

## 科学研究費助成事業 研究成果報告書

平成 26 年 6 月 10 日現在

機関番号：12601

研究種目：若手研究(B)

研究期間：2011～2013

課題番号：23740067

研究課題名(和文) 縮小型事前分布によるベイズ統計的推測の理論研究

研究課題名(英文) Bayesian statistics based on shrinkage prior

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

交付決定額(研究期間全体)：(直接経費) 3,400,000円、(間接経費) 1,020,000円

研究成果の概要(和文)：正規分布を一般化した球対称分布に従う誤差項を持つ線形モデルにおけるベイズ的な変数選択問題を研究した。事前分布として正規分布の尺度混合分布として得られる分離可能事前分布を想定した。その事前分布のもとでの事後オッズは、球対称性を持つ限りにおいて誤差項の分布に依存しないことを示した。この無依存性から、提案する方法はロバスト性を所有すると考えられる。さらに、提案する事後オッズはモデル選択に関する一致性を持つことが示された。また興味深い理論的結果として、この研究で導かれた分離可能事前分布は、かなり広い分布の集合の中でただ一つの存在であることも示された。

研究成果の概要(英文)：We study Bayesian variable selection in linear models with general spherically symmetric error distributions. We construct the posterior odds based on a separable prior function, which arises as a class of mixtures of Gaussians. The posterior odds for comparing among non-null models are shown to be independent of the underlying error distribution, provided that the distribution is spherically symmetric. Because of this invariance to spherically symmetric error distributions, we refer to our method as a robust Bayesian variable selection method. We demonstrate that our posterior odds have model selection consistency. Additionally we demonstrate that our class of prior function are the only ones within a large class of mixtures of Gaussians studied in the literature which are robust in our sense.

研究分野：数学

科研費の分科・細目：数学一般(含確率論・統計数学)

キーワード：ベイズ統計学

### 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、いくつかの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重要な位置を占めている。本研究では、理論的立場から線形回帰モデルにおける統計的推測問題を考える。

### 2. 研究の目的

(1) 線形回帰モデルにおける統計的推測問題において、特にその重要性・有効性が明らかになってきている縮小型事前分布を用いたベイズ統計的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本研究の大きな目的である。

本研究では主に、誤差項に正規性を仮定した正規線形回帰モデルを想定し、統計的推測として、推定、予測、変数選択を扱う。具体的には、有限個のデータが得られたもとで、以下の典型的問題

推定 回帰係数ベクトルの関数の推定、誤差分散の推定

予測 新たな有限個の説明変数が得られたときに、対応する目的変数に関する予測分布の構成

変数選択 無駄な変数を含む利用可能な全ての説明変数が与えられていると想定した場合の、データに照らして妥当な説明変数の組み合わせの選択

において、理論的にも実用的にも優れた推測方法の提案を目指す。

(2) その際に基礎となるのが Maruyama and Strawderman (2005, *Annals of Statistics*) の結果である。この論文では、適切な階層型縮小事前分布を想定すると、重積分で表現される目的変数の周辺尤度解析的に計算できることを示した。この exact な closed form である周辺尤度に基づいた統計的推測は、(推定量や変数選択規準などの) 関数の形から、容易にそれを用いる意味が解釈可能であり、また MCMC などの数値計算や Laplace 近似、経験ベイズ等の近似をする必要がない、という優れた特徴を持つので、応用場面で非常に有用であると考えられる。さらに、提案手法が理論的にも優れた性質を併せ持つことを示したい。

私、及び共同研究者の Pennsylvania 大学の Brown 教授と George 教授、Rutgers 大学の Strawderman 教授の知る限りにおいて、同じ理念に基づく取り組みや研究はなく、国際的な観点からも独創的であると考えられる。

### 3. 研究の方法

過去に受領した科研費では、本研究テーマと関連した研究テーマを扱った。その研究費を

使って、アメリカの複数の研究者を訪問して、共同研究を行ってきた。近年の業績の多くがその共同研究の成果である。本研究においても、そのスタイルでの共同研究を継続する。そのため研究費のかなりの部分を旅費に当てる。

### 4. 研究成果

(1) 平成 23 年度は、主にプロビットモデルにおけるベイズ統計的推測を考えた。プロビットモデルは、目的変数がバイナリーである場合に線形回帰モデルを拡張したモデルである。サンプル数が少ない場合は、最尤推定量及びそれに関連する漸近理論は、理論的根拠がなくなり、正確性を欠く。その場合に無情報事前分布を想定するベイズ理論は、オルタナティブとして、重要だと考えられる。

回帰係数のベイズ統計的推測を考えた時、これまでほとんど理論的な考察は行われていなかった。具体的には MCMC sampler により、乱数を発生させて、その標本統計量を用いて統計的推測を行うという計算機インセンティブな方法が、デフォルトとして広く使われて来た。理論的な結果がほとんど得られてこなかった理由は、同時分布関数が場合分けを含んでおり、数学的に取り扱いづらかったからである。

この度、私は同時分布を極座標を用いて表すことにより、取り扱いをかなり容易にした。無情報事前分布のもとでは、事後分布は同時分布関数に比例する。私は、極座標表現による同時分布関数の性質を利用して、

「この事後分布が 3 段階の階層構造を持つこと」「そのうちの 2 つは正規分布とカイ二乗分布であり、乱数の発生が容易なこと(もう一つは、やや取り扱いづらくチャレンジングな課題として残っている)」を示した。これにより、モンテカルロ法による乱数発生が可能となり、既存の MCMC サンプリングよりも効率的な乱数発生が期待される。さらに、回帰係数のベイズ推定量の明示的な表現が、既に述べた階層型の事後分布を用いて得られることも分かった。これらの結果をまとめた論文を、Arxiv にプレプリントとして置いた。

(2) 平成 24 年度は、誤差項が必ずしも正規分布とは限らない一般球面对称分布に従う場合の回帰モデルに対するベイズ型変数選択を考えた。Zellner の g-prior の尺度混合分布としての劣調和型事前分布を考えた。

結果として、誤差項が球面对称分布であれば、Bayes factor が正規分布のもとでのそれに一致することを示した。そのため、我々の方法は、ロバストなベイズ型変数選択方法と考えられる。

また、その Bayes factor は良い Bayes factor が持つべき性質である coherent であることも示される。さらに、誤差項が正規分布の場合において数多く研究されてきた Bayes factor のラプラス近似についても、発展的な

結果を与えた。

近似の結果として、Bayes factor は正規分布のもとでの BIC 型の Bayes factor を、「事前分布の超母数」「対応する部分モデルの決定係数」「対応する部分モデルの説明変数の数」の 3 つを引数とする単純な有理関数で修正した形になることが示された。

また、提案する Bayes factor は良い Bayes factor が持つべき性質である一貫性を持つことも示された。ここで列挙したロバスト性、一貫性、coherent 性を保証する事前分布は、本質的に劣調和型事前分布だけに限られることも示された。

数値シミュレーションや 2 つの実データセットに対して、提案する Bayes factor が BIC に比べて良いパフォーマンスを示すことも例証された。

(3) 平成 25 年度は、縮小型事前分布を用いたベイズ統計的推論のうち予測分布の構成に関して集中的に研究した。具体的には、正規分布に従う確率変数ベクトルから  $n$  個の標本が観測されている場合に、 $n+1$  番目に得られる同じ正規分布からの確率変数ベクトルの分布を推定する問題である。

最も単純なのは、 $n$  個の観測値による最尤推定量などの適当な推定量を密度関数にプラグインする方法であるが、それは決定理論の観点からは非許容的になり、右不変事前分布に基づくベイズ予測分布が優越することが知られている。正規分布の分散を既知とする場合には、深く研究されてきた。

Komaki (2001, *Biometrika*) は正規分布の次元が 3 以上の場合に、調和型縮小事前分布に基づくベイズ予測分布が、右不変事前分布に基づくベイズ予測分布を優越することを示した。これを一般化する形で、George et al (2006, *Annals of Statistics*) が分散既知の設定においては、平均ベクトルパラメータの推定と予測分布の推定が本質的に同等であることを示した。

正規分布の分散を未知とする場合は、Kato (2008, *AIMS*) が正規分布の次元が 3 以上のもとで Komaki (2001) に対応する結果を示した。私は、いくつかの観点から問題の構造が分散既知と未知の場合で本質的に違っており、次元が 1 や 2 の場合でも縮小型事前分布に基づくベイズ予測分布が優越するはずだと予想を立てた。

結果として、 $n=2$  という追加的な条件が必要であるが、その予想が正しいことを示した。この研究は、フランス INRIA の Boisbunon 研究員との共同研究であり、*Biometrika* に掲載されることになった。

## 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 5 件)

① Aurelie Boisbunon and Yuzo Maruyama (2014). Inadmissibility of the best equivariant predictive density in the unknown variance case. To appear in *Biometrika*.  
<http://arxiv.org/abs/1308.2765>

② Edward, I. George and Yuzo Maruyama (2014). Posterior odds with a generalized hyper-g prior. *Econometric Reviews*, 33, 251-269.  
<http://dx.doi.org/10.1080/07474938.2013.807181>

③ Yuzo Maruyama and William E. Strawderman (2013). Improved robust Bayes estimators of the error variance in linear models. *Journal of Statistical Planning and Inference*, 143, 1091-1097.  
<http://dx.doi.org/10.1016/j.jspi.2013.01.007>

④ Yuzo Maruyama and William E. Strawderman (2012). Bayesian predictive densities for linear regression models under alpha-divergence loss: some results and open problems. *IMS Collections*, 8, 42-56.  
<http://dx.doi.org/10.1214/11-IMSCOLL803>

⑤ Yuzo Maruyama and Edward, I. George (2011). Fully Bayes Factors with a Generalized g-prior. *Annals of Statistics*, 39, 2740-2765.  
<http://dx.doi.org/10.1214/11-AOS917>

[学会発表] (計 3 件)

① Yuzo Maruyama (2013.01.16) Posterior inference and model selection of Bayesian probit regression. International Workshop on Bayesian Model Selection, East China Normal University, China.

② Yuzo Maruyama (2013.01.09) A Bayes factor with reasonable model selection consistency for ANOVA model. International Workshop/Conference on Bayesian Theory and Applications, Banaras Hindu University, India.

③ Yuzo Maruyama (2011.06.14) Robust Bayesian variable selection with sub-harmonic priors. "O-Bayes11", the 2011 International Workshop on Objective Bayes Methodology, East China Normal University, China.

[図書] (計 0 件)

[産業財産権]

○出願状況 (計 0 件)

名称:

発明者：  
権利者：  
種類：  
番号：  
出願年月日：  
国内外の別：

○取得状況（計 0 件）

名称：  
発明者：  
権利者：  
種類：  
番号：  
取得年月日：  
国内外の別：

〔その他〕

ホームページ等

<http://home.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

### (2) 研究分担者

( )

研究者番号：

### (3) 連携研究者

( )

研究者番号：