

統計知能分野 (旧名称：数理システム論分野)

志望区分：シ - 5 (a)

教授 下平英寿 准教授 本多淳也 助教 中山優吾

統計学，機械学習，データサイエンスの手法と理論を探究

ビッグデータ、データマイニング、人工知能の流行を支える理論的基盤として統計学・機械学習は重要な役割を果たしています。とくにランダムネスを考慮してデータから帰納的推論を行う方法論を提供することが統計学の大きな特徴です。ベイズ統計学の事後確率、頻度論の p -値など、不確実性のもとで信頼度を定量化する試みは科学・工学・医学など様々な分野に普及しました。確率モデルを通してデータから推測、予測、決定を行うための様々な手法や概念、たとえば最尤法、モデル選択、ロバスト統計学、漸近理論、ブートストラップ、仮説検定などが生み出されてきました。近年ではデータ収集や処理が容易になり、ウェブやソーシャルメディア、生命科学や宇宙科学における大量のデータを扱える新しい方法論の必要性が増しています。

現実のデータにとりくんで、新たな理論を作る

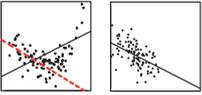
かつて遺伝学において R. A. Fisher が統計学を飛躍的に発展させたように、現実と向き合うことが方法論の発展をもたらします。我々の提案した信頼度計算の理論とアルゴリズムは DNA 配列解析（最近では**新型コロナ**の系統解析）、遺伝子発現解析でも使われています。機械学習の汎化誤差の理論、因果推論、複雑ネットワーク成長メカニズムの統計推測や、新しい情報統合の多変量解析法を提案してソーシャルメディアからの画像認識、文書データからの自然言語処理などの分野でも成果があります。動的意思決定アルゴリズムの構築にも取り組んでおり、新薬の開発や新商品の推薦のように試行錯誤を通して初めて情報が得られる現実の問題において、試行錯誤の対象を適切に選択するアルゴリズムを理論・実践の両面から研究しています。

数学とプログラミング、どちらも重要

研究で最も重要なのはアイデアとデータです。そして数学とプログラミングは力です。定理の証明とコーディングは似た作業ですね。数学に自信のある人、Python, R, C++のスキルがある人は活躍するチャンスがあるし、やる気さえあれば研究を通して実力はつくものです。

最近の活動は研究室ウェブサイトを御覧ください <http://stat.sys.i.kyoto-u.ac.jp>

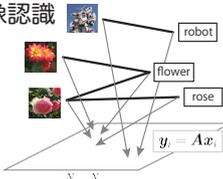
共変量シフト：予測分布と確率密度比


$$E_0 \left\{ \frac{q_1(x)}{q_0(x)} \log p(y|x; \theta) \right\} = E_1 \{ p(y|x; \theta) \}$$

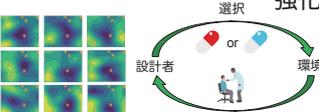
モデル選択 $AIC = -2 \log L(\theta) + 2m$

情報幾何学 $D(q, p) = - \int q(x) \log \frac{p(x)}{q(x)} dx$

情報量規準・汎化誤差 $\phi(A) = \sum_{i=1}^N \sum_{j=1}^N w_{ij} \|Ax_i - Ax_j\|^2$

画像認識 

バンディット問題の理論とアルゴリズム $\mu(x|\mathcal{H}_t) = (y_t - \mu(x_t))^\top K_t^{-1} k(x_t, x)$

強化学習 

統計学・機械学習

グラフ埋め込み $\Lambda^*(x) = \sup_{\lambda \in \mathbb{R}} \{ \lambda x - \log E[e^{\lambda X}] \}$

ディープラーニング

高次元・小標本の統計理論 

複雑ネットワークの統計解析 $P(k) \propto k^{-\gamma}$

仮説検定の高次漸近理論 $P(\sigma^2) = \Phi \left\{ \sigma \Phi^{-1}(Q(\sigma^2)) \right\}$

単語のベクトル表現 

自然言語処理

リサンプリング $P(k) \propto k^{-\gamma}$

分子進化系統樹と遺伝子発現解析

時系列解析