

戦略的創造研究推進事業 CREST
研究領域「現代の数理科学と連携するモデリング手
法の構築」
研究課題「先端的確率統計学が開く大規模従属性
モデリング」

研究終了報告書

研究期間 2014年10月～2022年3月

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)

§ 1 研究実施の概要

(1) 実施概要

確率過程の統計理論とそれを支える統計計算 (MCMC、分布の近似法) を開発、理論的な成果の計算機への実装 (YUIMA) を行い、そしてこれらの技術に基づき従属系のモデリングによる現象解析とイベント予測へアプローチした。さらに、応用の結果から、理論へ新たな課題が提起され、それを解決してゆくというサイクルを実行した。

たとえば、確率微分方程式に対して擬似尤度による推測ができるが、現実には統計量の値を計算するには、問題固有の最適化や積分計算における困難が伴う。擬似尤度解析 (Yoshida 2011 AISM) の理論で利用可能となったベイズ推定量を初期推定量として、推定量の系列からなる階段を上げて漸近有効な推定量を与えるハイブリッドマルチステップ推定量 (Kamatani and Uchida 2015 SISP) の提案によって、数値的に安定で正確な推定を行えるようになった。この推定量の漸近的性質が再び擬似尤度解析によって保証されている。

正確で高速なベイズ推定量の計算は MCMC の深い研究が必要になる。MpCN 法 (鎌谷法) の理論的解析でエルゴード的な意味での有効性 (Kamatani 2017 JAP) や高次元での有効性 (Kamatani 2018 Bernoulli) が示され、YUIMA に実装されている。このアルゴリズムはハイブリッドマルチステップ推定のようなベイズ推定量の計算にも利用できる。

モデル選択は統計的モデリングと同義であり、そのための情報量規準が重要である。我々はエルゴード的/非エルゴード的な従属性を持つ統計モデルに対する規準を必要としているが、伝統的な独立観測の場合に正当化されている情報量規準は一般に従属系においては正しくない。Eguchi and Masuda (2018 Bernoulli) は Schwarz 型の情報量規準 QBIC を従属系の擬似尤度に対して与えた。導出において擬似尤度解析が援用されている。

ビッグデータでは多くの共変量を現象の説明に利用できるいっぽう、どの共変量が本質的であるか見出すのは逆に困難になる。高次元パラメトリックモデルにおいて、パラメータのスパース性 (少数のパラメータのみが非零) を前提にモデル選択を行うのがスパース推定の概念である。LASSO は代表的なものであるがモデル選択の一致性の観点から良いものではない。Bridge や SCAD が知られているが、従属系に対して適用可能かどうかは検証の必要があることは、情報量規準ですでに経験したことである。擬似尤度解析を基礎に一般的な正則化法によるスパース推定が有効であることが数学的に証明された (Masuda and Shimizu 2017 MMS, Kinoshita and Yoshida 2019 arXiv, Suzuki and Yoshida 2020 JJSJS)。従属系に対する情報量規準やスパース推定法は超高頻度金融データの統計モデリングに使われた。

確率微分方程式の推定問題はいろいろな方向に発展した。ジャンプに対してロバストなボラティリティ推定のためのグローバルジャンプフィルター (Inatsugu and Yoshida 2021)、退化拡散過程の適合型推定 (Gloter and Yoshida 2021 EJS)、ハイブリッド型推定 (Kaino, Uchida and Yoshida 2017 BIC, Kaino and Uchida 2018 SISP, 2018 Metrika)、非エルゴード的従属性データのための非正規型擬似尤度の開発 (Masuda 2019 SPA) など多くの結果が得られた。確率偏微分方程式の推定は新しい課題である (Kaino and Uchida 2021 JSPI, 2021 JJSJS)。

超高頻度データであるリミットオーダーブックの統計モデリングを試みた。現象は連続的ではなく量子的になり、拡散過程 + マーケットマイクロストラクチャノイズという従来の価格変動の描像には限界がある。このような状況では点過程によるモデリングが有効になり、本研究ではリミットオーダーブックのモデリングを通じて、その方法を追求した。擬似尤度解析を点過程に対して構成し (Clinet and Yoshida 2017 SPA)、Cox 型モデルやマーク付き ratio model を、従属系に対する情報量規準や、本研究で与えた従属系に対するスパース推定法を適用して共変量を探索し、オーダーの生起の Bid/Ask サイド予測において、従来の Hawkes 過程における予測精度を改善できることを示した (Muni Toke and Yoshida 2017 QF, 2019 QF, 2020 arXiv:2001.08442)。

金融に関するモデリングにおいて、高頻度データに対するグラフィカル LASSO (Koike 2020 Entropy)、S&P500、VIX インデックス、Interexpectile Differences の関係性 (Bellini, Mercuri and Ryoji 2020 QF)、高頻度観測される金融資産価格間のマルチリードラグのウエー

ブレットの方法によるモデル化(Hayashi and Koike 2018 JFM)、先行遅行関係を内包する2証券から成る市場モデル(2次元の連続時間 Gaussian 確率過程)が無裁定性を有するための十分条件(市場の摩擦)の研究(Hayashi and Koike 2019 SPL)等の研究が進んだ。数理モデリングのテーマでは、さらに、エネルギー消費量予測、風速データの研究が行われた。

確率分布の精密近似に関して本質的な進展があった。非エルゴード的なマルチンゲール展開の応用が行われた:パワーバリエーション(Podoskij and Yoshida 2016 AAP)、プレアベレーシング推定量(Podolskij, Veliyev and Yoshida 2017 SPA)、Euler-Maruyama 近似の誤差分布(Podolskij, Veliyev and Yoshida 2020 AAP)。また、エルゴード的な状況でのWiener汎関数の漸近展開(Tudor and Yoshida 2019 SPA)、非エルゴード系でのSkorokhod積分の漸近展開(Nualart and Yoshida 2019 EJP)は極限定理における新しい結果である。

YUIMAは確率過程に対するシミュレーションおよび統計推測のためのRパッケージであり、確率微分方程式モデルの高速シミュレータに加え、擬似尤度関数や種々の統計量を効率よく計算するモジュールなど、関連する統計的計算のための多くの機能を備えた計算プラットフォームで、CRANで公開されている。いずれも数理的に保証された手法に基づいている。本課題ではYUIMA IIIとして、新たな統計理論の創出に応じて、日伊仏のチームが協力し、開発を進めた。YUIMAが従属性データに関係する諸分野でツールとして役立つことを願っている。

(2) 顕著な成果

<優れた基礎研究としての成果>

1. 擬似尤度解析の深化と展開

概要: Le Cam, Hajek 理論、Ibragimov-Hasminskii 理論を継承し、尤度比確率場の裾に関する大偏差評価の困難を一般的に解消したことが擬似尤度解析(Yoshida 2011 AISM)の意義である。端的な例として、非線形従属系に対してベイズ推定量の漸近挙動が明らかになった。本課題において、擬似尤度解析の水平および垂直展開が進んだ。確率微分方程式のパラメータ推定におけるハイブリッドマルチステップ推定法(Kamatani and Uchida 2015 SISP)、Eguchi and Masuda (2018 Bernoulli)の従属系に適用可能なベイズ型情報量規準(QBIC)の正当化、点過程の擬似尤度解析(Clignet and Yoshida 2017 SPA)、グローバルジャンプフィルターによるポラリティパラメータの安定的な推定(Inatsugu and Yoshida 2020 AISM)等に利用され、スパース推定における正則化法の研究においても擬似尤度解析が有用であることもわかってきている(Umezu, Shimizu, Masuda and Ninomiya 2019 AISM, Suzuki and Yoshida 2020 JJSDS)。

2. 非エルゴード的な従属性データのための非正規型擬似尤度の開発

概要: レビ駆動型モデルのトレンド、スケール、および局所安定指数は、非正規、非定常、非エルゴード的な大規模高頻度データ系列の主要な特徴量である。これら全パラメータを推定するための数理基盤を構築した(Masuda 2019 SPA)。とくに、モデルの長期間にわたる安定性の仮定を課すことなく、また外生変数・共変量過程の係数への取り込みも行いつつ、複雑従属性データを統一的に扱うことが可能となった。個体ごとのモデリングの深化のみならず、個体群の複雑動態・相互作用の統計解析など、今後さまざまな方面への発展・展開が期待される。

3. マルコフ連鎖モンテカルロ法の理論研究

概要: 従来手法では困難な、複雑な確率分布のモンテカルロ法を解析した。ハール測度を利用した、MpCN法(mixed preconditioned Crank-Nicolson algorithm)を考案し、高次元での有効性(Kamatani (2018 Bernoulli)、指数エルゴード性(Kamatani 2017 JAP)を示した。さらに、拡張研究も進められている(arXiv:2008.02906, 2005.05584)。確率過程の統計への応用としては、Kamatani and Uchida (2015, SISP), Kamatani, Nogita and Uchida (2016 BIC), Kaino

and Uchida (2018 Metrika), Kaino and Uchida (2018 SISP) がある。Jasra, Kamatani, Law and Zhou (2017 JNA)は効率的なマルチレベル粒子フィルターを提案し、すでに多くの論文で引用されている。

< 科学技術イノベーションに大きく寄与する成果 >

1. 超高頻度データの点過程によるモデリング

概要： データ計測技術とストレージ技術の進歩から、超高頻度データが利用可能になっている。たとえば金融データにおいては、証券価格形成の背後にあるダブルオークションシステムの超高頻度のスナップショットデータからなるリミットオーダーブックがあり、現象は連続的ではなく量子的になり、拡散過程+マーケットマイクロストラクチャノイズという従来の価格変動の描像には限界がある。このような状況では点過程によるモデリングが有効になり、本研究ではリミットオーダーブックのモデリングを通じて、その方法を追求した(Clinet and Yoshida 2017 SPA; Muni Toke and Yoshida 2017 QF, 2019 QF, 2020 arXiv:2001.08442)。

2. ハイブリッド推測法の開発

概要： 確率微分方程式のドリフトとボラティリティパラメータの推定量の収束率が異なることを利用して、収束が速いボラティリティパラメータを推定した後に、ドリフトパラメータを推定する適応型推定法は計算効率の観点から有効である。しかしながら擬似尤度関数の最適化を行う際の初期値の選定は難しい問題である。Kaino and Uchida (2018 SISP)は、エルゴード的拡散過程モデルに対して、縮約データを用いてボラティリティパラメータの初期ベイズ型推定量を導出し、間引きデータを用いてドリフトパラメータの初期ベイズ型推定量を算出して、擬似尤度関数に基づいた適応型推定量を効率よく計算するためのハイブリッド型推定法を提案した。

3. YUIMA プロジェクト

概要： 確率微分方程式の統計推測法、従属系に対するベイズ法や複雑な適合型アルゴリズム、多様な非正規型確率過程のシミュレーションや統計推測法、無限次元確率解析による高速期待値計算法など、我々が開発した最先端の確率統計学の方法は強力であるが、実装は容易ではない。ユーザー側のこの困難を解消し、さらには最先端の方法を包括的に利用する進んだ研究にも役立つように、チームの技術を集結して YUIMA を開発している。YUIMA は確率過程に対するシミュレーションおよび統計推測のための R パッケージであり、確率微分方程式モデルの高速シミュレータに加え、擬似尤度関数や種々の統計量を効率よく計算するモジュールなど、関連する統計的計算のための多くの機能を備えた計算プラットフォームで、CRAN で公開されている。いずれも数理的に保証された手法に基づいており、他に類を見ない。本課題では YUIMA III として、新たな統計理論の創出に応じて、日伊仏のチームが協力し、開発を進めて来た。確率過程の統計学自身が従属性データ解析のためのインフラであるが、YUIMAを通じて社会実装がさらに促進すると期待する。Iacus and Yoshida (2018 Springer) は YUIMA の基本構想と、その当時までに開発されていたモジュールについて解説している。

< 代表的な論文 >

1. Clinet, S., Yoshida, N.: Statistical inference for ergodic point processes and limit order book. Stochastic Processes and their Applications, Volume 127, Issue 6, 1800-1839 (2017).

概要： 長期観測された点過程にたいして擬似尤度解析を構成し、擬似最尤推定量および擬似ベイズ推定量の漸近正規性と積率収束を証明し、Hawkes 過程のエルゴード性に関して議論している。理論はマルコフ/非マルコフモデルおよび Hawkes 過程へ適用でき、超高頻度金融データで記述されるリミットオーダーブックの統計的モデリングへ応用できる。

2. Kaino, Y. and Uchida, M.: Parametric estimation for a parabolic linear SPDE model based

on discrete observations. Journal of Statistical Planning and Inference, Volume 211, 190–220. (2021).

概要：高頻度時空間データを用いた 2 階線形放物型確率偏微分方程式モデルの係数パラメータの推定法を開発した。さらに得られた適応型推定量の漸近正規性を証明し、大規模数値シミュレーションにより適応型推定量の漸近挙動を検証した。非エルゴード的拡散過程やエルゴード的拡散過程の統計的手法が 2 階線形放物型確率偏微分方程式のパラメータ推定に有効に応用されている。

3. Masuda, H.: Non-Gaussian quasi-likelihood estimation of locally stable SDE. Stochastic Processes and their Applications, Volume 129, Issue 3, 1013–1059 (2019)

概要：微小時間における安定分布近似は、広範な非正規レビ過程に対して成り立つ性質である。Masuda (2019) はこの性質を擬似尤度の構成へ応用した。有界期間での高頻度観測に基づく非エルゴードモデル設定において非線形なドリフト係数およびスケール係数をともに一致推定可能とする擬似尤度を構築し、推定量の理論的性質を導出した。ドリフト係数の一致推定は、拡散過程においては成り立たず、本質的に異なる現象を捉えている。駆動ノイズを局所安定レビ過程へ広げ、非正規安定型擬似尤度が有効にはたらく一つの数理的枠組みを与えた。

§ 2 研究実施体制

(1) 研究チームの体制について

① 「東京大学」グループ

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)

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・ネットワーク・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

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 准教授)2014.10-2020.10

内田 雅之(大阪大学大学院基礎工学研究科 教授)2020.11-2022.03

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

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)

研究項目

- ・確率統計理論研究
- ・ソフトウェア開発

(2) 国内外の研究者や産業界等との連携によるネットワーク形成の状況について

Ciprian A Tudor (Université de Lille)、Yuliia Mishura (Taras Shevchenko National University of Kyiv)、Mark Podolskij (Université du Luxembourg)、Veliyev Bezirgen (Aarhus University)

と漸近展開、David Nualart (Kansas University)と極限定理に関して、Arnaud Gloter (Université d'Évry Val d'Essonne)と漸近理論に関して吉田が共同研究を行なっている。また、増田は、レビ駆動モデルの確率解析的側面について、Alexei Kulik (ヴロツワフ工科大学)との共同研究を実施し、CREST 期間中に、Maud Delattre (INRAE: 国立農業・食糧・環境研究所)とレビ駆動型混合効果モデリングに関する意見交換が始まった。中長期的計画として、YUIMA 開発まで見込んだ共同研究を計画している。