

機関番号：17102

研究種目：若手研究 (B)

研究期間：2008～2010

課題番号：20700252

研究課題名 (和文) 特異モデルに対するモデル選択法の構築およびその応用

研究課題名 (英文) Model Selection for Singular Statistical Models and its Application

研究代表者

二宮 嘉行 (NINOMIYA YOSHIYUKI)

九州大学・大学院数理学研究院・准教授

研究者番号：50343330

研究成果の概要(和文):統計解析の不可欠な作業としてモデル選択がある。これは、モデル候補が複数あるときに、どのモデルが最もふさわしいかをデータから決める方法のことである。このモデル選択の最も基本的なものとして、AIC の利用がある。AIC はモデル内で推定される分布と真の分布の距離の推定量であり、これを最小にするモデルが最適なものとして選択される。しかし、特異モデルと呼ばれるものにおいては、実は妥当なAICがこれまで導かれていない。本研究の成果は、構造変化モデルや因子分析モデルといった特異モデルの具体例において、妥当なAICを導出したことにある。

研究成果の概要(英文):Model selection is an indispensable task in statistical analysis. It is a method to select the optimal model from data in several model candidates. To use AIC is one of the most basic ideas for such model selection. The AIC is an estimator of the distance between the true distribution and the estimated distribution in a model candidate, and so the model which gives the minimum AIC is regarded as the optimal model. For so-called singular statistical models, however, any reasonable AIC has not derived until now. The achievement of this research is to derive a reasonable AIC for a structural change model and a factor analysis model, which are examples of the singular statistical models.

交付決定額

(金額単位:円)

	直接経費	間接経費	合計
2008年度	1,200,000	360,000	1,560,000
2009年度	1,000,000	300,000	1,300,000
2010年度	1,000,000	300,000	1,300,000
年度			
年度			
総計	3,200,000	960,000	4,160,000

研究分野：数理統計学

科研費の分科・細目：情報学・統計科学

キーワード：計量経済学，計量心理学，情報量規準，漸近分布論，統計幾何学，非正則性，モデル選択，尤度比

1. 研究開始当初の背景

(1) 特異モデルについて

特異モデルとは、非正則性をもつ統計モデ

ル、つまり通常の統計理論を適用できない統計モデルの代表ともいえるものであり、それゆえに統計解析の基本であるモデル選択法さえも十分に開発されてこなかった。具体的には、非正則性を考慮しない理論

的に不当な方法か、単純な設定以外ではほとんど扱えない複雑な方法ばかりであった。応用上の需要が高いものに、構造変化モデルや折線回帰モデル、混合分布モデル、ARMA モデル、ランダム効果モデル、多層ニューラルネットワークモデル、因子分析モデル、隠れマルコフモデル(マルコフスイッチモデル)などがある。モデル選択では、尤度比が重要な役割を果たす。通常の(正則な)モデルにおいて、尤度比(正確には対数尤度比の2倍)は漸近的にカイ二乗分布に従い、それが情報量規準(AIC)の導出にも検定(尤度比検定)における閾値の導出にも用いられる。一方特異モデルにおいては尤度比の漸近分布はカイ二乗分布にはならないため、独自のモデル選択法が必要となる。

(2) 特異モデルに関する研究の動向

特異モデルに対する理論研究は古く1950年代からなされてきたが、それは個々の特異モデルに対する個別論であった。これに対し、最近是一般理論の構築が試みられつつある(Dacunha-Casttele and Gassiat, 2000; 福水・栗木, 2004)。尤度比の分布論については、統計幾何学(積分幾何学)を用いた理論が急速に発展し(Adler and Taylor, 2007)、陽には求まらない式や多重積分を含んだりするものの検定の枠組みで用いられている。その中で、特異モデルの非正則性を考慮した妥当な情報量規準は、ほとんど開発されていないといつてよい。

2. 研究の目的

より一般の特異モデルに対し、理論的に妥当かつ容易なモデル選択法を構築することを目的とした。道具として用いるのは、私が本課題に取り組む以前に開発してきた構造変化モデルに対する理論と、やはりそれまでに研究してきた統計幾何学、それから非正則性を扱うのではなく解消するために提案された理論である。特異モデルの一つである構造変化モデルにおいては、理論的に妥当かつ容易、さらに過去の不当な方法との差が顕著なモデル選択法を与える情報量規準の導出をおこなってきた。それをより一般的なものに拡張すること、つまり「特異モデルに対する情報量規準の導出」を第一の目標とした。それがかなわないときは非正則性を解消する方法を用い、非正則性が残る場合は統計幾何学の理論で対応すること、つまり「非正則性解消に基づく特異モデルのモデル選択法の確立」を

第二の目標とした。特異モデルは応用上重要なモデルを数多く含むため、統計解析に不可欠なモデル選択法を与えることは大きな意義をもつといえる。

3. 研究の方法

(1) 構造変化モデルの情報量規準に関する理論の拡張

特異モデルの情報量規準がなかなか開発されない中、私は特異モデルの代表例ともいえる時系列の構造変化モデルにおけるAICを導出した(Ninomiya, 2006)。具体的には、非正則性を無視すればAICは「 $-2 \times (\text{最大対数尤度}) + 2 \times (\text{変化数}) + 2 \times (\text{パラメータ数})$ 」となるわけだが、それは不当であり、実際は「 $-2 \times (\text{最大対数尤度}) + 6 \times (\text{変化数}) + 2 \times (\text{パラメータ数})$ 」となることを示した。この修正の適用範囲を広げる形で目的を達成することを計画した。

(2) 非正則性解消に基づく特異モデルのモデル選択法の確立

(1)で述べた拡張を一般の特異モデルに対しておこなうことが難しいことは、統計幾何学が与える結果の難しさからも想像できることであった。その中で目を引いたのが、Chen, Chen and Kalbfleisch (2001)による非正則性の解消であった。わずかな修正を尤度に加えることにより、その分布論を容易にするのである。しかし、それは単純な混合分布モデルに対する限られたものであったため、それを拡張する形で目的を達成することを計画した。

4. 研究成果

研究の方法で挙げた二つのテーマに関連する主成果として、以下の二つが挙げられる。「構造変化モデルの情報量規準に関する理論の拡張」に関連するのが「構造変化モデルに対するTICの導出」、「非正則性解消に基づく特異モデルのモデル選択法の確立」に関連するのが「因子分析モデルに対する非正則性解消およびモデル選択法の構築」である。

(1) 構造変化モデルに対するTICの導出

構造変化モデルに対する情報量規準の結果を、モデルの適用範囲ではなく、設定の適用範囲を広げるために拡張した。具体的

には、真の分布がモデルの中に入っているという仮定をはずしたもとの情報量規準、いわゆる構造変化モデルの TIC を導いた。指数型分布族に属する分布に独立にしたがう、あるいは自己回帰モデルにしたがう中でパラメータの値がシフトするという基本的な構造変化モデルにおいて、その TIC は陽な形で与えられることを示した。また、真の分布がモデルから少しずれている場合に、期待平均対数尤度と AIC, TIC を比較すると、AIC より TIC の方がはっきりと期待平均対数尤度に近いことを数値実験により確認した。そしてこの場合において、TIC によるモデル選択の方が AIC によるモデル選択に比べて妥当である、具体的に述べると AIC によるモデル選択は保守的すぎる結果を与えることも数値実験により確認した。

(2) 因子分析モデルに対する非正則性解消およびモデル選択法の構築

混合分布モデルより複雑な非正則性をもち、またこれまで需要の割に理論開発が進んでいなかった因子分析モデルに対し、非正則性解消を試みた。因子分析モデルは、計量心理学の分野で用いられる基本的なモデルである。そしてそのモデルを用いて情報を引き出すためには、因子数を推定することは不可欠な作業といえる。m 因子のモデルと (m+1) 因子のモデルを尤度比検定で比べることを繰り返すことによって因子数を推定する、という方法は最も自然な方法の一つであるが、因子分析モデルには特異モデルであるが故の非正則性があるため、実行は通常モデルのように容易ではない。具体的に述べると、尤度比検定統計量の漸近帰無分布が通常のようにカイ二乗分布にはならない。またこのようなケースにおいては、AIC や BIC といった情報量規準を形式的に用いることも有効ではない。そこで、非正則性を一部解消するために、まずはわずかな修正を尤度に施した。具体的には、因子負荷ベクトルは零ベクトルになってはいけないという実データ解析上は影響のない制約を加えた。この解消にもかかわらず、因子分析モデルはまだ非正則性を有している。そこで、他の特異モデルにおいて非正則性を扱うために開発された局所錐母数化の概念を因子分析モデルに対しても適用し、尤度比検定の漸近的な特徴付けを与えることをおこなった。そして、一般に特異モデルを幾何的に表現すると図 1 のようになるが、因子分析モデルを幾何的に表現すると図 2 のようになることを発見した。そしてこれを利用し、因子分析モデルにおける尤度比検定統計量は、他の

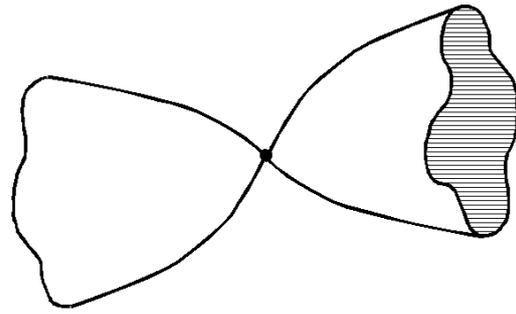


図 1：一般の特異モデルの幾何的表現

特異モデルのそれと比べるとシンプルなものに収束することを示した。この結果により、AIC を元来の定義に基づいて再評価することも可能となる。これに基づき、因子分析モデルのための新しい AIC を導出した。そして、従来の形式的な AIC や統計ソフトなどで実装されている Consistent AIC に比べ、新しい AIC は妥当な因子数の推定を与えることを、より具体的には真のモデルに Kullback-Leibler 距離の意味で近いモデルを与えることを、数値実験で示した。さらに、ある実データに対し、新しい AIC は最もふさわしいと言われる因子数を推定することを示した。このことは他の二つの AIC が異なる因子数を推定することを、つまり新しい方法が小さな改良を与えるわけではないこと意味している。

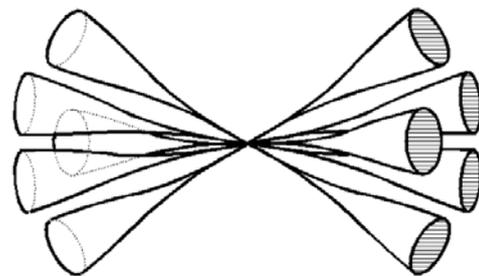


図 2：因子分析モデルの幾何的表現

5. 主な発表論文等

[雑誌論文] (計 3 件)

- ① 二宮嘉行・吉本敦, ベイズ法を用いた単木成長予測モデル, 森林資源管理と数理モデル, 査読有, 10 巻, 2011, pp. 333-349.
- ② Ninomiya, Y., Yanagihara, H. and Yuan, K.-H., Selecting the number of factors in exploratory factor analysis via locally conic parameterization, ISM Research Memorandum, Not refereed, No.1078, 2008, pp.1-34.

- ③ Yanagihara, H., Ninomiya, Y. and Yoshimoto, A., Analysis of grouped growth patterns in even-aged sugi forest stand within the framework of mixture model, FORMATH, Refereed, Vol.7, 2008, pp.39--60.

[学会発表] (計 5 件)

- ① Ninomiya, Y., A distribution theory for testing the existence of spatial clusters, Australian Statistical Conference, December 7, 2010, Perth.
- ② Ninomiya, Y., A conservative test for detection of spatial clustering. The International Symposium on Statistical Analysis of Spatio-Temporal Data, November 6, 2010, Kamakura.
- ③ 二宮嘉行, 構造変化を持つ時系列モデルに対する情報量規準, シンポジウム「非線形時系列に対する現象数理学の発展」, 2010年7月30日, 東京.
- ④ Ninomiya, Y., Yanagihara, H. and Yoshimoto, A., Bayesian method for predicting forest tree growth curve, FORMATH, March 14, 2010, Tokyo.
- ⑤ Ninomiya, Y., Yanagihara, H. and Yuan, K.-H., Selecting the number of factors in exploratory factor analysis via locally conic parameterization. Joint Statistical Meetings, August 5, 2009, Washington D. C.

[その他]

ホームページ:

<http://www.math.kyushu-u.ac.jp/teachers/view/51>

6. 研究組織

(1) 研究代表者

二宮 嘉行 (NINOMIYA YOSHIYUKI)
九州大学・大学院数理学研究院・准教授
研究者番号: 50343330