

Opening Remark にかえて: 趣旨説明と問題提起

統計数理研究所 伊庭幸人

プログラム上の制約でイントロが2日目になってしまい、少し変な具合だったが、以下の点について、スライド1枚ずつで短い話をした。

1 統計科学と他の分野の関係

非線形科学と統計科学

- 「非線形物理のモデルについて統計的にパラメータや関数形を推定する」という形の組み合わせを考えがちだが、非線形物理のモデルの多くは定性的な現象を再現するために開発されたものなので、本来の趣旨とは違う利用のされ方になっているおそれがある。また、推定した関数などが、「位相応答曲線」のような理論の内部的な概念だった場合、推定の結果が正しいかどうかは、長い推論や仮定なしには実証できないのではないか、という心配がある。
- 別の考え方としては、データ解析の手法の中に、非線形科学の概念やその理論の与える制約を組み込んでいくという方向がある。
- 「はじめにモデルありき」ではじまる非線形科学が複雑な系からのデータを解釈するために役立つという保証はない。極論すれば、「統計科学にとって非線形科学などは無用」という立場もありうる。

実験家と統計的方法

- 実験データの解析という要請に応えていくことはもちろん重要。
- 統計科学の側からすると、要請に応えるだけでなく、より積極的に対象の性質をあばく方法を提案していくことが望ましい。
- 一部の実験科学の分野では、理論モデルの研究がそれなりに尊重されているのに対し、統計科学的な研究は「統計ソフト屋」とみなされているように見える。統計科学の側にもそれなりの世界観があるということを知らせていく必要がある。

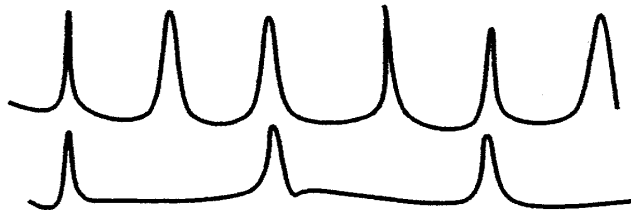
工学と科学

- 現代の統計科学，特に機械学習系の進歩は，工業プラントの制御やパターン認識など，工学系の応用に負う所が大きい．そこでは予測志向が強く，伝統的な「科学」の「真理」の追求とは相容れない面もあるように思われる．
- これについては，現代的な統計科学の中に「科学」でも受け入れられるような論理を求めていくという立場もあるし，むしろ「科学」のあり方の変容が求められている，という見方もありうる．
- 統計的手法のアウトプットとしては，「予測・判別」「分類」「検証」などがあるが，これらの適用範囲や相違点などについて，上記のことを踏まえた考察が必要だろう．

2 統計科学の内部での「非線形」

線形の解析手法の限界

- 原理的には，時系列の位相情報など線形手法では扱えないものはいろいろある．たとえば1 : 2の引き込みでの情報の流れはその一例．



- 実際には，純粋に非線形部分にのみ情報が含まれるというケースはそれほど多くないので，多変量 AR モデルなどでもかなりいける，という見方もある．
- たとえば，複雑なスパイクパターンに含まれる情報を取り出せる手法があればすばらしいが，複雑な情報を取り出す前に，まず単純でアナログな部分をモデル化しないといけないのが普通である．

ユニバーサルな手法か知識の表現か

- 旧式の統計学は「正規分布」＋「線形関数」の狭い範囲で「普遍性」を主張してきた．
- これを乗り越えようとする方向としては大きくわけて2つある．

- ひとつは、より広いクラスの普遍性を求めて、内在的な数理的根拠のある手法を追求するもの。「順序統計量に基づくノンパラメトリック検定」から、各種の「セミパラメトリック」「ロバスト」「非計量」などはこの方向のものと考えられる。
- もうひとつは、個別の知識を組み込める枠組みを重視し、普遍性は「個々のモデルを統制し検証する手段」によって実現しようという考え方。赤池らの情報量統計学はこの方向の先駆けである。「ノンパラメトリック平滑化」「階層ベイズ」「カーネル多変量解析」などもこの方向のものとして解釈できる。

解釈モデルと生成モデル

- 統計的なモデル化の大きな流れとしては、「生成モデル」と「判別モデル」（より一般には「解釈モデル」）の2つがある。
- 階層ベイズモデルや状態空間モデル・隠れマルコフモデルは生成モデルの典型。まずデータを生成するメカニズム全体を仮定して、ベイズの公式によって、与えられたデータを解釈する。
- 判別モデル（解釈モデル）では、与えられた目的に必要な最小限の部分を、なるべくベイズの公式などを必要としないようにモデル化する。
- 判別モデル（解釈モデル）の場合、および、生成モデルでも生成過程のモデルが充分信頼できないと感じられる場合は、モデルの検証や教師なし推定のために別の「生成モデル」を用意することが必要になる。「サロゲーション」や「ブートストラップ」はその例であると考えられる。
- 「サロゲーション」と「ブートストラップ」は分野による用語の違いと思ってもよいが、しいていえば、「サロゲーション」が「もしそれが無かったら」の世界を生成するのに対し、「ブートストラップ」は「もしそうだったら」の世界を生成する、という点が違うかもしれない。
- 以上は線形でも非線形でも同じであるが、非線形のモデル、非線形の現象を扱うことで、解釈と生成のギャップが大きくなる可能性がある。
- モデルの「リアリティ」の度合いということが、統計科学を眺めるひとつの軸になるのではないか。