

機関番号：15201

研究種目：基盤研究（C）

研究期間：2008～2010

課題番号：20500257

研究課題名（和文）

局所適合セミパラメトリック推測の新展開

研究課題名（英文）

New developments on local fitting semiparametric inference

研究代表者

内藤 貫太 (NAITO KANTA)

島根大学・総合理工学部・准教授

研究者番号：80304252

研究成果の概要（和文）：局所適合セミパラメトリック推測手法の理論的性質の解明と、生物統計への応用に貢献した。研究計画調書に沿って研究が進められた。局所適合セミパラメトリック平滑化の漸近理論の構築に進捗があった。モデル選択への応用についても成果を得た。更に、生物統計への応用として、線形写像とラディアル写像によるモデリングの元での歪曲度の分布論で新しい成果を得て、それをヒト胎児発生の調和度解析に応用した。

研究成果の概要（英文）：The research has contributed to the works to clarify theoretical properties of the local fitting semiparametric inferences and its application to biostatistics. The research has progressed according to the research plan previously submitted. The results of research include developing the asymptotic theory of semiparametric smoothing and the selection of smoothing parameters in the semiparametric generalized linear mixed model. As important results in the application to biostatistics, some distributional results of dilatation have been derived under linear and radial mapping models, which have been applied to the analysis of harmonized development of human fetuses.

交付決定額

（金額単位：円）

	直接経費	間接経費	合計
2008年度	1,300,000	390,000	1,690,000
2009年度	1,100,000	330,000	1,430,000
2010年度	1,000,000	300,000	1,300,000
年度			
年度			
総計	3,400,000	1,020,000	4,420,000

研究分野：総合領域

科研費の分科・細目：情報学・統計科学

キーワード：機械学習、生物統計、平滑化、

## 1. 研究開始当初の背景

科研費の支援を受け、局所適合平滑化の研究に一定の成果を挙げた後、平滑化に基づく新たな推測論の展開を着想するに至った。局所適合セミパラメトリック推測の新展開として、理論整備、機械学習との融合、高次元小標本での手法開発、そしてライフサイエン

スへの応用を柱として研究を開始した。

## 2. 研究の目的

研究は大きく三つに分類され、その三つ巴の中で進められた。

(1) 研究 A は理論・方法論の研究であり、局所適合セミパラメトリック推測のアイデアの具現を多変量解析手法の中で構築することを目的とした。重回帰におけるセミパラメトリック手法の提案、離散応答を持つ重回帰でのセミパラメトリック手法の提案とその平滑化パラメータ選択手法の提案が目的となった。

(2) 研究 B は融合研究であり、機械学習とセミパラメトリック平滑化の狭間で、特にブースティングによる密度推定法の構築を目的とした。セミパラメトリック一般化線形混合モデルにおける平滑化パラメータの効率的選択手法の提案も目的となった。

(3) 研究 C は応用研究であり、平滑化手法を用い、ヒト胎児発生過程の多次元スタンダード曲線の推定、歪曲度によるヒト胎児発生の調和度解析とそこで必要な推測理論の構築を進めることを目的とした。

### 3. 研究の方法

各年度ごとに研究、研究 B、研究 C にエフォート配分を行い研究を進めた。

(1) 研究 A では、パラメトリック回帰の残差をノンパラメトリック回帰することによって構成されるセミパラメトリック回帰手法の構築、そのための平滑化パラメータの選択手法の提案とその精度評価をおこなう。

(2) 研究 B では、U ダイバージェンスの逐次最適化による密度推定を提案し、非漸近的誤差限界を導出する。セミパラメトリック一般化線形混合モデルでの、計算量を軽減した平滑化パラメータ選択法を構築する。

(3) 研究 C では、多次元スタンダード曲線の発生生物学的に妥当なモデルの構築、歪曲度の等角写像の元での分布論的結果を得る。

### 4. 研究成果

成果を 3 つの研究それぞれで述べる。

(1) 研究 A の成果として、重回帰の枠組みにおいて、バイアス縮小をもつ核型平滑化推定量の平滑化パラメータ選択について議論し、成果を得た。雑誌論文⑥が該当する。この論文で議論された内容は、時系列の平滑化などにも拡張され、現在も論文が出版されている。

(2) 研究 B の成果として、セミパラメトリック一般化線形混合モデルにおける平滑化パラメータ選択の問題で成果を得ている。この

問題への接近としては、いわゆる一般化情報量規準 GIC が利用できる状況であるが、その計算コストが膨大であることが知られている。一方、セミパラメトリック回帰の議論において、罰則付き最小 2 乗法の平滑化パラメータが、誤差分散とランダム効果の分散の比で記述できることも知られている。この分散の比を用いた方法をモデル選択に用いるアイデアの具現がここでの成果であり、実際に計算時間が大幅に短縮され、かつ GIC で得られるモデルとほとんど差異のないモデルが提案手法で得られることが確認された。様々な実データへの適用において、その有用性を確認した。雑誌論文④が該当する。

機械学習との融合の部分では、U ダイバージェンスの逐次最適化による密度推定を議論し、非漸近的誤差限界を導出すると同時に、従来の核型推定量を改良することを数値実験で確認した。この議論は、2 乗距離に基づく方法では既に提案されていたものであるが、一般的なダイバージェンスにおいては議論が全くなかったものであり、極めて単純に言えば、1 つの凸関数と、辞書（既知の密度関数のクラス）さえ準備することで、多様な密度推定量が構成できるようになった。その成果は学会発表①②④の他、幾つかの国内研究集会、海外の大学でのセミナーでの講演で発表され、論文にまとめ投稿中である。

(3) 研究 C の成果として、まずヒト胎児発生過程の多次元スタンダード曲線の構築があげられる。出生後の乳児の成長曲線は母子手帳で見ることができる。一方で、出生前の胎内の胎児の発生については、未だ解明されていないことが多い。これまでの発生生物学での研究では、ヒト胎児標本の実測値データを用い、主に 1 次元でのスタンダードの構築が行われてきたが、発生過程を多次元的に見る必要があり、その場合のスタンダードの必要性から、この研究が存在する。ロジスティック型の関数を用いた非線形多変量回帰の枠組みで多次元スタンダードを推定する方法を提案したのが雑誌論文③である。この多次元スタンダードを用いることにより、ヒト胎児の発育異常を多次元的に早期発見できるようになる。

クラスタリングと多次元尺度法を用いてのヒト胎児発生 of 多次元パターン分析の成果は雑誌論文②にまとめられている。

更に、そもそもヒト胎児の調和的発生がどのようなメカニズムで行われるのかという疑問への接近として、胎児の各臓器部位間に写像の存在を想定し、その写像が等角写像になるように発生が進むというモデルを構築した。そして、そのモデルからの乖離を測るものとして、擬等角写像論で用いられる歪曲度を統計量として導入し、その歪曲度を用い

て発生の調和の度合を測ることを提案した。重要な成果として、線形写像モデル、ラディアル写像モデルにおいて、等角写像の元での歪曲度の分布論的結果を導出した。この結果により、調和の度合を  $p$  値を用いて測ることが可能となり、異なる臓器間の発生の調和の程度の比較が可能となった。ヒト胎児形態計測データに適用することで、発生生物学的にも興味深い、臓器間の調和的発生のが見られた。学会発表の主要なもの③⑤⑥⑦であるが、国内学会、研究集会、海外の大学でのセミナーにおいてもこの話題について講演した。論文としてまとめ、投稿中である。

(4) 研究 A・B・C には直接には含まれないが、深く関連する成果として、雑誌論文①⑤、図書①がある。

#### 5. 主な発表論文等

(研究代表者、研究分担者及び連携研究者には下線)

[雑誌論文] (計 6 件)

①Masaru Kanba and Kanta Naito, Selection of smoothing parameter for one-step sparse estimates with  $L_q$  penalty. To Appear in *Journal of Data Sciences*. (査読有)

②Jun Udagawa, Akira Yasuda, Kanta Naito and Hiroki Otani, Analysis of the harmonized growth pattern of fetal organs by multidimensional scaling and hierarchical clustering. *Congenital Anomalies*, 50 (2010), 175-185. (査読有)

③Kanta Naito, Jun Udagawa and Hiroki Otani, Multidimensional standard curve for the development process of human fetuses. *Statistics in Medicine*, 29 (2010), 2235-2245. (査読有)

④Takuma Yoshida, Masaru Kanba and Kanta Naito, A computationally efficient model selection in the generalized linear mixed model. *Computational Statistics*, 25 (2010), 463-484. (査読有)

⑤Inge Koch and Kanta Naito, Prediction of multivariate responses with a select number of principal components. *Computational Statistics and Data Analysis*, 54 (2010), 1791-1807. (査読有)

⑥Kanta Naito and Masahiro Yoshizaki, Bandwidth selection for a data sharpening estimator in nonparametric regression, *Journal of Multivariate Analysis*, 100 (2009), 1465-1486. (査読有)

[学会発表] (計 7 件)

①内藤貫太, 江口真透, U-ダイバージェンスによる密度推定, 科研費シンポジウム「計算機支援による統計手法, 理論・応用およびその周辺」2010年11月27日, 高知大学.

②Kanta Naito, Density Estimation with Minimization of U-divergence, Autumn Conference of Korean Statistical Society, International Session, Invited Talk, 6<sup>th</sup> November, 2010, Kyonggi University, Korea.

③Kanta Naito, On the Use of Dilatation in Multivariate Analysis, 73<sup>rd</sup> Annual Meeting, Institute of Mathematical Statistics, 13<sup>th</sup> August, 2010, Chalmers, Gothenburg, Sweden.

④内藤貫太, 江口真透, U-ダイバージェンスによる密度推定, 研究集会 第12回ノンパラメトリック統計解析とベイズ統計, 2010年3月29日, 慶応義塾大学.

⑤Kanta Naito, Statistical Analysis for Harmonized development of Systemic Organs in Human Fetuses, 国際シンポジウム「発生現象の解析」, 2009年10月30日, 島根大学医学部.

⑥内藤貫太, 宇田川潤, 大谷浩, Procrustes Analysis と歪曲度—ヒト胎児の調和的発生—統計関連学会連合大会, 2009年9月8日, 同志社大学(京田辺).

⑦Kanta Naito, Procrustes Statistics and Maximal Dilatation, International Symposium on Mathematical Analyses of Developmental Phenomena, 23<sup>rd</sup> November 2008, Shimane University.

[図書] (計 1 件)

①内藤貫太, 「医学統計学の辞典(丹後俊郎・小西貞則編集)」pp. 276-281, 朝倉書店(2010)

[その他]  
ホームページ等

6. 研究組織

(1) 研究代表者

内藤 貫太 (NAITO KANTA)

島根大学・総合理工学部・准教授

研究者番号：80304252

(2) 研究分担者

( )

研究者番号：

(3) 連携研究者

( )

研究者番号：