

## 科学研究費助成事業 研究成果報告書

平成 27 年 5 月 27 日現在

機関番号：32612

研究種目：若手研究(B)

研究期間：2011～2014

課題番号：23700341

研究課題名(和文) 多次元確率分布の主要点に関する研究

研究課題名(英文) A study on principal points of multivariate distributions

研究代表者

松浦 峻 (MATSUURA, Shun)

慶應義塾大学・理工学部・助教

研究者番号：70583368

交付決定額(研究期間全体)：(直接経費) 1,900,000円

研究成果の概要(和文)：確率分布のk個の主要点(k-principal points)とは、その確率分布に従う確率変数(ベクトル)と最も近い点との距離の2乗の期待値を最小にするk個の点の配置のことである。本研究では多次元確率分布のprincipal pointsの性質の解明と標本データからの推定問題に取り組み、主に以下の成果を得た。(1)主部分空間定理と呼ばれる、principal pointsが存在する範囲と主成分ベクトルが張る線形部分空間の関連性を示す定理の成立条件を従来よりも広い確率分布群に拡張した。(2)ある損失関数の下で最適性が保証されるなど、新たなprincipal pointsの推定量をいくつか導いた。

研究成果の概要(英文)：A set of k-principal points of a distribution is a set of k points that minimizes the mean squared distance between the random variable (vector) and the nearest point. In this research, the properties and estimation of principal points of multivariate distributions were studied. The main results are the following: (1) Conditions previously obtained for a principal subspace theorem, which ensures that principal points are in a linear subspace spanned by the first several principal components, were extended to more general multivariate distributions. (2) Several new estimators of principal points, some of which are proven to be optimal under some loss function, were derived.

研究分野：統計科学

キーワード：主要点, Principal points, 楕円対称分布, 球対称分布, 多次元混合分布, 主部分空間定理, 主成分ベクトル

## 1. 研究開始当初の背景

確率分布の  $k$  個の主要点 ( $k$ -principal points) とは、確率分布の次元数を  $p$  とおくと、 $p$  次元空間上の  $k$  個の点 ( $k$  個の  $p$  次元ベクトル) の配置を考えたとき、その  $p$  次元確率分布に従う確率変数ベクトルと最も近い点との距離の 2 乗の期待値 (これを平均 2 乗距離と呼ぶ) を最小にする  $k$  個の点の配置のことである。これは、確率分布の、 $k$  個の点への縮約の一種と考えることができる。また、1-principal point はその確率分布の平均値になり、 $k$ -principal points は確率分布の平均値の概念を  $k$  個の点に一般化したものという見方もできる。principal points は、定義はシンプルに与えられるが、その性質について、特に多次元確率分布の場合は未解明の部分が多い。

(1) その状況の中で、従来得られていた重要な結果の一つが Tarpey, Li, and Flury (1995) によって示された楕円対称分布の principal points における主部分空間定理 (principal subspace theorem) である。これは、楕円対称分布の principal points が、分散共分散行列の大きいほうの固有値に対応する固有ベクトル (主成分ベクトル) が張る (線形部分) 空間上に存在することを保証するものであり、principal points を実際に求める上で、また、principal points の性質を理解する上で重要な定理と位置づけられている。近年、Matsuura and Kurata (2011) によって、球対称分布の位置混合分布の principal points における主部分空間定理が導かれた。以上より、研究開始当初の段階では、主部分空間定理が示されているのは楕円対称分布の principal points および球対称分布の位置混合分布の principal points のみであった。

(2) また、現実には、想定している確率分布に未知なパラメータ (未知母数) が含まれていることが多く、その場合、その確率分布からの標本データを用いて principal points を推定する必要がある。この principal points の推定問題についても、特に多次元確率分布の場合の議論はほとんどなされておらず、Flury (1993)、Tarpey (2007) において多次元正規分布および多次元混合正規分布の下で、ノンパラメトリック推定量およびセミパラメトリック推定量と比較しての最尤推定量の良さが報告されているのみであった。

## 2. 研究の目的

上記の背景を踏まえて、本研究の主目的は以下のようにまとめることができる。

(1) 主部分空間定理が成立する分布の条件の拡張を中心に、多次元確率分布の principal points の性質の解明に取り組む。

(2) 未知母数を含む確率分布からの無作為標本による principal points の推定問題に取り組む。

以上の(1),(2)、さらに、それらの応用等に関連する研究に取り組み、多次元確率分布の principal points に対する総合的理解を深めることを目的とする。

## 3. 研究の方法

上記目的に対する研究の方法は以下の通りである。

(1) 多次元確率分布の principal points における主部分空間定理が成立する条件の拡張に取り組む。主部分空間定理が成立しない条件も合わせて調べる。また、principal points が一意に定まるときは、principal points を実際に求めることが非常に容易になることから、principal points の一意性の成立条件を調べることに取り組む。

(2) 未知母数を含む確率分布からの無作為標本が得られているときの principal points の推定問題について、何らかの損失関数の下での最適性が保証されるような推定量を導出することに取り組む。また、これまで推定問題が議論されていない多次元確率分布を取り扱い、その principal points の推定量を構築する。

## 4. 研究成果

(1) まず、 $p$  次元確率分布の  $k$ -principal points は、特に  $k$  の値が大きいときは、 $p$  次元空間全体上に存在するため、この場合は従来の主部分空間定理を適用することはできない。これを踏まえて、 $k$  個の点が存在する空間の次元数に制約を加えたうえで ( $m$  以下とする)、平均 2 乗距離を最小化する  $k$  個の点の配置を考えたとき、球対称分布の位置混合分布および楕円対称分布において最適な  $k$  個の点が第 1 ~  $m$  主成分ベクトルによって張られる線形部分空間上に存在することを示した。この結果は、Tarpey, Li, and Flury (1995) および Matsuura and Kurata (2011) で得られた主部分空間定理を含んで拡張したものになっている。

また、異なる楕円対称分布の混合分布の 2-principal points および一直線上に存在するときの  $k$ -principal points について、いくつかの条件の下、主部分空間定理が成立することを示し、さらに principal points の一意性の十分条件を導いた。

さらに、二次元正規分布の位置混合分布の 2-principal points について、第 1 主成分ベクトルとの位置関係を調べ、主部分空間定理が成立する (すなわち、2-principal points が第 1 主成分ベクトルが張る空間上に存在する) ための十分条件、および主部分空間定理

が成立しないための十分条件を示した。

(2) principal points の推定問題について、まず、確率分布に正規性を仮定し、推定された principal points による平均 2 乗距離の値の期待値を最小にする principal points の推定量の導出に取り組み、一次元正規分布および多次元正規分布において、(多次元正規分布の場合は principal points が一直線上に存在するという条件の下) 最適な推定量が t 分布の principal points を基にして得られることを示した。また、この推定量を用いることで従来の最尤推定量よりどの程度推定精度が改善されるかを調べた。さらに、これらの結果の location-scale family の分布および楕円対称分布への拡張を行った。

また、多次元確率分布の中でも、特に各値が 0 または 1 の 2 値のみを取る多変量 2 値分布の場合の principal points の推定問題に取り組み、多変量 2 値分布の記述に各変数の主効果および複数の変数間の交互作用効果で表される対数線形モデルを用いて principal points の最尤推定を行う方法を提案し、ノンパラメトリック推定量と精度を比較した。さらに、何次までの交互作用効果を取り込んだ対数線形モデルを用いるかについて、AIC, BIC, 尤度比検定を基に選択した場合の principal points の推定精度を比較した。また、変数の数が多いときの principal points の推定精度についても調べた。

さらに、多次元混合正規分布の principal points の推定について、混合数が既知という条件の下、無情報事前分布を用いて principal points のベイズ推定を行うときの推定精度を調べ、最尤推定量との比較を行った。また、混合数が未知の場合には、情報量規準と無情報事前分布を用いて principal points の推定を行う方法を提案し、その推定精度を調べた。

その他、多次元確率分布の principal points の適用・応用および他分野（主に統計的品質管理分野）との関連性を調べる研究に取り組んだ。

#### <引用文献>

Tarpey, T., Li, L., Flury, B. (1995), "Principal Points and Self-consistent Points of Elliptical Distributions", *Annals of Statistics*, Vol. 23, No. 1, pp. 103-112.

Matsuura, S., Kurata, H. (2011), "Principal Points of a Multivariate Mixture Distribution", *Journal of Multivariate Analysis*, Vol. 102, No. 2, pp. 213-224.

Flury, B. (1993), "Estimation of Principal Points", *Journal of the Royal Statistical Society. Series C (Applied Statistics)*, Vol. 42, No. 1, pp. 139-151.

Tarpey, T. (2007), "A Parametric k-means Algorithm", *Computational Statistics*, Vol. 22, No. 1, pp. 71-89.

#### 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6件)

Matsuura, S., Kurata, H., Tarpey, T., "Optimal Estimators of Principal Points for Minimizing Expected Mean Squared Distance", *Journal of Statistical Planning and Inference* (in press), 査読有.

Yamashita, H., Matsuura, S., Suzuki, H., "Estimation of Principal Points for a Multivariate Binary Distribution Using a Log-linear Model", *Communications in Statistics - Simulation and Computation* (in press), 査読有.

DOI: 10.1080/03610918.2014.992541

Yamashita, H., Matsuura, S., Suzuki, H. (2015), "Comparison of Model Selection Methods for the Estimation of Principal Points for a Multivariate Binary Distribution", *Total Quality Science*, Vol. 1, No. 1, pp. 22-31, 査読有.

DOI: 10.17929/tqs.1.22

Matsuura, S. (2014), "Effectiveness of a Random Compound Noise Strategy for Robust Parameter Design", *Journal of Applied Statistics*, Vol. 41, No. 9, pp. 1903-1918, 査読有.

DOI: 10.1080/02664763.2014.898130

Matsuura, S., Kurata, H. (2014), "Principal Points for an Allometric Extension Model", *Statistical Papers*, Vol. 55, No. 3, pp. 853-870, 査読有.

DOI: 10.1007/s00362-013-0532-z

Matsuura, S., Kurata, H. (2013), "Definition and Properties of m-Dimensional n-Principal Points", *Communications in Statistics - Theory and Methods*, Vol. 42, No. 2, pp. 267-282, 査読有.

DOI: 10.1080/03610926.2011.581791

[学会発表](計15件)

松浦 峻, Tarpey, T., "Estimation of Principal Points of Multivariate Normal Distributions When Principal Points Are Collinear", 2014 年度統計関連学会連合大会, 2014 年 9 月 16 日, 東京大学・本郷キャンパス(東京都・文京区)。

Yamashita, H., Matsuura, S., Suzuki, H. “Model Selection for the Estimation of Principal Points for a Multivariate Binary Distribution”, Asian Network for Quality Congress 2014, 2014 年 8 月 6 日, Singapore.

Matsuura, S., Kurata, H. “On 2-Principal Points of a Mixture of Two Bivariate Normal Distributions”, The 3rd Institute of Mathematical Statistics Asia Pacific Rim Meeting, 2014 年 6 月 30 日, Taipei (Taiwan).

山下 遥, 松浦 峻, 鈴木 秀男, “対数線形モデルを用いた 2 値型 principal points の推定に関する研究 - アンケートデータ解析への応用 -”, 日本品質管理学会第 43 回年次大会研究発表会, 2013 年 11 月 16 日, 大阪大学・吹田キャンパス (大阪府・吹田市).

松浦 峻, 倉田 博史, “正規分布における principal points の推定について”, 2013 年度統計関連学会連合大会, 2013 年 9 月 11 日, 大阪大学・豊中キャンパス (大阪府・豊中市).

山下 遥, 松浦 峻, 鈴木 秀男, “多変量 2 値分布における対数線形モデルを用いた 2 値型 principal points の推定”, 2013 年度統計関連学会連合大会, 2013 年 9 月 9 日, 大阪大学・豊中キャンパス (大阪府・豊中市).

Ueki, D., Matsuura, S., Suzuki, H., “Principal Points Estimation Using Mixture Distributions”, 2012 International Conference on Advanced Computer Science and Information Systems, 2012 年 12 月 1 日, Depok (Indonesia).

松浦 峻, 倉田 博史, “楕円対称分布の混合分布の 2-principal points”, 2012 年度統計関連学会連合大会, 2012 年 9 月 10 日, 北海道大学・札幌キャンパス (北海道・札幌市).

Matsuura, S., Kurata, H., “Principal Points of a Multivariate Distribution and Their Reduced-dimension Version”, The 2nd Institute of Mathematical Statistics Asia Pacific Rim Meeting, 2012 年 7 月 3 日, つくば国際会議場 (茨城県・つくば市).

松浦 峻, 倉田 博史, “多次元確率分布の主成分の principal points について”, 2011 年度統計関連学会連合大会, 2011 年 9 月 5 日, 九州大学・伊都キャンパス (福岡県・福岡市).

## 6 . 研究組織

### (1) 研究代表者

松浦 峻 (MATSUURA, Shun)  
慶應義塾大学・理工学部・助教  
研究者番号 : 7 0 5 8 3 3 6 8