されるようお願いします．

直線群の上に載る，という説が『レイ・ライン』と呼ばれるものです．英国が本場ではないかと思うのですが，日本が舞台のもあるようです．それでどうなのかは，私はよく知りません．やっぱり宇宙人がらみなのでしょうか．なんでよく知らない例を出したかというと，以前，ちゃんとした統計学の雑誌 (たしか Jounal of Royal Statistical Society A だと思います) に『レイ・ラインの研究』というのが載っているのをみかけておどろいたからです．確かに，仮説としてはずいぶん怪しげですが，きちんと反論しようとしたら，論理的にも数理的にも結構難しいのではないかと思います．数学としては幾何確率の問題ですね．

- 空のマッチ箱
  誰にでも起こりうるような状態，精神分裂病についての話です．ある種の状態では，われわれは，誰かにつけ狙われているとか，考えを注入されたり盗まれたりしているとか，毒を盛られているとか，むやみと思ってしまうことがあります．誰もいないのに，悪口や命令する声が聞こえたり，眠れなくなったりもします．これは脳の病気だと思われているわけですが，MRI とか血液の検査をやってはっきりとした異常がでるわけではないです．まあ，失恋とか恋愛とかだって脳の病的な状態でしょうし，症状がその人の個性がなくなる方向に出る点も同じです．ただ，分裂病の場合，放っておくと悪くなることがあるし，薬で症状を効果的にコントロールすることができるので，変な声とか聞こえる人[22]は一応病院に行ってみたほうがいいです．

  それで，原因はよくわからんとしても，こういう状態での症状を体系的に説明してみたい，というのがあります．精神病理学というやつですね．たとえば『変な考えが自然に出てくる』というのを中心にして『だから，不安になったり，寝れなくなったり，人付き合いが悪くなったりする』というのは，平凡だけど，ひとつの筋道の付け方です．本稿の立場から面白い考え方としては『『偶然』がなくなってしまうような病的状態だ』という解釈があります．この見方だと，ある種の知覚過敏のような症状と「変な考えが出てくる」ことを統合して考えられます．たとえば，机の上の空のマッチ箱をみて『俺がからっぽな人間だということをいうために誰かがここに置いたんだ』といって怒る人のことを考えてみます．ずいぶん変な結び付け方なので『脳がひどく混線してるのだろう』とか「むかしの経験・精神的外傷の反映だろう」とかいう解釈もできます．しかし，問題点は「世の中には無意味なことがある」ということを許容できないことにある，と考えたらどうでしょうか．特別な理由がなくて，そこにマッチ箱があるのに，無い「理由」をうんと考えるので，自分に結びつけた，変な考えになってしまうと考えるわけです．これは中井久夫の説，というかそのほんの一部ですが，以前に書いたことがあるように，とても統計学的な考え方だと思います．そして，ここで述べた他の話とも関係があります．

- 画像深層暗号
  画像深層暗号という分野があって，単行本も出ています．本当の暗号というよりも，隠し絵のようなものです．たとえば，画像をフーリエ変換して圧縮するときに，フーリエ係数の最後のビットを切り上げても切り下げても，見た目はほとんど変わらないでしょう．それで，最後のビットが 1 か 0 かに情報を入れておけば，それを知らない人には単なる画像にみえるのに，秘密のメッセージが入っているようにできるわけです．これは，本当に情報が隠されているような例です．

[22] 寝る寸前のやつは半睡時幻覚といって普通の人でもあるので大丈夫です．

こういう問題について,「危ない人」に反論するとなると,統計的検定とかが重要になってきます.もっとも,何日も寝てないでマッチ箱のことで怒っている人に数学的に反論して役に立つのかどうかは疑問ですし,「ビンラディンの手」などではどういう風にアンサンブルを組み立てて反論あるいは論証したらよいのかわかりません.また,画像深層暗号を規則を教えてもらわずに,画像だけから自動的に解読するのも大変そうです.だから,これが全部,統計学の問題とはいいませんが,情報の学問のスコープを示す例ではあるかもしれません.

ここで,「危ない人」という表現に抗議された方がいました.「どうにも困ってしまう人」とか「とんでもな人」という意味なのですが,そういうことをいうのはいかん,みんなと偏見なく話し合わなければ,というということのようです.しかし,どんな人とも本当に偏見なく話しあえるでしょうか.私には無理です.変な健康食品を売りつけようとする人,電車の中でひとりごとを言って手をふりまわしてる人,相対論は間違っていて重力のもとはガスの浮力だという人,そういう人にわざわざ近寄っていって友人になる気は私はしないです.アメリカの大精神科医,H.S. サリヴァン(ヘレンケラーの先生じゃないですよ)は,「どんな人がやってきてもびっくりしてタンスの上に避難しないことが精神科医の職業的義務だ」というようなことをいってます.私は精神科医ではないので,状況によってはタンスの上に避難してよいのです.その上,分裂病の専門家であるサリヴァン先生も,どうも別のある種の人は苦手のようで,著書で「こういう奴らはどうしようもない」などと嘆いてます.

情報の学問というのは,自分の前提としていること,偏見といっても良いですが,を,ひとつひとつ明示的に取り出していくことです.でも,これは終わりのない,難しいプロセスです.そして,その大前提は,「自分が全然公平なんかじゃないんだ」ということを自覚することでしょう.「こういう人とは話し合えない」とはじめから断定する(上ではしてますね;;)のは確かに間違いかもしれませんが,安易に「どんな人とも偏見なく..」などというのはもっと間違いだと思います.世の中には「私はアンサンブルではなく,個々のものを扱える科学をやる」などと発言する人もいますが,これでは「フランス語の優れている点は単語が人間がものを考える丁度その順番にならんでいることだ」と言ったフランス人と同じです.

9 共通の構造

いままでの例,前の章の例だけではなくて,情報圧縮や予測や乱数の検定,全部を通しての共通の構造は,図式化すると

データ = 雑音 + 構造
データ = 偶然 + 規則
データ = ある言語での文章 + その言語の辞書

のようになります. 2 つに分解したときのはじめのほうの部分が「驚きの源」「予期しないもの」としての「情報」ですが,それが,それを補完する「規則性」「共通理解」としての 2 番目の部分と裏表になっていることが重要です.この分解を決めているのが,モデルであり,表現であり,言語なわけです.AIC 最小化や MDL 原理などは,適応的にモデルを選ぶことで,分解の仕方を自動的に決めようという試みです.そこで決定的な点は,データの有限性でした.

重要なのは,しばしば,「意味」は「情報」のほうではなく,取り出された構造・規則,選ばれたモデル・言語のほうに入っているということです.象徴的に

情報量 = データ/モデル
情報量 = データ/表現

情報量 = データ/言語

とかけますが，取り出された「情報量」のみに注目すると，「既存の情報理論は意味を無視しているから，意味を扱えるような新しい情報理論が必要だ」ということになってしまいます．しかし，むしろ，「意味」は情報量の測定の前提となっている側に含まれているのです．そのことは，シャノンの情報理論ではいまいちはっきりしないのですが，AIC や MDL，「統計学」の「情報」という文脈での再評価を通じて明らかになってきました．この考えに立つと，「情報処理」については，

情報処理とは情報の捨て方の研究である．
情報処理とは「規則」をみつけることである．
情報処理とは「偶然」を定義することである．
情報処理とはアンサンブルを探すことである．

といえます[23]．

10 統計学・情報理論・情報学

今まで述べてきたようなことは，それぞれ個別の分野で対象ごと，状況ごとに研究したり考えたりしてもよいはずです．「情報に関する学問」では何を研究するのでしょうか? まず，統計学・統計科学・情報理論といった範囲の学問，狭義の数理的な分野については，

- モデルを表現しデータと結合する手法の研究
- モデルを選んだり評価するための一般論の研究
- モデルに関する分野共通の知識・考え方の研究

などが，守備範囲として考えられます．

さらに，より広い見地から，情報の学問とは何か，を考えてみると，

入れ物 (モデル・表現・言語)，たとえば
規格，書式，計算機言語，文字，貨幣，...
とそこに入れられるもののかかわり

を扱うのが，「情報学」なのではないでしょうか．モデル，表現，言語を決めることで，世界は，

雑音・偶然 | 構造・規則

あるいは，

入力データ | プログラム

特記事項・特注品 | 規格化された部分

[23] 一番はじめの表現は山崎眞見氏によると記憶しています．

と分けられるわけですが，この境界の揺らぐ様子，

を問題にするのが「情報学」ではないかと思います．

この全部が $-\sum_i p_i \log p_i$ の世界におさまるわけでも，確率モデルで扱えるわけではありません．たとえば，

- プログラミング言語の設計
- 文字コードの制定
- 診療カルテの書き方の研究
- 図書の分類
- 医療事故や航空事故における discommunication の組織的研究

なども，いま述べた意味では情報学の対象でしょう．ハードウエアの研究も，「プログラミング可能性」がディジタル機器の基本的特性だと考えれば，「器」と「中身」の関係として捉えられます．しかし，それでも，情報学の中で， 情報理論や統計学は相当重要な地位を占めるように思います．それは，確率とかアンサンブルという概念が，「可能なものの全体」「起こりうることの全体」を扱う上で基本的なものだからです．そして，語るためのことばを選ぶことは，暗黙のうちに，「考えられることの全体」を決めることになっているからです．

こういうと，何でも情報学になってしまうような気がします．情報学でないものは何でしょうか．たとえば，物理学では規則は神様が決めるので，その意味では情報学とは対極の分野かもしれません．どこまでが「初期条件」や「境界条件」で，どこまでが「法則」なのかということは，理論の進歩によって変わりますが，それは人間が実験に基づいて考えるので，それを考えるために情報学的，統計科学的な一般論というのはあまり有益ではなさそうです．

もっとも，特に統計物理では，例外的な局面で，情報学的な背景が突出してくるように思われたことがあります．古典統計力学での等分配則からのずれ，比熱のパラドックス はその例です．そこでは，「何が要素 (自由度) か」が問題になっています．これはある意味で，統計物理における「事前分布」の問題と言えなくもありません．もちろん，物理学が「究極理論」を背景にしていて，究極理論が，何が要素的な記号であるか，「真の表現」か，ということを含んでいるとすれば，それだけで，統計物理での事前分布の恣意性という問題は起きません．しかし，あらゆる力の起源などを含んだ「究極理論」が得られるまで，日常生活レベルの平衡統計物理が機能しないなら，実用上はとても困るでしょう．そういう心配なしに済む，微小な世界のことは日常生活レベルには効かない，というのが量子力学によって保証されるというのが，「比熱のパラドックス」の解決でした，これで，物理学は情報学的な悩みから救済されたのです．混合エントロピーについての「ギブスのパラドクス」は別の例ですが，これも量子論が救ってくれました．

11　どういうカテゴリーの学問なのか

それではそういう「情報学」というのは，どういうカテゴリーの学問なのでしょう．世の中には，「学問の種類別などナンセンス」という人もいれば，「物理学は実験の説明を通じて自然を説明するもので，科学である．数学とは全然違う，わかるかお前 ~ 」なんていう人もいて，どっちももっともなような気もするんですけど，ともかく考えてみます．

まず，科学 か，っていうのがあります．「情報科学」って言葉もありました．そういう要素は確かにありますが，普通の自然科学のように「決定的」な実験があるわけではないのも事実です．たとえば，極端な例かもしれませんが，ウィンドウズが最も売れたから，それが最も良いと証明されたとは言えないですよね．工学っていうのもあって，これはかなり良い線のような気もするんですが，じゃあ工学って何かというと良くわからなったりもします (特に「計数工学」とか「数理工学」とか「情報工学」とかは．)．「工学」が「科学」を応用して世の中に役立てることだとすると，「情報工学」のもとの「科学」は何でしょうか [24]．

数学 ? 実は，そう考えたい人も多いらしいのですが，定理や計算法は「考える材料」「考えを実行する方法」としては重要でも，それがすべてでは無いと思います．たとえば，数理的な議論に載せるためには，しばしば，[何かの量 $\to \infty$] の極限を考える必要がありますが，統計学においては，それが実際的か否かが常に問題になります [25]．ここで，判断規準が，「結果の豊かさ」のような内的なものだけではないところが狭義の数学との違いです．

それでは，これは哲学 でしょうか．私は実はかなりまじめにそう思っています．ただ，「哲学」としてどういうものをイメージするかによります．「○○さんは哲学があるから」とか「海を眺めているうちに哲学的気分になった」とか「人生いかに生きるべきかとの哲学的問い」とかいいます．最初のやつは，「一貫した方針がある」というような意味ですね．次のは，哲学 = 神秘という感じ．最後のは，人生について，何か積極的な指示を与えてくれるものとして「哲学」を捉えているようです．世間では，「哲学」はこの 3 つを混ぜたような内容で，とても難しいもの，というイメージでしょうか．

ここでいう「哲学」はそれとはちょっと違って，専門家の多くが考えているような哲学，その中でも批判哲学とか分析哲学というような系統のものです．いいかえれば，医者とか，カウンセラーとか，助産婦のメタファーで語られるような「哲学」です．分析哲学では，伝統的に数理論理学を基盤として重視します．言語学とかとも縁が深いかもしれません．これに対して，確率分布を用いてモデル化するなどというのは，あまり主流とはいえないでしょう．しかし，すでに詳しくみてきたように，アンサンブルを明示することによって隠された前提を暴くということは，「批判」「分析」の中心的部分だと，私は思います．

もちろん，本稿で論じたような内容を「哲学」とすると，それは哲学と呼ぶには，あまりにも実践的なものでしょう．なにしろ，会社の事業部の毎日の研究で問題になるようなことを論じているわけですから．しかし，そもそも「哲学」というのはそうあってしかるべきなのかもしれません．「究極の解決」と「実践的な解決」の間にはっきりした境界が無い，ということこそ現代哲学の教えではないのでしょうか．それで，私は『大人の哲学』としての情報学」なんてことを言ってみています [26].

もうひとつ，広い意味で「情報の学問」に関係する概念として，

[24] この原稿で考えてきたような面を無視すれば，ハードウエアを作るための物理学とか何かだということになるのかもしれません．案外，世間の認識はそういうものなのかも．

[25] さきにちょっと言及した AIC 対 MDL/BIC の議論はこの良い例です．証明の前にまず土俵が問われるのです，

[26] 「子供の哲学」といっている人がいるので，対抗して言ってみたのですが，本当は大人も子供もないのが「哲学」のように思います．ウィトゲンシュタインが言いたかったのも，そういう事なんじゃないかという気がします．私自身は，本来，とっても不安な「思春期の哲学」に属する人間ですが．

「批評」

ということがあります．私の中ではこのふたつは深い関係にあります．

12　意図と差別

この原稿に関連した講演をしたときに，

> 情報理論では，「私のいいたい『これ』」，というような「意味」の部分は取り扱えないでしょう[27]．そういうことが取り扱いたいのです．

というコメントがありました．書くとわかりにくいですが，いいたいことはわかるような気もします．私が話したかったのは，

> 「シャノンの情報理論では意味は取り扱えない[28]」といわれているが，それは一面的にすぎる見方なのではないか．

ということで，「私のいいたいこれ」が情報理論や統計学で論じられるとは主張していません．それに，「私のいいたいこれ」というのは，「意味」ともいえないことはないと思いますが，どちらかというと「意図 」(intention) のように思われます．

しかし，このコメントはそれなりに興味深い点を突いているような気がするので，これに関して少し論じます．まず，これは個人的な感覚ですが，「私がいいたいこれ」とか「私がやりたいこれ」とかいうことを私はあまり信じておりません[29]．それは，「個人」とか「自由意志」を信じないという意味ではなくて，もっとデリケートで現実的な問題です．あることをいったり，あることをやったりしたときに，私たちは本当にそれがどういう意味で，どういう意図であるかを理解しているとはいえません．それは，話を続け，行動し，他人とかかわってくるうちに，しだいにわかってくるものなのではないでしょうか．それから，これが大事なことなのですが，私が「自分はこういう意図だった」，「こういう意味でいった」と思っていることが，「本当」であるとは限らないと思うのです．

女性学のもっともわかりやすい成果のひとつに，次のようなものがあります．

> 辞典を一冊持ってきて，主語が女性の文と男性の文で，文の意味や使われている単語を比較すると，あきらかに偏りがある．

正確な結果は覚えていませんが，女性・男性の固定観念に一致した表現が知らず知らずに選ばれている，という結果です．いま，「知らず知らずに」と書きましたが，実際，辞典の編集者が意図してそういうことをやったとは思えません．辞書を使う人も，あらためてそう指摘されなければ気がつかないでしょうし，指摘された編集者は言いがかりだといって怒るかもしれません．そんなことは辞典の編集者にとって「私がいいたかったこれ」には入ってないからです．では，この指摘は意味が無いかというと，そうではないでしょう．指摘した人と同じ土俵にたって，彼らの気づいていな

[27]「脳」についていえば，「脳は情報理論では取り扱えない」と「脳が情報処理機械であるというテーゼには限界がある」の違いにも注意してください．脳が情報処理機械だと思わなくても，情報理論を使ってそれを扱うことはできるかもしれません．セメントキルンは情報処理機械ではないですが，それを制御するためのモデル化に情報理論や統計学を応用することはできます．

[28]シャノン自身もそれらしいことを言っているらしいのですが，本稿で論じたのは，むしろシャノン以後の発展，解釈の発展なのです．

[29]質問された方の趣旨は，もしかすると，「本当は無いはずの『私がいいたいこれ』が何故あるようにみえるのか，そのメカニズムを知りたい」という意味だったのかもしれません．そうすると話は大分違ってきますが，依然として，ここでいう「情報の学問」との態度の違いは存在して，以下の話はそれを理解するのに多少は役に立つのではないかと思います．

い要素や数学のまちがいを指摘することはできますが，著者が意識せずにやったという理由では，この研究の意義は減りません．むしろ，意識されなかった「前提」を暴露したということがポイントです．どうしてそれがわかったかというと，見慣れたものを，いままで考えられていなかった分類（女と男）で考えて，勘定してみたからです．これは，われわれの立場からは，「新しいアンサンブルを構成した」，あるいは「$-\sum_i p_i \log p_i$ の i として新しいものを考えた」と解釈できます．「私がいいたかったこれ」から出発せずに，それが本当は何なのかを考えていく人を支援するのが，統計学，情報学の立場だと思います．

「無意識」というと，もともと精神分析の概念ですが，精神分析も，ある意味では情報の学問です．それは，「無意識を記述することば」を作り出す試みといえるでしょう．こうした側面は，「影」とか「アニムス」のような用語を使ってお話を作っていくユング派でもっともはっきりしていますが，ほかの流儀も，本質的にはそう解釈できると思います．違う流派は違う言語を使うので，話が合わなくなってしまい，党派的だと非難されるわけですが，ある意味ではやむおえないことなのです．

それで，「分析していったその涯」には，何が残るのでしょうか．デヴィッド・ロッジの小説「小さな世界」には，分析されて，文体を奪われてしまった作家が出てきます．ある人が，この作家の作品をすべてデータベース化し，どのような用語法を好むかを分析して，本人に教えてあげるのですね．たとえば，「脂」（グリース）という単語が彼のお気に入りだというような，彼がまったく気づかなかったことが沢山発見されて，本人に伝えられます．次の日，彼は書こうとするけど，書けない．自分の癖をすっかり意識的に知ってしまったので，できるだけそれに逆らおうとするのですが，そうするとどんどん変になってしまうわけです．われわれの用語でいうと，彼はデータベースの情報に対応する p_i から計算した符号長 $-\sum_i p_i \log p_i$ が最長になるような文体で書こうとしたわけです．「驚き度」最大の文体．それで結局，小説家を廃業します[30]．

精神分析の精神，そして情報学・統計学の精神というのは，「分析していったその涯」に残る人間的なもの，善意とか親切といったもの，を信じるということなのだと思います．それはとても大変な，怖いことですが．

13　表現と近代とわたしたち

ある意味，ここまでずっと，「表現」“representation” ということについて書いて来たように思います．日本語は “re-” ではじまる言葉がしばしば苦手です[31]．“representation” は，「表現」と訳されますが，それが “presentation” に対応する，ということはあまり意識されません．“representation” は，含みのある言葉で，日本語の「表現」より広いものを意味しています．たとえば，“representive” といえば「議員」，議会制民主主義において民意を代表する人のことですね．これに対して，国家社会主義においては，指導者は「国家・民族の代表」ではなくて「国家・民族と等価であるような何か」です．そして，官僚や議員を排除することで，人民は，指導者＝国家と直接に，「表現」抜きで，結合するわけです．

西欧人にとってみれば，「近代の限界を超える」ことが「presentation 対 representation」の 2 分法を越えることだ，と意識されるのは，自然なことなのだろうと思います．2 分法を超えるには，どっちかに統一してしまうのが簡単ですが，「presentation」のほうに統一するわけにはいきません．「直接性の支配による近代の乗り越え」は，最近実験済みで，やばいことが証明されたからです．そこで，「ポスト構造主義」では representation のほうに統一しようということになって，「表現の自律

[30] 小説の最後で復活します．
[31] たとえば，“recognition” と “cognition” は，どう訳し分けたらいいのか，よくわかりません．

性」とか‚「『書いたもの』(representation) の『声』 (presentation) に対する優位」とかが強調されたわけです．

それで，それが，われわれに自然に理解できるかというと，理解できないです．こちらは近代人じゃ無いし，そもそも 2 分法なんてないのですから [32] ‚「美しい「花」がある、「花」の美しさといふ様なものはない」 [33]というのが，良くも悪くも，こちらの主流です [34] ．したがって‚「ポスト構造主義」から「クオリア」や「アフォーダンス」のような直接性優位の考えに乗り換えても，ぜんぜん平気．すべては情緒レベルの問題ということになるわけですね [35] ．

話は変わりますが，家の建築のことを書いた本 [36]を読んでいたら，

> 建築家の仕事は void を solid に翻訳することだ．

と書いてありました．void というのは‚「空間」ですが，むしろ人の生活する「場」とでもいったらいいかもしれません．void は設計できないから，それを，まわりの構築物，solid，に直せるような訓練を常にやることが，生活の場としての家を建築する人間としての基礎訓練だ，という論旨だと思います．表現ということですね‚「情報学」かどうかはわかりませんが，それっぽいアプローチだと思います‚「複雑系」の「構築的アプローチ」というのはこれとは違っています．家の場合でいうなら，まず適当に面白くて簡素な家を建ててみて，住んでみる．それで，どんどん試して，家作りを楽しもうというようなことになるのでしょうか．家作りで複雑系をやるのは，大金持ちでないとだめそうです．もっとも，さっきの本の著者は，空白の部分のある家を建てて，だんだん生長させるというのを実際にやってみたようです．また，別の「徹底調査・表現派」らしい建築家の建てたうちは「何年住んでも，家具の置き方が設計計画の通り」だそうですが，そういう家もちょっと住みにくいかもしれません．結局，ここでも，この 2 つの方向は相補的だということでしょうか．

「表現」が問われるような別の分野として，芸術としての写真があります．写真は「機械を介した表現」という意味では，複雑系のコンピュータシミュレーションとか CG による芸術とかと，共通点がありますね．むかし，土門拳という人がいまして，彼は「リアリズム写真」の提唱で有名でした．彼が対決しようとしたのが‚「雰囲気写真」だったということからもわかるように，彼の本当の目標は「リアル」を写し取るということよりも，2 元論的な「対象 対 表現」の構造を確立させることにあったのではないかと思います [37] ．表現する「相手」が断固として存在しなければ，表現しようがないんじゃないか，というふうに思いつめたのでしょう．土門拳は，当時，アマチュアの教祖的存在でしたが，カメラ雑誌の月例コンテストの選評が「写真批評」などの本になって残っていて，今読んでも，というか，今読むと，すごいです．むちゃくちゃ言ってますが，それは‚「近代」の不在に対する怒りというようにも読めます．土門拳のたんつぼ論，というのもありまして，駅にあった赤い痰壷が，駅が空襲で灰燼に帰してもひとり瓦礫の中にあるのを見た，これぞ実在である，というのですが，たんつぼ，壊れちゃったらどうするんでしょう．そんな根性の無いたんつぼは，実在とは認めてもらえないということでしょうか．

---

[32]「書いたものの優位」とか「主体の消滅」についていえば，中国文化圏では，文字の書いたものはそれだけで偉いことになっているようです．さらに，自然科学のコミュニケーションでは‚「主体としての著者そっちのけで，論文がお互いに引用し合い，再生産しあう」というポスト構造主義的状況がとっくに実現してるので，そういう分野の人にはなおさらわからないでしょう．

[33]小林秀雄の言葉ですが，おおもとは青山二郎だという話もあります．

[34]世界的にも，認知科学やロボット科学では，この手が主流になりそうなので，前近代人どもが喜んでいるわけです．実際‚「花の美しさ」を「美しい花かどうか判定するプログラム」と読みかえれば，小林秀雄は正しかったと認めざるを得ないのかもしれません．

[35]だいぶ前の方の脚注でも触れましたが，この解釈は違っているかもしれません．

[36]渡辺武信「住まい方の実践」，中公新書 (1997) ．

[37]もっとも，彼の標語は「対象とカメラの直結」なので，そのまま受け取ると「直接派」の筆頭なのですが．もっと考えなきゃだめかな．

いつまで話していてもきりがありませんね．結論をいいますと，情報の学問とは，西欧文明の到達点，あるいは

西欧文明の自意識の数理化

なのだと思います．これでは，日本が「情報に弱い」のも無理はないです．もっとも，これからの時代は，もはや 2 分論ではなく，サイボーグとか機械との合体の時代だという考えもあります．「攻殻機動隊」や「エヴァ」を見て育った人たちが研究の第一線に出て来るのも，そう遠い先ではないでしょうし，彼らはロボットとか人工知能というものをはじめからそういうものとして捉えていると思います．ずいぶん前に，西垣通は「ペシミスティック・サイボーグ」という本を書いて，「これからはサイボーグの時代になるけど，それは嫌だ」と言っていました．私はサイボーグでもクローンでもどんどんやればいいと思うのですが，来るべき政治的な「直接性の支配」の時代については，どうしたらいいのかわかりません．もはやそれに抗するすべはないように思われますが，せめて科学者や哲学者は思想の上でそれに協力しないで欲しいものです．

大体ひと通りやりました．どうもありがとうございます．