氏　　　　名	謝　剛強
学位（専攻分野）	博士（学術）
学位記番号	総研大甲第993号
学位授与の日付	平成１８年9月２９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	*M*-Decomposability and Elliptical Unimodal Densities
論文審査委員	主　査　教授　平野　勝臣 教授　中野　純司 教授　田村　義保 教授　栗木　哲 教授　清水　邦夫（慶應義塾大学）

M-Decomposability and Elliptical Unimodal Densities

In this thesis, we introduced the concept of M-decomposability in probability density functions and demonstrated its potential applications to statistical data analysis. To the best of our knowledge, there is no precedence to this line of research. The contributions in this work are two-fold: theoretical aspects of M-decomposability, and application of M-decomposability to statistical data analysis.

Firstly, we built the theoretical foundations of M-decomposability from ground up, starting with the definition, to the derivation of two theorems pertaining to M-decomposability. The concept of M-decomposability is closely related to modality of probability density functions and therefore should appeal to theoreticians and practitioners alike.

Secondly, we demonstrated the possible usage of the theoretical results derived here to scientific and statistical applications. One practical example is cluster analysis. We devised an algorithm based on M-decomposability to locate modes when a sample drawn from an unknown distribution is given. Our algorithm is non-parametric and only requires weak assumptions of approximate elliptical unimodality on the underlying clusters. We further demonstrated that with our algorithm, the number of clusters, which is assumed unknown, can be directly and automatically estimated. Another immediate application of M-decomposability is density estimation. We devised a strategy to refine kernel bandwidth to improve kernel density estimation.

The main theoretical results of the thesis are presented below. Let f, g, h be probability density functions defined on the d-dimensional space. If there exists α between 0 and 1 such that $f = \alpha g + (1-\alpha) h$, then we say that g and h are mixture components of f. We define M-decomposability as follows:

A density f is said to be M-decomposable if it is possible to decompose f into mixture components g and h such that the square-root of the determinant of the covariance matrix of f is larger than the sum of the square-root of the determinant of the covariance matrix of g, and the square-root of the determinant of the covariance matrix of h. Otherwise, we say that f is M-undecomposable.

We derived the following theorem based on the definition of M-decomposability.

(1) All elliptical unimodal densities with finite second moments are M-undecomposable.

This theorem applies to the class of elliptical unimodal densities, which encompasses a wide range of common pdf's including Gaussian, logistic, Laplace, Student's t (with degrees of freedom greater than 2), Von Mises, and many others. The theorem also serves as theoretical justification for using M-decomposability as criterion for cluster analysis.

Next, we derive a theorem relating M-decomposability with Kullback-Leibler divergence. The theorem provides further justifications for the implementation of M-decomposability.

As applications to the above two theorems, we provided several examples of cluster analysis and density estimation. We demonstrated that the concept of M-decomposability can have potentially far reaching impact on statistical data analysis.

論文の審査結果の要旨

データ解析やモデリングを行う際，複雑な構造をもつ確率密度関数を扱うよりも，より単純な構造をもつ確率密度関数の混合に分けることがよく行われる．本論文では，多峰確率密度関数を単峰密度関数に分解するための新しい概念，M-decomposability を導入し，その理論的性質を調べたものであり，あわせてデータ解析への適用を試みた論文である．タイトルの M は multimodal や mixture から来ている．

提出された論文は全 9 章 60 頁からなり，英語で執筆されている．第 1 章は序章で，M-decomposability の概念を導入する背景と論文の構成が述べられている．第 2 章では 1 次元確率密度関数に対する M-decomposability の定義と例が述べられている．標準偏差 σ_f をもつ確率密度関数 $f(x)$ が与えられたとき，適当な確率密度関数 $g(x)$, $h(x)$ と実数 $\alpha\,(0<\alpha<1)$ を用いて，それを混合分布 $f(x)=\alpha g(x)+(1-\alpha)h(x)$ に分解することができ，$\sigma_f>\sigma_g+\sigma_h$ となるような g, h, α が存在するならば，$f(x)$ は M-decomposable であると定義する．もしどのような g, h, α に対しても $\sigma_f \leq \sigma_g+\sigma_h$ であれば，$f(x)$ は M-undecomposable であると定義する．この章では，すべての一様分布は M-undecomposable であることを証明している．第 3 章では，台の集合が単調に減少する n 個の一様分布の混合分布を考え，単峰確率密度関数は $n\to\infty$ の極限としてあらわされることを証明し，第 4 章では対称単峰確率密度関数は M-undecomposable であることを証明している．

第 5 章では，第 6 章以降で展開される多次元の M-decomposability に関する議論に必要な事項と定理が簡潔にまとめられている．多次元確率密度関数 $f(x)$ の共分散行列を Σ_f, その行列式を $|\Sigma_f|$ とかく．一次元の場合の σ_f を $|\Sigma_f|^{1/2}$ で置き換えることなどにより，第 6 章で，d-次元確率密度関数 $f(x)$ の M-decomposability の定義を与えている．d-次元楕円体内で一定の密度をもつ楕円一様分布を考え，その性質が調べられ，楕円一様確率密度関数は M-undecomposable であることを証明している．第 7 章では，第 3 章と同様に有限個の楕円一様分布の混合分布を考え，多次元楕円単峰確率密度関数が M-undecomposable であることを証明している．

第 8 章では M-decomposability のデータ解析への応用を扱っている．ここでは，$f(x)$ が M-decomposable であれば，$KL[f||\tilde{f}]>KL[f||\alpha\tilde{g}+(1-\alpha)\tilde{h}]$ であることを証明している．ただし，$KL[\cdot||\cdot]$ は Kullback-Leibler divergence を示し，$\tilde{f}$ は平均と共分散行列が f と等しい正規確率密度関数を示す．この結果を用いて，クラスター分析と確率密度関数の推定への適用例を与えている．推定への適用例は 1 次元の場合である．

第 9 章では本研究のまとめと今後の課題が述べられている．

本研究において，分布の M-decomposability という新しい概念を導入し，それに関わるいくつかの重要な結果を証明している．合わせて統計的データ解析への応用を試みている．M-decomposability の概念を全く独自に考案した独創性とその性質を解明した数学的能力はともに評価できる．そしてこの概念が簡明なものであり，また，一様分布のひとつの特徴付けと考えられることは注目に値する．さらに，多次元データのクラスタリングの解析や確率密度関数の推定に利用できることを簡単な例で示すことによって，この概念が統計的データ解析への応用が可能であることを示している．これは理論の実用性を示唆するものと言え，優れた研究結果である．

以上のことから，審査委員会は委員全員一致して，本研究に対し博士の学位を授与するに十分値すると判断した．