

機関番号：17102

研究種目：若手研究 (B)

研究期間：2008～2010

課題番号：20700259

研究課題名 (和文) 解析的アプローチを用いた一般化線形モデルの Bayes 解析

研究課題名 (英文) An analytical approach to Bayesian generalized linear models

研究代表者

大西 俊郎 (OHNISHI TOSHIO)

九州大学・大学院経済学研究院・准教授

研究者番号：60353413

研究成果の概要 (和文)：

理論研究として主に次の3つを行った。

- (1) 最適な予測分布が満たす等式に着目し、Bayes 型モデル選択基準を論じた。
- (2) 共役事前分布が存在するという解析的な性質から Bayes 推定における双対性を導いた。
- (3) 共役事前分布をもつ location family について、線形混合モデルを研究した。

応用研究として Tweedie 分布をベースとする一般化線形モデルを用いてオーストラリア東部の降水量データの解析を行った。

研究成果の概要 (英文)：

The following three major topics were investigated.

- (1) Noting the equality satisfied by the optimal predictive density, we discussed a Bayesian model selection criterion.
- (2) We derived duality in the Bayesian estimation from the analytical property of having a conjugate prior.
- (3) We investigated a linear mixed model that is based on a location family having a conjugate prior.

As an application study we analyzed the rainfall data of East Australia using a generalized linear model whose underlying distribution is the Tweedie distribution.

交付決定額

(金額単位：円)

	直接経費	間接経費	合計
2008年度	800,000	240,000	1,040,000
2009年度	1,300,000	390,000	1,690,000
2010年度	800,000	240,000	1,040,000
年度			
年度			
総計	2,900,000	870,000	3,770,000

研究分野：統計科学

科研費の分科・細目：情報学・統計科学

キーワード：統計数学，一般化線形モデル，Bayes 統計学，共役事前分布，Tweedie 分布，降水量データ，国際情報交換，オーストラリア

1. 研究開始当初の背景

(1) 一般化線形モデル

指数型分布族をベースにした回帰モデルであり、経済学や生物統計学などの応用分野で広く用いられている。

(2) Bayes 推定

パラメータの次元が高いときに尤度推定より優れたパフォーマンスを示すことがよく知られている。高次元パラメータモデルである一般化線形モデルの Bayes 推定 は大きな発展が見込まれる重要な分野である。

(3) 共役事前分布

事前分布と事後分布が同一の分布族に属するという性質は解析的な性質といえる。指数型分布族が共役事前分布をもつことはよく知られているが、共役事前分布を持つ一般化線形モデルの存在はあまり知られていない。

(4) Tweedie 分布

指数型分布族に属し、理論的にも応用上も重要な分布である。特定のパラメータ領域において複合ポアソン分布に一致することが知られており、降水量データの解析や保険数理などで使われる。密度関数が複雑な形をしているが、共同研究者の Peter Dunn 准教授（サンシャインコースト大学、オーストラリア）は十分な精度の数値的評価が可能なプログラムを作成している。

(5) Tweedie 分布をベースにした一般化線形モデル

共役事前分布が存在し、解析的アプローチが可能であることを報告者が示した。興味深いことに、この共役事前分布は別の研究で導出された共役解析が可能な location family に属する。

2. 研究の目的

(1) Tweedie 分布が複合ポアソン分布に一致する場合

Tweedie 分布をベースにした一般化線形モデルにおける Bayes 型の推定問題と検定問題を研究する。推定問題では、報告者が提案した推定法の良さをシミュレーションにより明らかにする。検定問題では、検出力の高い検定方式の提案が目標である。応用として、近年史上最悪のかんばつに見舞われているクイーンズランド州の降水量データに適用し、どの因子が気候変化に影響を与えたのかを検証する。

(2) (1)以外の Tweedie 分布

Tweedie 分布はある分布のクラスであり、正規分布および逆ガウス分布も含んでいる。(1)以外の場合にも Bayes 型の推定問題および検定問題を研究する。正規分布および逆ガウス分布の場合には(1)の拡張が可能なことを個別に確認している。

(3) その他

共役事前分布が存在するという解析的な性質を持つ他の一般化線形モデルを探索する。

3. 研究の方法

(1) ピタゴラス関係

ある推定量が別の推定量を優越することを示すとき、

(ある推定量のリスク)

$$= (\text{別の推定量のリスク}) + (\text{非負の量})$$

というタイプの等式を導くやり方はスマートである。リスクの差が非負であること示せるだけでなく、リスクの差の挙動が明らかになるからである。このような等式をピタゴラス関係と呼ぶことにする。Stein 型推定が尤度推定を優越すること、および、Bayes 推定の最適性はピタゴラス関係に着目すると理解しやすい。

(2) 共役事前分布

共役事前分布が存在するモデルでは、

$$(\text{対数尤度比}) = (\text{損失})$$

というタイプの等式が成立する。これは興味深い等式である。なぜなら、左辺は尤度原理により大きければ大きいほどよい量であり、右辺はその性質上小さければ小さいほどよい量であるからである。

(3) 物理学的視点

多くの研究者によって情報学と物理学の結びつきが指摘されている。物理学的視点をもって研究を進めることは効率性の向上につながる。

(4) 共同研究

報告者が1人で行う研究であるが、Peter Dunn 准教授と積極的に意見交換・研究打ち合わせを行った。

4. 研究成果

(1) Saddlepoint 予測分布

下の研究(4)で扱った e-mixture 予測分布がキーンとなった。この予測分布が満たす等式に着目し、saddlepoint 予測分布という概念を導入した。Saddlepoint 予測分布とは対数尤度比と損失の2つの項がバラ

ンスするような予測分布である。研究の方法(2)で述べた等式と同様に、望大項と望小項をバランスさせるという面白い性質をもつ。

この新しい概念を通じて Bayes 型モデル選択基準との関連を論じ、Deviance Information Criterion (DIC) として知られているモデル選択基準に対してある種の正当化を与えることができた。

(2) 共役事前分布の存在と双対性

指数型分布族において共役事前分布が存在することおよび自然パラメータと平均パラメータの間の双対性があることはよく知られている。この研究では曲指数型分布族を題材にすることにより、共役事前分布の存在が双対性を誘導していることを明らかにした。

この研究は Tweedie 分布をベースにした一般化線形モデルの Bayes 推定において有用である。ある特定の location family と共通の尤度を持ち、この location family が曲指数型分布に属するからである。

(3) 予測信頼区間

下の研究(4)の応用である。e-mixture 予測分布を用いて予測信頼区間を構築し、新しい Bayes 型仮説診断法を提案した。

(4) e-mixture 予測分布

Bayes 予測問題において損失に応じて最適な予測分布が1つ定まる。この研究では e-ダイバージェンス損失の下での最適予測分布である e-mixture 予測分布に焦点を絞り、この予測分布が満たす等式を導いた。この等式が上記の研究(1)の原動力になった。通常文献などでは m-ダイバージェンス損失の下での最適予測分布である m-mixture 予測分布が用いられるが、この予測分布は上記の等式を満たさないことも明らかにした。

(5) 降水量データの解析

2010年にオーストラリアのクイーンズランド州は大洪水による甚大な被害を被ったが、その前年まで史上最悪の干ばつに見舞われていた。クイーンズランド州の降水量データを解析したのがこの研究である。Tweedie 分布をベースとする一般化線形モデルを用いた。

Tweedie 分布は降水量データなどゼロを多く含むデータの解析に非常に便利である。報告者はこの分布をベースにした一般化線形モデルは共役解析可能なモデルであることを証明した。Peter Dunn 准教授(サンシャインコースト大学、オーストラリア)の優れた数値評価プログラムと報告者

の理論研究を融合した。

(6) 線形混合モデル

研究(2)で述べたように、Tweedie 分布とある特定の location family は共通の尤度を持ち、したがって同一の共役事前分布を共有する。この location family をベースにした線形混合モデルを Bayes 法の枠組みで研究した。

Tweedie 分布は原点に mass をもち、正の領域では連続型であり、負の領域には密度を持たない。一方、上記の location family は実軸上の連続分布である。このように一見大きく異なる2つの分布であるが、尤度関数を共有する点は非常に興味深い。

(7) Location family における鞍点等式

Location family において対数尤度比および損失を上手く定義することにより、3種類の鞍点等式を導いた。具体的には尤度推定関数、Stein 型推定関数および Bayes 型推定関数が満たす鞍点等式を導き、研究集会で発表した。

この研究の続きは、報告者を研究代表者とする基盤研究(C)「鞍点等式を用いた Bayes 推測の新展開」(2011~2013年度)で行う予定である。

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 5 件)

- ① Takemi Yanagimoto, Toshio Ohnishi
“Saddlepoint condition on a predictor to reconfirm the need for the assumption of a prior distribution”
Journal of Statistical Planning and Inference 141, 2011, 1990-2000, 査読有. doi:10.1016/j.jspi.2010.12.011
- ② Toshio Ohnishi, Takemi Yanagimoto
“Duality induced from conjugacy in the curved exponential family”
Journal of the Japan Statistical Society 40, 2010, 23-43, 査読有.
<http://www.terrapub.co.jp/journals/jjss/abstract/4001/40010023.html>
- ③ Takemi Yanagimoto, Toshio Ohnishi
“Predictive credible region for Bayesian diagnosis of a hypothesis”
Journal of the Japan Statistical Society 39, 2009, 111-131, 査読有.
<http://www.terrapub.co.jp/journals/jjss/pdf/3901/39010111.pdf>

④ Takemi Yanagimoto, Toshio Ohnishi
“Bayesian prediction of a density function in terms of e-mixture”
Journal of Statistical Planning and Inference 139, 2009, 3064-3075, 査読有.
doi:10.1016/j.jspi.2009.02.005

⑤ 大西俊郎, Peter Dunn
“Tweedie 一般化線形モデルを用いたクイーンズランド州の降水量データの解析”
京都大学 数理解析研究所 講究録 1621, 2009, 135-152, 査読有.
http://www.kurims.kyoto-u.ac.jp/~kyod/o/kokyuroku/contents/pdf/1621-10.pdf

[学会発表] (計 11 件)

① 大西俊郎, 柳本武美
“Saddlepoint condition on a predictor and its implication”
RIMS 共同研究 “Statistical Information in Inference and Its Related Topics”
2011年3月8日 京都大学 数理解析研究所.

② 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”
科研費研究集会「計算機支援による統計手法, 理論・応用およびその周辺」 2010年11月26日 高知大学.

③ 大西俊郎, 柳本武美
“Location family における鞍点等式とその含意”
科研費研究集会「計算代数手法に基づく数理統計学の展開」 2010年11月16日 大宮ソニックシティ.

④ 大西俊郎, Peter Dunn
“Conjugate analysis of a linear mixed model in the location family”
科研費研究集会「統計科学の数理と応用」
2009年10月7日 岡山国際交流センター

⑤ 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”
統計関連学会連合大会 2010年9月7日 早稲田大学.

⑥ 大西俊郎, 柳本武美
“Location family における鞍点等式とその含意”
統計関連学会連合大会 2010年9月7日 早稲田大学.

⑦ 大西俊郎
“e-混合型予測分布による Bayes 予測およ

びそのモデル選択への応用”
統計数理研究所 共同利用研究集会「水産資源に対する観察データ解析のための統計推測」 2010年3月9日 統計数理研究所

⑧ Toshio Ohnishi
“Bayesian prediction in terms of e-mixture and its implication”
Institute of Statistical Mathematics, Indian Statistical Institute and Institute of Statistical Science, Academia Sinica Joint Meeting, 22nd January 2010, Indian Statistical Institute, Kolkata, India.

⑨ 大西俊郎, Peter Dunn
“Conjugate analysis of a generalized linear mixed model”
統計関連学会連合大会 2009年9月8日 同志社大学.

⑩ 大西俊郎
“Tweedie 一般化線形モデルの Bayes 解析”
統計数理研究所共同利用研究「水産資源に対する観察データ解析のための統計推測」
2008年9月25日 統計数理研究所.

⑪ 大西俊郎, 柳本武美
“Bayesian 最適予測子を用いた仮説の診断法の提案”
統計関連学会連合大会 2008年9月10日 慶應義塾大学.

6. 研究組織

(1) 研究代表者

大西 俊郎 (OHNISHI TOSHIO)
九州大学・大学院経済学研究院・准教授
研究者番号：60353413

(2) 研究分担者

()

研究者番号：

(3) 連携研究者

()

研究者番号：