

機関番号：12601

研究種目：若手研究（B）

研究期間：2009～2010

課題番号：21740065

研究課題名（和文）ある特別な縮小型事前分布を用いたベイズ統計的推測に関する理論的研究
研究課題名（英文）A theoretical approach to Bayesian statistical inference by a special shrinkage prior distribution

研究代表者

丸山 祐造（MARUYAMA YUZO）

東京大学・空間情報科学研究センター・准教授

研究者番号：30304728

研究成果の概要（和文）：

ANOVA モデルにおける新たなベイズ型モデル選択規準の構築に専念した。ANOVA において、水準数が大きくなる場合の一致性を考えると、最も標準的な BIC によるモデル選択は必ず間違ふという問題が指摘されてきた。これは BIC の導出において水準数が固定のもとで繰り返し数が多くなるという漸近的な状況の近似により導出されていることを考えると、自然な結果である。多くの研究者により、BIC の改良が考えられてきたが、如何なる漸近的な状況にも一致性を持つ規準は提案されていなかった。私は Maruyama and George (2010, Arxiv) で考えられた特別な縮小型事前分布を ANOVA モデルの場合に適用して、周辺密度を解析的に計算することにより、新たなベイズ型モデル選択規準を提案した。これは繰り返し数が大きくなる漸近的な状況だけでなく、水準数が大きくなる状況においても合理的な一致性を持つ。また繰り返し数が各水準で違うアンバランスなケースについて考察したこと、繰り返し数と水準数の両方が大きくなる漸近的な状況において、発散のオーダーの違いが一致性にどのような影響を及ぼすかを考察した。

研究成果の概要（英文）：

For the balanced ANOVA setup, we propose a new closed form Bayes factor without integral representation, which is however based on fully Bayes method, with reasonable model selection consistency for two asymptotic situations (either number of levels of the factor or number of replication in each level goes to infinity). Exact analytical calculation of the marginal density under a special choice of the priors enables such a Bayes factor. The most advantage of our Bayes factor over existing Bayes factors is its excellent closed form. Since many Bayes factors based on fully Bayes method involve the integral in the representation, they have to apply the Laplace approximation in practice. However, the answer to the question which type of the Laplace approximation is more appropriate, is obscure for some cases. On the other hand, our Bayes factor does not require thought and has a reasonable model selection consistency

交付決定額

(金額単位：円)

	直接経費	間接経費	合計
2009年度	1,800,000	540,000	2,340,000
2010年度	1,600,000	480,000	2,080,000
年度			
年度			
年度			
総計	3,400,000	1,020,000	4,420,000

研究分野：数学一般（含む確率論・統計数学）
科研費の分科・細目：
キーワード：数理統計学，統計科学

1. 研究開始当初の背景

線形回帰モデルは，多変量統計解析を行う上での最も基本的なモデルである．また，複雑なデータ構造を表現し得る非線形モデルの中にも，非線形関数の線形結合として表現される例は多い．従って，計算機が高度に発達した現代においても，線形回帰モデルは，依然として統計学の中で重要な位置を占めている．

2. 研究の目的

本研究では，主に統計的決定理論の立場から線形回帰モデルにおける統計的推測問題を扱う．このとき，縮小型の事前分布を用いたベイズ統計的アプローチが多くの場合に有効であることが，ここ 40 年ほどの研究で明らかになってきた．今回の研究においては，後で述べる特別な縮小型事前分布のもとで構成される推測手法に照準を絞り，その理論的性質を解明していくことが大きな目標である．特に，誤差項に正規分布を仮定した正規線形回帰モデルを想定して，そのモデルに対する典型的な問題として推定，予測，変数選択を扱う．

3. 研究の方法

この研究は理論的研究であり，自分の頭で考え，また問題を共有する共同研究者と議論を行う以外のことはしていない．共同研究者とのディスカッションのため，受領金額のおおくを使わせて頂いたことに感謝する．具体的に，議論した海外研究者は以下の通りである．

Larry Brown 教授（ペンシルバニア大）許容性について非常に多くの論文があり，統計的決定理論の権威である．研究全体について議論した．

Edward George 教授（ペンシルバニア大）予測問題における許容性の問題，ベイズの変数選択問題，及びリッジ回帰推定量を用いた予測について議論した．

Bill Strawderman 教授（ラトガース大）スタイン問題での階層型一般化事前分布の創始者である．本プロジェクトでは，予測に役に立つ事前分布について，詳細に議論した．

4. 研究成果

ここでは，Maruyama and Strawderman (2011, to appear in IMS Collections) 及び Maruyama and George (2010, Arxiv) で得られた研究成

果を詳細に述べる．

(1) Maruyama and Strawderman

p 個の説明変数を持つ線形回帰モデルから n 個の y が得られているとする．ここで新たに m 個の y が生じるときその予測分布を構成するという問題を考える．これは Komaki (1996, 2001, Biometrika) でその重要性，数理的な構造の面白さが見出された問題である．具体的には調和関数を事前分布とする縮小型ベイズ予測分布が，通常の自然な予測分布よりも良いことを証明した．また Kato (2008, Annals of ISM) では，誤差分布の分散が未知というより現実的な設定においても，Komaki の結果がそのまま継承されることを示した．

彼らは良さの基準として，カルバック・ライブラーのダイバージェンスを採用した．このダイバージェンスは非対称性がひとつの特徴であり，距離の公理を満たさない．これに対して，逆向きに測ったカルバック・ライブラーダイバージェンスまでを自然につなぐダイバージェンスのクラスとして，アルファダイバージェンスが知られている．統計的推測問題において，良さについての結果が採用するダイバージェンスによって，

「自然な予測分布が縮小型ベイズ予測分布によって改良される」という大枠が崩れるのは望ましくないので，別のダイバージェンスにおいても，Komaki あるいは Kato の結果が継承されることをチェックするのは意味があることである．

実は逆向きのカルバック・ライブラーダイバージェンスを採用するとき，予測問題は線形回帰モデルの未知母数である回帰係数ベクトルと誤差分散の同時推定の問題に帰着することが分かった．このことを利用して我々は新たな縮小型ベイズ予測分布を提案した．Komaki や Kato では，その良さの証明に推定問題で用いられた道具とは違うものを準備して巧妙に解いた．一方我々の証明は推定問題に帰着させたために，用いる道具に新規性があるわけではない．実はここで面白いのは，回帰係数ベクトルの次元が 1 の場合でも，ベイズ型予測分布の改良が証明されたことである．この種の縮小型ベイズ推測の良さを主張する文脈では，回帰係数の次元が 3 以上であることが通常必要とされる．つまり，今度はまた Kato の設

定（通常の向きのカルバック・ライブラーダイバージェンス）において、「回帰係数ベクトルの次元が低次元の場合にも、自然な予測分布が縮小型ベイズ予測分布によって改良されるのか？もし改良されるのであれば、その縮小型事前分布は具体的にどのような形をしているのか？」という興味深い問題が生じる。ここで問題が難しいのは、Kato は調和関数をもととする事前分布を用いたベイズ予測分布の良さを主張しているのだから、低次元の場合に彼のアイデアをそのまま移行できないことである。我々は推定問題及び予測問題において、既にその良さが主張されている縮小型事前分布を、階層的な構造を追うことによって詳細に検討して、良さを主張できる可能性が高い事前分布のクラスを構成した。残念ながら科研費の受領期間中に具体的にその証明を与えることは出来なかったが、引き続き Strawderman 教授との共同研究を継続している。

(2) Maruyama and George

正規線形回帰モデルにおける変数選択問題をベイズ統計学の枠組みで考える。ベイズ変数選択問題においては、各部分モデルに事前分布を与えたときの周辺密度を計算することが本質的である。多くの場合、数値計算による時間コストを削減するために、解析的に積分計算出来る共役事前分布を適切に使うことが肝要になる。しかし、完全に共役事前分布を設定してしまうと、客観性が担保できなくなるため、これまでの多くの研究では分散部分にハイパーパラメータを導入して、それにも分布を入れる、あるいは経験ベイズのような方法で、データから推定することが行われてきた。このとき、前者では、積分が残るため、ラプラス近似により、その積分を近似する必要がある。一方後者の経験ベイズ的方法においてはベイズの枠組みからやや外れるため理論的根拠が弱くなるという欠点がある。これらの問題を克服すべく、我々はハイパーパラメータの積分も解析的に積分計算出来るような新たな事前分布を見つけた。さらに既存の共役事前分布の弱点である、

1. 標本分布の分散が小さい成分をより大きく修正してしまう

2. 個体数が説明変数の数より多い場合に使えない

を克服するような一般化された共役事前分布を提案した。結果として、AIC や BIC のような既存の評価規準と対比できる形に書きなおしたとき、既存のものが決定

係数と説明変数の数の関数であるのに対し、我々のそれはさらに最小二乗推定量のノルムの減少関数になっている。最小二乗推定量のノルムは、多重共線性が大きい場合に、いくつかの係数が符号を打ち消し合うように大きくなり、それに応じてノルムが大きくなると考えられる。つまりノルムの大きいモデルは望ましくないと多くの研究者が考えてきた。しかし、これまで必ずしも理論的に正当化されていなかった。我々は適切な事前分布を設定することにより、モデルの評価規準の中にこれを取り込んだ。

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 1 件)

Yuzo Maruyama and William E. Strawderman (2011). Bayesian predictive densities for linear regression models under alpha-divergence loss: some results and open problems. IMS Collections 掲載決定. 査読あり.

[学会発表] (計 7 件)

[2010.11] gBF: A Fully Bayes Factor with a Generalized g-prior. Conference “The International Symposium on Statistical Analysis of Spatio-Temporal Data”, Kamakura, Japan.

[2010.09] Fully Bayes Model Selection with a Generalized g-Prior. Statistics department seminar, Rutgers University.

[2010.08] A Bayes factor with reasonable model selection consistency for ANOVA model. Joint Statistical Meetings, Vancouver.

[2010.03] (Poster Presentataion) Bayesian variable selection with sub-harmonic priors. Conference “Frontier of Statistical Decision Making and Bayesian Analysis”, The University of Texas at San Antonio.

[2010.03] A Bayes factor with reasonable model selection consistency for ANOVA model. 京都大学数理解析研究所「Statistical Experiment and Its Related Topics」研究会.

[2009.09] ANOVA モデルで水準数が増える場

合のベイズ型モデル選択規準の一致性について, 統計関連学会連合大会, 同志社大学.

[2009.06] (Poster Presentaion) Perfect consistency of a Bayes factor for ANOVA model. “O-Bayes09”, the 2009 International Workshop on Objective Bayes Methodology, The University of Pennsylvania, Philadelphia.

[その他]

ホームページ等

<http://www.csis.u-tokyo.ac.jp/~maruyama/>

6. 研究組織

(1) 研究代表者

丸山 祐造 (MARUYAMA YUZO)

東京大学・空間情報科学研究センター・准教授

研究者番号 30304728