

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 システム創成専攻、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 マス・フォア・インダストリ研究所、准教授)
- 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。扱いが困難な金融データに対するモデリングとその基礎となる確率統計の理論研究、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・サービスのデータは金融を含む様々な社会現象の統計解析に有望である。

### (1) 高頻度および超高頻度データの統計的モデリング

マーケット・マイクロストラクチャーの解明は課題の一つである。平成26年度は、リミット・オーダー・ブック(LOB)の統計的モデリング研究の基礎の整備を進めた。LOB およびプライスの点過程によるモデリングは最近の主流になりつつあり、それに対する統計解析理論の確立が急務である。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第一歩となる。

我々はこの問題に対して2つのアプローチを行っている。一つ目は、LOBの状態のエルゴード性を証明し、その上で伝統的な漸近決定論の形式(局所漸近正規, LAN)で統計量の漸近挙動を捉える方法である。我々は強度関数の一般的な条件の下でそれを証明している。2つ目の方法では、Hawkes型回帰も含む点過程回帰モデル(point process regression model)を構成し、有限観測時間においてマーケットのアクティビティが大きくなるという漸近近似によって統計的漸近理論を展開する。この方法では、非定常状況が扱え、共変量を自由に入れることができ、マルコフ性も必要ない。今後のモデリングにおいて様々な因果律が検証可能となると期待される。点過程に対する非エルゴード的疑似尤度解析(Non-ergodic Quasi Likelihood Analysis)の構成を行い、金融データへの応用も含めて、内外の集会で発表した。

超高頻度での強度関数の回帰にはリード・ラグあるいは causality の表現への期待がもとよりある。いっぽう、“高頻度”の時間スケールでのリード・ラグ計測の問題も重要な課題である。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や、微小リード・ラグ推定の基礎となる研究を行った。

### (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

有限時間離散観測に基づく伊藤過程の拡散係数(ボラティリティ)のパラメトリック推定は、典型的な非エルゴード的統計の問題であり、近年疑似尤度解析(QLA)が構成された(Uchida and Yoshida SPA2013)。スポット・ボラティリティの予測の観点から、スポット・ボラティリティ情報量規準 sVIC が導出できる。(論文 4)。非エルゴード性のため、従来の情報量規準は機能せず、数理統計学的にも新しい問題となる。正当化には、疑似尤度解析と Malliavin 解析を用いた高次漸近理論の方法が使われた。

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を実行するために、理論どおりの良い挙動をする推定量(計算法)を構成することは極めて重要である。しかしながら、大規模従属データなど推定困難な状況では、漸近有効な推定量の構成の足がかりになる初期推定量の導出すら困難である。従来想定していた初期推定量より、さらに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を示した(論文 6)。本手法は極めて普遍的なアイデアであり、確率微分方程式モデルに限らず一般の統計モデルに適用可能である。

九州大学グループは、Lévy 過程駆動型高頻度従属データ統計モデルのパラメータ推定に関して、具体的な構造を仮定しない Lévy 過程駆動型確率微分方程式の高頻度データに基づいたモデルの推定を考えた。係数パラメータおよび Lévy 測度の同時パラメータ推定において、係数と Lévy 測度の汎関数を分離型二段階で行う方法を提案し、推定量の漸近分布を導出した。とくに Lévy 測度の汎関数について、状態依存の係数の場合には推定量にバイアス補正が必要となることを明らかにし、その補正統計量を陽に求めた。関連して、**yuima** パッケージにおける Lévy 増分乱数生成関数の拡張、さらに内部関数 **qmle** について Lévy 駆動型への拡張および信頼区間・領域計算ルーチンに関して研究中である。

局所漸近二次モデルの **Schwarz** 型モデル相対評価基準の構成に関して、局所漸近混合正規性を包含する局所漸近二次モデルに対し、疑似尤度に基づいた **Bayes** モデル的観点から統一的に適用可能なモデル記述評価法を導出した。

### (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**

確率微分方程式のモデル選択や、推定における最適化のステップでより実用的な推定量の構成と解析および数値的検証が進んでおり、これらは近々 **YUIMA** へ実装される。今期の具体的な成果としては、ミラノ大学の **Stefano Iacus** 氏を中心に行われた、Lévy ノイズを持つ多次元 **CARMA(p,q)** モデルとシミュレーターの実装である。また、Lévy ノイズで駆動される **COGARCH(p,q)** モデルのシミュレーターのコーディングが進んでいる。