様式 C-19

科学研究費補助金研究成果報告書

平成 23 年 6 月 3 日現在

機関番号：13601
研究種目：基盤研究（C）
研究期間：2007～2010
課題番号：19530181
研究課題名（和文） 多変量分布の新たな視点に基づく解析

研究課題名（英文） Analysis of Multivariate Distribution from a new standpoint

研究代表者
椎名 洋 （SHIINA YO）
信州大学・経済学部・教授
研究者番号：80242709

研究成果の概要（和文）：
多変量分布に関して次のような新しい知見が得られた。１）ウィシャート分布の母集団分散共分散行列の固有根がグループ単位で限りなく広がった際に、標本固有根や固有ベクトルがどのような分布になるかについての考察。２）多変量正規分布の母集団分散共分散行列の固有根の推定量で、基準化された二乗損失に関して、許容的なものの導出。３）楕円分布の母集団固有根の寄与度に関して、従来使用されていた推定量の欠点を改良。

研究成果の概要（英文）：
On multivariate distributions, we have got following results; 1) For Wishart distributions, we derived the distribution of the sample eigenvalues and eigenvectors when population eigenvalues are block-wise infinitely dispersed. 2) For multivariate normal distributions, we derived an admissible estimator of population eigenvalues with respect to a normalized quadratic loss function. 3) For elliptically contoured distributions, we improved the classical estimator of the population contribution rates.

交付決定額

（金額単位：円）

	直接経費	間接経費	合 計
2007 年度	500,000	150,000	650,000
2008 年度	500,000	150,000	650,000
2009 年度	500,000	150,000	650,000
2010 年度	500,000	150,000	650,000
年度			
総 計	2,000,000	600,000	2,600,000

研究分野：社会科学
科研費の分科・細目：経済学・経済統計学
キーワード：多変量正規分布・ウィシャート分布・楕円分布・分散共分散行列

１．研究開始当初の背景

多変量分布の解析（分布論・推定論・検定論）に関しては、すでに多くの研究成果がえられていたが、いくつかの事項に関して未解決の問題がまだ多く残されていた（現在も同じ)。特に、標本固有根の分布は、非常に複雑で、Zonal 多項式によるエレガントな表現からさらに一歩踏み込んで、より具体的な推定・検定への応用になると、研究の多くは標本数が無限になった時の漸近分布を前提にしており、小標本の場合における正確な分布とそれに応じた推測方法は、開発の大きな余地のある問題であった。

２．研究の目的

本研究の目的は、上記のような背景を基に、小標本の下で、多変量分布に関して新たな知見を得ることを目的とした。特に、多変量正

規分布、あるいはそれを自然に拡張した楕円分布から得られた標本分散共分散行列の固有根の分布に関して、新しい知見を得ることを、第一の目的とした。

３．研究の方法
従来の多変量分布の小標本下の研究においては、主に解析的な方法が用いられてきた。端的に言えば、密度関数とその積分に関する操作を主な道具として、分布や推定・検定量の開発がなされてきた。本研究では、これとは違った道具、すなわち幾何的・代数的な道具を用いて多変量分布を扱うことによって新しい知見がえられることを期待して、幾何・代数の基本的な手法を理解し、その後に統計的な応用を図るという方法をとった。

４．研究成果
１）幾何的な道具
分布の集まりをリーマン多様体としてとらえ、微分幾何の手法を応用する、いわゆる情報幾何の基礎的な知識を習得した。また、この知識を実際の統計上の諸問題について応用するいくつかの論文を読んだ。その結果、研究目的につながるいくつかのアイデアを得たが、4 年間の間にこれを発展させて、具体的な形で成果をあげることはできなかった。今後の継続的な研究課題としたい。また、本来は代数的な手法も身につける予定でいたが、筆者の能力のせいもあり、十分に行えなかった。
２）本来目指した方法とは違い、従来の解析的な手法を用いて、以下のような具体的な成果をあげた。５．にあるように、これらの成果は全て査読付き雑誌に掲載されたか、あるいは投稿中である。

①ウィシャート分布は、多変量正規分布の標本分散共分散行列の分布であり、最も基本的な分布の一つである。この分布はパラメーター行列、すなはち、母集団の分散共分散行列によって規定される。この行列の固有根が無限に拡散するという特異な状態になったとき、標本固有根や固有ベクトルが、どんな分布になるかについては、すでに私と東京大学竹村教授との共同研究で一定の成果が得られていたが、これに関してさらに一般的な拡張を行った。

ウィシャート分布の次元をｐ個としたとき、母集団分散共分散行列の固有根はｐ個あるが、それらを大きい順に並べる。その際、一続きの固有根の集まりをブロックと呼ぶことにする。もし、ブロックごとに大きく固有根が拡散した場合、つまりブロック内の分散に比して、ブロック間の分散が非常に大きい場合に、標本分布、すなはち、ウィシャーと分布はどうなるか。これは、興味深いテーマである。日常さまざまな場面で、母集団分散共分散行列の固有根が、このようにブロック単位で拡散している状況は観察される。例えば、因子モデルにおいて、因子のもたらす変動は、誤差のそれに比して非常に大きく（大きいから、因子と呼べると考えてもよい。)、結果としてｐ個の固有根は、因子の数からなる大きな固有根のブロックと、それ以外の誤差の影響のみからなる小さな固有根のブロックに分かれる。昨今、盛んに研究が行われている超高次元のデータについても、高次元ゆえに自然にブロック単位の固有根の分散は生じる。

ここで、ブロック単位の固有根の無限拡散を正確に定義すると以下のようになる。各ブロックごとに、スケールパラメーターを考える。そして、各固有根は、自己の属するスケールパラメーターとブロック内の相対的な大きさを示すパラメーターの積として表される。この時、ブロック内のパラメーターは、固定して、ブロックのスケールパラメーターの比（隣接するブロックごとに比をとり、大きな固有根のグループのそれを、小さな固有根のグループのそれで割る。結果として、この比の数は、ブロックの数より一個少ない）が全て無限大になることを、ブロック単位の固有根の無限拡散の定義とする。

我々の分析の結果、母集団分散共分散行列が、ブロック単位で無限に拡散した場合、標本行列（ウィシャート行列）の固有根、固有ベクトルは次のような分布になる。固有根のブロック分割（こちらは一次元）に対応して、自然に標本行列にもブロック分割（こちらは、二次元）が入ることに注意しておく。その際、同じブロックナンバーを行と列にもつ、（二次元）ブロックを主ブロック、そうでないものを副ブロックと呼ぶことにする。さらに、標本ベクトルからなる直交行列についても同様の分割を考える。
１）主ブロックをそのスケールパラメーターで割って標準化すると、その標準化したブロックはそれぞれ独立なウィシャート分布になり、その自由度は（元の全体のウィシャート分布の自由度）－（最下層の主ブロックから自己のブロックまでのすべてのブロックの次元の和）になる。
２）標本ベクトルからなる直交行列の副ブロックはゼロに確率収束する。これを適当に基準化すると（基準化のやり方は、若干複雑なので省略する。）正規分布に収束する。
３）上に登場するすべてのブロックは漸近的に独立である。

②多変量正規分布の分散共分散行列の固有根を推定する場合、標本分散共分散行列の固有根を使用するのが一般的であるが、先述の通り、この分布は極めて複雑である。分布に

関するこれまでのほとんどの成果は、標本数が無限大になった場合の漸近分布に関するものであり、小標本の場合の成果も、母集団分散共分散が単位行列（或いは、その定数倍）という特殊な状況に限られていた。（先に①で述べた成果は、母集団分散共分散の固有根が無限に発散した場合に関するものであるが、これは、母集団分散共分散が単位行列とちょうど正反対の状態ととらえることができる。）

標本固有根の分布が小標本下で正確に把握できない以上、それを使った推定量の分布はさらに理解が難しいというのが現状である。一般に推定論において、許容的な推定量があるかどうか、また存在するとしたらどのような推定量かという問題は、基本的な問題であるが、この問題も上のような事情で長らく未解明であった。ここで言う、許容的とは、以下のことを意味する。ある損失関数（真のパラメーターと推定量の差を測るものさし）の、真のパラメーターの下での平均値をリスクと呼ぶが、このリスクが少ないほど優れた推定量と言える。もし、推定量Ａのリスクが、推定量Ｂのそれよりも、一様に、つまり真のパラメーターがどんな値であっても小さい時、ＡはＢをドミネイト（優越）すると言う。残念ながら、ある一つの推定量が他の全ての推定量をドミネイトすることは、通常無い（すなはち、一様最少リスク推定量は存在しない）。逆に、保守的に考えて、どんな推定量にもドミネイトされることがないという推定量も重要である。このような推定量を許容的という。

筆者と竹村教授（東京大学）は共同でこの問題に取り組み、ガンマ分布のスケールパラメーターの許容的な推定量を導く Ghosh & Singh(1970)の手法を拡張して、基準化二乗損失関数に関して、許容的な推定量を開発することに成功した。Ghosh＆Singh(1970)は、母数スケールパラメーターの空間に測度を考える、一般ベイズ推定量に似た推定量の許容性を証明しているが、これを一般の次元に拡張するためには、様々な工夫が必要になる。特に、標本固有根が複数存在しているために、その間の関係が重要になり、中でも、標本固有根が無限に拡散した時の推定量の形を調べる必要がある。しかし、これは上記①で述べた研究の応用であり、①の成果は、この研究においても、重要な役割を果たしている。

③多変量正規分布の分散共分散行列の固有根の寄与率（個々の固有根の大きさを、固有根の総和で割ったもの）は、主成分分析・因子分析、あるいはより一般的な共分散構造モデルの次元決定において、もっともよく使用される指標の一つである。これに関して従来、ほとんどすべての場合に使われてきた推定量は、標本分散共分散行列の固有根の寄与率であった。標本固有根はMLEであり、当然幾つかの望ましい漸近的な性質を持っている。しかし、小標本の下ではこれよりも優れた推定量が存在することが、多くの研究によって示されてきた。（この推定問題は、母集団分散共分散行列の直交不変な推定量と密接な関係があり、母集団分散共分散行列の直交不変推定量としては、いわゆるスタイン型の縮小推定量を始めとして、様々なタイプの推定量を開発されている。（詳しくは、T.W.Anderson(2003)、R.J.Muirhead(1992)等の教科書を参照)）

古典的な推定量の一番の欠点は、バイアスがかなり大きいことであり、その結果として、次元推定の道具としては、次元を過小評価してしまう傾向がある。因子分析、主成分分析等において、次元を過小評価することは、過大評価するよりはるかに深刻な問題をもたらすので、この欠点を改良することは重要なテーマである。

今回の研究では、統計的決定理論の立場から、一定の損失関数に関するリスクについて、古典的な推定量よりも良いパフォーマンスを示す推定量を開発した。直接、寄与率の推定量のパフォーマンスを二乗損失等の損失関数に関して計測するのは、数学的に困難なので次のようなアプローチをとった。まず、標本分散共分散行列の寄与率の分布は、母集団のそれのみに依存することを証明した。これは、母集団分散共分散行列のスケール変換に関して、標本寄与率が不変であることを意味するので、母集団分散共分散行列の固有根の総和が一に等しいと想定しても一般性を失わないことになる。この状況では、母集団分散共分散行列の固有根＝母寄与率になるので、以下のような推論が成り立つ。もし、ある寄与率推定量が良い推定量なら、それを標本分散共分散行列の固有根と置き換えて作った母集団分散共分散行列の直交不変推定量（以下、これをプラグイン推定量と呼ぶ）も、良い推定量になるのではないか。この考えに基づいて、プラグイン推定量のパフォーマンスを尤度損失関数（スタインの損失関数、或いはカルバックライブラー型損失関数とも呼ばれる）を使って計測することで、直接、寄与率推定量のパフォーマンスを見ることへの代替とするアプローチをとった。

母集団分散共分散行列の直交不変推定量に関しては、尤度損失に関するリスクの不偏推定量という強力な道具があり、これを用いることによって、古典的な寄与率推定量から作ったプラグイン推定量をドミネイトするプラグイン推定量を開発した。そして、このプラグイン推定量の固有根が新しい寄与率推定量となる。この新しい推定量の性質は以下の２つである。

１）この寄与率推定量を、直接二乗損失関数に関するリスクで評価することは、解析的には困難なので、シミュレーションを行った。その結果、古典的推定量のリスクを大きく減らしていることが判明した。新しい推定量が損失関数のタイプに依存しないことがある程度保障されたと考えられる。
２）一定の因子モデルの下で、乱数を発生させ、それに基づいて新旧推定量を使って次元推定を行い、真の次元との比較を行った。先に述べたように、古典的推定量はバイアスが大きく、次元を過小評価することが多いのに対し、新しい推定量は、むしろ過大評価をしがちであることが判明した。これにより、次元の過小評価という問題が改善されていることが確認された。
　以上の全ての結果は、多変量正規分布に基づく議論だが、実際の論文では分布をより一般的なクラスである楕円分布に広げてもまったく同じ結果が得られることが証明されている。本論文の成果は、正規分布にとらわれないという意味でロバストな結果になっている。

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３件）
①Yo Sheena, “Modified estimator of the contribution rates of population eigenvalues”, arXiv 1104.1019, 2011, 査読無.
②Yo Sheena & A.Takemura, ”Admissible estimator of the eigenvalues of the variance-covariance matrix for multivariate normal distributions”, Journal of Multivariate Analysis, 査読有, Vol.102, 2011, 801-815.
③Yo Sheena & A.Takemura, “Asymptotic Distribution of Wishart Matrix for Block-wise Dispersion of Population Eigenvalues”, Journal of Multivariate Analysis, 査読有, Vol.99, 2008, 751-775.

〔学会発表〕（計１件）
①Yo Sheena, “Admissible estimator of the eigenvalues of the variance-covariance matrix for multivariate normal distributions”, The 7th conference on multivariate distributions with applications, 2010/8/8-13, Maresias, Brazil.

６．研究組織
(1)研究代表者
椎名　洋 (SHIINA YO)
信州大学・経済学部・教授
研究者番号：80242709

(2)研究分担者
（　　）

研究者番号：

(3)連携研究者
（　　）

研究者番号：