

機関番号：12601

研究種目：基盤研究（B）

研究期間：2008～2010

課題番号：20300097

研究課題名（和文） 多重性を考慮した予測分布の構成法の開発と応用

研究課題名（英文） Construction methods for predictive distributions under multiplicity

研究代表者

駒木 文保（KOMAKI FUMIYASU）

東京大学・大学院情報理工学系研究科・教授

研究者番号：70242039

研究成果の概要（和文）：ベイズ統計学における予測の問題では、未知パラメータと予測したい量とのデータに対する条件付相互情報量を最大化することが有効であることを明らかにし、これに基づいて多重性をともなう予測問題について考察した。ベイズ予測分布の範囲で多重性の問題を解決する予測分布が構成できることを示した。また、一般には最良の事前分布は複雑になり最良の事前分布を厳密に求めることは困難になる。近似が必要になるため、数値的な手法を開発した。

2項分布モデル・多項モデルに関する予測問題の研究に加え、線形回帰モデル、時系列モデル、Wishart モデル等のより複雑なモデルでの予測問題の研究を行った。これらのモデルを用いた予測を考える際に必要となるモデルの情報幾何学的な性質について明らかにした。

研究成果の概要（英文）：It has been proved that priors maximizing the mutual information between the parameter and the quantity to be predicted given the observation play an essential role. By using this fact, prediction problems with multiplicity have been studied. It has been shown that predictive distributions resolving the multiplicity can be constructed as Bayesian predictive distributions. The best prior distributions become complex and difficult to obtain. Numerical approximation methods have been developed. In addition to binomial and multinomial models, prediction problems concerning linear regression models, time series models, and Wishart models have been investigated. Information geometrical properties of these models have been studied and applied to construct superior prediction.

交付決定額

（金額単位：円）

	直接経費	間接経費	合計
2008年度	2,500,000	750,000	3,250,000
2009年度	2,200,000	660,000	2,860,000
2010年度	2,000,000	600,000	2,600,000
総計	6,700,000	2,010,000	8,710,000

研究分野：総合領域

科研費の分科・細目：情報学・統計科学

キーワード：ベイズ統計，情報量

1. 研究開始当初の背景

ベイズ統計の方法は、さまざまな応用分野で有用であることが認識され、最近、理論・応用両面から注目されている。しかし、予測したい量に適応した事前分布の構成、特に多

重性を考慮した予測のための事前分布の構成については、有効な方法がほとんど知られていなかった。

研究代表者は、統計理論を予測分布の視点から統一的に扱う研究を進め多様体の大域的な微分幾何学的性質を調べることにより、

従来良いとされてきた予測（例えば、最尤推定に基づく予測やジェフリーズ事前分布に基づくベイズ予測）を優越する予測を構成する方法を開発した。ベイズ推測理論と情報幾何の特に大域的な性質を本質的に結び付ける視点は従来ほとんど無かった。その方法をさらに推し進めることにより、予測の目的に依存した事前分布を利用した予測分布の構成や、特に多重性を考慮した実用的な予測分布の構成法を与えることが重要となっていた。

2. 研究の目的

ベイズ統計法において、予測したい量に適応した事前分布の構成や、多重性を考慮した予測のための事前分布の構成については、有効な方法が知られていなかった。本研究では、予測分布の性能を評価するという視点から、この問題を解決することを目指した。

統計モデルには、多項分布やそのサブモデルのようにパラメータ空間がコンパクトであるものと、正規モデルのようにパラメータ空間がコンパクトではないものがある。ベイズ理論においては両者には大きな違いがある。パラメータ空間がコンパクトな場合には、多重性を考慮した予測分布が原理的には存在する。しかし、最良の予測分布は大変複雑になるので、適切なクラスの中から最良の予測分布を十分近似できるものを選ぶものを選ぶ必要がある。そのために、事前分布のクラスとして、混合分布モデルや階層ベイズモデルなどの扱いやすいクラスを利用して、その中で一番性能の良いものを選ぶ手法を開発する。

パラメータ空間がコンパクトでない場合には、理論的な取り扱いが難しくなり、ミニマックス性を要請しただけでは予測法が決まらなると予想される。例えば、回帰モデルや、時系列モデル、Wishart モデルなどでは、モデルの情報幾何的な性質を考慮することが必要になる。これらの実用上重要なモデルの幾何学的な性質を調べるとともに、予測の構成につなげることを目指す。

ベイズ統計において事前分布の設定は常に問題になるため、多くの研究がなされてきた。情報理論の視点から相互情報量に基づいて事前分布の設定をすることを提案した重要なものに reference prior の理論がある。しかし、予測性能の評価の立場から、条件付相互情報量の期待値を評価することにより、性能の良い予測分布が構成できることが最近の研究代表者の研究により示された。この方法は従来の考えられていた方法と本質的に異なるものである。さらにこのアプローチは多重性を考慮した予測にも有効であることが予想される。この研究を発展させることによ

り、今までにない多重性の取り扱いの方法が開発する。

3. 研究の方法

予測分布の構成法についての理論的研究、ベイズ予測分布の事前分布を数値的に求める手法の開発を行った。

ベイズ統計学において、無情報事前分布をデータと未知パラメータとの相互情報量の最大化に基づいて構成するという考え方は Bernardo (1979), Berger and Bernardo (1989) の論文以来、多くの研究がなされ、reference prior の理論として発展してきており、多くの研究がなされている。一方、相互情報量最大化では適切な事前分布が構成できない問題があり、解決のためには、データと未知パラメータとの相互情報量を最大化するのではなく、未知パラメータと予測したい量とのデータに対する条件付の相互情報量を最大化することが適当であることが、研究代表者により示されている。これは、reference prior とは異なる定式化ではあり得られる事前分布も異なるものになる。本研究での定式化により、reference prior が何をしているのかも明確に理解することが可能になる。情報理論の専門家である大濱靖匡教授と連携して、この方面の理論的な研究を進める。

一般には、最良の事前分布は大変複雑になるので、近似が必要になる。例えば2項分布モデルの場合には各パラメータにベータ分布を仮定したりする方法や数値的なアプローチが必要にある。2項モデル、多項モデルに対しては、単純な共役事前分布ではなく、階層ベイズモデルや、多数の点（粒子）を用いた分布の表現などにより構成できるより多様な事前分布を用いて予測を改良する数値的な方法を開発する。

また事前分布を設定したときの条件付相互情報量は、解析的に求まらないため、数値的に評価する必要がある。オペレーションズリサーチ、特に乱数や準乱数の専門家である諸星穂積教授と連携をとり、相互情報量を乱数や準乱数を用いた数値積分を用いて評価し、それを最大化する手法を研究する。

より複雑なモデルに対してもモデルの情報幾何学的な性質を調べることにより、性能の良い予測を構成する研究を行う。

研究成果の発表と情報交換のため、予測理論と応用に関する研究集会を2009年には「ベイズ統計への情報理論的アプローチとその周辺」と2010年には「ベイズ統計・量子統計の新展開」というテーマで開催した。連携研究者の大濱教授、諸星教授をはじめとする多数の参加者と意見交換を行った。2回の研究集の日時、場所、発表者、発表内容は以下

のとおりである。

研究集会「ベイズ統計への情報理論的アプローチとその周辺」

日時：2009年12月4日（金）

場所：東京大学 本郷キャンパス 工学部6号館 3階 セミナーB

発表：

- 1) 田中冬彦（科学技術振興機構さきがけ/東京大学）「優調和事前分布に基づいたスペクトル密度のベイズ推定」
- 2) 駒木文保（東京大学）「相互情報量とベイズ統計」
- 3) 金森敬文（名古屋大学）「対数尤度の変形による事前知識の導入とブースティングへの応用」
- 4) 鈴木大慈（東京大学）「Multiple Kernel Learningの効率的な計算手法とその拡張について」
- 5) 竹内純一（九州大学）「最尤符号とベイズ符号」
- 6) 大濱靖匡（徳島大学）「多端子情報源符号化と統計学との接点について」

研究集会「ベイズ統計・量子統計の新展開」

日時：2010年12月3日（金）

場所：東京大学 本郷キャンパス 工学部6号館 3階 セミナー室C

発表：

- 1) 諸星穂積（政策研究大学院大学）「RQMCによる高次元数値積分と応用」
- 2) 田中冬彦（科学技術振興機構さきがけ/東京大学）「スペクトル密度の推定における漸近展開と事前分布の選択」
- 3) 廣瀬善大（東京大学）「compressive samplingとモデル選択に関するレビュー」
- 4) 杉山太香典（東京大学）「推定誤差確率に基づく量子測定装置の性能評価と、非正則モデルとしての量子力学」

4. 研究成果

ベイズ統計学において、無情報事前分布をデータと未知パラメータとの相互情報量の最大化に基づいて構成するという考え方に基づいた多くの研究がなされ、reference priorの理論として発展してきている。一方、相互情報量最大化では適切な事前分布が構成できない問題があり、解決のためには、データと未知パラメータとの相互情報量を最大化するのではなく、未知パラメータと予測したい量とのデータに対する条件付の相互情報量を最大化することが適当であることが、研究代表者により示されている。

本研究では、この事実に基づいて多重性をとまなう予測問題について考察し、ベイズ予測分布の範囲で多重性の問題を解決する予

測分布が構成できること、ミニマックス性の意味で最良の事前分布が決まること、多重性に代表される困難な問題を解決する最良の事前分布が情報理論的な構成が原理的に可能なことを示した。また、一般には最良の事前分布は複雑になり何らかの近似が必要になるため、各パラメータに共役事前分布であるベータ分布を用いる方法と数値的なアプローチを研究した。特に、2項モデル、多項モデルに対しては、単純な共役事前分布ではなく、階層ベイズモデルや、多数の点（粒子）を用いた分布の表現などにより構成できるより多様な事前分布を用いて予測を改良する方法を開発した。

2項分布モデル・多項モデルに関する多重性を持つ予測問題の研究に加え、線形回帰モデル、時系列モデル、Wishartモデル、経験尤度に基づくノンパラメトリックなモデル等のより複雑なモデルでの予測問題の研究を行った。これらのモデルを用いた予測を考える際に必要となるモデルの情報幾何学的な性質について明らかにした。

さらに、従来は予測分布の良さに関する評価基準としてKullback-Leiblerダイバージェンスを用いる研究が主であったが、Kullback-Leiblerダイバージェンス以外のより一般的な評価基準のもとでの予測分布の性能に関する理論的な研究を行った。

事前分布を設定したときの条件付相互情報量は、解析的に求まらないため、数値的に評価する必要がある。条件付き相互情報量最大化による事前分布の構成についてのアルゴリズムについてマルチコアワークステーションを用いて実装するプログラムを作成して数値的な検証を行った。

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7件）

- ① Tanaka, F. and Komaki, F. Asymptotic expansion of the risk difference of the Bayesian spectral density in the autoregressive moving average mode, Sankhya Series A, 掲載決定.
- ② Hirose, Y. and Komaki, F. (2010). An extension of least angle regression based on the information geometry of dually flat spaces. Journal of Computational and Graphical Statistics, vol. 19, 1007-1023.
- ③ Suzuki, T. and Komaki, F. (2010). On prior selection and covariate shift of β -Bayesian prediction under α -divergence risk, Communications in Statistics - Theory and Methods, vol. 39,

1655-1673.

- ④ Inagaki, K. and Komaki, F. (2010). A modification of profile empirical likelihood for the exponential-tilt model. *Statistics & Probability Letters*, vol. 80, 997-1004.
- ⑤ Komaki, F. (2009). Bayesian predictive densities based on superharmonic priors for the 2-dimensional Wishart model, *Journal of Multivariate Analysis*, vol. 100, 2137-2154.
- ⑥ Kobayashi, K. and Komaki, F. (2008). Bayesian shrinkage prediction for the regression problem, *Journal of Multivariate Analysis*, vol. 99, 1888-1905.
- ⑦ Tanaka, F. and Komaki, F. (2008). A superharmonic prior for the autoregressive process of the second order, *Journal of Time Series Analysis*, vol. 29, 444-452.

[学会発表] (計 10 件)

- ① 廣瀬善大・駒本文保 Least Angle Regression の情報幾何学的拡張による高次元データ解析. 2010 年度統計関連学会連合大会, 9月7日, 早稲田大学.
- ② 近藤健司・駒本文保 ベイズ経験尤度法における優調和事前分布の利用とその有限標本理論. 2010 年度統計関連学会連合大会, 9月7日, 早稲田大学.
- ③ 駒本文保 ベイズ理論の現在. 2010 年度統計関連学会連合大会, 9月5日, 早稲田大学.
- ④ Hirose, Y. and Komaki, F. An extended least angle regression for contingency tables, *International Conference on Trends and Perspectives in Linear Statistical Inference*, July 27, 2010, Tomar, Portugal.
- ⑤ 廣瀬善大・駒本文保 双対平坦空間における Least Angle Regression と情報量規準. 2009 年度 統計関連学会連合大会 9月8日 同志社大学 京田辺・今出川キャンパス.
- ⑥ 駒本文保 ノンパラメトリック回帰モデルにおけるベイズ推定. 2009 年度 統計関連学会連合大会 9月8日 同志社大学 京田辺・今出川キャンパス.
- ⑦ Komaki, F. A predictive approach to prior construction in Bayesian inference. *The 1st Institute of Mathematical Statistics Asia Pacific Rim Meeting*, 29 June 2009, Seoul National University, Korea.
- ⑧ 駒本文保 Arimoto-Blahut アルゴリズムの拡張と Bayes 統計への応用. 2008 年度

統計関連学会連合大会 9月10日 慶應義塾大学・理工学部 矢上キャンパス.

- ⑨ 廣瀬善大・駒本文保 compressive sampling 問題における最適化問題とアルゴリズムに関する研究. 2008 年度 統計関連学会連合大会 9月8日 慶應義塾大学・理工学部 矢上キャンパス.
- ⑩ 稲垣和久・駒本文保 経験尤度法における修正プロファイル尤度. 2008 年度 統計関連学会連合大会 9月8日 慶應義塾大学・理工学部 矢上キャンパス.

[図書] (計 0 件)

[産業財産権]

- 出願状況 (計 0 件)
- 取得状況 (計 0 件)

[その他]

6. 研究組織

(1) 研究代表者

駒木 文保 (KOMAKI FUMIYASU)
東京大学・大学院情報理工学系研究科・教授
研究者番号: 70242039

(2) 研究分担者

(3) 連携研究者

大濱 靖匡 (OOHAMA YASUTADA)
徳島大学・大学院ソシオテクノサイエンス研究部・教授
研究者番号: 20243892
諸星 穂積 (MOROHOSI HOZUMI)
政策研究大学院大学・大学院政策研究科・教授
研究者番号: 10272387