

吉田 朋広

東京大学大学院数理科学研究科
教授

先端的確率統計学が開く大規模従属性モデリング

§1. 研究成果の概要

連続時間非線形時系列に対するモデリングと解析を可能にする普遍的な数理科学的方法の発見と体系化を試みている。大規模な従属性のあるデータに対するモデリングとその基礎となる確率統計理論の研究、および解析ツールとして役立つソフトウェア開発を行う。高頻度データに基づく統計的モデリング、確率統計理論と計算機統計学的方法の発見、および統計モデリングを可能にする YUIMA 開発を軸に研究が進んだ。

リミットオーダーブック (LOB) は観測間隔が極めて短く従属性の高いビッグデータを与える。その統計的モデリングの過程で得られる方法は普遍的であり、確率過程の統計学に新たな発展をもたらすと期待している。LOB の非定常高頻度従属性データに対して、オーダーの生起確率の評価が可能になる **ratio model** を提案し、擬似尤度解析の構成によって統計量の性質を導いた。確率過程に対して正当化された情報量規準やスパース推定の方法を用い、オーダーの予測に役立つ共変量の探索を進めた。

確率過程の汎関数の分布の正確な近似が統計推測理論を構築するための基礎となる。

Malliavin 解析による漸近分布理論において漸近展開は新たな課題である。Wiener 汎関数の漸近展開がエルゴード系に対して導出された。これは、フラクショナルブラウン運動で記述される証券価格変動のモデリングに繋がる研究である。

高頻度観測される金融資産価格間のマルチリードラグをウェーブレットの方法によってモデル化した。また、Wiener 汎関数の最大値の近似に関する結果を与え、高頻度データへ応用した。

深層学習の汎化誤差について、統計的自由度を用いたバイアス-バリエンスのトレードオフを明らかにした。スパース推定の正則化法など、学習理論と確率過程の統計学の融合へ向けて研究が進行している。

2017 年度の成果であった MpCN 法の高次元解析を行った。MpCN 法はベイズ統計学に幅広く使える統計計算手法であり、YUIMA にも実装されている。2018 年度は高次元の統計モデル

への応用を念頭に、次元数と計算コストの関係、すなわち次元の呪いの解析を行った。MpCN 法は、裾の軽い場合、重い場合ともに、次元の呪いの影響が、一般的な手法よりオーダーが一つ少なく、計算効率が高いことを示した。同じ性質を持つ統計計算手法は他に知られていない。また、パラメータの自動チューニングを行う頑健なモンテカルロ法の解析や、時系列解析を念頭に行列値のマルコフ連鎖モンテカルロ法の解析を行った。行列値モデリングは、計算が困難であることから応用研究が少なく、マルコフ連鎖モンテカルロ法によってその困難を減少させることを試みた。

観測ノイズ付きの高頻度データに基づく拡散過程モデルのドリフトおよび拡散係数パラメータの適応型推測法を提案して、その数学的正当性を証明した。

Masuda (Stochastic Processes and their Applications 2019) において、非正規型の確率微分方程式 (SDE) モデル族に対する有効かつ統一的に機能する推測方式が導出された。本結果は、多様な局所安定型の駆動レビ過程に適用でき、しかも既存の推定手法の精度を本質的に改善するものである。

これまで対象としていなかった、駆動レビ過程の分布の段階的推測法を考案した。非正規ノイズの数学的構造を適切に捉えた回帰モデリング理論の開発・実装が進んでいる。YUIMA 開発についても、新たな YUIMA クラスとの融合によって SDE モデル構造の表現を豊かにし、モデル評価や最適化をより多機能かつ効率的に行うための戻り値や内部機能の拡充も進めた。

【代表的な原著論文】

1. Kengo Kamatani, "Efficient strategy of Markov chain Monte Carlo method for high-dimensional heavy-tail distribution", Bernoulli, Volume 24, Number 4B, 3711-3750, 2018.
2. Hiroki Masuda, "Non-Gaussian quasi-likelihood estimation of SDE driven by locally stable Lévy process", Stochastic Processes and their Applications, 129:(3), 1013-1059, 2019 March.
3. Ciprian Tudor and Nakahiro Yoshida, "Asymptotic expansion for vector-valued sequences of random variables with focus on Wiener chaos". Stochastic Processes and their Applications, online 2018

§2. 研究実施体制

(1)「東京大学」グループ

- ① 研究代表者: 吉田 朋広 (東京大学大学院数理科学研究科 教授)
- ② 研究項目
 - ・高頻度時系列データへの統計的モデリングとデータ解析
 - ・統計モデリングの基礎となる確率過程の統計学の研究とその応用
 - ・確率過程に対する統計解析およびシミュレーションのためのソフトウェア YUIMA III の開発
 - ・ソーシャル・ネットワーキング・サービスの情報を使ったデータ解析とイベントの予測

(2)「大阪大学」グループ

- ① 主たる共同研究者:鎌谷 研吾 (大阪大学大学院基礎工学研究科 講師)
- ② 研究項目
 - ・確率微分方程式に基づく統計的モデリングと高頻度データ解析
 - ・大規模時系列データ解析の基礎となる計算統計理論とその応用
 - ・大規模時系列データに有用なモンテカルロ法の開発

(3)「九州大学」グループ

- ① 主たる共同研究者:増田 弘毅 (九州大学大学院数理学研究院 教授)
- ② 研究項目
 - ・確率統計理論研究
 - ・ソフトウェア開発