

令和 5 年 5 月 1 日現在

機関番号：12501

研究種目：基盤研究(C)（一般）

研究期間：2019～2022

課題番号：19K11851

研究課題名（和文）セミパラメトリック関数推定に基づく統計解析の新たな展開

研究課題名（英文）New Expansion of the Statistical Analysis Based on Semiparametric Function Estimation

研究代表者

内藤 貫太 (Naito, Kanta)

千葉大学・大学院理学研究院・教授

研究者番号：80304252

交付決定額（研究期間全体）：（直接経費） 3,400,000円

研究成果の概要（和文）：研究期間を通じて、3つのサブテーマを意識して研究を推進してきた。これら3つのサブテーマを俯瞰して振り返ると、本科学研究費による研究は、「方法の構築 理論の確立 応用の展開」がサイクルとして繰り返されていた。実質科学としての統計科学に携わる者として、このサイクルの繰り返しにより、自らの研究スタイルが確立された。各サブテーマで念頭に置いた学術的問いに対しても、全てではないものの、かなりの解答を得た。それらの解答を含む実績として、期間中に論文11編が採択・出版になり、更に3編が投稿中である。科研費の支援のもと、本研究は十分な成果を得たものと判断する。

研究成果の学術的意義や社会的意義

3つのサブテーマ全てにおいて成果を得た。サブテーマ「理論的拡張と深化」では、ダイバージェンスや再生核ヒルベルト空間を用いた推測など、数理的な研究での成果を得た。また、サブテーマ「様々な関数推定法の評価」では、高次元、正則化、リスク最小化アルゴリズムなど様々な関数推定の枠組みにおける手法を理論的に評価できた。更に、「新たな応用の開拓」においては、歪曲度の応用展開について考察すると共に、ロバスト、同時信頼領域と言った観点での推測法、そして損保数理への応用で研究成果を得ている。このような成果は全て、実データ解析手法として利用できるもので、研究成果は「方法提供」という形で社会的意義となっている。

研究成果の概要（英文）：The research has been set forwarded by three subthemes during the supported period. By looking down all three subthemes, this research has been implemented with a cycle of "development of methodology", "development of theory" and "applications of the methods". As a researcher of statistical science as a substantial science, my style of research has been developed well by a repeat of this cycle. The answers have been provided to each academic question in the subtheme by this research. Including these answers, 11 papers have been published or accepted in journals, and 3 papers have been submitted, during the supported period. This records mean that sufficient results have been obtained by the support to the research.

研究分野：数理統計学

キーワード：関数推定 ノンパラメトリック ダイバージェンス カーネル法 サポートベクター回帰 パターン認識 誤差限界 機械学習

科研費による研究は、研究者の自覚と責任において実施するものです。そのため、研究の実施や研究成果の公表等については、国の要請等に基づくものではなく、その研究成果に関する見解や責任は、研究者個人に帰属します。

様式 C - 19、F - 19 - 1、Z - 19 (共通)

1. 研究開始当初の背景

データから未知の“関数”の推定をする。これを関数推定と呼ぶ。統計科学における問題の多くは、広い意味では関数推定に帰着される。教師なしデータでは分布関数(データのしたがう確率分布)の推定が代表的問題であり、教師ありデータの場合は回帰分析、判別分析などが関数推定の枠組みに入る。科研費のサポートを受けたこれまでの研究経歴から、H31年度からの3年間では、3つのサブテーマを掲げ、それらの研究が糾う中で、関数推定の研究を更に太く発展させたいと考えるに至った。サブテーマと、学術的「問い」としては以下のように定めた。

(1) サブテーマ「新たな応用の開拓」では、いわゆる対応のある多変量データの解析に関して新たな方法の導入を試みた。歪曲度と呼ばれる統計量を新たに導入し、この歪曲度に基づく新たな多変量解析手法の構築とその応用領域の開拓に取り組んだ。サブテーマ「新たな応用の開拓」での研究開始当初の学術的「問い」として、『歪曲度を用いて対応のある多次元データが持つどのような情報を、どこまで抽出できるのか?』を定めた。

(2) サブテーマ「様々な関数推定法の評価」においては、学術的「問い」として、『関数推定におけるセミパラメトリック手法で真に有効なものはどれなのか?』を定めた。先行研究も含め、これまで幾つか提案され議論されてきたセミパラメトリック手法のうち、どの手法が有用なのかについては議論がほとんど見当たらなかったため、この問いへの解を得ることを目指した。正規化を含めた、様々な関数推定手法について、その精度評価の確立と、問題点・改良点の発見も目指した。

(3) サブテーマ「理論的拡張と深化」では、教師なしデータで重要な、確率分布に関する仮説検定である適合度検定に焦点を当てた。特に、学術的「問い」として、『実数空間で構築された適合度検定は、関数空間でどのように拡張できるのか?』を設定し、適合度検定の数学的拡張を目指した。近年、画像データなど扱うデータのスタイル自体も多様になっており、データを抽象的な関数空間のランダムエレメントと捉えた方が都合がよく、そうすると関数空間での分布型の検定が重要となる。このような背景から上記の問いが導かれた。

2. 研究の目的

各サブテーマでの明確な研究目的を定めた。

(1) サブテーマ「新たな応用の開拓」における目的として、

高次元データを含め、様々な種類の対応のある多次元データに対して歪曲度を用いた解析手法を開発し、従来の多変量解析手法との比較をしつつ、その有用性に関する知見獲得を目指す。

また、重要な応用として損保数理における関数推定の問題にも取り組み、従来の方法の精密化が可能かどうかの検討を行う。

多次元空間に埋め込まれた1次元曲線の推定とその同時信頼領域の構築を目指す。

スピアマン・ランク行列と呼ばれる、空間ランクに基づく行列から導かれるロバストな主成分分析の性質を、影響関数の導出を通して深く調べる。

(2) サブテーマ「様々な関数推定法の評価」では、

セミパラメトリック密度推定および回帰に関して、既存手法の理論比較を目的とする。

また、高次元パターン認識に有効な判別関数の開発についても取り組むことを目指す。

更に、機械学習のアルゴリズムに基づく回帰関数推定の方法構築も目指す。特に、従来の漸近論的考察と併せて、非漸近論的考察も試みる方針を立てた。

L1 正則化によるノンパラメトリック回帰の問題点を抽出し、問題点を克服する手法の確立と精度評価を目指す。

サポートベクター回帰の平滑化パラメータを、罰則尤度の方法で決定する新しいロバストな回帰分析手法の確立とその理論的精度評価を目指す。

(3) サブテーマ「理論的拡張と深化」においては、

関数空間における適合度検定で有効なものを開発すると同時に、その検定の性質を調べ明らかにすることを目的とした。特に、従来のユークリッド空間における適合度検定が、再生核ヒルベルト空間に拡張可能かどうかを明らかにする。

更に、局所ダイバージェンスに基づく密度推定、回帰推定といった関数推定法の構築とその理論的性質の解明も目的とした。

3. 研究の方法

各サブテーマで実際に取り組まれた研究方法は以下の(1)(2)(3)にまとめられる。

(1) サブテーマ「新たな応用の開拓」では、

歪曲度の推定をノンパラメトリック関数推定手法を用いて行うことで、より柔軟に歪曲度が推定可能となり、適用できる多次元データの対象が格段に増えることを示した。従来の多変量解析手法との比較を通して、歪曲度が抽出する情報の有用性を示した。

また、保険数理における関数推定に関しては、破産確率という関数に着目し、漸近展開を用いることで、破産確率をより精度よく推定できるかどうかを理論面・数値面両方から調べていくことを行った。

ノンパラメトリック核型平滑化を用いて、多次元空間に埋め込まれた1次元曲線の推定を行い、直交ベクトル場を用いて同時信頼領域を構築する方法を確立した。

スピアマン・ランク行列に基づく主成分分析のロバストネスについて、影響関数の導出と評価、感度分析、シミュレーション、実データ適用をとおして調べた。また、スピアマン・ランク行列以外のロバストな共分散指標についても考察を与えた。

(2) サブテーマ「様々な関数推定法の評価」においては、

パラメトリックな手法で関数推定を行い、その残差、推定し残された部分をノンパラメトリックに推定し補完する、パラメトリック手法とノンパラメトリック手法をブレンドしたセミパラメトリック手法について、その漸近的性質を調べた。

また、高次元パターン認識に有効なアプローチとして、ナイーブベイズと呼ばれる線形判別関数の高次元版に着目し、その性質を調べた。

ステージワイズ最小化と呼ばれる機械学習のアルゴリズムに基づく回帰分析手法を開発し、その性質を調べた。手法の精度評価として、非漸近論的評価を確立した。

L1 正則化によるノンパラメトリック回帰の既存手法は、無駄な制約が含まれており、関数推定の精度が損なわれていることを明らかにし、それを克服する手法を提案した。

ラプラス型密度を導入することで、サポートベクター回帰の平滑化パラメータの選択を、罰則付き尤度の観点から実装できることを示した。

(3) サブテーマ「理論的拡張と深化」では、

ノンパラメトリック推測で用いられるV統計量の漸近理論、およびヒルベルト空間における漸近理論を用いることで、関数空間における適合度検定の漸近分布を導出する。画像データ、遺伝子データなど、いわゆる高次元小標本の枠組みでの適合度検定の実装まで行った。

また、ダイバージェンスに基づく統計推測に、核関数を介在させ、それを局所化することの意義を、関数推定の枠組みで調べた。

4. 研究成果

各サブテーマで研究成果を得た。

(1) サブテーマ「新たな応用の開拓」では、

歪曲度のノンパラメトリック推定手法を確立し、その漸近的性質を調べた。対のある多変量データの新しい関連性指標が提案されたことになる。等角写像からの乖離を測るものとなっており、従来の多変量解析手法では抽出できない情報の獲得が可能となり、有用性を示した。当初の学術的「問い」に対しては、「歪曲度は正準相関などの線形な関連性を越えた関係性を示すことがわかった」が解答となる。統計関連学会連合大会での講演に繋がった。

保険数理における資産過程として、古典的なクラメル・ルンドバーグ・モデルにおける破産確率というリスクに焦点を当てた。破産確率に関しては、適当なスケールリングの下での極限分布が知られている。一方で、極限分布で近似し、その近似を改良することを目的とした破産確率の漸近展開については、研究が見られなかった。破産確率の漸近展開を与え、その精度を理論面・数値面の両方から調べ、実際の損保数理の場面においても、リスクヘッジの参考指標として用いることができることを示した。統計関連学会連合大会で講演すると共に、成果をまとめた論文を投稿した。

多次元空間に埋め込まれた1次元曲線のノンパラメトリック核型推定量に基づき、直交ベクトル場を構成して同時信頼領域を構築するアルゴリズムの提案とその精度を調べ、よい精度で推定できることがわかった。学会・研究集会での発表を重ね、成果は論文にまとめ投稿された。

空間サイン、空間ランクといった多変量概念から派生する統計量に基づく主成分分析の性質を調べ、論文1編としてまとめ出版された。スピアマン・ランク行列に基づく主成分分析のロバストネスについて様々な角度から調べた結果を学会・研究集会で発表し、得られた知見をまとめたものを論文としてまとめ投稿した。

(2) サブテーマ「様々な関数推定法の評価」においては、

セミパラメトリック手法の比較を漸近的枠組みで行った結果として、漸近分布はどの方法も同じになることが示された。したがって、漸近的により高次の評価が必要になることがわかった。当初の学術的問いに対しては、「十分な解答が得られなかった」となるが、研究自体は今後も継続される。

高次元におけるパターン認識、すなわち判別分析に有効な手法の1つとして、ナイーブベイズと呼ばれる線形判別関数の高次元版に焦点を当てた。その理論的性質を漸近理論を用いて導出すると共に、シミュレーションを通して数値的にもその良い挙動を調べた。この成果は論文1編として出版された。更に、高次元における期待値近似を用いたスムーズ・ナイーブベイズと呼ば

れる手法を新たに提案し、高次元での理論的性質を調べた。この成果についても論文1編として出版されるに至った。

重回帰分析の枠組みにおいて、ステージワイズ最小化と呼ばれるアルゴリズムに基づく回帰分析手法を開発し、その性質を調べ、様々なデータへ適用し、その有用性を明らかにした。一般化線形回帰や擬似尤度といった手法を一般化して考えることで、より柔軟な手法が開発されたのと、理論的精度評価として、アルゴリズムの性質を駆使して、非漸近論的評価を確立したことが重要な成果と言え、論文1編として出版された。ノンパラメトリック核型密度推定の枠組みで、同様のアルゴリズムを構築し、方法を確立した。成果の論文1編を投稿し、採択された。

既存手法の問題点を克服する手法を提案し、その精度評価を与えると共に、既存手法に対しての優位性を様々な観点から与え、成果を論文1編にまとめ、出版された。

サポートベクター回帰のパラメータ推定と、含まれる平滑化パラメータの選択を同時に実行できる罰則付き尤度を導入し、従来法と同じ精度を持ちつつ、計算時間が大幅に短縮されることを示した。成果をまとめた論文1編が出版された。

(3) サブテーマ「理論的拡張と深化」では、

ユークリッド空間における適合度検定の理論的拡張として、再生核ヒルベルト空間における適合度検定、特に正規性の検定についてその理論的性質を深く調べた。深遠な理論計算に基づき、帰無分布の有効な近似法を考案した。シミュレーションと高次元実データへの適用を通して、その実際の性能を調べた。また、適合度検定の帰無分布が正規分布になるような、ある種の変換統計量について、再生核ヒルベルト空間でも検討し、再生核ヒルベルト空間においても同様の変換による漸近正規性が得られることを証明した。一連の研究は、当初の学術的問いへの解答を与えている、即ち、「実数空間で構築された適合度検定は、再生核ヒルベルト関数に拡張でき、有効に機能する」が解答となる。これらの成果は論文2編として出版された。

また、ダイバージェンスに基づく統計推測に、核関数を介在させ、それを局所化することの意義を、関数推定の枠組みで調べた。局所化が有効に機能することを密度推定と回帰分析の枠組みで示し、論文2編として出版された。

期間を通じて、論文11編が出版・採択となり、3編が投稿中である。科研費の支援により本研究は十分な成果が得られたと判断する。

5. 主な発表論文等

〔雑誌論文〕 計11件（うち査読付論文 11件 / うち国際共著 1件 / うちオープンアクセス 1件）

1. 著者名 Uemoto Takumi, Naito Kanta	4. 巻 174
2. 論文標題 Support vector regression with penalized likelihood	5. 発行年 2022年
3. 雑誌名 Computational Statistics and Data Analysis	6. 最初と最後の頁 107522 ~ 107522
掲載論文のDOI (デジタルオブジェクト識別子) 10.1016/j.csda.2022.107522	査読の有無 有
オープンアクセス オープンアクセスではない、又はオープンアクセスが困難	国際共著 -
1. 著者名 Matsushima Yuki, Naito Kanta	4. 巻 34
2. 論文標題 Improvement on LASSO-type estimator in nonparametric regression	5. 発行年 2022年
3. 雑誌名 Journal of Nonparametric Statistics	6. 最初と最後の頁 964 ~ 986
掲載論文のDOI (デジタルオブジェクト識別子) 10.1080/10485252.2022.2085700	査読の有無 有
オープンアクセス オープンアクセスではない、又はオープンアクセスが困難	国際共著 -
1. 著者名 Inge Koch, Lyron Winderbaum, Kanta Naito	4. 巻 11
2. 論文標題 Principal component analysis of standard and spherical covariances from the population and random samples to real and simulated data	5. 発行年 2022年
3. 雑誌名 American Journal of Theoretical and Applied Statistics	6. 最初と最後の頁 122-139
掲載論文のDOI (デジタルオブジェクト識別子) 10.11648/j.ajtas.20221104.13	査読の有無 有
オープンアクセス オープンアクセスではない、又はオープンアクセスが困難	国際共著 -
1. 著者名 Nishida Kiheiji, Naito Kanta	4. 巻 -
2. 論文標題 Kernel density estimation by stagewise algorithm with a simple dictionary	5. 発行年 2022年
3. 雑誌名 Computational Statistics	6. 最初と最後の頁 -
掲載論文のDOI (デジタルオブジェクト識別子) 10.1007/s00180-022-01303-7	査読の有無 有
オープンアクセス オープンアクセスではない、又はオープンアクセスが困難	国際共著 -

1. 著者名 Kanta Naito and Spiro Penev	4. 巻 15
2. 論文標題 Regression using localised functional Bregman divergence	5. 発行年 2021年
3. 雑誌名 Electronic Journal of Statistics	6. 最初と最後の頁 6544-6585
掲載論文のDOI (デジタルオブジェクト識別子) 10.1214/21- EJS1947	査読の有無 有
オープンアクセス オープンアクセスとしている (また、その予定である)	国際共著 該当する

1. 著者名 Makigusa Natsumi, Naito Kanta	4. 巻 156
2. 論文標題 Asymptotic normality of a consistent estimator of maximum mean discrepancy in Hilbert space	5. 発行年 2020年
3. 雑誌名 Statistics & Probability Letters	6. 最初と最後の頁 108596 ~ 108596
掲載論文のDOI (デジタルオブジェクト識別子) 10.1016/j.spl.2019.108596	査読の有無 有
オープンアクセス オープンアクセスではない、又はオープンアクセスが困難	国際共著 -

1. 著者名 Makigusa Natsumi, Naito Kanta	4. 巻 180
2. 論文標題 Asymptotics and practical aspects of testing normality with kernel methods	5. 発行年 2020年
3. 雑誌名 Journal of Multivariate Analysis	6. 最初と最後の頁 104665 ~ 104665
掲載論文のDOI (デジタルオブジェクト識別子) 10.1016/j.jmva.2020.104665	査読の有無 有
オープンアクセス オープンアクセスではない、又はオープンアクセスが困難	国際共著 -

1. 著者名 Tamatani Mitsuru, Naito Kanta	4. 巻 48
2. 論文標題 High dimensional asymptotics for the naive Hotelling T2 statistic in pattern recognition	5. 発行年 2019年
3. 雑誌名 Communications in Statistics - Theory and Methods	6. 最初と最後の頁 5637 ~ 5656
掲載論文のDOI (デジタルオブジェクト識別子) 10.1080/03610926.2018.1517217	査読の有無 有
オープンアクセス オープンアクセスではない、又はオープンアクセスが困難	国際共著 -

1. 著者名 Takuma Yoshida and Kanta Naito,	4. 巻 134
2. 論文標題 Regression with stagewise minimization on risk function.	5. 発行年 2019年
3. 雑誌名 Computational Statistics and Data Analysis	6. 最初と最後の頁 123-143
掲載論文のDOI (デジタルオブジェクト識別子) 10.1016/j.csda.2018.12.011	査読の有無 有
オープンアクセス オープンアクセスではない、又はオープンアクセスが困難	国際共著 -

1. 著者名 Kenta Kawamura and Kanta Naito	4. 巻 47
2. 論文標題 Asymptotic theory for local estimators based on Bregman divergence.	5. 発行年 2019年
3. 雑誌名 Canadian Journal of Statistics	6. 最初と最後の頁 628-652
掲載論文のDOI (デジタルオブジェクト識別子) 10.1002/cjs.11516	査読の有無 有
オープンアクセス オープンアクセスではない、又はオープンアクセスが困難	国際共著 -

1. 著者名 Inge Koch, Kanta Naito and Hiroaki Tanaka,	4. 巻 61
2. 論文標題 Kernel naive Bayes discrimination for high-dimensional pattern recognition.	5. 発行年 2019年
3. 雑誌名 Australian and New Zealand Journal of Statistics	6. 最初と最後の頁 401-428
掲載論文のDOI (デジタルオブジェクト識別子) 10.1111/anzs.12279	査読の有無 有
オープンアクセス オープンアクセスではない、又はオープンアクセスが困難	国際共著 -

〔学会発表〕 計13件 (うち招待講演 0件 / うち国際学会 0件)

1. 発表者名 山添滉弥、内藤貴太
2. 発表標題 ノンパラメトリック平滑化による埋め込み1次元曲線の信頼領域
3. 学会等名 統計関連学会連合大会
4. 発表年 2022年

1. 発表者名 山添滉弥、内藤貴太
2. 発表標題 多次元空間における埋め込み1次元曲線の同時信頼領域
3. 学会等名 科研費研究集会「統計科学の開拓」
4. 発表年 2022年

1. 発表者名 渡邊宏太、内藤貴太
2. 発表標題 スピアマンランク行列による主成分分析のロバストネス
3. 学会等名 統計関連学会連合大会
4. 発表年 2022年

1. 発表者名 渡邊宏太、内藤貴太
2. 発表標題 スピアマンランク行列による主成分分析のロバストネス
3. 学会等名 科研費研究集会「統計科学の開拓」
4. 発表年 2022年

1. 発表者名 内藤貴太
2. 発表標題 Nonparametric Dilatation Analysis
3. 学会等名 統計関連学会連合大会
4. 発表年 2022年

1. 発表者名 岡山大成, 内藤貴太
2. 発表標題 ガンマクレームによる資産過程における破産確率の漸近展開
3. 学会等名 統計関連学会連合大会
4. 発表年 2021年

1. 発表者名 岡山大成, 内藤貴太
2. 発表標題 ガンマクレームによる資産過程における破産確率の漸近展開
3. 学会等名 科研費シンポジウム「予測モデリングの理論と応用」
4. 発表年 2021年

1. 発表者名 渡邊宏大, 内藤貴太
2. 発表標題 スピアマンランク行列によるロバストな主成分分析
3. 学会等名 科研費シンポジウム「予測モデリングの理論と応用」
4. 発表年 2022年

1. 発表者名 山添滉弥, 内藤貴太
2. 発表標題 埋め込み1次元曲線のカーネル密度推定による信頼領域
3. 学会等名 科研費シンポジウム「予測モデリングの理論と応用」
4. 発表年 2022年

1. 発表者名 Kanta Naito and Spiridon Penev
2. 発表標題 Regression with localized functional Bregman divergence
3. 学会等名 統計関連学会連合大会 於富山大学（オンライン）
4. 発表年 2020年

1. 発表者名 Kanta Naito and Spiridon Penev
2. 発表標題 Regression with localized functional Bregman divergence
3. 学会等名 科研費研究集会「統計科学の革新にむけて」於金沢大学サテライト・プラザ
4. 発表年 2020年

1. 発表者名 Kanta Naito
2. 発表標題 Asymptotics for density estimator with local divergence
3. 学会等名 統計関連学会連合大会
4. 発表年 2019年

1. 発表者名 内藤貴太
2. 発表標題 Semiparametric regression with local divergence
3. 学会等名 科研費シンポジウム「統計的推測および確率解析に関する総合的研究」
4. 発表年 2019年

〔図書〕 計0件

〔産業財産権〕

〔その他〕

-

6. 研究組織

	氏名 (ローマ字氏名) (研究者番号)	所属研究機関・部局・職 (機関番号)	備考
研究 分担者	西田 喜平次 (Nishida Kiheiji) (50631652)	兵庫医療大学・共通教育センター・講師 (34533)	
研究 分担者	玉谷 充 (Tamatani Mitsuru) (80749846)	島根大学・学術研究院理工学系・助教 (15201)	

7. 科研費を使用して開催した国際研究集会

〔国際研究集会〕 計0件

8. 本研究に関連して実施した国際共同研究の実施状況

共同研究相手国	相手方研究機関		
オーストラリア	University of New South Wales	University of Western Australia	