

2016 年度理工学術院総合研究所 「アーリーバードプログラム」 申請書

2016 年 4 月 20 日

理工学術院総合研究所長 殿

申請区分 (いずれかに○)	A・ <input checked="" type="checkbox"/> B			
	A：博士後期課程 1 年および 2 年（助成上限：40 万円） B：博士後期課程 3 年および学位取得後 3 年以内のポストドク（助成上限：80 万円）			
フリガナ	リュウ ゲン			
氏名（年齢）・性別	劉 言 <input checked="" type="checkbox"/> (満 27 歳) <input checked="" type="checkbox"/> 男・女			
所属	基幹理工学部・研究科 応用数理学部・専攻（谷口研究室）			
学年・資格	<input type="checkbox"/> 博士後期課程 年 <input checked="" type="checkbox"/> 助手 <input type="checkbox"/> 助教 <input type="checkbox"/> 次席研究員 <input type="checkbox"/> 研究助手 博士号取得年（取得者のみ記入）：2015 年			
連絡先	内線 73-5746	e-mail yan.liu@aoni.waseda.jp		
専門分野	数理統計・時系列・金融工学			
推薦者	<input checked="" type="checkbox"/>			
過年度応募	<input type="checkbox"/> 初めて <input type="checkbox"/> 再応募（過去不採択） <input checked="" type="checkbox"/> 再応募（2013 年度採択）			
略歴 (高校卒業以降)	<div> <div>学歴</div> <div> 2008 年 3 月 土浦日本大学高等学校 卒業 2008 年 4 月 早稲田大学基幹理工学部 入学 2011 年 3 月 早稲田大学基幹理工学部応用数理学部 単位取得退学 2011 年 4 月 早稲田大学数学応用数理専攻修士課程 飛び級入学 2013 年 3 月 早稲田大学数学応用数理専攻修士課程 修了 修士（理学） 2013 年 4 月 早稲田大学数学応用数理専攻博士課程 入学 2015 年 3 月 早稲田大学数学応用数理専攻博士課程 修了 課程博士（理学） </div> </div> <div> <div>職歴</div> <div> 2014 年 4 月 日本学術振興会 特別研究員（DC2）～2015 年 3 月 2015 年 4 月 日本学術振興会 特別研究員（PD）～2016 年 3 月 2016 年 4 月 早稲田大学基幹理工学部応用数理学部 助手（現在に至る） </div> </div>			
学振・科研費・各種助成 等外部資金の 応募・採択状況	研究費の名称等	申請年	研究テーマ	採択・不採択
	学振特別研究員 ・特別研究員奨励費	2014	非有限分散時系列データに対する 頑健な統計量の開発に関する研究	採択
	科研費 ・若手研究（B）	2016	非有限分散時系列モデルとその解 析手法の妥当性に関する統計的理 論の新展開	不採択
	各種助成金 ・アーリーバード研究奨励金	2013	自己基準化法による新裾指数推定 量の性質に関する研究	採択

1. 研究計画

研究課題 (60 文字以内)	欠損値を含む時系列データの新解析手法とその漸近理論の構築	申請者 氏 名	劉 言
-------------------	------------------------------	------------	-----

本研究の①研究の背景、②研究の目的・内容、③研究の特色・独創的な点について記入
(過年度採択者は、過年度の研究成果・研究の進捗状況と今後の展開も明記する)

①研究の背景

ビッグデータの時代といわれる昨今、情報技術の進歩に伴い、大規模・高頻度の金融データが容易に入手できるようになった。このような金融データは、往々にして非正規性を示し、時として非有限分散性を示す。非有限分散データの解析法は、データの分布の裾指数によって決まる。申請者は、過年度の研究では、この裾指数をどのように推測できるかについて自己基準化法という新たな方法を提案した。その成果は、投稿論文 Liu (2013)で発表した。大規模・高頻度金融データについて、大きな変化をもたらすジャンプ過程の存在を仮定したモデルの解析が盛んに行われている。一方で、大規模・高頻度データの中には、欠損値の存在も実証されているが、殆ど着目されていない。図 1 では、欠損値を含む時系列データとその元データを示しているが、欠損値のある左図はジャンプの様相を示している。このように、欠損値を含む時系列データの解析は、大規模・高頻度データ解析の観点から、急務となっている。申請者は、一般的な時系列データに対し、真のモデルとの距離を縮小する新しい頑健な統計手法を過去の研究で提案した。この手法は既に数回に渡り、学会発表している。本研究では、この手法の欠損値データへの応用を考える。また、ニューサウスウェールズ大学の Dunsmuir 教授と連携を取り、本研究を進める予定である。

図 1 欠損値を含む時系列データ(左)と自己回帰モデルから生成される元データ(右)

②研究の目的・内容

欠損値を含む時系列データは、振幅変調モデル(Amplitude modulated model)によってモデリングされる。元の定常時系列を $\{X(t)\}$ とすれば、欠損値を含む時系列データは、 $Y(t) = a(t) X(t)$ で表現される。但し、系列 $\{a(t)\}$ は、観測値が存在する時、 $a(t) = 1$ 、観測値がない時、 $a(t) = 0$ として定義され、バイナリ過程と呼ばれる。欠損値を含む時系列データの統計的解析は、定常過程に対する最も有効である Whittle 尤度による解析で解決される(Dunsmuir and Robinson (1981))とされていた。申請者は、非有限分散時系列データに対する頑健な統計量の研究をきっかけに、頑健統計量の欠損値時系列データへの応用を考え、2016 年 3 月の熊本国際シンポジウムで、研究結果を Dunsmuir 教授の前で発表した。この発表を契機に、Dunsmuir 教授と文通を始めた。Dunsmuir 教授の長年バイナリ過程やそれに纏わる補間問題についての研究成果によると、Whittle 尤度による欠損値データの解析は最適ではないことがわかった。本研究の目的は、以下となる：

- (1) 漸近理論が成り立つ欠損値を含む時系列データの枠組みの構築。バイナリ過程が観測系列と独立で定常または漸近的に定常である時に、漸近理論が成り立つことが知られている。バイナリ過程に対する新たな解析法の下でも統計量が漸近的に正規性を示しているため、新たな枠組みを構築する必要がある。
- (2) 上記の枠組みの下では、Whittle 尤度による統計量は最早、最適な統計的手法ではない。欠損値データを含む時系列データに関する最も有効な手法を構築し、申請者の提案した頑健性を有する統計量を有機的に関連付ける。特に、この二者が密接に関連していることは、申請者の結果からわかっている。

③研究の特色・独創的な点

本研究は、ニューサウスウェールズ大学の Dunsmuir 教授と共同に進める。Dunsmuir 教授は、欠損値を含む時系列データを研究する第一人者であり、Whittle 尤度による解析を統計学の代表的ジャーナル *Journal of the American Statistical Association* で発表した後、長年、欠損値データに深く関わるバイナリ過程の性質について研究してきた。Dunsmuir 教授とのコラボレーションを通して、欠損値を含む時系列データ解析の里程標となる統計的解析手法の開発と最適推測理論の構築を目指す。新しい統計的理論の展開という側面はもとより、応用も広範囲に渡る。ビッグデータといわれる今では、外的要因により欠損値が存在する環境時系列データの解析、薬と患者の具合による因果効果の統計的解析、更に大規模・高頻度で不規則に観測される金融データの解析等、膨大な応用が考えられる。アーリーバードプログラムの異分野交流を通して、本研究を学際的な研究へ展開し、アーリーバードのメンバーたちと新たな共同課題に挑戦する。

Dunsmuir, W. and Robinson, P. M. (1981), Estimation of Time Series Models in the Presence of Missing Data. *Journal of the American Statistical Association*, 76, 560-568.

2. 研究費支出計画

奨励研究費額 (総計) 800 千円	支 出 内 訳									
	設備備品費		消耗品費		旅費交通費		謝金・補助員費		その他(図書等)	
	260 千円		0 千円		450 千円		0 千円		90 千円	
支出内訳	品名	金額	品名	金額	事項	金額	事項	金額	事項	金額
※品名及び事項は主な項目のみ記入してください。	MacBook Pro	260 千円		0 千円	日本数学会(関西大学)	100 千円		0 千円	書籍(statistical Analysis with Missing Data, Applied longitudinal Analysis)	40 千円
	(13 インチ、Core i7 (3.4 GHz)、16GB メモリ、512GB フラッシュストレージ)				日本統計学会(金沢大学)	100 千円				
					University of New South Wales で共同研究	250 千円			論文投稿料	50 千円

※A : 上限 40 万円、B : 上限 80 万円

3. 研究業績

学会発表

Liu, Y. Robust parameter estimation for irregularly observed stationary process. 熊本国際シンポジウム. 2016 年 3 月 4 日.

Liu, Y. Box-Cox transformation for variance stabilization of dependent observations. 日本数学会. 2016 年 3 月 19 日.

Xue, Y., Liu, Y. and Taniguchi, M. Minimax extrapolation error of predictors. 日本数学会. 2016 年 3 月 19 日. (ほか多数)

投稿論文

Liu, Y. (2016). Optimal portfolio of the Government Pension Investment Fund based on the systemic risk evaluated by a new asymmetric copula. *ASTE Special Issue, Waseda University*, Vol. B13, 19-33.

Suto, Y., Liu, Y. and Taniguchi, M. (2016). Asymptotic theory of parameter estimation by a contrast function based on interpolation error. *Statistical Inference for Stochastic Processes*, 19(1), 93-110.

Akashi, F., Liu, Y. and Taniguchi, M. (2015). An empirical likelihood approach for symmetric α -stable processes. *Bernoulli*, 21(4), 2093-2119.

Liu, Y. (2013). Asymptotic moments of symmetric self-normalized sums. *Scientiae Mathematicae Japonicae*, 77(1), 59-67.

(ほか 2 編)

4. 本事業への参加意欲

今回のアーリーバードプログラムに参加する目的は、自分自身の研究の遂行と、異分野の研究者との情報交換や勉強会を通して、研究者としての視野を広げ、研究者同士で新たな研究テーマに挑戦することである。このように考えているのは、過去に本事業に採択され、異分野の若手研究者との交流の大切さを実感できたからである。2013年度のアーリーバードのメンバーに採用された時、まだ博士課程の一学生であった。アーリーバード事業の支援を受けたお蔭で、研究者として最初の論文(Liu (2013))を上梓した。その当時の研究は、Hill 推定量という統計学の中で裾推定を行う上では常識と思われていた統計量のほかに、何か新しい方法で裾を推定し、非有限分散時系列データの解析に役立てないかというシンプルなアイデアであった。しかし、年度末の統計研究セミナーでは、アメリカやドイツから来た著名な研究者たちから好評を受けた。更に、論文を出版した後、アメリカの研究者から共同研究の申し出も受けた。一人前でもない私にとっては、アーリーバードの支援の力で、研究者として最高の船出が出来た。メンバー間の交流では、お互いの研究紹介を前期の定例会で行った。私の発表に対しては、その当時のメンバーである亀崎允啓さんに、分布の裾推定以外に、分布の頂点の厚さについての推定はないかという質問が聞かれた。当時未熟な私にとって、質問は難しかった。実は、時系列モデルの中で、長期記憶過程のスペクトル分布は原点周りで発散する為、それに対する推定量の構築はかなり興味深い問題である。この発想は、去年、時系列の予測問題のスペクトル分布の minimax 解を考えることになり役立った。後期では、「研究推進」と「キャリアパス」のディスカッションに加え、産業技術総合研究所のイノベーションスクール意見交換会と施設見学会を通して、その当時に行っていた他分野の興味深い研究の動向を、身をもって実感でき、また将来研究者になる決心がついた。

本事業に参加し、参加者たちとの交流を通して、本研究課題の遂行に加え、参加者たちとお互いの研究を学際的な研究に発展させていきたい。応用数理学科の一員としては、現実にある現象に数理理論を活かし、またその現象から新たな数学の問題を見出し、新しい数学を作っていく。ここで、数理理論を活かしていくことも、また数学の問題を見出していくことも、学内外の研究者や企業等の異分野との交流が不可欠である。今回の研究では、欠損値を含む時系列データの解析手法を新しく提案する。これまでの統計学では、時系列の全データの観測を想定することが多く、それに対する漸近理論や最適推測理論が展開されてきた。しかし、世間でビッグデータの時代といわれている今では、大規模・高頻度のデータが観測されており、大きな問題点の一つは、欠損値の存在である。これまでの統計的最適理論を用いて、欠損値を含む時系列データの解析を行っていても、最も良い結果が得られるわけではない。また、現場のデータ解析でもしばしば、欠損値の存在を無視したり、都合よく欠損値を解釈したりする。そこで、今こそ欠損値を含む時系列データの解析が新しい数学の枠組みで行われる必要性が生まれた。本研究は、欠損値を含む時系列解析の基礎研究ではあるが、応用面でも、共同研究者の Dunsmuir 教授と、ボストンのケンモア広場の一酸化炭素による汚染データに対し、新しい手法による解析を目指している。本事業の研究助成を受けることとなれば、グローバルで最先端の研究に関わることができる。更に、この時代だからこそ、汚染データに限った応用ではない。医学・金融・環境等、広範な応用が考えられる。異分野の研究者との交流を通して、研究が連鎖的に進んでいくことは容易に想像できる。例えば、里山の放射能の動態や流出データを基に、どのような汚染傾向があるかを分析することが可能である。この解析の為には、どのような環境ロボットやモニタリング技術を用いれば良いか考えることができる。このように、データ科学から、医学・金融・環境へ研究が発展していく。今回の事業参加は二回目となるため、メンバーの中でリーダーとなって主体的に定例会・意見交換会・見学会や成果報告会に参加する。また、本研究の成果をもって、次年度以降の科研費や助成金・外部資金に応募する。