

氏 名 MD NURUL HAQUE MOLLAH

学位（専攻分野） 博士（統計科学）

学位記番号 総研大甲第 901 号

学位授与の日付 平成 17 年 9 月 30 日

学位授与の要件 複合科学研究科 統計科学専攻
学位規則第 6 条第 1 項該当

学位論文題目 Multivariate Analysis to Explore Latent Structure by
Minimum Beta-Divergence Method

論文審査委員	主 査	助教授	南 美穂子
		教授	池田 思朗
		助教授	藤澤 洋徳
		教授	江口 真透
		教授	村田 昇（早稲田大学）

論文内容の要旨

申請論文は 8 章からなる。第 1 章は全体の構成を説明し、第 2 章では主成分分析、因子分析、射影追跡法など線形変換による古典的な多変量解析について 2 次情報（分散構造）のみを使う手法、高次情報を使う方法に分けて総合的なレポートをまとめている。第 3 章は独立成分分析について、その基本概念、これまでに提案された諸々の手法、これらの手法の問題点などの総合報告をしている。

次の 3 章が Mollah さんの独自の研究をまとめた本論文の主要部分であり、いずれの章においても β ダイバージェンスを用いた手法を提案している。 β ダイバージェンスは分布間の距離を測る尺度であるが、分布から外れたデータ点（外れ値）が存在するときその影響をあまり強く受けにくいという特徴がある。最小 β ダイバージェンス(推定)法はデータの分布とモデル分布との β ダイバージェンスを最小にするパラメータの値を推定値とする方法であり、推定値は外れ値の影響をあまり受けにくい。第 4 章ではこの特徴を活かした頑健な推定方法、第 5、6 章ではこの特徴を発展的に活かし、局所的な構造をその構造に含まれないデータ点の影響を受けずに柔軟に推定する手法を提案している。

第 4 章は、FastICA などの独立成分分析手法で必要とされる前処理のホワイトニング(球状化)に対して改良の提案をしている。従来の方法では外れ値があるとその影響を受けてホワイトニングが適切に行われず復元行列がうまく推定できないという問題があった。本論文では最小 β ダイバージェンス法を用いて平均ベクトルと分散行列を推定しホワイトニングする方法を提案し、その影響関数を求め頑健性を理論的にも示している。最小 β ダイバージェンス法におけるチューニング・パラメータ β の値は 0 に近いほど有効性が高いが、大きいほど頑健性が上がる。与えられたデータに対し、できるだけ有効性を失わずかつ頑健性を持つ適切な β の値を選択するために K 重交互検証法を使って β の値を適応的に求める方法を提案した。数値的な比較実験においてその有用性を確認し、音声データの解析結果においても優れた性能を示した。

第 5 章は、データ空間の潜在的な構造が大域的ではなく強い空間依存性を持っている場合の主成分分析法を提案している。空間内に複数の潜在構造があるモデルとしては Tipping and Bishop (1997, 1999) による mixture PCA モデルとしてのアプローチがあるが、本論文で提案する方法は β ダイバージェンスと空間局所性の感受性を高めるために局所尤度の考えを援用し、逐次的に局所的な主成分構造を捉える。 β ダイバージェンスは外れ値の影響を受けにくいという特徴があるが、本章の問題設定においては、局所的な潜在構造に従わないデータ点は外れ値と同様にその構造の推定に影響を与えないよう低いウェイトが与えられる。逐次的な探索を停止するための基準にはウェイトの合計がある一定の値以上であるデータ点の割合、つまりいずれかの構造に含まれると判断されるデータ点の割合を用いることを提案している。また、各局所構造を推定する際に重要となる局所尤度の中心と位置パラメータの初期値の選択に対しても工夫がされている。コンポーネント数を予め与えずともよく、どの成分にも含まれない外れ値がある場合でも影響を受けずに主成分構造が捉えられるなど、従来の方法にはない利点があり、音声データ、高次元データ、コンポーネント数の多いデータに対する数値実験によってもこれらを示し、提案する方法の有用性を示した。

第 6 章は、第 5 章と同様に潜在構造が局所的な場合における独立成分分析手法を提案し

ている。Te-Won Lee et al. (2000) は mixture ICA モデルとして混合正規モデルに対する EM アルゴリズムと似た方法を提案したが、コンポーネント数の推定などに困難な点が指摘されている。本論文では次のような探索的で柔軟な方法を提案している。最初に、適当な観測ベクトルを 1 つ取り、それを原点として頑健な方法を用いて混合行列を推定する。このとき、その原点回りの構造が強ければ、原点に近いデータだけをもとに学習することになる。次に外れ値として判断された残りのデータの中から、観測ベクトルを 1 つ取って原点とし、同様の解析を行う。この操作を繰り返し、ウェイトによる前章と同様の規準が満たされるまで続ける。これにより、局所的な潜在構造が逐次的に探索される。また、あらかじめコンポーネント数を与える必要もない。頑健な方法としては最小 β ダイバージェンス法を用い、 β の値は前章と同様に K 重交互検証法によって適応的に求めている。数値実験では音声データや高次データを含むいくつかのデータを用いて従来の方法と比較し優れた点があることを示した。

第 7 章では FastICA と同様に逐次的に独立成分を推定する頑健な推定法を含め、今後の研究課題をいくつか上げ、第 8 章で全体のまとめを述べている。

論文の審査結果の要旨

申請論文は 8 章からなる。第 1 章は全体の構成を説明し、第 2 章では主成分分析、因子分析を含む古典的な多変量解析について総合的なレポートをまとめ、第 3 章は独立成分分析についての総合報告をしている。次の 3 章が本論文の主要な部分である。

第 4 章は、FastICA などの独立成分分析手法で必要とされる前処理のホワイトニング(球状化)に対して改良の提案をしている。従来の方法では外れ値があるとホワイトニングが適切に行われず復元行列がうまく推定できないという問題があった。本論文では β ダイバージェンスによって平均ベクトルと分散行列を推定しホワイトニングする方法を提案している。チューニング・パラメータ β の値は 0 に近いほど有効性が高いが、大きいほど頑健性が上がる。与えられたデータに対し、できるだけ有効性を失わずかつ頑健性を持つ適切な β の値を選択するために K 重交互検証法を使って β の値を適応的に求める方法を提案した。数値的な比較実験においてその有用性を確認し、音声データの解析結果においても優れた性能を示した。

第 5 章は、データ空間の潜在的な構造が大域的ではなく強い空間依存性を持っている場合の主成分分析法を提案している。空間内に複数の潜在構造があるモデルとしては Tipping and Bishop (1997, 1999) による mixture PCA モデルとしてのアプローチがあるが、本論文で提案する方法は β ダイバージェンスと空間局所性の感受性を高めるために局所尤度の考えを援用し、逐次的に局所的な主成分構造を捉える。コンポーネント数を予め与えずともよく、どの成分にも含まれない外れ値がある場合でも影響を受けずに主成分構造が捉えられるなど、従来の方法にはない利点があり、数値実験によってこれらを示した。

第 6 章は、第 5 章と同様に潜在構造が局所的な場合における独立成分分析手法を提案している。Te-Won Lee et al. (2000) は mixture ICA モデルとして混合正規モデルに対する EM アルゴリズムと似た方法を提案したが、コンポーネント数の推定などに困難な点が指摘されている。本論文では次のような探索的で柔軟な方法を提案している。最初に、適当な観測ベクトルを 1 つ取り、それを原点として頑健な方法を用いて混合行列を推定する。このとき、その原点回りの構造が強ければ、原点に近いデータだけをもとに学習することになる。次に外れ値として判断された残りのデータの中から、観測ベクトルを 1 つ取って原点とし、同様の解析を行う。この操作を繰り返し、提案された規準が満たされるまで続ける。これにより、局所的な潜在構造が逐次的に探索される。また、あらかじめコンポーネント数を与える必要もない。頑健な方法としては最小 β ダイバージェンス法を用い、 β の値は前章と同様に K 重交互検証法によって適応的に求めている。数値実験では音声データや高次データを含むいくつかのデータを用いて従来の方法と比較し優れた点があることを示した。

第 7 章は今後の研究課題、第 8 章で全体のまとめを述べている。

申請論文は、外れ値の影響を受けにくいという β ダイバージェンスの特徴を発展的に生かし、潜在構造が局所的であって全空間に複数の構造が存在する場合に逐次的に局所的な構造を捉えるという興味深く実用的でもある方法を提案している。よって博士論文審査委員会は Mollah さんの博士申請論文は学位授与に十分値する水準にあると判断した。