

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。ソーシャル・ネットワーキング・サービスのデータの統計解析も進めている。

(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モデリングの問題に対して確率統計学によるアプローチを試みている。平成27年度は、リミット・オーダー・ブック(LOB)の統計的モデリングにおいて進展があった。金融高頻度データには、観測の非同期性、非正則性、マイクロストラクチャー・ノイズの問題がある。点過程によって板情報をそのままにモデリングすることによって、これらの困難が解消される。

LOB の点過程によるモデリングは研究の新たな流れになりつつあり、統計解析理論の確立を目指している。統計的モデリングの成否は、情報量規準やスパース推定等のモデル選択のための統計学的技巧にかかっており、そのためには、非常に一般的な統計モデルを扱える漸近決定理論の枠組みの構築が必要となる。我々はこの問題に対してエルゴード的統計と非エルゴード的統計の2つのアプローチを行っている。このような一般的なスキームの準備の上で、実際のデータから現象を探ることによってLOBのモデリングが可能になる。LOBの強度過程を状態空間の関数を用いてモデリングし、データより共変量の探索を行っている。提案したモデルがリミット・オーダー・ブックのダイナミクスを説明し、実データに対して従来のモデルより良いフィッティングを与えることを示し、マーケット・シミュレーターの構築を試みた(Muni Toke and Yoshida 2016)。

リード・ラグ問題に関して、林(2015)は国内高速 3 株式市場間の価格形成の先行遅行関係の研究を発展させた。

(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

有限時間高頻度観測によるボラティリティ推定は典型的な非エルゴード統計の問題である。擬似尤度解析は Uchida and Yoshida(SPA 2013)によって与えられている。Yoshida(SPA 2013)のマルチンゲール展開の方法は非エルゴード系の高次統計推測論へのアプローチを可能にした。この方法を基礎に、Podolskij and Yoshida(2016)は p -変動の漸近展開を与えた。マイクロストラクチャー・ノイズのある場合に、それを除去するプレアゲレーシングを用いたボラティリティ推定量に対して漸近展開の研究が進んだ。ボラティリティのパラメータ推定量の誤差分布の漸近展開は推測統計で基本的な問題である。擬似尤度解析とマルチンゲール展開の方法を用いることで、その漸近展開の研究が進んだ。

さらに、リード・ラグの新しいモデリング、ベイジアンテンソル推定量の収束率(Suzuki 2015)、時間内生性とボラティリティ推定(Koike 2016)、ジャンプとノイズがある場合の非正則観測によるセミマルチンゲールの共分散推定(Koike 2016)において成果があった。

確率微分方程式による統計的モデリングと高頻度データ解析のためには、計算機統計学

的課題が数多くある。大阪大学グループは、拡散型確率過程のパラメータ推定の新しい構成法を計算統計の観点から考案した。大規模従属データでは、漸近有効推定量の構成に使う初期推定量の導出すら困難である。その解決のために、従来想定していた初期推定量より低い収束レートの初期推定量から逐次的に精度を上げていき、最後に漸近有効な推定量を導出する multi-step 推定量が有効であることが知られている (Kamatani and Uchida SISP 2015)。Kamatani, Nogita and Uchida(2016)は、multi-step 推定法を非エルゴード的拡散型確率回帰モデルのボラティリティパラメータ推定に適用し、導出した推定量の漸近混合正規性およびモーメントの収束を示した。

大規模時系列データに有用なモンテカルロ法の開発に進展があった。歪んだ尤度関数を持つ複雑な事後分布解析のため、適合的に MCMC 法を微調整する技術について、鎌谷は Beskos, Jasra と解析を行った。さらに、Jasra, Kamatani, Law, Zhou は、複雑な時系列データのフィルタリングのための、逐次モンテカルロ法にマルチレベルモンテカルロ法を組み合わせることで計算負荷を大きく軽減する方法を提案した。

局所漸近二次モデルのベイズ型モデル評価法として、Eguchi and Masuda (MI preprint series 2015) の情報量規準 QBIC を、本年度はとくに、非エルゴード連続過程モデルの解析を行い、QBIC によってエルゴード的拡散過程の場合に、段階的にモデルを一致推定できることを示した。

一般の局所安定レビ過程の局所漸近正規性を示し、とくに駆動レビ過程の増分が微小時間極限でも歪んでいる場合にはトレンドとスケールの推定の漸近直交性が破綻し、収束率が本質的に変わることを明らかにした (Ivanenko, Kulik and Masuda 2015)。

また、駆動レビ過程の変動指数が既知である下で、一次元安定疑似最尤推定量の正規尺度混合分布による近似理論を導出した。

高頻度従属データに基づくレビ過程駆動型モデルの正規型疑似尤度推定について、Masuda and Uehara (2016) の改良としてスケール係数とドリフト係数をも段階的に推定する可能性を見出した。

(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

点過程の強度過程を表現するオブジェクトの作成、Rcpp を用いた MpCN 法の導入およびモンテカルロ法を用いた時系列データの解析手法のコーディング、摂動モデルの汎関数の期待値計算のための高次漸近展開法の整備、レビ過程シミュレータ関数の C 言語化を行った。微小変動の正規近似と棄却法を併用したレビ測度入力形式の開発、レビ駆動型 qmle 関数を介した段階推定法、QBIC その他モデル評価規準、安定疑似尤度推定法などの実装を行っている。

(4) Twitter データ解析

このプロジェクトは、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、クローラーをセットアップし、ツイートの収集が始まった。テスト期間ののち、平成 27 年 8 月から日本語の県別の Twitter データの収集と解析を始めた。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。