博士（工　学）　清　水　信　夫

学位論文題名

データ解析における主要点の特性に関する研究

学位論文内容の要旨

「最適施設配置問題」は、ある地域において最適な施設の配置を見出す問題である。施設の数や種類が急速に増え、配置対象となる地域における状況が複雑化している今日においては、社会のあらゆる分野において、この問題を効率的に解く重要性が高まっている。その重要性に呼応して、最適施設配置問題を解くための方法も数理的手法を軸にいろいろと開発され、「最適配置の理論」として体系付けられるようになってきた。一方、最近では計算機の急速な進歩と普及によって、計算幾何学的手法も広く採り入れられるようになり、精度の高い解を従来よりも短い時間で導出できるようになってはいるが、問題の複雑化に応じた高精度かつ高速な解法の開発や、それを支える理論の構築は現在も重要な課題である。

Flury(1990) により提案された主要点 (Principal Points) は、確率分布における密度関数を k個の領域に分割する際に最適となるような各領域のある種の代表点として定義される。数学的にはクラスター分析における k-means 法と同等のアルゴリズムにより導出されるが、主要点の概念は、確率分布を k個の代表点により表すという観点で、「最適配置の理論」の一般化された基礎理論となり得る。

期待値に関して対称な 1 変量確率分布における k個の主要点は期待値に関して点対称となる (対称性をもつ) ことが予想されるが、この予想は常に正しいわけではない。このテーマに関しては、これまでに数多くの研究が行われてきたが、2 点の場合においては、様々な分布における主要点の値や、期待値に関して非対称な主要点が存在する確率分布の例が示されている。しかし、3 点以上の場合における対称性についてはこれまであまり考察されていなかった。一方、2 変量が互いに独立な正規分布における k個の主要点の配置について、一方の分散が他方よりも大きい場合に k個の主要点 ($k \leq 5$) が一直線上に並ぶ配置となる場合があることが計算機シミュレーションにより示されている。しかし、分布の分散共分散行列と主要点の配置との間の関係や、主要点の数が多い場合の配置については未解明の部分が多く残されていた。

本論文は、主要点の特徴的性質を理論的およびシミュレーションを含む計算幾何学的な立場から論じたものであり、以下の 6 章から構成されている。

第 1 章では、以上に示されているような本論文の背景および主要点のもつ意味について触れ、本研究の目的および構成について述べた。

第 2 章では、k個の主要点に関する従来の研究について詳述した。まず、主要点の定義を示した上で、対称な 1 変量分布における期待値に関して対称な 2 点について行われた理論的な考察について示し、対称な 2 点がそれらの点からの 2 乗距離の期待値 (目的関数) の極小値を与えるときの必要条件や、その条件を満たさない場合において 2 個の主要点が非対称となる例についての研究を示した。また、数値計算により得られている、標準正規分布における k個の主要点 ($3 \leq k \leq 5$) の値についても示した。次に、多変量楕円分布における k個の

主要点について行われた理論的な考察を示した。さらに、2変量が互いに独立な正規分布におけるk個の主要点についても、分布の分散共分散行列が対角行列 $\mathrm{diag}(\sigma^2, 1)$ となる場合に関して、いくつかのσにおける計算機シミュレーションにより得られている値を示し、σの値が大きいときにk個の主要点の配置が一直線上に並ぶ配置となることを示した。また、本研究で使用しているk-means法を援用したk個の主要点の導出アルゴリズムを示し、さらに、対称な1変量分布におけるk個の主要点の対称性に関する既知の十分条件について述べた。その一方で、主要点の概念を応用した解析の例として、村木・大瀧・水田(1996,1998)による天気図の解析を紹介した。

第3章では、対称な1変量確率分布が与えられた場合における3個の主要点について述べた。まず、正規分布における3個の主要点の値について、期待値に関して対称な3点が目的関数の極小値を与えるときの必要条件を理論的に求め、さまざまな分布において対称な3点が必要条件を満たすかどうかを考察した。

第4章では、対称な1変量分布におけるk個の主要点の対称性に関して、Chow(1982)の定理に基づく新しい定理を導出し、Li & Flury(1995)におけるk個の主要点の対称性に関する定理の誤りを示した。また、Trushkin(1982)の定理を密度関数に適用することにより新しい十分条件を示した。さらに、正規分布やロジスティック分布、両側指数分布など、以上の十分条件が成立する分布の例を示し、これまで理論的には解明されていなかったこれらの各種分布においてk個の主要点の対称性が成立することを示した。これにより、k個の主要点の対称性が成立する確率分布族を拡張し、確率分布によって対称性の有無をある程度判断できることが確かめられた。

第5章では、2変量が互いに独立な正規分布におけるk個の主要点の配置について述べた。まず、分布の分散共分散行列が対角行列 $\mathrm{diag}(\sigma^2, 1)$ となる場合について、計算機シミュレーションを用いることにより、k個の主要点の配置がk角形から直線に変わるσの境界値を$k \leq 5$におけるそれぞれの場合において求めた。これにより、Flury(1990)が示した未解決問題の解が得られることを示した。また、2変量標準正規分布$(\sigma = 1)$におけるk個の主要点$(k \leq 12)$の配置に関してもシミュレーションを行った結果、主要点となる配置についてはいずれのkに関しても、$k \geq 5$においては少なくとも1本以上原点を通る線対称軸をもつ凸なl角形(ただし$l < k$)の内側に、原点に関して点対称もしくは少なくとも1本以上原点を通る線対称軸をもつ$(k - l)$個の点の集合が存在するような配置が得られることが示された。さらに、主要点となる配置もしくは局所的最適配置のうち特徴的な幾何的性質をもついくつかの配置については理論的考察も行った。

最後に第6章では、本論文における結論を述べた。また、今後の課題として、対称な1変量確率分布のうちt分布、Johnson's S_u分布、混合正規分布など、k個の主要点の対称性に関する十分条件を満たさない分布の存在について触れ、これらの分布における主要点の対称性の有無に関する条件の導出の必要性について言及した。さらに、最適施設配置問題への拡張や、クラスター分析および主成分分析などの各種多変量解析への応用の可能性についても述べた。

学位論文審査の要旨

主　査	教　授	佐　藤　義　治
副　査	教　授	新　保　　　勝
副　査	教　授	伊　達　　　惇
副　査	教　授	水　田　正　弘

学　位　論　文　題　名

データ解析における主要点の特性に関する研究

近年, 最適施設配置問題は, 利用者の分布や環境の状態を考慮することによって, ますます複雑化しており, 理論的な解析よりもむしろ計算機による近似的な解法が中心である. 一方, この種の問題に関する理論的な考察が, データ解析における主要点 (principal points) と関連して注目されている.

主要点 (Principal Points) とは, ある確率分布に従う領域内に k 個の点を配置したとき, 領域内の点から最も近く配置された点へのユークリッド距離の 2 乗の期待値が最小となる点の配置として定義される. 主要点に関する従来の研究では, 確率分布がたとえ期待値に関して対称な 1 変量分布であるとしても, 主要点は必ずしも対称とはならないことが示されている. このことは, 従来の配置の理論にも大きな影響を与える.

一方, 互いに独立な 2 変量正規分布における k 個の主要点の配置について, 一方の分散が他方よりも大きい場合には k 個の主要点 ($k \leq 5$) が一直線上に並ぶことが示されている. しかし, 分布の分散共分散行列と主要点の配置との関係や, 主要点の数が多い場合の配置については未解明の部分が多く残されていた.

以上のような背景に基づき, 本論文は主要点の特徴的性質を理論およびシミュレーションを含む計算幾何学的な立場から論じたものであり, 以下の 6 章から構成されている.

第 1 章では, 本論文の背景および本研究の目的について述べたものである.

第 2 章では, k 個の主要点に関する従来の研究について述べている. 期待値に関して対称な 1 変量確率分布における 2 個の主要点が必ずしも対称ではない例および対称であるためのいくつかの十分条件を紹介している. また, 数値計算により, 1 変量標準正規分布における k 個の主要点の値や, 互いに独立な 2 変量正規分布における k 個の主要点の配置の変化について Flury(1990) により提示された未解決問題を紹介している. さらに, 主要点の概念を応用した解析の例を示している.

第 3 章では, 期待値に関して対称な 1 変量確率分布において, 主要点の候補となり得る対称な 3 点が目的関数を極小にするときの必要条件および十分条件を理論的に導出している. さらにいくつかの分布に関して本条件を検討し、3 個の主要点の対

称性について議論している.

第 4 章では, 期待値に関して対称な 1 変量分布における k個の主要点の対称性について, Chow(1982) の定理に基づく新しい定理を導出している. さらに, Li & Flury(1995) における定理の誤りを示している. また, Trushkin(1982) の定理を密度関数に適用することにより k個の主要点の対称性に関する新しい十分条件を示している. 以上の十分条件を各種分布に適用することにより, あらゆる kに関して k個の主要点の対称性が成立する確率分布族を拡張し, 多くの場合、確率分布の種類によって対称性の有無を判断できることを示している.

第 5 章では, 計算機シミュレーションを用いて, 互いに独立な 2 変量正規分布における k個の主要点の配置が直線となる分散共分散行列の値の範囲を求め, これにより Flury(1990) によって提示された未解決問題を解決している. また, 2 変量標準正規分布における k個の主要点の配置をシミュレーションにより求めた. さらに特徴的な幾何的性質をもついくつかの配置については理論的考察を行い, シミュレーションの値の正当性を裏付けている.

最後に第 6 章では, 本論文における結論として, 本研究の成果がクラスター分析および最適施設配置問題の基礎的な理論となり得ることを述べている. また, 今後の課題として, 対称な 1 変量確率分布のうち本研究で示した k個の主要点の対称性に関する十分条件を満たさない分布における主要点の対称性の有無に関する条件の導出の必要性について言及している.

これを要するに, 著者は, 確率分布における主要点の対称性に関するいくつかの有用な定理を導出するとともに, 主要点の配置パターンの特性を解明し、データ解析に関する多くの実用的な新知見を得ており, 情報解析学, 情報処理工学の発展に寄与するところ大なるものがある.

よって著者は, 北海道大学博士 (工学) の学位を授与される資格あるものと認める.