

研究種目：基盤研究 (B)

研究期間：2006～2009

課題番号：18300092

研究課題名 (和文) 高次元小標本におけるデータ解析の数理統計学的基礎とその応用

研究課題名 (英文) MATHEMATICAL STATISTICS FOR DATA ANALYSIS IN HIGH DIMENSION,  
LOW SAMPLE SIZE CONTEXT AND ITS APPLICATIONS

研究代表者

青嶋 誠 (AOSHIMA MAKOTO)

筑波大学・大学院数理物質科学研究科・教授

研究者番号：90246679

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：多変量解析、機械学習、生体生命情報学、マイクロアレイ、高次元データ

#### 1. 研究計画の概要

マイクロアレイデータ等の高次元小標本データに対する高次元漸近理論を新しく構築し、高次元小標本特有の各種推測問題の解決に応用することを目的として、次の3つのテーマを掲げた。

- (1) 高次元小標本の幾何学的構造の解明と、標本数決定を伴う高次元漸近理論の構築。
- (2) 高次元空間に内在する固有空間の推定。
- (3) パターン認識の諸問題への応用。

#### 2. 研究の進捗状況

(1)については、昨年度までの研究によって、ほぼ解決を見た。高次元小標本データの幾何学的構造は、従来までは、正規性もしくは類する仮定の下でしか解明されておらず、現実のデータの有り様を反映したものになっていなかった。本研究課題では、これを非正規の下で解明することに成功した。高次元小標本の潜在空間はノイズに埋もれ、固有空間に不定性が生じる。そのため、推定の不一致性が問題になる。そこで、本研究課題では、推定の不一致性を解決するために、標本数決定を伴う高次元漸近理論を構築した。これを用いると、高次元小標本の標本数が、次元に関するある比較的緩いオーダー条件を満たすとき、推定は一致性を有するものになる。

(2)については、今年度までの研究で、一つの解決が得られた。高次元データに内在する潜在空間の次元を小標本で推定するには、ノイズの影響を如何に篩にかけるかが問題になる。上記(1)の成果によって、非正規の一般的な状況において、一致性を有する有効次元の推定を構築することに成功した。また、近年、多くの研究者が注目する、高次元小標

本における主成分分析の次元の呪い現象に、理論的な解釈を与えることができた。高次元小標本における主成分分析の不一致性を、非正規の一般的な場合で解決するための理論を構築することに成功した。固有空間の推定と主成分スコアの推定は、理論的に一致性をもち、高精度のパフォーマンスをもつことが実験でも明らかになった。

(3)については、高次元小標本における一般化固有値問題として、特有の射影法の開発と、高次元小標本における情報量規準の再考が必要になる。現在、情報量の概念とその役割について精査を行い、理論とパフォーマンスの両面について研究を続行している。

#### 3. 現在までの達成度

②おおむね順調に進展している。

(理由)本研究課題の一番の核となる高次元漸近理論を概ね完成させることに成功したことが、本研究計画を順調に進展させている理由である。

#### 4. 今後の研究の推進方策

(3)を中心に研究を推進する。高次元小標本におけるパターン認識の諸問題に、特有な射影法を開発して、理論とパフォーマンスで性能を確認する。その際に、AIC、BIC、クロスバリデーション、ブートストラップ版などのモデル選択法や LARS-Lasso 推定による予測モデルの構築を考える。いくつかの情報量等を導入して、それらの相互関係に基づいて推定量の一致性、漸近効率等について考察する。