

科学研究費助成事業 研究成果報告書

平成 28 年 10 月 1 日現在

機関番号：22301

研究種目：若手研究(B)

研究期間：2012～2015

課題番号：24740069

研究課題名(和文) ベイズ型情報量基準の理論と応用

研究課題名(英文) Theory and application of a Bayesian information criterion

研究代表者

宮田 庸一 (Miyata, Yoichi)

高崎経済大学・経済学部・准教授

研究者番号：10514250

交付決定額(研究期間全体)：(直接経費) 2,700,000円

研究成果の概要(和文)：ベイズ型情報量基準を導くためには統計モデルの周辺尤度と呼ばれる積分に対してラプラス近似を行う必要がある。そしてその近似を行うためには、被積分関数もしくはその主要な部分を最大にする値(ここではモードと呼ぶ)が必要になる。しかし、パラメータの個数に対して標本の個数が十分に大きくない場合には正確なモードを求めることが難しい時がある。本研究においては、正確なモードに近い動きをする漸近的なモードを用いてラプラス近似を行ったときの妥当性および漸近誤差を明らかにした。またその漸近的なモードを構築するためには、パラメータの推定量の収束の速さが大きく影響することを明らかにした。

研究成果の概要(英文)：When applying Laplace's method to approximate an integral, we require an exact mode of its integrand. We derive a Laplace approximation using an asymptotic mode of an asymptotic mode instead of the exact mode, and present some ways to establish asymptotic modes with a desirable order. The class of the asymptotic modes is wide and includes consistent estimators with some convergence rate and a true parameter.

研究分野：統計学(漸近理論)

キーワード：ラプラス近似 ベイズ統計学 情報量基準

1. 研究開始当初の背景

- (1) 金融工学, 遺伝子解析, 医学などの多くの分野において, データが取得されており, 有益な情報を抽出するための統計モデルが議論されている. 不適切な統計モデルを選ぶことは, データを生じている構造に偏りのある推定, 将来の観測値の予測において大きな予測誤差を生じさせる可能性が高い. このため適切な統計モデルを選ぶことが必要となり, そのためには観測値に対してフィットさせる統計モデルに対する評価が重要となる. 統計モデルの評価法の1つとして, 統計モデルのある種の予測誤差最小化原理に基づく赤池情報量規準(Akaike information criterion, AIC)と統計モデルの周辺尤度最大化原理に基づくベイズ型情報量規準(Bayesian information criterion, BIC)がある. 研究開始当初, 周辺尤度を近似するための別の手法としてはモンテカルロ法があり, その汎用性の広さから現在でも多くの分野で使用されている. しかしながら周辺尤度は一般的にはとても小さい値を取るため, モンテカルロ法を用いて近似するためには非常に多くの乱数を生成させる必要があり, 高速に計算できるという点において, 漸近的な近似手法を用いた AIC, BIC もいまだに様々な分野に対して多くの貢献がある.

- (2) BIC を導出するためには, 通常, 積分の形で表される周辺尤度に対してラプラス近似を行う必要がある (Schwarz, G. 1978). そしてその近似を行うためには, 被積分関数もしくはその主要な部分を最大にする値(ここではモードと呼ぶ)が必要になる. しかし, 周辺尤度における被積分関数がパラメータの個数に対して標本の個数が十分に大きくない場合, 被積分関数がパラメータの非線形関数になっている場合には正確なモードを数値的に求めることが難しい時がある.

2. 研究の目的

- (1) 統計モデルの周辺尤度における積分に対して, 正確なモードの代わりに漸近的なモードを用いてラプラス近似を行い, ベイズ型情報量規準を導出する. その際, ラプラス近似の漸近誤差を評価し, かつそれにより導出されたベイズ型情報量規準の漸近分布, モデル選択における一貫性を明らかにする.
- (2) 漸近的なモードを構築するための手法を明らかにする.
- (3) さらに正規線形過程, Autoregressive conditional heteroscedasticity (ARCH) モデル, Generalized ARCH(GARCH)モ

デルなどの非線形モデルに対してラプラス近似を適用し, それにより導出される情報量規準の有効性, 妥当性を検証する.

- (4) (1)で導出されたベイズ型情報量規準の有効性をシミュレーションにより確認する.

3. 研究の方法

ベイズ型情報量規準, およびラプラス近似に関する先行研究のサーベイを行う. その上で上述した研究の目的(1)においては, Miyata(2004)で与えられたオーダー n^{-1} の漸近的なモードを用いたラプラス近似を, より広いクラスの推定量を用いたラプラス近似へと拡張を行う. ここでオーダーとは, 漸近的なモードが被積分関数の正確なモードから漸近的な観点からどのくらい乖離があるのかを示す指標である. また, 研究の目的(2)については, ラプラス近似において好ましい漸近誤差を持つような漸近的なモードの構築方法の検討を行う. 研究の目的(3)においては, 正規線形過程, GARCH モデルが, (1)で与えられたラプラス近似の正則条件を満たすことを確認する.

(1)で得られた展開式の高次の項を含めたベイズ型情報量規準の良さを確認するためにシミュレーションを行う.

4. 研究成果

- (1) n は標本の大きさとする. 周辺尤度に対して $O_p(n^{-\alpha})$ (ただし $\alpha > 1/2$) の漸近的なモードを用いてラプラス近似を行った場合, その漸近誤差(相対誤差)がラージオーダー $O_p(n^{-\min(\alpha, 1)})$ になることを確認した. 即ち, 周辺尤度で表される積分に対して, その被積分関数の正確なモードにある程度近い値(漸近的なモード)を用いてラプラス近似を行っても妥当性は失わないということである. しかしその漸近誤差のオーダーは, 漸近的なモードがどのくらいその正確なモードに近いかに依存する. より具体的に述べると, 周辺尤度を含んだ

$$I_n = \int_{\Theta} \lambda(\theta) \exp\{-nh_n(\theta)\} d\theta$$

という形の積分を考える. ここで $h_n(\theta)$ はこの積分における主要な項であり, 対数尤度関数を含んだものになっている. 例えば $h_n(\theta) = -(1/n) \log\{p_n(y|\theta)\pi(\theta)\}$ となる. $p_n(y|\theta)$ は観測ベクトル $y=(y_1, \dots, y_n)$ の確率密度関数とし, $\pi(\theta)$ をパラメータ $\theta \in \mathbf{R}^d$ の事前分布とする.

$|A|$ は行列 A の行列式を表す記号とし, $D^2 h_n$ を $h_n(\theta)$ のヘッセ行列を漸近的なモード $\hat{\theta}$ で評価したものとする. 同様にして, 他の関数についても, 漸近的なモードで評価したものは, $\hat{\theta}$ を省いた形で

表すことにする．記号の簡略化のため， $C_n = \{(n/2)(Dh_n)'(D^2h_n)^{-1} Dh_n\}$ とし， $\alpha^* = \min\{2\alpha, 4\alpha - 1, 6\alpha - 2, \alpha + 1, 2\}$ とする． $\hat{\theta}$ を $O_p(n^{-\alpha})$ (ただし $\alpha > 1/2$) の漸近的なモードとすると，ある正則条件の下で，以下の展開式を導出した．

$$\begin{aligned} \mathcal{I}_n &= (2\pi)^{d/2} |nD^2h_n|^{-1/2} \exp\{-nh_n\} C_n \\ &\times \left\{ \lambda + \frac{a_{1,n}}{n^\alpha} + \frac{a_{2,n}}{n^{3\alpha-1}} + \frac{a_{3,n}}{n} + \frac{a_{4,n}}{n^{\alpha^*}} \right. \\ &\quad \left. + O_p\left(\frac{1}{n^2}\right) \right\} \end{aligned}$$

ここで $a_{1,n}$, $a_{2,n}$, $a_{3,n}$, $a_{4,n}$ は $O_p(1)$ の項であり， $a_{1,n}$, $a_{2,n}$, $a_{3,n}$ の具体的な形は Miyata[1] で与えられている．結果的にはベイズ型情報量規準を導出の妥当性の観点からは，必ずしも最尤推定量を用いなくてもよいことが確認された．

- (2) オーダー $O_p(n^{-\alpha})$ (ただし $\alpha > 1/2$) の漸近的なモードの構築方法を明らかにした．具体的には，収束レートが $O_p(n^{-\alpha})$ の推定量 (即ち $n^\alpha(\hat{\theta} - \theta) = O_p(1)$) は オーダー $O_p(n^{-\min(\alpha, 1/2)})$ (ただし $\alpha > 1/2$) の漸近的なモードとなることを明らかにした．また オーダー $O_p(n^{-1})$, $O_p(n^{-2})$ の漸近的なモードを構築するためのいくつかの方法を明らかにした．特に真の値 θ_0 に対するルート n 一致推定量からの $-h_n(\theta)$ 最大化のためのワンステップ推定量は オーダー $O_p(n^{-1})$ の漸近的なモードとなることを明らかにした．この理論の詳細は，Miyata [1]，宮田 [2] で与えられている．またこの [1]，[2] の研究結果を利用することにより，ベイズ推定量などの事後平均に対して閉じた形の完全指数型近似 (詳しくは Tierney and Kadane 1986 を参照のこと) を与えることができた．
- (3) GARCH モデル，一般化線形モデルにおいて研究成果 (1) で与えた漸近展開の妥当性を保証するための正則条件を満たすことを確認した．特に GARCH モデルにおいては，Kristensen and Linton (2006) の閉じた形で表すことができる推定量からのワンステップ推定量が オーダー $O_p(n^{-1})$ の漸近的なモードとなり，これを用いて周辺尤度に対するラプラス近似が可能であることが明らかになった．
- (4) 実際のデータに対して統計モデルの選択を行う場合には，候補となる統計モデルは全て誤特定された (misspecified) されたと見るべきであることが知られている．今回の研究により，Lv and Liu (2014) の条件に事前分布の連続性に関する条件を加えることで，誤特定された一般化線形モデルの周辺尤度に対するラプラス近

似を厳密に導出することが可能であることを示した．詳しくは Miyata[3] で与えられている．またそれにより得られた高次の導出項を含めたベイズ型情報量規準を提案し，一致性が成り立つための条件を確認し，およびシミュレーションを行うことによりそのパフォーマンスを確認した．注意すべき点としては，あるガウス事前分布を用いて高次の項を含むベイズ型情報量規準を用いた場合，Schwarz の BIC の持つ真のモデル，もしくは真のモデルに最も近いモデルよりも小さいモデル (即ちパラメータ数が少ないモデル) を選んでしまう弱点を改善していることが確認された．

参考文献

[1] Miyata, Y. Laplace approximations using n^α -consistent estimators, (2016年6月中旬に投稿予定)

[2] 宮田庸一，ラプラス近似のベイズ統計学への応用とその周辺，数学，日本数学会邦文誌，岩波書店 (現在，改訂中)

[3] Miyata, Y. Laplace approximations and Bayesian Information criteria in misspecified models (2016年7月に投稿予定)

[4] Kristensen, D. and Linton, O. (2006). A closed-form estimator for the GARCH(1,1) model, *Econometric Theory*, 22, 323-337.

[5] Lv, J. and Liu, J. S. (2014) Model selection principles in misspecified models. *J. R. Statist. Soc. B*, 76, Part 1, 141-167.

[6] Tierney, L. and Kadane, J.B. (1986). Accurate Approximations for Posterior Moments and Marginal Densities, *J. Amer. Statist. Assoc.*, Vol. 81, 82-86.

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 0 件)

現在，4. 研究成果において挙げた **参考文献** の [1], [3] を投稿に向けて準備をしている．

〔学会発表〕(計 5 件)

ここで から は国際シンポジウムもしくは国際学会になり，は 国内のシンポジウムとなる．

Yoichi MIYATA. (2015年11月). Some asymptotic properties of Bayesian

information criteria in misspecified models, Waseda International Symposium " High Dimensional Statistical Analysis for Spatio-Temporal Processes & Quantile Analysis for Time Series " ,Waseda University.

Yoichi MIYATA. (2015 年 8 月). The validity of Bayesian information criteria in misspecified models, Joint Statistical Meetings, Washington State Convention Center, 800 Convention Place, Seattle, USA.

Yoichi MIYATA. (2015 年 3 月). The validity of Bayesian information criteria in misspecified models, Waseda International Symposium "Asymptotic Sufficiency, Asymptotic Efficiency and Semimartingale", Waseda University.

Yoichi MIYATA. (2014 年 8 月). Asymptotic Properties of Bayesian Type Estimators When It Is Not Assumed the Hessian Matrices of Contrast Functions Converge, Joint Statistical Meetings, Boston Convention and Exhibition Center, Boston, USA.

一般化 BIC における最近の動向とその漸近的な性質について, 科研費シンポジウム「統計的推測の理論的基礎とその応用」, 筑波大学筑波キャンパス.

〔図書〕(計 0 件)

4. 研究成果において挙げた**参考文献**の[2]の原稿を 2016 年 4 月に提出し, それに対してレフリーからいただいたコメントに従い, 改訂を行っている.

6. 研究組織

(1) 研究代表者

宮田 庸一 (MIYATA, Yoichi)

高崎経済大学・経済学部・准教授

研究者番号: 10514250