

記念講演会要旨

統計科学の将来

統計数理研究所 赤池 弘 次

(平成 2 年 11 月 2 日, 統計数理研究所 講堂)

はじめに

この講演では, 統計科学研究における先人の業績と講演者自身の経験とを通じて, 統計的方法の応用研究と基礎研究との相互作用を明らかにし, 統計科学の発展のための条件を論じることにはしたい。

1. 統計科学の定義

統計科学は英語の Statistical Science あるいは Science of Statistics に相当する。Statistics という英語そのものの内容は, 社会集団の学の定義に始まり, 今世紀初頭には確率論を利用する観測データの処理方法の学を意味するように広がっている。その本質は, 集団の特徴の把握とその有効利用を目指すものと考えられる。統計科学は, データの現われ方に当面の対象の動きを支配する法則性が反映されるとみなし, データを生み出した仕組の特徴をデータを通じて把握, その有効利用の方法を研究する科学と定義することができる。日常の新聞紙面を賑わす社会経済統計データの動き, あるいは台風の進路予想等は, いずれも統計科学の具体的な適用場面を示す事例である。

2. K. ピアソンと R.A. フィッシャーの業績と研究環境

計量生物学の創始者として, 統計学の数理的な形成に大きく寄与したのは K. ピアソンである。ピアソンははじめに F. ゴールトンの著述 Natural Inheritance を通じて相関の概念に触れ, これに大きな興味を抱いたといわれる。ロンドン大学において同僚の動物学教授 W.F.R. ウェルドンとの接触を通じて, 計量生物学の提唱と, 雑誌 *Biometrika* の創刊を実現し, 進化論に関する考察を展開した。

数理的側面から見ると, 大量データの特記述のための頻度曲線のあてはめの方法と, その適合度の検定のためのカイ 2 乗検定の導入が最も著しい貢献であり, これらはいずれも統計理論の発展の歴史に残るようなものである。動物学という具体的な研究分野の研究者ウェルドンとの共同研究と, 大量データの処理という具体的な課題とが, 統計科学的に見て著しい貢献に導いたものといえよう。

R.A. フィッシャーは K. ピアソンの進化論に関する業績に刺激を受けたといわれるが, ローザムステッド農事試験場に職を得て, 農業実験に関するデータ処理の方法とその理論を展開した。試験場での最初の仕事として, 過去の穀物収量の実験データの解析を行ない, その過程を通じて分散分析あるいは実験計画などの手法の確立に進んだ。

フィッシャーの統計科学に対する理論面での貢献で最も著しいものは、尤度概念の導入とその積極的利用の展開である。これに基づいて推定理論が組織的に展開された。さらにフィッシャーは検定統計量の分布の導出を行ない、小標本の精密な分析法の実用化を推進した。小標本の視点の重視は、農事試験関係のデータが、環境条件の変動を考慮すると、きわめて限られた数のものとならざるを得ないという事実から発したものである。

尤度概念の導入は、ピアソン以来の頻度曲線のあてはめの問題に対し、一般性のある解法を与えようとしたものである。これは、統計科学研究の歴史的な発展の自然な産物ともいえる。実際に、尤度の導入を初めて行なったのはフィッシャーの統計に関する最初の論文においてであり、ケンブリッジ大学における学生時代の業績である。このように見ると、今世紀における統計に関する数理的研究を方向づけたとみなされるフィッシャーの業績も、深くその研究環境に依存していることが明らかとなる。

3. 情報量規準の視点とベイズモデルの利用

フィッシャーの穀物収量の年次変動データの解析では、トレンドの除去のために直交多項式のあてはめが行なわれている。この場合あてはめるべき多項式の次数が与えられれば、残差項が独立に同一の正規分布に従うことを想定して、係数の最尤推定値は単純な最小2乗法によって求められる。しかし最尤法の考えは、直接多項式の次数の推定には適用できない。ここに、未知パラメータ以外の構造が確定している場合にしか適用できないという最尤法の限界があった。

演者が導入した情報量規準 AIC は、尤度を対数変換して得られる対数尤度の符号を反転させたものの平均値が、各パラメータ値が決定する分布から真の分布がどの程度離れているかを示す、カルバック・ライブラー情報量に対応することに着目することによって得られた。AIC はパラメータ数の増加分に対するペナルティを表す項を含み、最小の AIC を与える次数を次数の推定値とすることにすれば、フィッシャーが扱った多項式の次数決定問題の実用的な解が得られる。

AIC 導入の主な貢献は、情報量概念に基づく情報量規準の視点の導入にある。AIC の考えは、実用上のモデル決定の必要性から得られたものであり、最小 AIC 法は未知パラメータの数を合理的に制限する手順を与えた。しかしながら、未知パラメータが一定の確率分布(先験分布)に従うものとするベイズモデルを利用すれば、未知パラメータの数を制限する必要はなくなる。実用上の問題は先験分布をどう定めるかであったが、先験分布の表現に少数の未知(超)パラメータを導入し、情報量規準の視点から検討することにすれば、容易にベイズモデル実用化の指針が得られることが明らかとなった。こうして、ベイズモデルの利用に対する理論的なよりどころが与えられ、全く新しい統計的情報処理法の実用化が可能になった。応用研究に出発した基礎的研究の結果が、新しい応用分野の開拓に有効な視点を与えたわけである。

4. 統計科学の将来

現在、科学的観測あるいは経営管理等の分野で、急激なデータの蓄積が生じており、有効な情報の抽出法の必要性は無限に増大しつつある。したがって、統計科学の将来は基本的に明るいものであることは確かである。ただ、前述のような歴史的事例について見ると、統計科学の研究を成功的に発展させるためには、

- ・ 具体的な問題の分野の研究者
- ・ 統計的方法の研究者
- ・ 数値的方法あるいは計算機利用の研究者

の密接な協力が必要であることが分かる。

我が国の場合、統計科学の研究教育体制は、先進諸国に比し驚くほどに組織化がおこなわれている。統計数理研究所は、大学共同利用機関として、上記の条件を充たす形で共同研究を推進することについて経験を蓄積してきている。その活動が、我が国における統計科学研究体制の確立に寄与するものとなれば、研究所にとってまことに幸なことであると考えられる。