

統計科学とは何だろう

統計数理研究所 名誉教授 赤 池 弘 次

1. はじめに

ここでは、統計学についての混乱した認識を正し、統計概念の積極的な利用を可能にする研究あるいは教育の体制づくりに寄与することを目的として、統計科学の定義を論じる。このために、まず統計学の定義の変遷に関する簡単な歴史的考察から始めて、統計あるいは統計学についての古いイメージを拂拭しつつ、一般の科学研究に関連する知的活動として際立った特徴を持つ、データに依存して合理的な推論を展開する際の情報処理の本質を探る科学として統計科学を捉える。更にこれをデータに基づくイメージ構成過程の科学と見ることによって、より具体的な視野が開けることをも指摘する。

2. 統計学の定義を廻る混乱

統計科学とは、日本語の通常の使い方から言えば、統計の科学、あるいは統計に関する科学、ということになる。ここで早々に問題となるのは、統計という語の意味である。統計は英語の statistics の訳であるが、statistics にはもともと学問としての統計学と、大量のデータに基づく数値としての統計値とのふたつの意味があり、これが我が国における統計という語の利用にもそのまま反映されている。これだけでも混乱の原因となるのに十分であるが、統計学を表わすものとしての statistics 自体がドイツ語の Statistik の英訳であり、更に、もとの Statistik はイタリア語に語源を取る新造語で、その語源についても諸説が生じるような状況があった。このような由来から見れば、統計あるいは統計学の定義を廻って、さまざまな誤解や混乱が生じるであろうことは容易に想像される。

統計概念の発生が、国家の経営と結びついたものであることは確かである。人口調査、財産調査等の調査が国力の測定に関連して行なわれたことは、歴史上の古い時代から知られている。このようなデータから、社会の動きを支配する法則を捉えようとの科学的な試みを行なった最初の例として、英国における J. Graunt による生命表の解析 (1662 年) が知られている。これは天文学者であり数学者でもあった E. Halley のような科学者等の貢献により、やがて政治算術の名前で知られる研究分野を構成するようになった。

Graunt とほぼ同じ時期に、ドイツでは行政のための科学として一国の国勢の記述を論じる学問分野が誕生し、G. Achenwall がこれに対して Statistik という名称を与えた。J. Sinclair は、スコットランドの統計書を発行するに際して、ドイツにおいて国勢に関する調査が良く行なわれていることに言及し、一般の人々の注意を惹くことを目的として国の状態を示す数値の集まりを statistics と呼ぶことにしたと述べている。やがて、本来記述的な国勢学としての Statistik と数量的な観察と分析に重点を置く政治算術としての statistics の間で正統性を廻る争いが生じたが、後者が勝利を収め、社会集団の法則性の数理的な解析を目的とする科学として statistics の名称が定着した。この辺の歴史的な展開の詳細は、*Journal of The Statistical Society of*

Londonの第1巻(1839年)の序論,第46巻(1883年)656-679頁に掲載されたV. Johnの統計学の語源の歴史的検討に関する独語論文の英訳, Guy (1885), Yule (1905), Pearson (1978), Hill (1984)等によって見ることができる。

このように定義された統計学は,その対象により,人口統計,社会統計,経済統計,政治統計等の分野に専門化するが,そこに用いられる方法としては同一のものがあること,更に同様な研究方法が,医学,心理学,生物学,物理学等の分野においても必要とされることが認識され,やがてその数理的側面を研究する学として数理統計学 (mathematical statistics) が成立することとなる。観測データの処理に確率論的考察を利用するという数理面に関係する歴史的展開の詳しい記述は Stigler (1986)に見られる。

統計的方法の一般性が認識されるとともに, statistics をあらためて方法の学と捉える見方が今世紀初頭にほぼ定着したことは, 1910年に初版が発行された G.U. Yule の著書 (Yule (1929)) によっても明らかに認められる。これにより, 個別科学に属する固有の対象領域に関連する研究は, 夫々の分野の名称を附加して人口統計学, 社会統計学等のように呼ばれるか, あるいは計量心理学, 計量経済学等の名称を持つ研究分野に包含されるようになり, 現在の世界の大勢としては, 統計学は統計的研究方法を研究する学を意味するものと理解されるようになっていたのである。

しかしながら, 辞典等の記述に見られる現在の我が国の社会通念に於ては, 統計学は未だにある場合には古い定義による社会集団の法則性の研究をするものと捉えられ, またある場合には数学的形式的側面のみ注目されて数学の一分科と見なされたりしている。このような我が国における統計学の受け取られ方は, 最近迄の文部省科学研究費補助金の分科細目に典型的に示されていた。平成元年度に至る迄, 経済学の部に分科統計学が置かれ, 理学の部の分科数学の細目数学一般が確率論・統計学を含むものとされて来た。これは, 統計学は科学としては社会集団を研究対象とするものであるが, この科学に固有の手続き及び観察の方法の研究にも学として考慮すべきものがあると捉えていた19世紀後半の西欧における統計学観にほぼ対応しているように見える。

穂積(1980)は, 明治3年2月発布の大学規則にスタチスチックスを訳して国勢学としたことに触れ, これは2百年前の欧州における用例を, その後の沿革も知らずにそのまま採用したのであろうとしている。これに代わって明治7年頃から統計学がスタチスチックスの訳として一般的に用いられるようになったのであるが, 今度は社会集団を研究する学としての意味に固定化され, 世界の大勢が統計学を方法の学と捉える方向に転換したにも拘わらず, 我が国においては統計学の諸科学の中における位置付けの積極的な見直しがなされないままに経過したのである。我が国における統計あるいは統計学の初期の導入の状況については, 昭和36年に総理府統計局から発行された「政表課誌」, あるいは高野岩三郎(1942年)による「杉享二博士と本邦の統計学」と題する論文に詳しい記述が見られる。

3. 数理統計学の発展とその問題点

国勢学としての Statistik に対して政治算術が statistics としてその優位性を確立するようになることを可能にしたのは, 後者における頻度分布の概念の数学的取り扱いの持つ一般性であろう。これにより, 社会集団を支配する法則性の研究を通じて得られた集団の特性値として定義される諸概念が, 広く一般の科学研究においても有効性を持つことが次第に認識されるようになった。熱力学的エントロピーの統計的あるいは確率的解釈を通じて展開された気体分子運動論の研究成果が, M. Planck による量子論の導入に導いたのはこのような発展の方向を示

す最も顕著な例である。

社会集団の法則性の追求は、観測データに基づく推論の数学的な展開の必要性を明らかにしたが、これにはデータに基づいて必要な分布の形を決定することが基本的な問題となる。このことから、分布の形を数学的に表現することと、データを用いてその形を決定することが、方法としての統計学の主要な課題となる。今世紀初頭頃における K. Pearson の計量生物学に関連する研究は、この方向の研究の展開の典型的なものを示している。彼は更に分布の型の想定 の妥当性を確率論的に検証するための検定の方法を導入し、R.A. Fisher は利用できる観測データの量が極めて限られる状況の下で分布の決定が求められる場合を考え、分布がいくつかの変数値 (パラメータ) を除いては決定されているものとして未知のパラメータの値をデータに基づいて推定するという形に分布の決定の問題を定式化し、推定の理論を展開した。検定あるいは推定の方法と理論の展開には高度の数学的知識が要求され、ここにこれらの研究を中心とした数理統計学が成立し、不確実な推論の合理的な実現を目的とする、いわゆる帰納的推論の組織的な展開を目指す研究が展開されるようになった。

第2次世界大戦に際し、数理統計学的手法については、生産ならびに品質管理、OR、情報分析等の分野での利用が目覚ましく発展し、戦後英米両国を中心として数理統計学関係の学科が多数の大学に開設されるようになった。これらの多くが単に department of statistics と呼ばれていることから、今世紀半ばには単に統計学という場合には方法の学としての統計学を示すものであるとの考え方が完全に定着していたことがうかがえる。

このようにして方法の学として定着した統計学は、果して成功的に展開されて来たであろうか。米国における状況を見るに、戦後の大学における種々の具体的分野に関する知識あるいは経験を持った第一世代の教育者等は、多くの魅力ある問題を提示するとともに他の関連分野の教育研究にも参加協力していた。しかし具体的な研究対象を持たないままに方法の精密化のみを追求する形の研究が主流となるにつれ、統計学の研究は次第に孤立化の傾向を示し、統計学的方法を必要とする個々の分野の研究者等は自らの力で統計的な問題の解決を進めるという状況が目立ち始めた。これが Shafer (1990) が問題として取り上げている現在の米国の大学における統計学科の置かれた状況である。なお、この論文については筆者もコメントを求められ、論文と共に印刷されている。

Shafer が問題とするような統計学の研究教育体制を採る限り、同様な問題はどこでも発生する。過日 (平成5年11月4日) 日本学術会議統計学研究連絡委員会の主催で行なわれたシンポジウム「大学教育における統計学」においては、新しい方向を目指す積極的な議論もあったが、この危惧が我が国においても現実のものとなっていることを示す議論があったことに注意しなければならない。

4. 統計科学の定義

種々の概念的な混乱を避けるために、統計学に代わるものとして統計科学を考える。長々しい定義を避けて、統計科学を“統計的な物の見方を研究対象とする科学”と一応定義してみよう。統計的な物の見方とは何かが問題になるが、その議論は後回しにして、まず統計科学という語のこれ迄の使われ方を見ることにする。

統計科学を英訳すると statistical science となる。英国王立統計協会の前身たるロンドン統計協会の会誌を見れば、statistical science は science of statistics の同義語として二つの表現が交互に用いられていることが認められる。統計科学には statistical sciences という複数型の英訳もある。これは、統計科学の研究が方法あるいは対象について種々の専門的な研究分野に

細分化することを想定した表現とみなされる。

米国数理統計学会の *IMS Bulletin*, Vol. 18, No. 2, 1989 に掲載された編集者宛の手紙で、McGill 大学の A.M. Mathai は、statistical science と statistical sciences を自分の新造語と考える旨を述べ、彼自身 1969-71 年に statistical sciences をランダム現象の組織的な研究と定義して 1971 年に Statistical Science Association of Canada を設立したものであり、これらの語がそれ以前に使われていたかどうかを知りたいと述べている。前述のロンドン統計協会雑誌は 1883 年に出版され、我が国においても統計科学研究会 (Research Association of Statistical Sciences) が 1941 年に既に設立されていたことを考えると、Mathai 氏がこれらの語を用いた最初の人でないことは明らかである。ランダム現象の研究と言う定義は、歴史的に発展して来た統計学の定義の自然な延長からはやや外れているように思われる。

昭和 63 年の総合研究大学院大学の開学に際して、数物科学研究科に統計科学専攻が設置され、科学的仮説の構築・検証等、合理的な推論あるいは予測の実現を目的とするデータの有効利用に関する教育研究を行なうこととなった。また、日本学術会議統計学研究連絡委員会による平成 2 年 12 月 21 日附の同委員会報告「統計学研究教育体制の整備のための具体的方策について」では、“もともと集団現象の研究に端を発した統計概念の情動的側面に注目しつつ、その深化発展と有効利用を広く諸科学との関連において展開することを目指す科学”として統計科学を定義している。統計科学の定義に関係するこれらの記述には筆者自身関与したものであるが、本章の始めに与えた定義の内容をより具体的に敷衍したものとなっている。

科学研究費補助金の分科細目についても、経済学の部における分科統計学は平成 3 年度より経済統計学に変更され、平成 5 年度より複合領域の部に分科統計科学が新設された。これらの動きを総合的に眺めると、現在の我が国における統計関連の研究教育を取り巻く環境は、ようやく旧来の統計学の定義を廻る混乱の影響から脱し、統計科学の名のもとに新たな発展段階を迎えていることが認められる。

5. 統計的な物の見方とその科学

ここで統計的な物の見方とは何かについて考えてみよう。統計的な物の見方とは、ひとつ、あるいは一群、の観測値(データ)が与えられた場合に、これを適当な集団に結びつけることによってその意味を理解するような物の見方である。この集団の構造がデータの意味を決定する為に用いられる。簡単な例として、ある人の血圧の測定値が与えられた場合、これを同年代の同性の人の血圧の同じような条件の下での測定値の集団と対比することにより、血圧の正常異常についてのひとつの判断が可能になるが、同じ測定値を本人の過去の測定値の集団と対比して見ることにより、別の意味での判断がなされる場合もある。このように、対比させて考える集団の選び方によってデータから取り出される意味が変わる、すなわちデータの見方が決定されるのである。この対照の目的の為に設定する集団(仮に対照集団と呼ぶことにする)を如何に構成するかが、統計的な物の見方の成否を決定することになる。

当然のことながら、この対照集団の選択は、データ利用上の目的と、利用者が持つ知識、経験に依存してなされる。データの持つ情報はこれに応じて取り出されるわけであり、単なる観測値あるいはデータだけでは本来情報の名には値しない。かくして、統計的な物の見方を有効に展開するためには、様々の角度からデータを眺める力を持たなくてはならないことが分かる。したがって、統計的な物の見方についての科学的な研究を効果的に進めるには、様々な状況におけるデータ利用の実態についての理解が必要であり、ここから統計科学の研究が広く各種の研究分野にわたる具体的利用から離れては推進し得ないものであることが結論される。

データに基づく推論の目的を予測あるいは制御の実現とする場合には、現在のデータを一定の確率的な構造から生み出されたものとみなして可能な観測値の集団と結びつける方法の有効性が広く認められている。このときには、確率的な構造として種々のものを想定し、データに基づいて此等の構造の適切さを情報量の名で呼ばれる評価規準に従って比較検討することにより、データから有効な情報を抽出する方法を実現することが可能になる。このようなデータの利用方法は、極めて広い適用可能性を持ち、また情報量規準が誰にも同様に理解できる客観性、あるいは間主観性を持つということから、統計科学の科学としての特性を明示するひとつの固有の方法を与えるものとなっている。

6. イメージ構成過程の科学としての統計科学

これ迄の統計学の理論的研究の成果について見ると、与えられた条件の下での最適化という形式での問題の精密化の議論が多い。現実の問題の処理という立場においては、このような接近は高々既存の方法の枠内での修正改良にしかつながらない。制御の立場との類比で見れば、現在の対象システムの構造と動きを前提とするフィードバック制御の改良に類するものとなる。統計的方法の研究がこの段階に止まる場合には、実用上既存の手法を超える著しい改善への寄与は期待し難くなる。

統計的方法のより積極的な利用は、問題の新しい捉え方の可能性を探る形での利用である。制御との類比で言えば、対象システムの構造の改変かあるいは既存のシステムの構造を利用しつつもその動きに積極的な変更を導入するフィードフォワード的な制御を考えることになる。これは人間の積極的な知的活動を表わすものであり、通常試行錯誤の過程を通じて実現される。ここでは対象に関する新しい見方を積極的に投げ込むことが必要であり、そのための努力を継続する意識的活動が要求される。統計的な枠組みで考えれば、これは種々の期待あるいは仮説を積極的に構成提示し、その適否を検討するという過程によって実現される。この過程は C.S. Peirce によって *abduction* と呼ばれているものであり、このような過程と情報量規準による統計的モデルの評価との関係については既に別の所で論じてある (Akaike (1992))。

英国王立統計協会の *RSS NEWS*, Vol. 21, No. 3, 1993 によれば、スポーツ関係行事の数の増加とともにスポーツ統計に関心が高まり、新しくスポーツ関連の統計学の雑誌の発刊が議論されているという。この場合についても消極的に単なる過去の得点記録等の解析に止まることなく、より積極的に競技者の動きそのものに関係するスポーツ科学の問題と取り組むことにより、やがて心身問題というような統計的視点の展開と本質的にかかわる問題との接触の端緒が得られるようになるであろう。競技者の動きはある種のイメージによって導かれる。従って、ここでは種々のデータに基づく適切なイメージの構成が問題となる。統計科学の定義を、あらためて“データに基づくイメージ構成過程の科学”としてみるとどうであろう。この定義は、これ迄の統計的手法の展開における図的表現重視の根拠を明らかにすると同時に、イメージの構成という、人間の「こころ」あるいは脳の最重要機能との結びつきをも明らかにする (シャングジュ (1989))。更に、人が眼で物を見る場合に、ひとつひとつの光子 (光の粒子) の到着の刺激から連続的な像を作り上げる過程を考えると、統計的なイメージの構成過程がまさに言葉通りに「物の見方」にかかわっていることが分かる。

このような立場から見ると、統計科学の研究者があらゆる個別科学の分野の研究者と密接な連繋を保つことの必要性は明らかである。統計学研究連絡委員会の報告で与えられた統計科学の定義の妥当性があらためて確認される。多分野の研究者の交流によって実現される統計科学的接近による具体的問題の成功的な解決例の集積のみが統計科学の発展を支えて行くのであ

る。このような状況にあって、統計科学が独立の科学としてその固有の内容を豊かにし新しい方向への展開をつづけるには、ひとつひとつの具体的問題を深く掘り下げるという努力とともに、更に多数の応用的な問題の集団に共通の特性に注目するという高次の統計的視点に立った研究が進められなくてはならないのである。

参 考 文 献

- Akaike, H. (1992). Implications of informational point of view on the development of statistical science, presented at the First U.S./Japan Conference on the Frontiers of Statistical Modeling, An Informational Approach (to be included in the Proceedings Vol. 3 (ed. H. Bozdogan), Kluwer, Dordrecht).
- ジャンジュール, J.-P. (1989). 『ニューロン人間』(新谷昌宏 訳), みすず書房, 東京.
- Guy, W.A. (1885). Statistical development, with special reference to statistics as a science, *Jubilee Volume of the London Statistical Society*, 72-83.
- Hill, I.D. (1984). Statistical Society of London—Royal Statistical Society, The first 100 years: 1834-1934, *J. Roy. Statist. Soc. Ser. A*, **147**, 130-139.
- 穂積陳重 (1980). 『法窓夜話』, 岩波書店, 東京.
- Pearson, E.S. (1978). *The History of Statistics in the 17th and 18th Centuries, Lectures by Karl Pearson given at University College London during the Academic Sessions 1921-1933*, Griffin, London.
- Shafer, G. (1990). The unity and diversity of probability, *Statist. Sci.*, **5**, 435-462.
- Stigler, S.M. (1986). *The History of Statistics: The Measurement of Uncertainty before 1900*, Harvard University Press, Cambridge, Massachusetts.
- Yule, G.U. (1905). The introduction of the words "statistics", "statistical" into the English language, *J. Roy. Statist. Soc.*, **68**, 391-396.
- Yule, G.U. (1929). *An Introduction to the Theory of Statistics*, 9th ed., revised, Griffin, London.

Statistical Science, What It Means

Hirotsugu Akaike

(Emeritus Professor of the Institute of Statistical Mathematics)

In this paper the development of the meaning of the word statistics is briefly reviewed and the development of the concept of statistical science is discussed to correct the previous confused image of statistics. Finally, the interpretation of statistical science as the science of the process of image construction based on data is introduced to provide a view of its relation to intelligent human activities in general.