

特集 / 統計科学の最前線

統計科学の新展開

甘 利 俊 一

統計学は古くから発展した伝統的な科学の一つである。しかし、近年コンピュータが発達し、大規模な計算ができるようになった結果、従来の枠を破る手法が次々と現れた。さらに、人工知能、脳、遺伝子、金融、心理など多くの分野への応用が広がり、新しい統計科学として情報分野の基本的な方法論の一つとなり、新しい発展を遂げつつある。

1. 統計学から統計科学へ

統計学は歴史の古い学問である。部族や国家など、社会組織が整備されるにつれ、人口、収穫量、税、貿易額などのデータが集められ、これを集計する国勢学が生まれる。ただデータを収集するだけでなく、その代表値や分布の広がりなど、統計値が必要になってくる。こうして統計学が誕生した。

こうした記述統計学を一変させたのは、R.A. Fisher である。Fisher は、データの背後にそれを生成するメカニズムとしての確率分布があり、測定されたデータはその確率分布からの標本であると考えた。こうすると、データをただまとめるのではなく、データからその背後に隠れた確率分布すなわちデータを生成するメカニズムについての推論を行う、という姿勢が明確にできる。こ

れは確率論に基礎を置く、推論の科学である。

統計学では、データを発生する確率分布に対して、未知のパラメータを含む分布の族を想定する。このどれかからデータが生成されたものとして、データからそのパラメータの値を推定すれば、分布が推定できる。このとき、どのような推定量が良いのであろうか。また、分布の推定はどの程度まで良くできるのであろうか。

Fisher はデータを発生する確率をパラメータの関数と考えた確率尤度に着目し、これを最大にする最尤推定が一番良いと信じた。たしかに、データ数の多い漸近理論にあっては、これは最良のものの一つである。また、推定に関連してデータが持っている情報量を考察することができる。これは Fisher 情報量（多次元の場合は情報行列）である。Cramer-Rao の定理が、情報量を用いて推定の良さの限界を明らかにする。このほか、それ自体では情報を持たない補助統計量の役割、条件付推論など、古典推定理論の枠組みが完成していく。

もう一つの重要な枠組みは検定である。データを発生する分布に対してある仮説 H_0 を置き、これが受容できるか、それとも棄却されるべきかを確率に基づいて決めるのである。このために、対立仮説 H_1 を置く。しかし、検定は H_0 か H_1 かを決めるのではない。 H_0 が受容できるか、棄却されるべきかを決めるのであって、両仮説は対称に扱