数理システム論分野　　志望区分：シ－５

教授　下平英寿　　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求

統計学が注目されています。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は重要な役割を果たしています。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。一方で、ウェブやソーシャルメディア、または生命科学や宇宙科学では大量のデータが主導す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。研究室で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（例えば**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、最近では因果推論，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。このような応用研究の経験をふまえて 2018 年度は数理的な研究成果として、ニューラルネットワークと内積によって表現できる関数のクラスを数学的に明らかにして、さらにそれを大幅に拡張する手法（擬ユークリッド空間への埋込）とその理論を提案しました。これを自然言語処理の単語埋め込みに応用した修士 1 年の研究は NLP2019 にて若手奨励賞と最優秀ポスター賞を受賞しました。機械学習の転移学習ではデータの確率分布が学習時から変わる共変量シフト (covariate shift) という問題設定を提起し、これがいまでは深層学習の加速として広く利用される batch normalization につながっています。

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`