

「現代の数理科学と連携するモデリング手法の構築」
平成26年度採択研究代表者

H28 年度 実績報告書

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

§ 2. 研究実施の概要

本研究では、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。

(1) 高頻度および超高頻度データの統計的モデリング

高頻度金融データの解析で障害になるのが、観測の非同期性、非正則性、マイクロストラクチャーノイズの問題であるが、リミット・オーダー・ブック(LOB)の統計的モデリングが包括的なアプローチを可能にする。統計解析のためには、共変量を自由に取り込める点過程回帰モデルに対して擬似尤度解析(Quasi Likelihood Analysis)を構築することが基本となる。Yoshida (Annals of Institute of Statistical Mathematics 2011)に基づき、点過程回帰モデルに対し、統計的確率場の多項式型大偏差不等式を導き、最尤型推定量、ベイズ型推定量の漸近分布および分布の裾確率の精密評価を与え、擬似尤度解析を構成した。マルコフ型点過程に対してエルゴード性の保証が課題となり、とくに Hawkes 型モデルに対してそれがなされた。Hawkes 型点過程は近年高頻度金融データに対するモデリングで盛んに用いられているが、擬似尤度解析が適用可能になった (Clinet and Yoshida, Stochastic Processes and Their Applications, on-line)。点過程に対する擬似尤度解析を基礎とし、リミット・オーダー、マーケット・オーダー、キャンセルションのタイミングに対する点過程の強度過程を状態空間の関数を用いて表現し、共変量の探索によって、LOB のダイナミクスを説明するモデルを構成した (Muni Toke and Yoshida, Quantitative Finance 2017)。スプレッドとリミット・オーダー・ボリュームの強度過程への影響に関して、我々の提案したモデルは、従来モデルより良くデータを説明できる。

リード・ラグ現象が金融データや SNS データに現れる。モデリングと統計的性質の解析を進めている。

(2) 大規模従属性データ解析を支える確率統計理論の研究

非エルゴード系の高次統計推測論が進展している。マルチンゲール展開 (Yoshida, Stochastic Processes and Their Applications 2013) を用い、Podolskij and Yoshida (Annals of Applied Probability 2016) は p -変動の漸近展開を与えた。さらに、マイクロストラクチャーノイズのある場合に、Podolskij, Veliyev, Yoshida (Stochastic Processes and Their Applications, to appear) はプレアベレージングによる2次変動の推定量の漸近展開を与えた。

モデル選択の方法は、データと統計モデリングをつなぐ鍵である。情報量規準, LASSO やより一般の正則化法の確率論的評価を、従属性モデルを含んだ一般的な形で行った。スパース推定における AIC 型統計量が正当化され、エルゴード的拡散過程からの大規模高頻度データの場合に、ドリフト係数および拡散係数の同時スパース推定が可能になった (Masuda and Shimizu, Mathematical Methods of Statistics, to appear)。ここでも従属構造での汎用性を持つ擬似尤度解析の方法が使われている。さらに、点過程の推定や有限時間高頻度観測下でのボラティリティ推定において正則化法の漸近的性質の研究が進んでいる。

上原と増田は、レビ過程で駆動される確率微分方程式(SDE)の高頻度従属データに基づいた正規型擬似尤度推定について、スケール係数とドリフト係数の段階的推定量を考案し、漸近分布を見出した。

MCMCの研究において、MpCN法のエルゴード性を解析し、従来手法よりも遥かに広いクラスで幾何エルゴード性が得られることを示した (Kamatani, Journal of Applied Probability 2017, to appear).

(3) 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 YUIMA III

R パッケージ YUIMA の開発を行っている。

共変量過程を容易に導入できる点過程回帰モデル(point process regression model)の強度過程をオブジェクト化し、推定およびシミュレーションの関数が適用可能になった。

qmle 関数に Rcpp を導入し、数倍の高速化を実現した。ベイズ手法である adaBayes も Rcpp で高速化し、高速なマルコフ連鎖モンテカルロ法である MpCN 法も導入した。

レビ過程で駆動される確率微分方程式の正規型擬似尤度推定のための qmleLevy 関数を実装した。多変量非正規レビ過程シミュレータ関数ライブラリ、およびレビ SDE のサンプルパスの擬似生成関数を、C 言語を介し高速化した。擬似ベイズ情報量規準 QBIC を YUIMA に実装した。

(4) Twitter データ解析

Google クラウド上にバーチャルマシンとデータベースを開設し、Twitter データの収集を行っている。日本語の県別の Twitter データをもとに、センチメントアナリシスの結果を日本地図に毎日表示している。

【今年度の代表的論文】

1. Podolskij, M., Yoshida, N.: Edgeworth expansion for functionals of continuous diffusion processes, Annals of Applied Probability, Volume 26, Number 6 (2016), 3415-3455.
2. Kimura, A., Yoshida, N.: "Estimation of correlation between latent processes", Jan Kallsen, Antonis Papapantoleon (eds) Advanced Modelling in Mathematical Finance: In Honour of Ernst Eberlein. Springer, 131-146 (2016)