統計数理研究所彙報
第29巻 第1号 (1981)

統計学の新しい兆し*

——データ解析志向としての——

統計数理研究所 林 知己夫

(1981年8月 受付)

兆しは，既に「そこにある」ものではなく，人が見出すものである．見出すものである以上，個人の視点というものに深く係り合いがある．この意味で主体的であるとも言えるし主観的なものでもある．従って，「私の」体験と言うものが多く物を言うことになる．これから始めよう．

1. 私の統計に対する体験

統計学を始めて学んだ時，統計的検定論，推定論が中心であった．今からみると，フィッシャー流の言葉を使って，ネイマン流の理論を講ずるの感があった．ここに，理解を阻むものがあった．しかし，当時これに気が付く由もなかった．第一種の過誤，第二種の過誤，これはそれなりに解る．しかし，ある検定をしたとき，とどのつまり，自分はどれだけの過誤を犯しているかとなると何も答えてくれないのである．両者勘案すると言っても，第二種の過誤は対立仮説のあり方により非常に大きなものにもなり，most powerful と言っても使いものにならなくなる．正に空虚であると感じたし，標本数を大きくすれば棄却されない仮説は存在しないのではないかと感じた．片端から，統計数値表に有る検定法を使ってみるとその道の「常識」で——測定誤差範囲や意味のある差異はどの位かを考慮に入れた意味——は同じと見えるものが，標本数を大きくすると「同じ」という仮説は棄却されてしまうのである．計算の桁数を一つよけいにとったために——$\bar{x}=5.31$，$\bar{y}=5.31$ としたとき桁を一つ余計にとって（これは全く実質的には意味のないことである場合），$\bar{x}=0.531$，$\bar{y}=0.534$ とすること——，検定で「同じ」という母集団仮説が棄却されるということを経験した．こうなると検定論は何なのか．標本数が大きくないとき，その道の常識でおかしくないから使えるし，標本数が大きいときはおかしくなる，では論理として不可思議と言わざるを得ない．標本数を大きくした時の漸近理論などあるが，検定論にとって意味のあることか．この標本数と検定の関係は深刻なものであった．標本数は大きくとることが情報は多いのにきまっている——そうでないとしたらおかしな科学である——のに，無意味なものが検出されてくるのであっては困却せざるを得ない．

そのころ，伏見康治氏の確率論及統計論（河出書房，昭和17年3月）というのを読んでいたがその379頁にウェルドン (Weldon) の骰子の実験というのが載っている．12個の骰子を同時に振って，5の目か6の目かの出たものの数を読む．この実験を26306回繰返してみた．12個の骰子を振って i 個5の目か6の目の出る確率は

$$p_i=\binom{12}{i}\left(\frac{1}{3}\right)^i\left(1-\frac{1}{3}\right)^{12-j}$$

---

* 統計数理研究所37周年記念講演会によるもの

となる筈である（一つの目の出る確率は 1/6 と等確率と考える）.

これと実験のデータを比較してみると次のようになる.

i	実験相対頻度	p_i
0	0.007	0.008
1	0.044	0.046
2	0.124	0.127
3	0.208	0.212
4	0.233	0.238
5	0.197	0.191
6	0.117	0.111
7	0.051	0.048
8	0.015	0.015
9	0.004	0.003
10	0.000	0.000
11	0.000	0.000
12	0.000	0.000
	26306回	

通常の意味ではよく一致している．小数点以下3桁目せいぜい6の違いである．つまり $\max|データ-p_i|=0.006$ である．ここで χ^2 の値を出しているが $\chi^2=40.8$ であり，自由度は 11 である．$\Pr\{\chi^2\geqq 40.8\}=0.00003$ であるとし，「これは驚くべき程小さい確率で我々は殆ど確実に上の理論の誤りであることを推断しても差支えない．ウェルドンの骰子は“正しく”作られていなかったと言える」と書いてある．これを読んで我が意を得たりという感じであった．26000回もデータをとって，合うモデルがあろう筈はない．なぜならば，それだけの精度でこのような実験が出来る筈はないからである．世の骰子で普通に振って，この精度は出ない．こうなると骰子は確率の代用ではなくなってくる．骰子ばかりではなく，統計数値表の乱数表でも 0〜9 の数字を入れ変えているではないか（しかも 15000 の数字ではなく，1000個宛に区切って）．凝似乱数であれば（あるいは非常に優れた物理乱数でなければ）100,000 も作れば恐らく，χ^2 検定でまず等確率（もしくは χ^2 の値の分布）や独立性に基く仮説は棄却されよう．χ^2 検定にかからぬものはまず世の中にはあるまい．χ^2 検定は

$$\chi^2=\sum_i^k\left(\frac{n_i-np_i}{np_i}\right)^2 np_i=n\sum_i^k\frac{\left(\frac{n_i}{n}-p_i\right)^2}{p_i}$$

というわけで，データの相対頻度の近さの函数に標本数を乗じたもので，よほど $\frac{n_i}{n}$ が p_i に近付かないと n に負けて大きな数になる．$\frac{n_i}{n}$ が p_i に近くなるなり方は実験条件により限度があるのでそれほど縮まらない．これが実験や調査測定の実情であって，数学としての確率論との差異である．確率論にもとづく統計量の理論を実際に適用するときの問題を無視しては統計学はないと感じたものである.

ここに，実質的な差を評価しておかなくては，少くとも，検定論や推定論はそのもやもやした理屈を認めたとしても trivial なものとなってしまう.

私としては，検定論の効用は必然的に少い標本数に根ざすフィッシャー流の統計学が科学的な思考として有用なものであると感じた．もっとも，フィッシャー流の exact distribution（他の情報を入れずにそれだけの条件で厳密にきまるという意味）には，応用上疑問百出であるが，科学的立場としては筋が通っているように思った．ネイマン流のは数学的であるにしても科学としてはあまりにも空しいという感覚であった．このあたりを勘案して，検定論，推定論をデータ解析という科学的方法の中に位置付けるフィロソフィーの必要を痛感したのである．それだけではなく，当該科学に際して，「実質的」に差があるとは何か，をフォーミュレイトして，これを妥当性を以て表現することを統計学は志す必要があり，このための方法論を作ることは焦眉の問題であると感じたわけである.

さて，検定論の標本数と数学的表現を持つ仮説のたて方の問題について論じたが——これは

数理統計学の範囲外とする意見もあるがこうなれば空虚な理屈で，統計学を使う人はあるまい――，もとへもどってその核心の論理に対する疑問はどうする術もない．

後になって，第一種，第二種過誤関係のことはワルドによる統計的決定理論が発展してきて一応の解決が出てきてほっとしたことを覚えている．推定の問題についても，区間推定について，理屈は解るにしてもなにか釈然としないものが残った．推定の幅をつくっても，その手にした土の幅の端の点には意味がなく，幅の付け方による信頼度というところに，腹に応えぬところがあり，割り切れなさを感じた．しかし，推定の方は幅の大小が一応の精度の目安となるので「科学的」意味は汲みとれた．

確率論にしても，ルベーグ測度の確率論に疑問を持った．ルベーグ測度0は一体実際に何を意味するものであるか．データなるものを念頭の基礎とするとき，如何なるものであるか，という問題である．これも後になって，M. Fréchet は測度0はあまりにも大きすぎる，この段階分けは出来ないものかという研究を始めたのを知ったが同感を禁じ得なかった．しかし，残念ながらこの「稀なこと」(raréfaction) の研究は失敗に終ったという論文を発表していた．

話をもとへ戻そう．0, 1 という標識で表現される区間 [0, 1] 間の実数をもとに置いて考えたとき，コレクティブ確率系列をなすものは実数の濃度で存在するということがあるが，我々が普通手にする系列は前にも述べたように標本数を大にする，つまり系列の長さを大きくすると検定すれば外れるというおかしなことになる．実数の濃度の確率1で確率系列にぶつからないのである．しかし，我々の現実に手にする系列は殆どがランダム性が現れることなく，何等かの規則性（きれいな規則性は珍らしく汚い規則性が多い）を示すものばかりであるのにその生起する確率は0というのであっては，――我々の行為は有限，せいぜい可附番と考えるのが至当である，この上に立っての議論であるから，我々の打ち当る系列は殆ど無規則性を示すものである筈である――，測度0の現象的意味は一体何なのか．この測度0の概念は曖昧なもの，混迷に満ちたものと思われてくる．無限に続ければ，無規則になるという理屈もあり得ようがこれを認めるとすれば有限の下でする我々の行為に対して無意味である．

我々としては，有限系列の確率論，フォン・ミーゼズのコレクティフ論，試行系列と対応の付く測度に基く確率論，主観確率というところに落ち着かざるを得ないような気持になっていた．

これらあたりが私の統計における最初の出会いであり，こういう気持から統計学を学び，役に立つ統計学を探し求めるという旅に出たわけである．

2. 私の考えている統計数理

私としては，データというものを中心に据える，つまり，いかにデータをとり，いかに分析して，情報を剔抉するか，ということが核心になる．データによる現象解析である．こういうわけで，これまでの統計学と異った立場から統計学を考えるという意味で，統計数理というものを考えた．これがどんなものかは既知のことであるからここには触れない*．いずれにしても統計学の基礎，そのフィロソフィーを確立し，統計学を進めるあるいはデータを分析して情報を引き出すための方法論を明確にすることを研究対象とした．これに加えて，標本調査法，社会調査法，質的なものを科学的意味で数量化して取扱おうとする数量化の方法，現象予測の方法を研究の範囲と定めたわけである．

統計学では確率変数 X といきなり書くが，X として，いかに表現するかが解らなければ，また X にもとづく統計的分析が意味あるものでなければ，空虚なことになる．数量化はここ

* 未知の方は，林，「数量化の方法」，また「データ解析の考え方」ともに東洋経済，1974，1977，参照．

から始まる．しかし，これは全く顧みられていなかったところで，ここから手を付けるべきであると感じたのである．

そのころ，私の知ったのはアメリカ（イスラエル）のガットマン (Louis Guttman) である．一対比較法による質的データの数量化，スケログラムアナリシス（尺度構成，質的なものから物差しを作る），というものを知った．正に私の考えていた方向であった．ガットマンは社会学，数学，統計学の専攻の人であり，類例のない考え方を示しており，肯綮に当るものがあった．彼とてもいずれの分野でも主流ではなく，自ら評して「こうもり」の如きものと言われていると言っていた．

なお統計学に関係ある分野で operation research（作戦研究）が抬頭し*，新しい息吹きを与え，その出発点のフィロソフィーに私も大いに感銘を覚え期待するところが多かったが，思わぬ方向に進展し，形骸化してしまった．日本においては，戦前より北川敏男は統計科学を旗印とし，統計学の新しい発展を期し，推測統計学，情報科学**と展開してきており，情報に関する包括的，総合的，統一的方法論の完成を志向しているが数理統計学的色彩が濃い．

これまでが私の体験・立場というものであって，こうした人間が統計学の中に新しい兆しを見出してきたということである．

3. 新しい兆し その1 exploratory data analysis データ解析の将来

統計学で有名なトゥキイ (J.W. Tukey) が (The Future of Data Analysis, AMS, Vol. 33, 1962) という論文を発表し，これまでの統計学を厳しく批判し，新しい行き方を示唆している．多くの統計学者に評価されたと思えたのであるが何らかの影響を実質的に与えたとは考えられない．一向に新しい兆しがそれから出た形跡はない．私の方からみていると，そこに提示されていることのかなりの部分は，私の考えていた統計数理・数量化の考えで片が付くので，それほどの新鮮味はなかったが，数理統計学が新しい方向に踏み出したとの感があった．

1977年に発見的，探索的データ解析 (exploratory data analysis) という本を出し，統計量中心の統計学からの脱皮を宣言し，データを素直に眺めるための視点，アルゴリズム，それに基くグラフ化から情報を取り出すことの重要性を述べている．私どもにとって，正にその通りであって，いつも普通行っていることなのであり今更ことあらためてデータを眺める重要性を説かねばならぬ程形式化，数学化されてしまったアメリカにおける統計事情に目をみはることになった．もっとも Tukey の "exploratory" という立場が理解されていないことは，薄々気が付いていた．ある統計学の大家が，exploratory という言葉を使って講演をしたとき，Tukey が「何が exploratory か」と質問した．その答えは「もっと数学的に精密にすれば……」というものであった．Tukey はこれを皮肉り，exploratory は「数学的に精密」とは全く異質のものだと述べているところをみると，その大家は全く exploratory を違ったものに考えていたと思わざるを得ない．数学的に精密にきれいな形にすればする程現象のデータの質から遊離してしまうのが通例である．

1979年ヴェルサイユのデータの解析 (Analyses des Données) と情報学 (Informatique) の国際会議で発表した論文 Approaches to Analysis of Data That Concentrate Near Intermediate-Dimensional Manifold (Ed. E. Diday et al., North-Holland Pub Comp., 1980, 3-13頁)—— J.H. Friedman, P.A. Tukey と共著——により一層進んだ形が出てお

* 戦中から行われていたが——これは日本においても同様であった——目についてきたものとして P.M. Morse and E.G. Kimball, Methods of Operation Research, Wiley, 1951.

** 北川敏男，統計学の認識，白楊社，1948：情報科学の視座，共立出版，1970 など参照

り PRIM'79 という分析プログラムのことに言及している．

要点はデータを素直に眺め，この中から有用な情報をとり出すまでの考え方，視点，データ処理を示すものと考えてよい．どんな結果が出るかについて「何等の予定された知識を持たずに多次元データ（高次元空間内の点の雲と言ってよい）の性格を探索してその形を追究したい，そしてその点の雲を低次元の中に収約して核心となる情報をより取り出し易くすることを主眼としたい」と述べている．この中に言われていることは，私どもが統計数理，数量化という考えの下に展開してきたフィロソフィーや方法と随分近いのであり，ここに述べられていることの多くは，すでに我々が実用化して用いていたものである．ただ，Tukey には，測定された数量的 x をそのまゝ用いるという立場を固執し，スケーリング（質を数量化する，数量を目的に応じ測定に付与する）の考え方がないので甚だ窮屈に感じるのである．数量化の方法に依れば簡単に行くものを，大変苦労している感じがする．彼の言う所を私の方の言葉で書き直してみよう．

1. 精密化より，荒削りでも局面の転換する理論・方法を志向する
2. うまいグラフ表示，フレクシブルな立場よりするデータのグラフ化——非線型，よぢれの表示，グラフの形の認識
3. 集団分割（partitioning と名付けている）の方法とそれに基く部分集団での成分分析（線型構造化），逆にその構造が出るように集団分割する（複雑なものは我々のいう各種方法による集団分割——クラスタリング——とその中での数量化による構造探索，クラスタリングと構造の同時探究）
4. 部分集団の構造把握とそれを全体的に総合して拡大して全体の構造の露呈
   （上述と同じく，集団分割の仕方と構造変容との関係の総合により全貌把握，一度全体を壊して細分し明瞭な構造を得て再構成し全貌を知る）
5. コーディング
   データの核心を単純化しコード化（集団分割後のコーディングを含む）して本質を把握する——モデル化——
   （データのもつ特性を鋭敏にする．モデル化して本質を把握する．また濃淡強調して特色を把えこれをもとに分析し，重要な性質を浮き立たせる）
6. 測定変動と実際の変動（measurement variation と actual variation）を見越してデータをみて行く．中間的な目の粗さの見方が必要である．
   （我々のいう中梯理論の重要性，目的とデータ・モデルの目の粗さの兼ね合い，釣合い）

いずれにしても，データに対してあまりにも楽観的でなく懐疑的でもないという態度が読みとれ，慎重に事を運びつつ兆候を見のがさない態度であると私には感じられる．

私流に読み過ぎているかも知れないが，これまでの統計学になかった感覚で議論が進められていることは確かである．しかし，これは，データによる現象解析を志向し，データをいかに取り分析するか，と言うことを素直に考え，真剣に取り組み，重要と思うところを解決して行くと言う態度を採れば，当然こうなる筈のものだと思うのである．

フィッシャーが，農業・生物現象のフィールドを踏え，彼は彼なりに重要と思うところを取扱うために統計学を考えたことに敬服する．しかし，これが多くの数理統計学者の手に掛り歪んだ形——私の立場から見ての話である——に展開し，勿論重要ないくつかの問題が明快に取扱われたとは言え，大勢は「有らぬ」方向に進んでしまい，多くの人々の眼を奪い統計の本質が遠ざかってしまったのではないかと思う．誰彼の罪に帰せるべきでなくこれを鵜呑みにした統計学者の自らの問題であった．Tukey は警鐘を打ち，自ら問題解決にのり出した第一人者

であると言えよう．

Tukey （私の会った経験から言って F. Mosteller もこの一派と思ってよい）がこうした動きに出ている一方，フランスではベンゼクリ (Jean-Pierre Benzécri) が新しい方向に踏み出していた (1973)．これについては後述するが，フィネイ (D.J. Finney) が data と number は違う，統計学の detective な態度を強調する Royal Statistical Society の会長就任講演をしたのも同じ方向の動きと言うことが出来る*．

また，Tukey の exploratory データ解析のほか統計学の領域でもカテゴリカル・データ分析（対数線型，ロジット，ロジスティックモデル等による)**，AIC (Akaike's Information Criterion) も出てきているし，正統的な統計学以外のところで，データの分析を志向する方法が出てきている．後述するベンゼクリのアナリーズ・デ・ドネ，多次元尺度解析 (MDS) に関するもの——多重線型 (multilinear)，非線型モデルを含める——グラフィック方法，クラスター分析，パタン認識，ファジィ論理，システム科学的方法，等々のものがあり***，統計学にこうしたものを取り込んで脱皮する可能性があることを感じている．つまり統計学は常識をとりもどす兆候を臭わせ出したと言ってよいような気がする．

4. 新しい兆し その 2 analyses des données

数量化と同じ様な動きが，ヨーロッパ，特にフランス，イタリー，東欧圏において盛んになってきている．1973 年以来のことである．この方法を analyses des données（データの解析）というが，英語の data analyses とは全く様相を異にする．日本流に意訳すれば，数量化と言うのが一番当っている．

1979 年に Analyses des Données et Informatique という国際会議がパリ近くの INRIA (L'Institut National de Recherche d'Informatique et d'Automatique) で E. Diday, L. Lebart, J.P. Page, R. Tomassone を組織委員会として開催された．また，この一派の人達が昨年の7月イタリーのナポリで，Data Analysis and Informatics, Meeting The French School という会合を開き，analyses des données の研究発表をやり，また講習会を開いた．この方面の主な人物を参考に書いてみると，仏では E. Diday, Y. Escoufier, J.P. Fenelon, M. Jambu, L. Lebart, Y. Lechevallier, A. Morineau, G. Saporta, R. Tomassone, J.P. Bouroche という人達で，30 才の後半から 40 才の前半位で大体同じ年令層である．伊では，V. Amato, A. Gili, N. Lauro, E. Marubini, A. Rizzi などが出ていたが，このほか，数量化の主の様な A. Herzel (1974 年以降の発表，quantificazione という言葉を盛んに使っている）がいる．

さてこの一派の総師は J.P. Benzécri（ベンゼクリと読む．バンゼクリではない）パリー第6大学の教授である．日本の人々には Benzécri の名は珍らしかろうが，フランスでの数量化というよりパタン分類（III 類）の元祖である．correspondance の名の下に 1973 年頃から精力的に仕事をし，本を書くやら雑誌を出すやら大変な活躍ぶりである．これまで私もいくつか本や雑誌をみていたが幻の人物であった．つまり国際会議など名前は出ているが本人は一向出席しない．一時は「寄せ名」ではないかと疑ったが，最近実在の人物とわかり，会いたいと思っていた．

---

* D. J. Finney, Problems, Data and Inference, Journal of Royal Statistical Society, Series A, Vol. 137, 1974, 1-22 頁，或は Numbers and Data, Biometrics, Vol. 31, 1975, 375-386 頁

** 後注参照

*** カタストロフィー論やフラクタル論は，数学志向で，データ分析志向とは言い難い．いわゆる情報理論のうちデータ解析を志向するものは上記の方法に含まれよう．

パリ大学統計学部には二人の教授がおり一人は Dugué（先年統数研に二ヶ月ばかり滞在した）で，一人は Benzécri である．一方は全くの数理統計学，他方はコレスポンダンス（データ解析志向）という訳である．

Benzécri 教授は見るからに変り者で，風貌正に教祖そのものであった．弟子は大学内外で200人を下らず，頂点の帷りの中に籠り，弟子がいろいろ活躍するという形である．弟子は各地におり，アフリカ諸国にも信奉者がいるという．彼は統計学者か，数学者かといろいろの人に尋ねてみると，Benzécri est Benzécri という訳で笑話に終る．しかし，なかなか面白い表現で印象に残る．

ベンゼクリ教授の統計学に対する考え方が面白い．彼が analyse des correspondances を考えた動機は，数量化の発生と甚だ似たところがある．従来の数理統計学への反旗であった．彼の研究の基盤としての「プリンシプル」をあげているが，その冒頭に，統計は確率ではない (statistics is not probalility) と宣言している．1930～1960 年の間高度にこ難しい数理統計学の理論は，すべて確率の匂い付けをした屋上屋ともいうべき精密化の仮説に基いて発展してきている．私の同僚数学者達は，自分達がいかに出来るかを誇示してきた．しかし，その偉大な枠組に適合するデータは殆どないのだ．だからそのようなものは忘れる方がよい (it is better to forget about it.)．

次のプリンシプルとしてあげていることを示してみよう．モデルは想像から来るものでなくて，データから来べきものである．数理統計学の悲しむべき情況は，煩しい検定理論の展開である．そして検定論理の不毛性を説き，大事なことは検定ではなく，合理的な仮説を作るために役立つようなものであるべきだと結んでいる．

出来る限り多くの情報は同時に用いるべきである*．50 の点がもともと 10 次元空間の点であると考えられているのに，大体円周上に並んだとすれば，それが人工的なデータでない限り，「有意」として検定する必要はあるまい．これが第 3 のプリンシプルである．

このあとコンピュータとの関係に言及している．いずれにせよ，こうした立場から analyses des données を考え出しデータの中から意味のある形を認識すること (reconnaissance des formes) を追究しているのである．

外的基準のない場合，データの山から靄をつきぬけて妥当な情報を浮かび上らせることだという私の主張に近い．私としてはその類似性に驚いているが，根本の行き方・思想になにがしかの差があり，データ解析の発展方向・拡がりに違いが出ているのである．

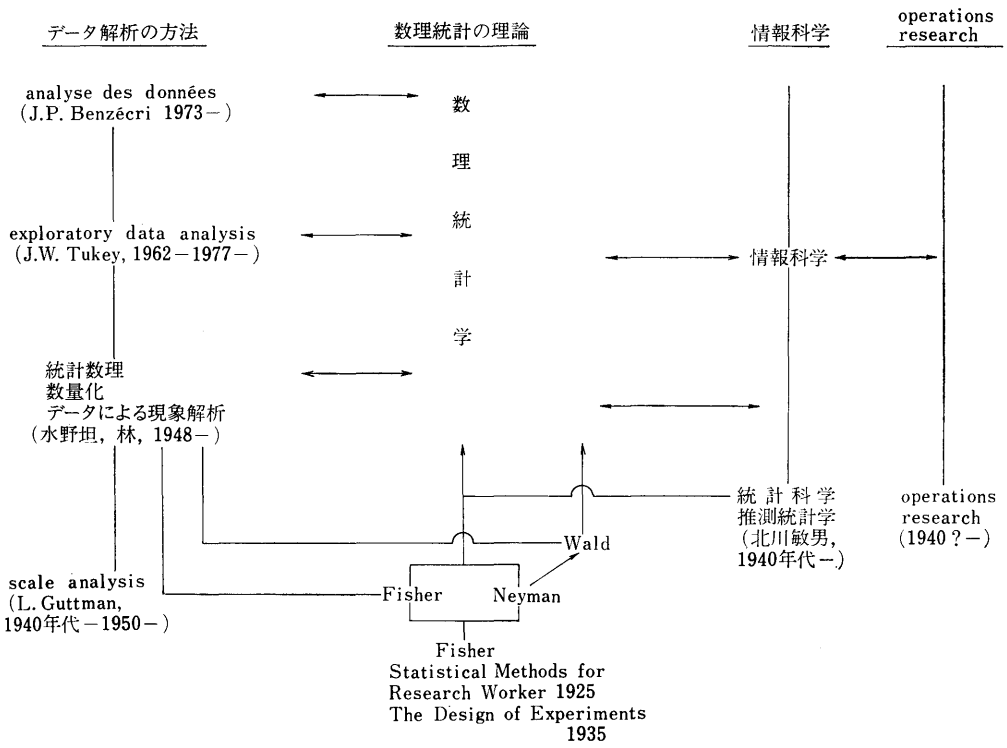
ヨーロッパにおけるこうした動きは，以上のほか，英国にも波及して興味を惹き出している（数値分析法のSneath，農業統計の Nelder を始め，D.R. Cox も関心を示している）．

以上の様な流れを一応の図表にしてみると次頁の様になる．これに記載しない多くの流れがあるが，私の立場から纏めてみた大略の図柄である．

このように見てくると米国の一部とフランス，イタリーを中心とするヨーロッパに私の考えていたような統計学が動き出していることは喜びに堪えない．統計学が新しい方向に踏み出している底流のあることは，「創立 35 周年記念号発刊に当って」（統数研彙報第 27 巻，第 1 号，1980）にも一部触れておいた．

---

* この次に 1 次元の数値で「0.7≠0.5 という不等関係が，検定して有意に出ても，実際の文脈において意味のある差だと確信のもてる人はあるまい」と読める文章が続いている．



5. 終　り　に

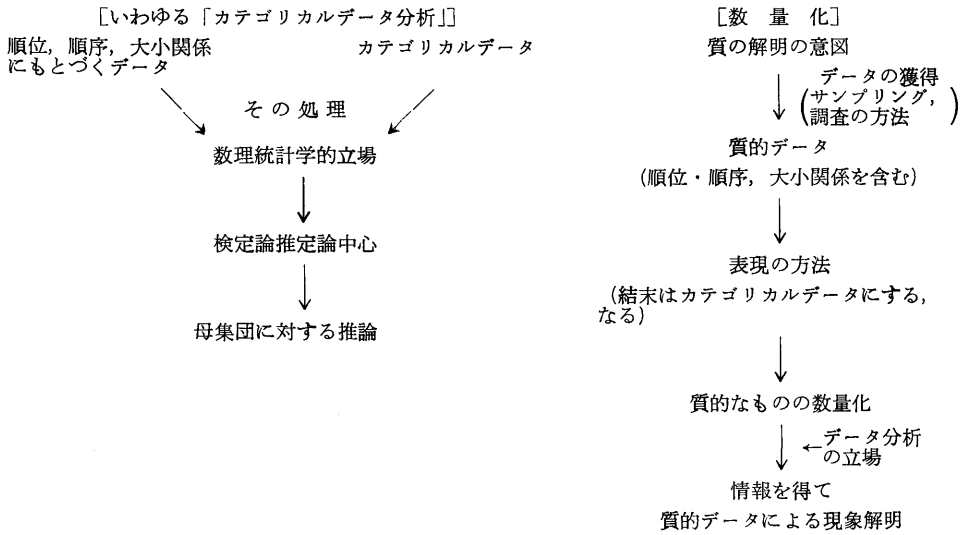
統計的方法を通り一遍に応用するというような単純な問題ではなく複雑な対象に切り込む方法が望まれている．前述の様に exploratory（探険的，探索的，危険を避けて情報を探り出す戦略・戦術というようなことの意訳的立場），detective（探偵の如しの意訳的立場，嘘，誤魔化し，情報隠し，を含むデータから，真実を探り出す方策）などどう考え，どう切り込んで行くかというような視点，考えの立場の設定はもとより，兆候発見のために，データをいかにとるか (design of data)．データをどう分析して行くかという形なきものに形を与えて行くフィロソフィー・方法の定式化も重要な問題となってくる．これまで述べてきたデータ分析の諸方法が深められ，拡大され，脱皮を重ねると共にコンピュータの一層の有効な活用が期待されることになる．

これまで取扱うことの難しかった大規模系，不確定要素の多い複雑系のものを適切に取扱う――本来曖昧なものを曖昧なままにそれなりに厳密に取扱って必須な情報を露呈させて行くことも重要である――方法論の開発も望まれてくる．仮説から仮説へ進みつつ，必要な情報を取り出して行く過程そのものの取扱い，前記の系の制御のための，データによる現象解析のためのダイナミックな方法が射程距離の中に入ってきた感じがある．つまり，正気を取り戻しつつある統計学と言おうか．

後　　注

カテゴリカル・データ分析と数量化の考え方の差異はやはり論じておいた方がよい．数量化もカテゴリカル・データを取扱うのであるが全く異ったものになっている．対比表を次にあげておく．

表　カテゴリカルデータの取扱い



数理統計学で言うカテゴリカル・データ分析はやはり統計的検定論が中心であって対数変換を重用しているが，私はここに疑問を感じる．誤差のない数学的関係が成立するのであれば，単調増加の関数の変換は数学的仮説（母集団統計量の等しいこと）を検定するのに差支えないがたとえば対数変換は小さい0と1との間を0と$-\infty$の間に拡大し――一方大きいところは縮めてしまう――て行くのであるから誤差がどの当りでどの程度であるかによって意味が大きく違ったものになる．また，なまの数が大切のとき対数変換その他の変換によるデータ処理は何を行っていることになるか，$x\,(>0)$の小さいとき，または大きいとき$\log x$を基にする推定の誤差はずい分拡大されてくるのではないか，（中位のxのみを問題にすれば線型と大差なくなる）などの問題点を感じる．変換はデータ処理の数学的立場からすべきではなく，問題の本質に応じてすべきと私は考えている．このあたり二つの立場の差は大きい．

二つの立場は，同じカテゴリカル・データを扱うと言いながら異ったものになり，私の言う数量化の方法は，きわめてソフトな方法（方法論，考え方）を含むもので，これにデータ処理の数理方法が加わるというもので，トゥキイの exploratory データ解析と軌を一にする．

The Germination of New Statistical Methods as the Logic of Analysis of Data

Chikio Hayashi
(The Institute of Statistical Mathematics)

The germination of new logic, methodology and philosophy for analysis of data has been recently found in the sphere of statistics. We can mainly mention here Prof. Tukey's exploratory data analysis and Prof. Benzécri's "analyses des données—analyse de correspondance". These are quite similar to the idea of "statistico-mathematical theory" or theory of quantification of qualitative data which we have developed since 1948.