

データサイエンス

**keyword**

- 統計科学
- モデル選択
- 関数データ解析
- スパース正則化

---

**松井 秀俊**  
Hidetoshi Matsui

データサイエンス学部  
准教授

**【プロフィール】**

- ・2004年 九州大学 理学部数学科卒業
- ・2009年 九州大学大学院 数理学府博士後期課程 修了
- ・2016年～ 滋賀大学データサイエンス教育研究センター 准教授  
科学技術振興機構 さきがけ研究員
- ・2017年～ 滋賀大学データサイエンス学部 准教授

**【専門分野】**  
統計科学

**【主な社会的活動】**

- 所属学会
  - ・応用統計学会
  - ・日本統計学会
  - ・統計科学研究会
  - ・計算機統計学会

**【研究テーマ履歴】**

- ・経時測定データに対する統計モデルの構築およびスパース正則化法に基づく推定
- ・がん細胞システムへの影響解明のための高次元統計的モデリング手法の開発

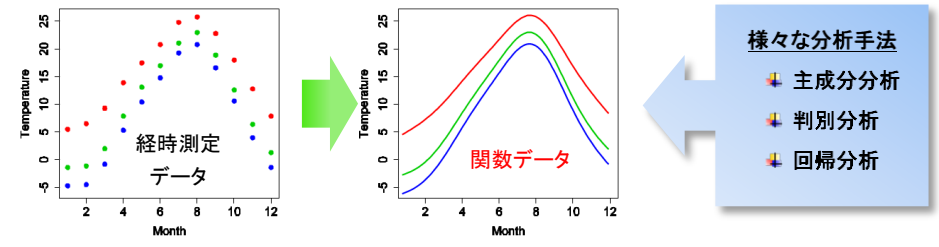
**【代表的な研究テーマ】**

- 経時測定データに基づく統計的モデリング
- スパース正則化を用いた変数選択法の構築

**課題解決に役立つシーズの説明**

近年の計測技術、計測機器の発展によって、これまでは計測することが困難であった大規模なデータが得られるようになってきた。そしてそれに伴い、データの構造が多様化、複雑化の傾向を見せている。本研究では、このようなデータから有効な情報を集約して抽出し、新たな知見を探り出すための統計的モデリング手法の開発を行っている。

データ構造の一つに、各個体が時間の経過に応じて繰り返して計測される形式がある。例として、下図のように日本の3地点(個体)の月別平均気温の推移を記録したデータがある。この他にも、複数名の患者に対する治療後の症状の経過や株価の変動などが挙げられる。このようなデータは経時測定データとよばれ、1時点のみのデータに比べてはるかに豊富な情報が内包されている。しかし、このような形式のデータを計測する際、ある個体がある時点で計測値を得られなかったり、個体ごとに計測時点がバラバラになってしまう状況が多い。このような場合、これらのデータに対して主要な統計手法を直接適用することが困難となってしまう。そこで、これらのデータは時間の経過と共に連続的に推移しているものと考え、曲線を当てはめて得られた関数を改めてデータとして処理する解析手法が提案された。この方法は**関数データ解析**とよばれ、これにより上で述べた問題点を解決し、比較的容易に経時測定データを解析できるようになる。



- 関数データ解析を用いる利点:**
- ✓ 経時測定データの観測誤差(ノイズ)を除去した解析が可能
  - ✓ 観測時点数が多い場合、その次元を削減
  - ✓ 観測時点・観測時点数が個体ごとに異なっても容易に解析可能
  - ✓ 関数データの微分情報を利用可能

また、高次元データから効率的に情報を縮約する手法について研究している。いま、ある都市の住宅の価格と、これに関連しそうな要素(間取り、築年数、駅からの距離など)の候補となる数多くの情報がデータとして得られたとき、住宅価格に実際に関連(因果関係)がある要素を選択したいという問題を考える。これは**変数選択**問題とよばれ、住宅価格の例だけでなくあらゆる分野で古くから直面する問題である。変数選択のための方法としては、候補となる要素の組合せの中から最適なものを探索する方法が多く用いられてきたが、要素の数が多いと組合せの数が膨大となり探索が困難となる。これに対して近年提唱された**スパース正則化**は、統計モデルを推定すると同時にこれらの要素の選択を可能にする方法で、近年の統計学の研究で最も注目を集めている分野の一つである。本研究では、様々な形式のデータを解析するための統計モデルに対してスパース正則化を応用することで、データ間の複雑な関係性を明らかにすることを目的としている。

**企業・自治体へのメッセージ**

本研究では、上記の手法を生命情報科学(バイオインフォマティクス)や化学計量学(ケモメトリックス)、農学など、様々な分野のデータ分析への応用も行っています。本研究で開発している統計手法は、分野を問わず汎用的に用いられるものです。経時測定データといった複雑な構造を持ったデータから、有効な情報を新たに抽出するための統計モデルの開発にご協力いただける企業・自治体を探しています。