

機関番号：32641

研究種目：基盤研究(C)

研究期間：2008 ～ 2010

課題番号：20500259

研究課題名(和文) ベイズ法の発展的適用による回帰分析の展開

研究課題名(英文) Expanding the regression analysis through the innovative applications of Bayesian methods

研究代表者

柳本 武美 ( YANAGIMOTO TAKEMI )

中央大学 理工学部 客員教授

研究者番号：40000195

研究成果の概要(和文)：研究成果の概要(和文)：Bayes モデルでは母数を超母数として扱うよりも、可能ならば informative な事前分布の仮定が望ましい。この観点から平滑化を中心とした多次元平均ベクトルの同時推定を基本に手法の開発を行った。対立軸として近年注目されていた DIC を中心に据えて研究した。また研究のシーズとして e-混合型の予測子と二つの不偏性の性質から Bayes モデルの評価規準を考察した。その結果、e-混合交叉検証が良い性能を示すこと通じて今後の新しい研究の展望が開けた。

研究成果の概要(英文)：

The assumption of a prior information for a parameter contained in the sampling density is essential, and the use of a proper prior density is obviously desired. This view allows us to pursue fundamental subjects such as the smoothing method. I began with exploring the deviance information criterion (DIC). Our tools for this challenging problem are the use of the e-mixture predictor and the two notions of the unbiasedness of a potential function. Then I attempted to explore a cross-validation criterion (CVC). The induced CVC fortunately shows good performance, and the reasons of performance becomes clear. The results are expected to stimulate our future studies of this subject.

交付決定額

(金額単位：円)

	直接経費	間接経費	合計
2008年度	1,000,000	300,000	1,300,000
2009年度	800,000	240,000	1,040,000
2010年度	600,000	180,000	780,000
総計	2,400,000	720,000	3,120,000

研究分野：

科研費の分科・細目：情報学・統計科学

キーワード：(1) 交叉検証法 (2) 平滑化 (3) 不適切事前分布 (4) Bayes 法 (5) DIC  
(6) GCV

## 1. 研究開始当初の背景

18世紀に始まる Bayes 法が大きく展開したのは、1980 前後である。時系列解析における平滑化がその嚆矢であった。その有用性は瞬く間に広く認識された。しかし Bayes 法では、データから真理を探る発見的文脈では問題にならないが、実験の設計では標本サイズの決定には不都合で、結果の解釈にも不都合である。頻度論者にも受け入れられる Bayes 法の必要性は早くから指摘されてきた(Breslow, 1990 JASA; Efron 1990, Stat. Sci)。誰もが気づくこの研究が進まないのにはそれなりの理由がある。Bayesian は急速に複雑化するモデルへの対応に忙しく、頻度論者は、現場に必要な手法の細かい改善に没頭している。この問題は、分布論からの近似法の研究が乏しく、成果が得にくいおそれがある。この点を克服するためには、この問題に切り込むための明確な視点とそれを実現する道具が必要である。

形式的な Bayes 法を柔軟に適用することにより新しい発展が見込める。この発展のためには様々な従来の常識を越えた柔軟な Bayes 法の展開が必要である。より実用的には線形混合モデルのように事前分布を変量効果因子として扱い、分散分析あるいは一般化線型モデルとして展開する試みがなされてきた。しかし、Bayes 法としての定式は、より高い視点から統一的に眺めることを可能とする。変量効果モデルでは、複雑な現象を記述するために誤差分布を簡単にせざるを得ないので、推測方式がその場限りになる。一方、Bayes 法の研究ではモデルの選択と数値計算にのみ関心があって、革新的な推測法式の展開が等閑視された状態になっている。

## 2. 研究の目的

具体的なアイデアとして、1) 推測の枠組みは Bayes 法としながら、事前分布の仮定は柔軟に行う、2) 周辺尤度比の代わりに KL 分離度を用い

る、2) 周辺密度と標本密度を使い分けに格別の注意を払う、そして最後に 3) 愚直にシミュレーションを実行してその中から有用な不変量を見出す、の3点がある。特に 1) については準備的な結果を得ていた。この着想を具体的に議論している既存研究はない。しかし、この点は甘利らの幾何学的考察から見ると極めて自然に着想される。KL 分離度と Neyman-Pearson 補題との関連あるとの指摘が文献上にも見られる。ここで、超母数の推測には専ら周辺尤度に基づくとの暗黙の了解は格別の裏付けが無いことに注意する。実際、Bayes モデルを柔軟に与えると母数と超母数は必ずしも明確には区別できない。

頻度論の規準から見て優越する期待される Bayes 法の手法を開発し、実際に優越することを確認する。この試みが成功すると期待される分野として、多くの層が存在してモデルに層別母数と共通母数が存在する場合がある。この場合は、従来取り扱われて来た共通オッズ比の推定に止まらず、これまでには平滑化として扱われてきた問題をも多く含んでいる点で極めて一般的である。しかも、Bayes 法が最尤法に対して極めて優位になる条件を満たしている場合でもある。そこで、回帰モデルなど実用性を考慮して、絞った対象モデルについて計算方法を含めた具体的な手法の提示と、その優越性の提示を行う。先ず初めには、ある程度の解析的な扱いが期待できる誤差分布の下で研究を始める。その上で、一部の研究結果が適用できるより実際的なモデルに勤める。

リスクを評価規準とした Bayes 推定量は、従来には許容性についての研究しかなかった。その良さや改善の量としての評価は注力されてこなかった。そこで、Bayes 法が良い性能を示すことが見込まれるモデルの中で比較的シンプルな構造をしたモデルについて、その良さを量的に評価する。許容性は理論的な関心はあっても、実際に適用する場合には、改善の程度が明確でないといく

い。この点は、解析的方法と愚直なシミュレーションの繰り返しを通じて達成できると見込まれる。

### 3. 研究の方法

Bayes 法を頻度論の枠組みで評価すると伝統的な Bayes 法の枠組みでは展開が限られる。KL 分離度の利用である。まず、頻度論的な評価をする、つまりすべての母数は超母集団からの標本ではない、とすると周辺尤度比は概念的に受け入れ難い。推定したモデルと仮定したモデルとの距離を、推定量との関連で定義した方が考えやすい。また、Bayes モデルでは無い場合にも、最大化尤度比を評価するより、KL 分離度で評価した方が性能がよい。簡単な例として、標準的な平均と分散未知の正規分布における分散の検定（信頼区間）の場合を挙げることが出来る。また、以前に Bayes 法に関連して考察を加えた際に、KL 分離度を用いる方がより発展性に富むとの結果を得ていた。この点を数値的あるいは特殊な例での考察により検討を進める。尤度比は最尤法と不可分であるが、KL 分離度では、推定量の選択とは別個にその妥当性を検討できる意味で、基本的に柔軟なアプローチである。

標本分布に含まれるすべての母数に事前分布を仮定するのではなく、大部分に仮定して注目する母数には仮定しない hybrid 法はデータ解析以外にも試験の計画を重視する頻度論には受け入れ易い。この仮定は、Bayes 法の推論形式にも修正を促すことになる。実際、推定方程式からのアプローチが有力な代替案になる。周辺尤度の振る舞いは、通常の漸近分布理論が想定する関数とはかなり異なるからである。

モデルの母数は、あくまでも固定されていて推定的手段として Bayes 法を適用する。この立場は、Bayes 法の頻度論的理解では大前提となるはずであるが、なかなか採用されていない。提案する推定量は通常のリスクで評価し、検定量(信頼区間)は周辺分布ではなく標本分布の下で評価する。

Bayes 法あるいはより実践的な混合モデル法との根本的な違いがある。この違いは比較を困難にするが、混合モデルとは究極の目的では共有するので、導出の規準については比較検討を行う。その意味で先ず比較する対象は、様々に工夫された尤度法である。その工夫には、対の補正・条件付尤度など修正尤度法が含まれる。

### 4. 研究成果

複雑なモデルを Bayes モデルとして扱う場合には事前情報をどのように設定するかがポイントであることを明確に指定した。科学的推論に伴う社会的制約があれば、その制約が優先される。科学的な尤もらしさを事前分布として用いることはできない。そうすると hybrid 型の事前分布が有用になる（学会発表 6, 11）。じっさいに回帰分析で傾向母数に事前分布を仮定しなくても従来法が改善できた（発表論文 4）

一方でそうした制約がない場合には、利用するような設定にする方が望ましい。安易に母数を超母数として扱うよりも、可能ならば informative な事前分布そうでなくても無情報事前分布を仮定する方が、論理的整合性の面からもまた手法の性能の面からもより望ましいに違いない（発表論文 5）。これが Bayes 推論の本質でもある（学会発表 5, 7）。この着想から平滑化を中心とした多次元平均ベクトルの同時推定を基本に手法の開発を行った。既存の方法が抱える本質的な困難とそ

の理由を探ることにあつた。幸いこの問題に関する基本的な技法は直実に進展させてきている(発表論文 6)。

先ず、対立軸として近年注目されていた DIC を中心に据えて研究することにした。また研究のシーズとして統計的な量に広く観察される双対構造と頻度法から見た Bayes 法の評価であった。通常用いられる m-混合型ではなくて e-混合型の予測子から Bayes モデルの評価規準を考察した(発表論文 1-3)。e-混合予測子に基づいてモデル評価について着実な展開を示すことが出来た(発表論文 7)。特に e-混合交叉検証法を提案すると共に、モデル評価の統計量の正則条件として gradient の不偏性を考察した。これ迄は評価規準を potential としてのみ考察されていたからである。推定方程式の理論の分野では当然の要請が、モデル評価の分野では等閑視されていることも確認できた。その結果、e-混合交叉検証が良い性能を示すことをその根拠と共に得ることが出来た(投稿準備中)。この成果は単に e-混合交叉検証法についての成果に止まらずに新しい展開が期待できる。

#### 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 7 件)

- 1) Eguchi, S. and Yanagimoto, T.: Asymptotical improvement of maximum likelihood estimators on Kullback-Leibler loss. J. Statist. Plann. Inf., 138, pp3502-3511, 2008.
- 2) Yanagimoto, T. and Ohnishi, T.: Bayesian prediction of a density function in terms of  $\theta$ -mixture. J. Statist. Plann. Inf., 139, pp3064-3075, 2009.
- 3) Yanagimoto T. and Ohnishi T.: Predictive credible region for Bayesian

diagnosis of a hypothesis with applications. Japan. J. Statist. Assoc., 39, pp111-131, 2009.

4) Minoda, Y. and Yanagimoto, T.: Estimation of a common slope in multiple gamma regression models: An empirical Bayesian approach. Comput. Statist. Data Analysis, 53, pp4178-4185, 2009.

5) Minoda, Y., Kamakura, T. and Yanagimoto, T.: Improving Bayesian estimation of the end point of a distribution. J. Japan. Soc. Comput. Statist., 22, pp79-91, 2009.

6) Ohnishi, T. and Yanagimoto, T.: Duality induced from conjugacy in the curved exponential families. J. Japan. Statist. Soc., 40, 23-43, 2010.

7) Yanagimoto, T. and Ohnishi, T.: Saddlepoint condition on a predictor to reconfirm the need for the assumption of a prior distribution. To appear in J. Statist. Plann. Inf., 41, 2011.

[学会発表] (計 12 件)

- 1) 柳本 武美, データ科学の 4 要素が示す統計教育の展開 連合統計大会, 平成 20 年 9 月 8 日, 慶応大学.
- 2) 柳本 武美, 追加データによる仮説診断のベイズ予測法 「高度情報抽出のための統計理論・方法論とその応用」, 平成 22 年 11 月 20 日, 九州大学.
- 3) Minoda, Yuta, T. Kamakura and Yanagimoto, T. Improving Bayesian estimation of the end point of a distribution IASC(国際計算統計協会)2008 大会 20 年 12 月 8 日 横浜インターコンチネンタルホテル.
- 4) 柳本 武美, 鞍点予測子から見た informative 事前分布の利用, 連合統計大会, 平成 21 年 9 月 8 日, 同志社大学.
- 5) 柳本 武美, 鞍点予測子から見た informative 事前分布の有用性, 研究集会 「統計的推測・確率解析とその周辺の話題の理論と応用」, 平成 21 年 12 月 12 日, 秋田大学.
- 6) 柳本 武美・柳本正勝, 計量と実証の視点から見た食品の安全性評価, 連合統計大会 平成 22 年 9 月 8 日, 早稲田大学.
- 7) 柳本 武美・大西俊郎, 事前分布の仮定を促進する経験 Bayes 法の展開, 連合統計大会, 平成 22 年 9 月 8 日, 早稲田大学.
- 8) 渡部大志・鎌倉稔成・柳本 武美, 経験 Bayes 平滑化における次数の選定, 連合統計

大会，平成 22 年 9 月 9 日，早稲田大学。

9) 柳本武美・大西俊郎，事後密度にのみ基づいた経験 Bayes 法の利点，「計算機支援による統計手法，理論・応用」シンポ，平成 22 年 11 月 26 日，高知大学。

10) 柳本武美，事前分布の評価における DIC，交差検証，周辺尤度，九州大学・数理学・統計科学セミナー，平成 22 年 12 月 10 日，九州大学。

11) 柳本武美，事前分布が仮定できるための要件 - 利害関係者の視点，シンポ「医薬品の研究開発における統計的課題」，平成 22 年 1 月 18 日，大阪大学。

12) 大西俊郎・柳本武美： Saddlepoint condition on a predictor to reconfirm the need for the assumption of a prior distribution. 京都大学数理解析研究所共同研究集会，平成 23 年 3 月 8 日，京都大学

## 6. 研究組織

(1) 研究代表者 柳本，武美

(Yanagimoto, Takemi )

中央大学・理工学部・客員教授

研究者番号：40000195