様式Ｃ－１９

科学研究費助成事業（科学研究費補助金）研究成果報告書

平成２４年　６月９　日現在

機関番号：１２６０１
研究種目：基盤研究（C）
研究期間：2009～2011
課題番号：21500272
研究課題名（和文）　ユークリッド距離行列の順序構造とその多次元尺度構成法への応用
研究課題名（英文）　Ordering structure of Euclidean distance matrices with applications to statistical multidimensional scaling

研究代表者
倉田　博史　（KURATA　HIROSHI）
東京大学・大学院総合文化研究科・准教授
研究者番号：50284237

研究成果の概要（和文）：研究結果の概要は以下の４点にまとめられる。
（1）multi-spherical なユークリッド距離行列の特徴付けに関して新たな数学的事実を導出した。
（2）非負値定符号行列の固有値の間にマジョライゼーションの意味での順序関係が成り立つとき、緩やかな条件の下で、対応する Euclid 距離行列の固有値の間にもモジョライゼーションの意味での順序関係が成り立つことを示した。
（3）セル行列という応用上しばしば用いられる特殊なユークリッド距離行列の作る集合の性質について調べた。また、セル行列の固有値に関するマジョライゼーションの意味での順序関係について数学的事実を導出した。
（4）多次元分布のプリンシパルポイントの存在範囲について幾つかの数学的事実を導出した。特に分布が幾つかの球面対称分布の位置混合分布で与えられている場合を考察し、プリンシパルポイントが存在する部分空間を特定化するための定理を導いた。

研究成果の概要（英文）：The results of this research are summarized as follows:
(1) Some new results on the characterization of multi-spherical Euclidean distance matrices were derived.
(2) The authors showed that if the eigenvalues of a centered positive semidefinite matrix majorizes those of another centered positive semidefinite matrix, then the same inequality holds for the eigenvalues of the corresponding two Euclidean distance matrices.
(3) The authors investigated some properties of cell matrices, which are special Euclidean distance matrices.
(4) The authors derived some new results on the linear subspace in which a set of principal points of a multivariate mixture distribution.

交付決定額

（金額単位：円）

	直接経費	間接経費	合　計
2009 年度	1,400,000	420,000	1,820,000
2010 年度	1,100,000	330,000	1,430.000
2011 年度	900,000	270,000	1,170,000
年度			
年度			
総　計	3,400,000	1,200,000	3,600,000

研究分野：情報学
科研費の分科・細目：統計科学
キーワード：マジョライゼーション、ユークリッド距離行列、多次元尺度構成法、置換行列群。

１．研究開始当初の背景

多次元尺度構成法は、データの視覚的表現に優れている一方で、点配置の解釈が分析者の主観に強く依存するという短所がある。この問題は、例えば点配置の経年変化を調べる場合などのように、多数の点配置を比較しなければならない状況ではより大きなものとなりうる。ところが、研究開始当初の既存研究においては、ユークリッド距離行列を比較するための基礎研究が十分とは言えない状況であった。

２．研究の目的

本研究の目的は、Euclid 距離行列の数理的構造、特に順序構造を明らかにし、その多次元尺度構成法への応用において従来よりも正確かつ客観的な手法を提供することでる。

倉田は、Kurata and Sakuma(2007)において、「群によって誘導される順序(group-induced ordering 、 group majorization ordering)」なる概念を用いて Euclid 距離行列の順序構造について調べた。本研究は、これらの理論をより発展させ、多次元尺度構成法をより精緻化することを目的とする。

より具体的な研究目的（研究目標は）は、研究の方法とも関連するため次項で述べる。

３．研究の方法

研究の方法は 4 つのステップからなり、それらは次の通りである。

(1)Euclid 距離行列の空間の順序構造をより明確にすること、

(2)対応する点配置の空間(それは非負値定符号行列の空間の部分錘になる)の順序構造をより明確にすること、

(3)上記の順序を効率的に表現する不変量や共変量を見つけること、

(4)上記の多次元尺度構成法への応用を扱い、順序構造に関する理解をより深めること。

以上４通りのアプローチによって研究する。

４．研究成果

まず前項の（１）と（２）についての件研究成果として、multi-spherical 構造の特徴付けについて考察し、multi-spherical ユークリッド距離行列全体の陽的表現を導出した、というものが挙げられる。その結果は、Kurata and Tarazaga (2010)として”Linear Algebra and its Applications”誌に採択された。同時に、EDM の零空間の構造に関する幾つかの定理も得た。より詳しく述べれば、次の通りである。ユークリッド距離行列は、spherical なユークリッド距離行列と non-spherical なものに分けられる。Spherical なユークリッド距離行列とは、対応する点配置が適当なユークリッド空間内の超球面上に存在するようなものである。それ以外のものを non-spherical と言う。Spherical なユークリッド距離行列は特殊なものであるかのような印象を与えるが、実はユークリッド距離行列の集合を対称行列のなす線形空間の部分集合とみたときの内点は全て spherical となる。一般にユークリッド距離行列は非負値定符号行列と１対１に対応する。特に、Spherical なユークリッド距離行列は正値定符号行列と対応する。本研究では、non-spherical なユークリッド距離行列の中で spherical に近い構造を持つものとして、multi-spherical なユークリッド距離行列に注目した。ユークリッド距離行列が multi-spherical であるとは、対応する点配置が複数の超球面上に存在するようなものである。本研究では、multi-spherical なユークリッド距離行列を、対応する非負値定符号行列の構造によって特徴付ける定理を導いた。

また、（３）の問題と関連して、ユークリッド距離行列の固有値が、対応する非負値定符号行列の固有値とどのような関係を持つかについての研究を行った。研究結果は論文"Majorization for the eigenvalues of Euclidean distance matrices" (Kurata and Pablo Tarazaga)として Linear Algebra and Its Applications 誌に掲載された。この研究の成果は次の通りである。即ち、２つの非負値定符号行列を考え、それらの固有値の間にマジョライゼーションの意味での順序関係が成り立つとする。このとき、対応するユークリッド 距離行列の固有値の間にもモジョライゼーションの意味での順序関係が成り立つか、という問題を考察した。ここで、２つの非負値定符号行列の固有値の間にマジョライゼーションの意味での順序関係が成り立つということは、2 種類の点配置の散らばりを比較していることに対応する。そして一方の点配置が他方のそれよりも散らばりが大であることを仮定していることに相当する。このような関係が成り立っているとき、ユークリッド距離行列の固有値の間にも同様の関係が成り立っているか否かは自明ではない。つまり、より散らばった点配置に対応するようなユークリッド距離行列の方が、そうでないものよりも固有値の散らばりが大きいかどうかは明らかではない。本研究では、そのような関係が成り立つための十分条件を導いた。

この他、セル行列(cell matrix) の構造を持つユークリッド距離行列について考察し、その固有値の順序構造などについて明らかにした。nXn セル行列とは、ｎ次元の非負ベクトルが存在して、個体ｉとｊの距離がその

非負ベクトルの第ｉ成分と第ｊ成分との和で表されるようなユークリッド距離行列を指す。これは spherical なユークリッド巨利行列となる。本研究では、2 つの非負ベクトルの間にマジョライゼーションの関係が存在するとき、対応するセル行列のの固有値の間にもマジョライゼーションの関係が存在することを示した。また、所与のユークリッド距離行列に（フロベニウスノルムの意味で）最も近いセル行列を求める問題についても新しい結果が得られ、最も近いセル行の陽的表現を与える定理を導いあ。結果は、Pablo Tarazaga 教授との共著として公表準備中である。

多次元尺度構成法は、考察対象の母集団に複数の群が混在するような状況、即ち母集団が幾つかの多次元分布の混合分布である場合に用いられることが多い。混合分布を要約するのに有用な概念として、プリンシパル・ポイントがある。プリンシパル・ポイントとは、確率分布を（ある種の平均２乗距離の意味で）最も良く近似する点の集合のことであり、その点の個数ｎに応じて n-プリンシパル・ポイントと呼ばれる。因みに、1-プリンシパル・ポイントはその分布の平均であり、これは分布に依らない。多次元分布のプリンシパル・ポイントの理論的導出は難しく、2-プリンシパル・ポイントであっても、殆どの分布でその陽的表現は得られていない。そのため、プリンシパル・ポイントが存在する範囲を求める問題、具体的には、それが存在する部分空間を特定しようとする研究が 90 年代から行われている。本研究では、関連したトピックとして、多次元混合分布のプリンシパル・ポイントの存在範囲についての結果も得ている。具体的には、分布が球面対称分布の位置混合分布で与えられているときに 2ープリンシパル・ポイントや n-プリンシパル・ポイントの存在範囲を求めた Kurata and Qiu (2011), Matsuura and Kurata (2011, 2012 等である。

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 10 件）

Hiroshi KURATA, A theorem on the covariance matrix of a generalized least squares estimator under an elliptically symmetric error, Statistical Papers, 51 (2010), no. 2, 389-395.

Hiroshi KURATA and Pablo TARAZAGA, Multispherical Euclidean distance matrices, Linear Algebra and Its Applications, 433 (2010), no.1, 534-546.

黒田佑次郎・岩瀬哲・岩満優美・山本大悟・梅田恵・川口崇・坂田尚子・倉田博史・佐倉統・南雲吉則・中川恵一，「乳癌患者の更年期症状が QOL に与える影響について」， 総合病院精神医学，22 (2010), no1, 27-34.］

Kazumasa MORI and Hiroshi KURATA, The MSE of an adaptive ridge estimator in a linear regression model with spherically symmetric error, Scientiae Mathematicae Japonicae, 72 (2010), no.1, 1-9.

Shun MATSUURA and Hiroshi KURATA, A principal subspace theorem for 2-principal points of a general location mixture of spherically symmetric distributions, Statistics and Probability Letters, 80 (2010), 1863-1869.

Shun MATSUURA and Hiroshi KURATA, Principal points of a multivariate mixture distribution, Journal of Multivariate Analysis, 102 (2011), 213-224.

Hiroshi KURATA and Dingxi QIU, Linear subspace spanned by principal points of a mixture of spherically symmetric distributions, Communications in Statistics-Theory and Methods, 40 (2011) issue 15, 2737-2750.

Shun MATSUURA and Hiroshi KURATA, Definition and properties of m-dimensional n-principal points, Communications in Statistics-Theory and Methods (to appear).

Hiroshi KURATA and Pablo TARAZAGA, Majorization for the eigenvalues of Euclidean distance matrices, Linear Algebra and Its Applications, 436, (2012) 1473-1481.

松浦峻・倉田博史， 異なる球面対称分布の位置混合分布の principal points の性質について,京都大学数理解析研究所講究録 1758, (2011) 60-80.

〔学会発表〕（計３件）
統計関連連合大会 2009 年、2010 年、2011 年

〔図書〕（計１件）
倉田博史・星野崇宏『入門統計解析』新世社 2009 年

６．研究組織
(1)研究代表者
倉田博史 (KURATA Hiroshi)
東京大学・大学院総合文化研究科・准教授
研究者番号：50284237

(2)研究分担者
( )

研究者番号：

(3)連携研究者
( )

研究者番号：