

# 一般化された局所特定化過誤モデルにおける 信用集合と信頼集合の一致について

吉 村 有 博

## 要 旨

本論文は弱い特定化過誤を伴う滑らかな有限次元モデルにおけるパラメータの信用集合と信頼集合の一致の可能性について検討する。Hjort and Claeskens (2003) の局所特定化過誤を一般化した枠組みにおいて、我々は標準的な事前分布の仮定の下で Bernstein-von Mises 定理のバージョンを示す。このとき漸近的な情報行列等式が成立することを示し、最尤推定量のバイアスが標準誤差より大きい場合でさえ、信用集合と信頼集合は漸近的に互いに近づくことを示す。モンテカルロ研究からは、バイアスによりいずれの集合も真値の被覆に失敗する場合でさえ、信用/信頼集合は互いに近づくことが示唆される。結果は特定化過誤が弱い状況において信用集合に信頼集合と同程度の被覆性質を持つとする解釈を与える正当化になる。さらに、結果は信用/信頼集合の計算に互いの計算手法を代替させることの正当化になる。この結論はベイジアンと頻度論のいずれの立場にとっても有益である。得られた結論は既存研究に含まれない新規的知見であり、モデル不確実性下の統計的推論の発展への貢献が期待される。

キーワード：Bernstein-von Mises 定理、一般化された局所特定化過誤の枠組み、ベイジアン推測、信用/信頼集合、情報行列等式

## 1. 序論

近年の統計学や計量経済学において、Bernstein-von Mises 定理（以後、BvM 定理）の研究に大きな興味を持たれている。von Mises (1919) と Bernstein (1927) らによる定理によれば、滑らかなモデルにおける有限次元パラメータの結合事後分布は漸近的に多変量正規分布に従い、その平均は最尤推定値、分散共分散行列は Fisher 情報行列の逆行列になる (Le Cam and Yang (1990), van der Vaart (1998) など)。定理が示唆する興味深い点は、ベイジアンと頻度論それぞれの立場に基づくパラメータの信用集合と信頼集合が、任意の信頼水準に対して、漸近的に一致することである。このとき、ベイジアンにとって導入が不可欠な事前分布は、二つの推論の違いに本質的な影響をもたらさないことが知られる。定理によりベイジアン信用集合に頻度論の意味での被覆性質を持つとする解釈を与えられることは重要である。さらに、定理により信用/信頼集合の計算に互いの計算手法を代替させる根拠を与えられる。例えば、信頼集合の計算が容易でない場合において、マルコフ連鎖モンテカルロ法 (MCMC) で計算された信用集合が信頼集合を代用しうる (Chernozhukov and Hong (2003))。また一方で、ブートストラップにより計算された信頼集合が信用集合を代用しうる (Kitagawa, Montiel Olea, Payne

and Velez (2020)). 二つの推論の立場には哲学の違いが明確に存在するが、信用 / 信頼集合の一致の可能性はいずれの立場にとっても有益であるように思われる。

BvM 定理の研究の進展において、特に近年、定理の無限次元モデルへの拡張が盛んに研究されてきた。とりわけ、興味のあるパラメータが無限次元の場合には、信用 / 信頼集合の一致が一般には成立しない事実が Cox (1993), Freedman (1999) 等によって指摘された。このようなネガティブな結果は、BvM 定理の拡張とベイジアン-頻度論の関連に関する基本的な限界であるかのように見える。しかしながら、より最近になって、一部のノンパラメトリックモデルや興味あるパラメータが有限次元のセミパラメトリックモデルを中心として、多くのポジティブな結果も報告されてきた。これには、例えば、Kim and Lee (2004), Kim (2006), Castillo and Nickl (2013, 2014), Shen (2002), Chernozhukov and Hong (2003), Bontemps (2011), Bickel and Kleijn (2012), Castillo (2012), Rivoirard and Rousseau (2012), Kato (2013), Castillo and Rousseau (2015), Leahu (2011), Knapik, van der Vaart and van Zanten (2011), Szabó, van der Vaart and van Zanten (2015), Norets (2015), Chae, Kim and Kleijn (2019) を挙げることができる。Ghosal and van der Vaart (2017) は定理の無限次元への拡張に関する近年の展開をよく整理している。

モデルパラメータの次元に関する議論の一方で、信用 / 信頼集合の不一致をもたらす重要な条件がモデルの特定化過誤 (misspecification) である。実際、特定化過誤が存在する場合、有限次元モデルでさえ、事後分布の分散は最尤推定量の標本分布の“サンドイッチ形式”の分散 (例えば White (1982)) と一致しない。このとき、一般に信用集合と信頼集合は一致しない。この事実は比較的古くから知られていたようだが (Hartigan (1983), Geweke (2005) など)、近年再び、ベイジアンと頻度論の対比がより明確な枠組みにおいて Kleijn and van der Vaart (2012), Müller (2013) が取りあげた。とりわけ前者は基礎的な漸近論的枠組みにおける特定化過誤 BvM 定理を示し、信用集合と信頼集合は一般には一致しないことを指摘した。この結果は事前分布の影響が漸近的に消えゆく状況にも関わらず生ずることに注意されたい。彼らの結果の意義は、ベイジアンと頻度論それぞれの立場から見える推論の不確実性が根本的に異なることを、特定化過誤を含めた一般かつ共通の枠組みにおいて改めて明確にした点にある。彼らの結果に照らせば、古典的な BvM 定理と集合の一致は、あくまで特定化が正しいとする極めて理想的な仮定の下での特殊例であることが明瞭になる。統計学や計量経済学において特定化過誤の脅威は普遍的に存在するため、特定化過誤は独自に重要な視点である。また、ノンパラメトリックベイズが特定化過誤バイアスの回避に部分的に動機づけられた点を踏まえれば、特定化過誤はモデルパラメータの次元とも決して無関係ではない (ノンパラメトリックベイズの動機は Ghosal and van der Vaart (2017) など)。

Kleijn and van der Vaart (2012) により特定化過誤下の集合の一致に関する基礎が確立された一方で、計量経済学では特定化の誤りの程度が“弱い”場合の研究が重要である。複雑な計量経済モデリングにおいて潜在的な特定化過誤は常に存在するものの、それに伴うバイアスはい

つでも深刻であるわけではない。実際、応用上のモデル特定化の手続きは優れた統計的性質をもつ検定やモデル選択法に基づきデータドリブンに行われる。このため、実際の最終モデルは真の分布の近似として決して悪いものではなく、モデリングバイアスは残るにせよ、その程度が深刻とは限らない。経験的に見ても、これまで応用で用いられた多くの最終モデルは比較的単純な有限次元モデルであり、この事実は特定化手続きで生ずるバイアスが深刻でないことを部分的に反映したものと思われる。特定化の誤りの程度が弱い状況下では集合の不一致の程度は幾らか緩和される可能性があるが、その全体像は既存研究で明らかにされていないようである。特に、バイアスが深刻な場合に焦点を当てた Kleijn and van der Vaart (2012) の枠組みは、この点に対して自明な解答を持たないように思われる。加えて、彼らが指摘したのは Kullback-Leibler ダイバージェンスの最小値あるいは“疑似真値”の集合の不一致であるが、これは計量経済学で重要な真値に関する示唆を含まない。Freedman (2006) が強調したように、もし特定化過誤バイアスが深刻ならば、疑似真値の推論はさほど大きな意味を持たないかもしれない。Müller (2013) は特定化過誤下の頻度論のサンドイッチ形式を事後分散として人工的に用いることで頻度論性質を改善する方法を提案したが、これも疑似真値に焦点を当てたものである。バイアスが深刻ではない場合、疑似真値は真値と大きく乖離するわけではないため、真値に関する集合の性質の解明に繋がる可能性がある。計量経済学では潜在的な特定化過誤を部分的に許容しながらも、真値の推論に関する応用上有益な結果が期待されている。とりわけ Chernozhukov and Hong (2003), Norets (2015), Chae, Kim and Kleijn (2019) は部分的な特定化過誤を許容しながら真値の推論に関するポジティブな結果を導出した点で、この方向性と関連していると思われる。

本論文は弱い特定化過誤を伴う有限次元モデルにおける信用/信頼集合の一致の可能性について検討する。弱い特定化過誤として、我々は Hjort and Claeskens (2003) の局所特定化過誤の枠組みと、バイアスの収束レートを一般化した局所特定化過誤の枠組みに焦点を当てる。特に、局所特定化過誤は統計学や計量経済学で現れる弱い特定化過誤あるいはモデル不確実状況の近似として有用であり、集中情報量規準 (Claeskens and Hjort (2003), Zhang and Liang (2011) など) や頻度論モデル平均 (Hjort and Claeskens (2003), Liu (2015) など) の基礎である。事実、この枠組みは考慮に入れる最大のモデルに含まれる幾つかのパラメータが極めてゼロに近い状況の近似と見なせる。このとき、バイアスが少し出ることをあえて許容しつつ、ゼロに近いと思われるパラメータをモデルから除外することでより単純なモデルを選択することが正当化される。このような特定化過誤の弱い状況はとりわけ計量経済モデリングで典型的に現れる。さらに、我々はバイアスの収束レートが一般化された局所特定化過誤における信用/信頼集合の一致の可能性を検討する。この設定は最尤推定量のバイアスが標準誤差より遅いレートでゼロに収束する場合を含む。そこでの集合の一致は自明でないため、この設定の検討には明確な新規性がある。この枠組みは特定化過誤が存在しながらも真値に関する問題を考慮に入れ

られる十分に一般的な枠組みであり、疑似真値に焦点を当てた Kleijn and van der Vaart (2012) には含まれない問題の検討を可能にする。

本論文の貢献は、弱い特定化過誤下における真値の信用 / 信頼集合の一致に関する条件を明らかにしたことである。特に、多くの応用例を持つ局所特定化過誤を含む一般化された局所特定化過誤モデルがその重要な例であることを指摘する。我々はこの枠組みにおいて BvM 定理のバージョンが成立することを示し、バイアスの収束レートは漸近的な結果に影響しないことを指摘する。さらに、漸近的な情報行列等式が成立することを示し、この結果は疑似真値のための信用 / 信頼集合が一致することを示唆する。このとき、十分に標本サイズが大きければ、これらの集合は真値のための集合として解釈可能である。またモンテカルロ研究からは、小さなバイアスが原因でいずれの集合も真値の被覆に失敗する状況でさえ、信用 / 信頼集合は互いに近づくことが示唆される。この結果は、真値の被覆に関わらず、弱い特定化過誤下においていずれの立場の推論も似た結論を導く可能性を示唆する。さらに、この結果は特定化過誤が存在するにも関わらず信用集合に信頼集合と同程度の被覆性質を持つとの解釈を与えられること、そして信用 / 信頼集合の計算に互いの計算手法を代替させることの正当化になる。これらの結論は既存研究に含まれない新規的知見であり、ベイジアンと頻度論のいずれの立場にとっても有益なものとして期待される。

本論文の構成は以下である。第2章で一般化された局所特定化過誤の枠組みを提示し、ベイジアン-頻度論間の信用 / 信頼集合の一致の可能性を検討する。第3章でモンテカルロ研究のデザインと結果を示す。第4章で結論付ける。補論で仮定と定理の証明を提供する。

## 2. 弱い特定化過誤下の BvM 定理と情報行列等式

### (1) レビュー

本論文が基礎とする Kleijn and van der Vaart (2012) の枠組みと結果を以下で概観する。 $Y^n = (Y_1, \dots, Y_n)$  は真の密度関数  $f_{true}(y)$  に従う無作為標本とする。パラメータ空間  $\Theta$  を  $\mathbb{R}^k$  の適当な部分集合とし、 $\theta$  の関数として適切に滑らかなモデルを  $f_\theta(y)$  とする (詳細は Kleijn and van der Vaart (2012), van der Vaart (1998))。ここでは真の密度関数がモデルに含まれない、すなわち、 $f_{true} = f_\theta$  となる  $\theta \in \Theta$  が存在しないという意味での特定化過誤を考える。Kullback-Leibler ダイバージェンス  $KL_\theta = E_{true}[\log(f_{true}/f_\theta)]$  は有限として、これを最小にする  $\theta^* \in \Theta$  が一意に存在するとして、これを“疑似真値”と呼ぶ。ただし、 $E_{true}$  は  $f_{true}$  に関する期待値を表わす。 $\ell_\theta(y) = \log f_\theta(y)$  として、1階導関数のベクトル  $\dot{\ell}_\theta(y) = \frac{\partial}{\partial \theta} \ell_\theta(y)$  と2階導関数の行列  $\ddot{\ell}_\theta(y) = \frac{\partial^2}{\partial \theta \partial \theta^T} \ell_\theta(y)$  に対して、 $J_\theta = -E_{true}[\ddot{\ell}_\theta]$ 、 $I_\theta = E_{true}[\dot{\ell}_\theta \dot{\ell}_\theta^T]$  とする。

ベイジアンと頻度論者は共通モデル  $f_\theta$  を用いて、それぞれの方法に基づき  $\theta$  の推論を行う。ベイズ推論では、 $\Theta$  上の事前密度  $\pi$  が  $\theta^*$  の近くで連続かつ正值と仮定して、ベイズの定理に

に基づき、 $Y^n$ を固定した下での事後密度  $\pi_n(\cdot|Y^n)$  を考える。ここで事後密度に従う“確率パラメータ”を  $\vartheta$  とする。一方、頻度論では対数尤度関数を最大にする最尤推定量  $\widehat{\theta}_n \in \Theta$  が一意に存在するとして、 $Y^n$  を固定することなく、その標本分布を考える。このとき、Kleijn and van der Vaart (2012) の特定化過誤 BvM 定理に基づく事後分布と、最尤推定量の標本分布 (Huber (1967), White (1982)) は、 $n$  が大きいとき、それぞれ次で近似される：

$$\vartheta|Y^n \sim N_k(\widehat{\theta}_n, J_{\theta^n}^{-1}/n), \quad \widehat{\theta}_n \sim N_k(\theta^*, J_{\theta^*}^{-1} I_{\theta^*} J_{\theta^*}^{-1}/n). \quad (1)$$

ただし、 $N_d(\mu, \Sigma)$  は  $d$  変量正規分布で平均  $\mu$ 、分散共分散行列  $\Sigma$  とする。この結果は古典的な BvM 定理 (van der Vaart (1998) の定理 10.1, Ghosal and van der Vaart (2017) の定理 12.1) による示唆とは対照的に、いずれの多変量正規分布の分散共分散行列も Fisher 情報行列の逆行列ではなく、かつ互いに異なるものである。特に、 $\theta_j$  の区間推定について、それぞれの分布に基づく  $\theta_j$  の (対称な) 信用/信頼区間はいずれも最尤推定値を中心とするが、半径は異なる。従って、特定化過誤下では、信用区間と信頼区間の一致は一般に成立しない。ここで、事前分布の存在が集合の不一致をもたらすわけではないことに注意したい。特定化過誤下の有限次元モデルでは、事前分布の存在は漸近的なベイズ推論に本質的な役割を果たさない。言い換えれば、事前分布の選択により分散共分散行列を人工的に互いに一致させることは期待できない。

Kleijn and van der Vaart (2012) の集合の一致に関するネガティブな結果は、特定化過誤下の統計的推論の基礎である。彼らの結果の意義は、ベイジアン-頻度論間の推論でそれぞれの立場が見ている不確実性は根本的に異なることを改めて明確にした点にある。この結果において、古典的な BvM 定理と信用/信頼集合の一致は、特定化が正しいとする極めて理想的な仮定の下での特殊例であることが明確になる。しかし、彼らの研究は以下の二点については検討がなされていない：(1) 信用/信頼集合の不一致の指摘は疑似真値の集合に関するものであるが、真値の推論に関する示唆はあるのか、(2) 特定化過誤下においてなお集合が一致する重要なモデルは存在しないのか。特に、彼らの結果は特定化過誤の程度が“弱い”場合の真値の推論に関して自明な解答を含まないようであるが、この点は真値の推論に興味のある計量経済学において重要である。

## (2) 一般化された局所特定化過誤の枠組み

弱い特定化過誤として我々が焦点を当てるのは、以下に定義する一般化された局所特定化過誤の枠組み (generalized local misspecification framework, GLMF) である。我々が GLMF に焦点を当てる根拠は、モデル不確実性の基礎的枠組みである Hjort and Claeskens (2003) の局所特定化過誤 (LMF) を特殊な場合として含むこと、パラメータの真値で評価された漸近理論を考察できること、そしてバイアスの明示的な収束レートに着目できることが挙げられる。特に二つめは重要であり、この枠組みは疑似真値に焦点をあてた Kleijn and van der Vaart

(2012) で自明でなかった真値に関する性質の検討を可能にする。

無作為標本  $Y^n = (Y_1, \dots, Y_n)$  が次の分布から得られるとき、標本は GLMF に従うと定義する：

$$f_{true}(y) = f_{\theta_0}(y)\{1 + r(y)n^{-\alpha} + R(y, n^{-\alpha})\}. \quad (2)$$

この枠組みにおいて、モデルは  $\int f_{\theta}(y)dy = 1$  を満たす  $f_{\theta}$ 、興味あるパラメータは  $\theta_0 \in \Theta \subset \mathbb{R}^k$  とする。 $r(y)n^{-\alpha}$  はモデル特定化の誤りの度合いを決定する部分と解釈される。 $\alpha \geq \alpha_0 > 0$  は所与のバイアスコントロールパラメータであり、 $\alpha_0 < 1/2$  を許容する。一般には、 $r(y)$  には適当な条件を満たす未知関数を許容する。 $R(y, t)$  を  $o(\|t\|)$  を想定する近似の剰余項とする。 $Y_i = Y_{ni}$  は厳密には triangular array であるため、(2) は  $n$  に依存するが、以降で  $n$  は省略する。

データ生成過程 (2) に対して、ベイジアンと頻度論者は共通モデル  $f_{\theta}$  に基づき推論を行う。所与の  $\alpha$  と有限の  $n$  に対して  $f_{true} \notin \{f_{\theta} : \theta \in \Theta\}$  であるため、モデルは必ず特定化過誤である。 $n \rightarrow \infty$  につれて特定化過誤の度合いは小さくなるが、所与の  $\alpha$  が小さいほどその収束は遅い。Kullback-Leibler ダイバージェンス  $KL_{\theta, n} = E_{true}[\log(f_{true}/f_{\theta})]$  は有限として、その最小値  $\theta_n^*$  が一意に存在するとする。 $E_0$  を  $f_0 = f_{\theta_0}$  に基づく期待値とし、 $J_0 = -E_0[\ddot{\ell}_0] = -E_0[\ddot{\ell}_{\theta_0}]$  とする。

GLMF は漸的にモデルの特定化がほぼ正しいと見なせる設定として十分に一般的な枠組みである。一般に  $r(y)$  には未知関数が許容されるため、 $f_{\theta}$  と  $f_{true}$  は同じ分布族である必要はない。さらに、この枠組みは Hjort and Claeskens (2003) の 10.1 節で考えられた特定化過誤よりも一般的であり、 $\theta_0$  に対するモデリングバイアスは  $O(n^{-\alpha})$  が許容される。これは標準誤差よりもバイアスの収束レートが遅い場合である  $\alpha < 1/2$  を含む。GLMF の利点は真値とそれに対するバイアスの収束レートに着目できる点にあり、この点において Kleijn and van der Vaart (2012) の枠組みとは異なる。事実、Kleijn and van der Vaart (2012) は特定化過誤が弱い状況を扱っておらず、真値やバイアスの収束レートに関する議論を含まない。また、後述するように、もし  $f_{true}$  が 1 次元の局外パラメータ  $\gamma$  を含み、 $\ell_{\theta, \gamma} = \log f_{\theta, \gamma}$  をある  $(\theta_0, \gamma_0)$  で展開できるなら、(2) において  $\alpha = 1/2$ ,  $r(y) = \frac{\partial}{\partial \gamma} \ell_{\theta_0, \gamma_0}(y) \delta$  とする場合が LMF に相当する。この意味で GLMF は LMF の一般化である。一般の  $r(y)$  では (2) の  $\theta_0$  はもはや真値と解釈できないが、 $r(y) = \frac{\partial}{\partial \gamma} \ell_{\theta_0, \gamma_0}(y) \delta$  のときは  $\theta_0$  を真値と解釈できる。

もし  $\sum_{i=1}^n \ell_{\theta}$  を最大にする最尤推定量  $\widehat{\theta}_n$  が一意に存在するなら、適当な条件の下で、 $\widehat{\theta}_n$  は  $\theta_0$  に対して一致性を持つ。実際、もし  $\ell_{\theta}$  が凹関数なら  $\widehat{\theta}_n \rightarrow_{P_{true}} \theta_0$  を示せる (補論 A (3) の命題 1)。この結果は  $\alpha < 1/2$  の場合でさえ成立することに注意したい。次の BvM 定理のバージョンは、GLMF における  $\widehat{\theta}_n$  で中心化された  $\vartheta$  の事後分布の漸近正規性である。定理のために  $\widehat{\theta}_n \rightarrow_{P_{true}} \theta_0$ 、事前分布の連続性と正值性、そして  $\ell_{\theta}$  の一様性を仮定する。 $\|f\|_{TV} = \int |f(y)| dy$  を密度関数  $f$  に対する全変動ノルムとする。ただし、全変動距離に関する密度関数の収束は、通常の意味での分布収束を意味する (例えば van der Vaart (1998) の 2.9 節)。

**定理 1.**  $Y^n$  はある  $\alpha \geq \alpha_0 > 0$  に対して (2) に従うとする. A1-A7 を仮定する. このとき

$$\|\pi_n(\sqrt{n}(\vartheta - \widehat{\theta}_n) | Y^n) - N_k(0, J_0^{-1})\|_{TV} \rightarrow_{P_{true}} 0.$$

定理 1 の仮定と証明は補論 A (1) で提供される. 証明は本質的に Bickel and Doksum (2001) の定理 5.5.2 の証明ステップに従うものであるが, GLMF に合わせて修正されたものである. 特に, 全変動距離に関する確率収束の議論が追加されている. 定理 1 は  $\alpha = 1/2$  のとき LMF に相当する後述の結果 (系 1) を含む.

全変動ノルムは位置とスケールの変更に関して不変より, 定理 1 は次を意味する

$$\|\pi_n(\vartheta | Y^n) - N_k(\widehat{\theta}_n, J_0^{-1}/n)\|_{TV} \rightarrow_{P_{true}} 0.$$

よって,  $\vartheta$  の結合事後分布は多変量正規分布で近似され, その平均は (バイアスを含む) 最尤推定値, 分散共分散行列は  $J_0^{-1}/n$  である. Kleijn and van der Vaart (2012) の定理 2.1 は疑似真値で評価された分散共分散行列のみを導出したが, GLMF に焦点を当てた定理 1 は興味ある  $\theta_0$  で評価された分散共分散行列である. 重要なこととして定理 1 の結論は  $\alpha$  に依存しない. すなわち,  $\widehat{\theta}_n$  が  $\theta_0$  に対して一貫性を持つ限り, 定理はバイアスの収束レートは事後分布の漸近正規性に影響を与えないことを示唆する.

$\alpha < 1/2$  の GLMF ではバイアスが標準誤差より大きいいため,  $\theta_0$  で中心化された最尤推定量の標本分布を導出することは困難である. それにも関わらず, 対数尤度関数の最頻値に近い  $\theta_n^*$  は,  $n$  が大きくなるにつれて, 遅いレートで  $\theta_0$  へ近づく. 実際, もし  $\ell_\theta$  が凹関数なら  $\theta_n^* \rightarrow \theta_0$  を示せる (補論 A (3) の命題 2). さらに, 追加条件の下でその収束のレートは明示的に  $O(n^{-\alpha})$  であることを示せる (補論 A (3) の命題 3). このとき, 漸近的には  $\theta_0$  に対するバイアスは遅いながらも消えてゆき, “漸近的な情報行列等式” が成立する. 具体的には,  $\theta_n^* \rightarrow \theta_0$ , (2) の  $r, R$  に関する優越条件を仮定した下で, 次の定理が成り立つ.

**定理 2.**  $Y^n$  はある  $\alpha \geq \alpha_0 > 0$  に対して (2) に従うとする. A1-A3, A8-A14 を仮定する. このとき

$$-E_{true}[\ell_{\theta_n^*}^{\ddot{}}] = E_{true}[\ell_{\theta_n^*}^{\dot{}} \ell_{\theta_n^*}^{\dot{T}}] + o(1).$$

定理 2 の仮定と証明は補論 A (2) で提供される.

定理 2 は,  $n$  が大きいとき,  $\theta_n^*$  で評価された  $J_{\theta_n^*}$  を  $I_{\theta_n^*}$  で置き換えることを正当化する. これより, 最尤推定量のサンドイッチ形式の分散共分散行列は  $J_{\theta_n^*}^{-1} I_{\theta_n^*} J_{\theta_n^*}^{-1} / n \approx I_{\theta_n^*}^{-1} / n$  と近似される. このとき (1) に照らせば, 最尤推定量の漸近正規性の条件 (例えば Kleijn and van der Vaart (2012)) の下で, 定理 1 と定理 2 は有限の  $n$  における分布に関して, 次の近似を意味する

$$\vartheta | Y^n \sim N_k(\widehat{\theta}_n, I_{\theta_n^*}^{-1}/n), \quad \widehat{\theta}_n \sim N_k(\theta_n^*, I_{\theta_n^*}^{-1}/n).$$

このとき、事後分布と標本分布のそれぞれの平均は互いに“逆転”したものとなり、いずれの分散共分散行列も近い状況が実現する。したがって、 $\theta_n^*$ が $\theta_0$ に対して持つバイアスは小さくならないにも関わらず、GLMFの下で、 $\theta_n^*$ に関する信用/信頼集合は漸近的に近づくことが示唆される。応用上の示唆として、これは信用集合の計算にブートストラップ等を用いた信頼集合を代用すること、そして、信頼集合の計算にMCMCを用いた信用集合を代用することを正当化する。

定理1と定理2は $\theta_n^*$ のみならず $\theta_0$ に関する信用/信頼集合の漸近的な一致をも示唆する点で重要である。実際、 $\theta_n^*$ が $\theta_0$ の近傍にいくらかでも近づくならば、 $\theta_n^*$ を中心とする集合は $\theta_0$ を中心とする集合として解釈できる。このとき、十分大きな $n$ において、信用/信頼集合のいずれも $\theta_0$ を中心とする集合として近づくことが示唆される。この結論はGLMFに焦点を当てることで明確になる点である。しかし、 $\alpha < 1/2$ の場合、バイアスが標準誤差を凌駕するため、所与の名目水準を達成するような真値の被覆性質はおそらく期待できない。これは、 $n$ が大きくなるにつれ、先に標準誤差が小さくなりつつ遅れてバイアスがゼロに近づくためである。この状況における $\theta_0$ の被覆性質と集合間の一致可能性をより詳しく見るために、後述するモンテカルロ研究が助けになる。

### (3) 局所特定化過誤の枠組み

ここでGLMFの重要な例としてLMFに焦点を当てる。LMFは弱い特定化過誤あるいはモデル不確実性の基礎的枠組みとして多くの応用例を持つことに加え、バイアスを直接推定することで真値の被覆を期待できる信用/信頼集合を明示的に構成できるため有用である。特に、この枠組みは集中情報量規準 (Claeskens and Hjort (2003), Zhang and Liang (2011) など) や頻度論モデル平均 (Hjort and Claeskens (2003), Liu (2015) など) の基礎として重要である。例えば、LMFは考慮に入れる最大のモデルに含まれる幾つかのパラメータが極めてゼロに近い状況の近似と見なせる。このときバイアスが少し出ることをあえて許容しつつ、ゼロに近いパラメータをモデルから除外することでより単純なモデルを選択することが正当化される。実際、LMFで生ずるバイアスは標準誤差と同じ $O(1/\sqrt{n})$ ゆえ、 $\theta_0$ で中心化された最尤推定量の漸近分布を導出することが可能となり、平均二乗誤差を最小にするモデルを選択できる (Claeskens and Hjort (2003))。このような弱い特定化過誤的状況は実際のモデリングで典型的に現れる。

基本的なLMFは無作為標本 $Y^n = (Y_1, \dots, Y_n)$ が次の分布から得られるものと定義される：

$$f_{true}(y) = f(y; \theta_0, \gamma_0 + \delta/\sqrt{n}). \quad (3)$$

ただし、 $\theta_0$ は興味ある $k$ 次元パラメータ、 $\gamma_0$ は1次元の既知定数、 $\delta$ は未知パラメータとする。パラメータの真値は $(\theta_0, \gamma_0 + \delta/\sqrt{n})$ である。 $\gamma$ を無視して $\theta$ のみを推定する場合、モデルは $\gamma = \gamma_0$ とした $f_{\theta, \gamma_0}(y) = f(y; \theta, \gamma_0)$ であり、有限の $n$ で $f_{true} \notin \{f_{\theta, \gamma_0} : \theta \in \Theta\}$ である。ただし、モデリン



グバイアスが  $O(1/\sqrt{n})$  という意味で局所的な特定化過誤である。

LMF は GLMF の特殊な場合に相当する。ここで  $\ell_{\theta,y}(\mathbf{y}) = \log f_{\theta,y}(\mathbf{y})$  を  $(\theta_0, \gamma_0)$  で展開できると仮定する。  $\dot{\ell}_0 = \frac{\partial}{\partial \theta} \ell_{\theta_0, \gamma_0}, \dot{\ell}_{\gamma,0} = \frac{\partial}{\partial \gamma} \ell_{\theta_0, \gamma_0}$  として、  $I_0 = E_0[\dot{\ell}_0 \dot{\ell}_0^T], I_{\gamma,0} = E_0[\dot{\ell}_0 \dot{\ell}_{\gamma,0}^T]$  とする。このとき、  $f_{true}$  はある  $R(\mathbf{y}, t)$  について

$$f_{true}(\mathbf{y}) = f_0(\mathbf{y})\{1 + \dot{\ell}_{\gamma,0}(\mathbf{y})\delta n^{-1/2} + R(\mathbf{y}, n^{-1/2})\}$$

と書き直せるが、これは GLMF の  $\alpha = 1/2, r(\mathbf{y}) = \dot{\ell}_{\gamma,0}(\mathbf{y})\delta$  とした場合である。

Hjort and Claeskens (2003) の補題 3.2 より、(3) と正則条件の下で、 $\theta_0$  で中心化された  $\theta$  の最尤推定量の分布は次で近似される

$$\sqrt{n}(\widehat{\theta}_n - \theta_0) \sim N_k(I_0^{-1}I_{\gamma,0}\delta, I_0^{-1}).$$

特に、これは  $\widehat{\theta}_n \xrightarrow{P_{true}} \theta_0$  を意味する。また、追加条件の下で  $\theta_n^* = \theta_0 + I_0^{-1}I_{\gamma,0}\delta/\sqrt{n} + o(n^{-1/2})$  を示せるため、最尤推定量の分布は  $\widehat{\theta}_n \sim N_k(\theta_n^*, I_0^{-1}/n)$  と見なせる。このとき漸近分散は  $I_0^{-1}/n$  であり、サンドイッチ形式ではない。実際、このとき定理 2 と同様に漸近的に情報行列等式が成立し、特に  $J_0 = I_0$  である。したがって、この結果は  $\theta_n^*$  の信用 / 信頼集合の漸近的な一致の可能性を示唆する。これを確認するベイジアンの結果が、次に示す定理 1 の系である。

**系 1.**  $Y^n$  は (3) から生成されたとする。  $r(\mathbf{y}) = \dot{\ell}_{\gamma,0}(\mathbf{y})\delta$  に対して A1-A7 を仮定する。このとき

$$\|\pi_n(\sqrt{n}(\vartheta - \widehat{\theta}_n)|Y^n) - N_k(0, I_0^{-1})\|_{TV} \xrightarrow{P_{true}} 0.$$

Hjort and Claeskens (2003) の仮定 C1-C4 の下で、A2, A7 は自動的に満たされる。よって、系の仮定は制約的ではなく、典型的な LMF において成立する。

全変動ノルムの性質より、系 1 は  $\|\pi_n(\vartheta|Y^n) - N_k(\widehat{\theta}_n, I_0^{-1}/n)\|_{TV} \xrightarrow{P_{true}} 0$  を意味する。したがって、 $\vartheta$  の結合事後分布は多変量正規分布で近似され、平均は局所バイアスを含む最尤推定値、分散共分散行列は  $I_0^{-1}/n$  である。以上の漸近的な結果を合わせれば、  $j = 1, \dots, k$  に対して、  $(\theta_n^*)_j$  の  $(1 - \alpha)$ -信用 / 信頼区間は一致し、ともに  $(\widehat{\theta}_n)_j \pm z_{\alpha/2} \sqrt{(I_0^{-1})_{jj}/n}$  である。ただし、  $z_\alpha$  を標準正規分布の  $100\alpha\%$  分位点とする。

LMF は真値に対する区間の一致が明示的に示される点において重要である。前述の結果より漸近的に  $\theta_n^* \rightarrow \theta_0$  ゆえ、疑似真値を基に構成された区間は真値の被覆を期待できる区間として互いに近づく。すなわち、この結果を真値に関する区間として信用 / 信頼区間が一致すると解釈できる。さらに、LMF では局所バイアスを推定可能なため、より明示的に真値の区間を考えることもできる。これを見るため、ここで  $I_0^{-1}I_{\gamma,0}\delta$  を推定されたものとみなす (推定方法は Hjort and Claeskens (2003)。ただし、 $\delta$  は不偏推定できるが一致推定できない)。前述の結果より、バイアス修正された最尤推定量の分布は  $\widehat{\theta}_n - I_0^{-1}I_{\gamma,0}\delta/\sqrt{n} \sim N_k(\theta_0, I_0^{-1}/n)$  で近似される。

このとき、 $(\theta_0)_j$ の信頼区間は

$$(\widehat{\theta}_n)_j - (I_0^{-1}I_{\gamma,0}\delta)_j/\sqrt{n} \pm z_{\alpha/2}\sqrt{(I_0^{-1})_{jj}/n}.$$

一方、 $\theta_0$ の推定のため、推定可能な項でバイアス修正された $\vartheta$ の事後分布は、系1と全変動ノルムの性質より、次で近似される

$$\vartheta - I_0^{-1}I_{\gamma,0}\delta/\sqrt{n} \sim N_k(\widehat{\theta}_n - I_0^{-1}I_{\gamma,0}\delta/\sqrt{n}, I_0^{-1}/n).$$

いま $\widehat{\theta}_n - I_0^{-1}I_{\gamma,0}\delta/\sqrt{n} \rightarrow_{P_{true}} \theta_0$ より、修正された $\vartheta$ の事後分布は確かに $\theta_0$ を中心とした分布である。また、 $Y^n$ を固定した下で $\widehat{\theta}_n - I_0^{-1}I_{\gamma,0}\delta/\sqrt{n}$ は計算可能である。結果として、解析に基づく $(\theta_0)_j$ の信用区間は $(\widehat{\theta}_n)_j - (I_0^{-1}I_{\gamma,0}\delta)_j/\sqrt{n} \pm z_{\alpha/2}\sqrt{(I_0^{-1})_{jj}/n}$ であり、信頼区間と再び一致する。

### 3. モンテカルロ研究

GLMFにおける信用/信頼集合の真値の被覆確率と集合の一致可能性を検証するため、モンテカルロ研究を行う。計量経済学でポピュラーな正規線形回帰に基づくデータ生成過程 (data generating process, DGP) を考える：

$$DGP1: Y_i \sim N(\theta_0 + \theta_1 X_{1i} + \theta_2 X_{2i} + \delta_1 n^{-\alpha} X_{3i}, \sigma^2).$$

$$DGP2: Y_i \sim N(\theta_0 + \theta_1 X_{1i} + \theta_2 X_{2i} + \delta_1 n^{-\alpha} X_{3i}, \sigma^2 + \delta_2 n^{-\alpha} X_{3i}^2).$$

ただし、 $X_j \sim N(1,1), (j=1,2,3)$ ,  $X_j$ 間の共分散を1/2とする。 $\theta = (\theta_0, \theta_1, \theta_2)^T = (1,1,1)^T$ ,  $\delta = (\delta_1, \delta_2)^T = (1,1)^T$ ,  $\sigma^2 = 1$ とする。実験に用いるサンプルサイズは $n \in \{20,50,100,1000\}$ , バイアスコントロールは $\alpha \in \{1,1/2,1/4\}$ とする。

これらのDGPに対して、変数に $X_i = (1, X_{1i}, X_{2i})^T$ のみを用いた均一分散の回帰モデル $Y_i \sim N(\theta_0 + \theta_1 X_{1i} + \theta_2 X_{2i}, \sigma^2)$ を想定する。このとき、DGP1は回帰関数に関する特定化過誤、DGP2は回帰関数に加えて $X_3$ の実現値に依存する不均一分散に関する特定化過誤である。 $\alpha = 1$ は特定化過誤が極めて弱い場合、 $\alpha = 1/2$ はLMF、 $\alpha = 1/4$ は特定化過誤が強い場合に相当する。

Kleijn and van der Vaart (2012)に従い、興味ある有限次元パラメータを $\theta_1$ とする。 $\theta_1$ の95%信用区間 (CrI) として、 $\theta, \tau = \sigma^2$ の正規-ガンマ事前分布 $\pi(\theta, \tau) = N(\theta | \underline{\Omega}, \tau^{-1} \underline{\Omega}^{-1}) \text{Gamma}(\tau | \underline{a}, \underline{b})$ に基づく対称な信用区間を以下とする (Bernardo and Smith (1994))：

$$\text{CrI: } (\theta_{post})_1 \pm t_{0.025}(2a_n) \sqrt{a_n^{-1} b_n \left( \underline{\Omega} + \sum_{i=1}^n X_i X_i^T \right)_{11}^{-1}},$$

ただし、 $t_a(k)$ を自由度 $k$ の標準 $t$ 分布の100 $a$ %分位点とし、

$$\theta_{post} = \left( \underline{\Omega} + \sum_{i=1}^n X_i X_i^T \right)^{-1} \left( \underline{\Omega} \underline{\theta} + \sum_{i=1}^n X_i Y_i \right),$$

$$a_n = \underline{a} + n/2, \quad b_n = \underline{b} + \sum_{i=1}^n (Y_i - X_i^T \theta_{post}) Y_i / 2 + (\underline{\theta} - \theta_{post})^T \underline{\Omega} \underline{\theta} / 2.$$

ここで、 $\theta_{post}$ は解析に基づく $\theta$ の事後平均である。ハイパーパラメータは $\underline{\theta} = (0, 0, 0)$ 、 $\underline{\Omega} = \text{diag}(1, 1, 1)/100$ 、 $\underline{a} = 1$ 、 $\underline{b} = 1$ とする。異なるハイパーパラメータに基づく場合と代替的な事前分布として漠然事前分布  $\pi(\theta, \tau) \propto 1 \times \tau^{-1}$  を用いる場合を検討したが<sup>5</sup>、結果の傾向に違いはほぼ見られなかったため省略する。通常、信用区間の構成には最高事後密度区間等が用いられるが、この設定では事後分布の近似的な正規性が期待されるため対称な信用区間に焦点を当てる。一方、 $\theta_1$ の95%信頼区間(CI)として、Kleijn and van der Vaart (2012)に基づき、最尤推定量  $\widehat{\theta}_n = (\sum_{i=1}^n X_i X_i^T)^{-1} \sum_{i=1}^n X_i Y_i$  (最小二乗推定量) とサンドイッチ形式の標準誤差に基づく以下とする：

$$\text{CI: } (\widehat{\theta}_n)_1 \pm z_{0.025} \sqrt{\widehat{\text{Var}}(\widehat{\theta}_n)_{11}/n},$$

ただし、 $\widehat{\text{Var}}(\widehat{\theta}_n) = \left( \frac{1}{n} \sum_{i=1}^n X_i X_i^T \right)^{-1} \left( \frac{1}{n} \sum_{i=1}^n (Y_i - X_i^T \widehat{\theta}_n)^2 X_i X_i^T \right) \left( \frac{1}{n} \sum_{i=1}^n X_i X_i^T \right)^{-1}$ 。ここで、 $\widehat{\text{Var}}(\widehat{\theta}_n)$ は不均一分散ロバスト推定量(White (1980))と関連するが<sup>6</sup>、これはあらゆる特定化過誤を想定した下で導かれたサンドイッチ形式であり、不均一分散を考慮するために用いるわけではない。

このモンテカルロ研究では、各設定に対して仮想的な繰り返し標本における真値  $\theta_1 = 1$  の被覆確率と区間の長さの中央値(Len)を報告する。繰り返し回数は5000とする。表1、表2はそれぞれDGP1、DGP2に対する被覆確率と区間の長さを報告したものである。

まずDGP1の結果について、 $\alpha = 1$ なら、 $n$ が大きいほどに信用区間の被覆確率は0.95に近づいており、通常のBvM定理と整合的な結果を確認できる。この結果は95%信用区間が頻度論の意味の95%で真値を被覆することを意味する。このとき、 $n$ が大きいほど二つの区間の被覆確率と長さが近いため、区間の境界そのものが互いに近いことが示唆される。LMF( $\alpha = 1/2$ )でも似た傾向を見てとれて、系1により示唆された真値を中心とした区間の一致の可能性を確認できる。被覆確率は0.95よりやや低い、これはおそらく標準誤差と同程度のバイアスの存在によりしばしば真値を被覆できないためであると思われる。 $\alpha = 1/4$ では、いずれの区間の被覆確率も名目水準を著しく下回ることが分かる。 $n$ が大きいほど、むしろ真値の被覆に失敗する傾向がある。これはおそらく、標準誤差よりバイアスの収束が遅いことで真値から離れた領域に分布が集中するためであると思われる。重要な点は、 $n$ が大きいほどに二つの区間の被覆確率は同じ値に近づき、区間の長さも近いことである。この結果は、漸近的な区間の一致が $\alpha$ に依存せずに成立することを示唆した定理1と定理2と整合的である。つまり、無視できないバイアスにより二つの区間は真値の被覆に頻繁に失敗しつつも、最終的な区間推定

表 1 : DGP1 における 95 % 信用区間 (CrI) と 95 % 信頼区間 (CI) の被覆確率と長さの中央値

n	$\alpha = 1$				$\alpha = 1/2$				$\alpha = 1/4$			
	CrI	Len	CI	Len	CrI	Len	CI	Len	CrI	Len	CI	Len
20	0.934	1.058	0.877	0.902	0.923	1.072	0.859	0.920	0.895	1.122	0.834	0.973
50	0.943	0.648	0.920	0.610	0.940	0.651	0.914	0.613	0.882	0.675	0.855	0.640
100	0.946	0.457	0.936	0.441	0.932	0.458	0.923	0.445	0.857	0.471	0.840	0.456
1000	0.950	0.143	0.950	0.143	0.942	0.143	0.941	0.143	0.653	0.145	0.648	0.144

表 2 : DGP2 における 95 % 信用区間 (CrI) と 95 % 信頼区間 (CI) の被覆確率と長さの中央値

n	$\alpha = 1$				$\alpha = 1/2$				$\alpha = 1/4$			
	CrI	Len	CI	Len	CrI	Len	CI	Len	CrI	Len	CI	Len
20	0.930	1.101	0.869	0.954	0.924	1.243	0.875	1.089	0.903	1.455	0.846	1.284
50	0.941	0.664	0.922	0.623	0.930	0.736	0.915	0.698	0.908	0.863	0.894	0.823
100	0.943	0.460	0.936	0.446	0.942	0.501	0.932	0.487	0.880	0.590	0.872	0.577
1000	0.949	0.143	0.948	0.143	0.944	0.148	0.942	0.148	0.717	0.168	0.722	0.169

は互いに似る可能性が示唆される。このとき、信用区間には信頼区間と同じ被覆確率が認められる点は興味深い。真値の推論に失敗するときでさえ、信用区間には信頼区間と同程度の（頻度論的）信頼性が保たれることが示唆される。

DGP2の結果においても同様の傾向を見てとれて、無視できないバイアスをもたらす $\alpha = 1/4$ でさえ、ベイジアンと頻度論は似た推論を導く可能性を確認できる。このときDGP1に比べて被覆性質がやや改善するよう見えるが、これはおそらく信頼区間が用いたサンドイッチ形式が不均一分散ロバスト推定量と関連することが部分的な要因であると思われる。

図1は $[0,1]$ を20分割する $\alpha$ に対する被覆確率のプロットである。上段がDGP1、下段がDGP2、それぞれ左から $n = 20, 100$ の場合である。DGP1のいずれの $n$ の場合も、 $\alpha$ に応じて連続的かつほぼ単調的に被覆確率が変化する傾向を見てとれる。 $n = 20$ なら、事前分布の影響等のため二つの区間の被覆確率に乖離が見られるが、 $n = 100$ なら被覆確率は $\alpha$ に一樣に近づいている。このとき、 $\alpha \geq 1/2$ なら被覆確率はいずれの区間も0.95に近いが、特に $\alpha \leq 1/4$ なら被覆確率はいずれの区間も著しく名目水準を下回る。それにも関わらず、 $n$ が大きくなれば二つの区間の性質は $\alpha$ に一樣に近いことが示唆される。再び、この結果は $\alpha$ に依存せずに区間の一致を示唆した定理1と定理2と整合的である。DGP2も同様の結論が示唆される。

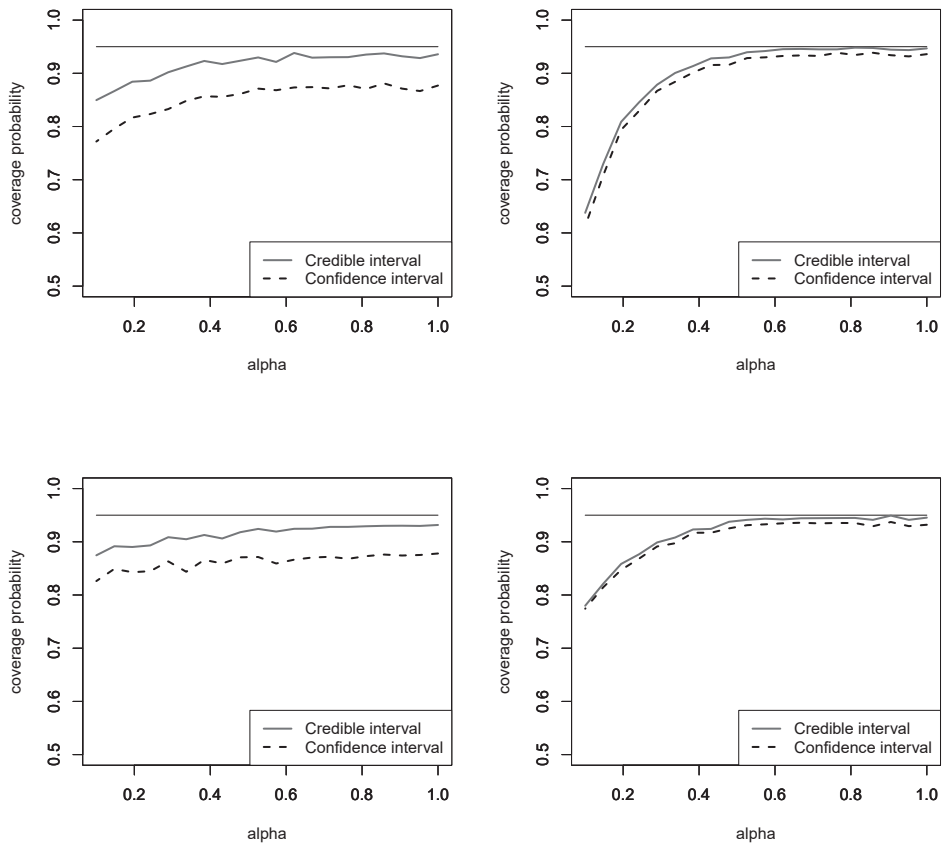


図1：95%信用区間 (CrI) と95%信頼区間 (CI) の真値の被覆確率。横軸は $\alpha$ ，縦軸は被覆確率。上段はDGP1，下段はDGP2，左は $n = 20$ ，右は $n = 100$ 。水平の黒線は名目水準95%。

#### 4. 結論

本論文は一般化された局所特定化過誤モデルにおける真値の信用 / 信頼集合の一致の可能性を検証した。特に、近似的に情報行列等式が成立するとき、バイアスは残ったままで信用 / 信頼集合は近づくことが明らかにされた。このとき、バイアスの収束レートの遅さは集合の一致の結論に強い影響を及ぼさない。モンテカルロ研究からは、いずれの集合も真値の被覆に失敗する場合でさえ、集合は互いに近づくことが示唆された。この結果は、特定化過誤が弱い状況において、信用集合に信頼集合と同程度の被覆性質を持つとの解釈を与えられることを示唆する。さらにこの結果は、真値の被覆性質に関わらず、信用集合の計算にブートストラップを用いる、あるいは信頼集合の計算にMCMCを用いることの正当化になる。この結論はベイジアンと頻度論のいずれの立場にとっても有益なもの期待される。

## A. 補論

補論では定理1と定理2の仮定と証明, そして一貫性と収束レートに関する3つの命題を示す. 以下では, 記述の簡略化のため  $k=1$  の下で仮定を置くが, 固定された一般の  $k < \infty$  に対する仮定の下で, 全く同様に示すことができる. 以下のすべての確率収束  $\rightarrow_p$  は,  $P_{true}$  に関する収束とする.  $a_n = O_p(b_n)$  を, 任意の  $\epsilon > 0$  に対してある  $M > 0$  が存在して  $P(|a_n| > M|b_n|) < \epsilon$  ( $P$  の下での確率有界),  $a_n = o_p(b_n)$  を  $a_n/b_n \rightarrow_p 0$  とする.

### (1) 定理1の仮定と証明

A1.  $\Theta \subset \mathbb{R}$  は凸集合として,  $\theta_0$  は  $\Theta$  の内部に含まれるとする. 任意の  $\theta \neq \theta_0$  に対して  $f(y, \theta) \neq f(y, \theta_0)$  を満たすとする.

A2.  $\ell_\theta(y) = \log f_\theta(y)$  は  $\theta_0$  の近傍で2回連続微分可能とし,  $E_0[\ddot{\ell}_0] < \infty$  とする.

A3.  $E_0[\ddot{\ell}_0 r] < \infty$ ,  $E_0[\ddot{\ell}_0 R(\cdot, t)] = o(|t|)$ .

A4. 任意の  $\gamma > 0$  に対して,  $\epsilon(\gamma) > 0$  が存在して,  $n \rightarrow \infty$  につれて,

$$P_{true} \left( \sup_{|\theta - \theta_0| \geq \gamma} \frac{1}{n} \sum_{i=1}^n (\ell_\theta(Y_i) - \ell_0(Y_i)) \leq -\epsilon(\gamma) \right) \rightarrow 1.$$

A5. 任意の  $\epsilon_n \rightarrow 0$  に対して,  $\sup_{|\theta - \theta_0| \leq \epsilon_n} \left| \frac{1}{n} \sum_{i=1}^n \ddot{\ell}_\theta(Y_i) - \frac{1}{n} \sum_{i=1}^n \ddot{\ell}_0(Y_i) \right| \rightarrow_{P_{true}} 0$ .

A6. 事前密度  $\pi(\theta)$  は任意の  $\theta \in \Theta$  で連続で,  $\theta_0$  において正値とする.

A7.  $n \rightarrow \infty$  につれて,  $\widehat{\theta}_n \rightarrow_{P_{true}} \theta_0$ .

仮定の妥当性に関して言及する. A3はGLMFにおけるHjort and Claeskens (2003) のC2, C3に相当する. A4は $\theta_0$ の識別可能性と関連する条件である. もし $\theta_0$ における平均対数尤度が十分大きな $n$ で高確率で一意的に高いなら条件は満たされる. A5は対数尤度関数の $\theta_0$ の近傍での一様連続性に相当する. A6はBvM定理に必要な仮定であり, 特に事前密度は $\theta_0$ の近傍で必ず正値でなければならない. もし事前密度が $\Theta$ の広い範囲に正値を与える確率密度なら仮定は正当化される. ただし, 事前密度は既知のパラメトリックな確率分布である必要はない. A7はハイレベルであるが, 例えば対数尤度関数が凹関数ならGLMFにおいて正当化される(命題1). よく知られるように, 対数尤度の凹性は線形回帰やプロビットモデルを含む計量経済学の多くの例で成立する (Hayashi (2000), Newey and McFadden (1994) など).

**定理1の証明.** 証明の簡略化のため  $k=1$  で示すが, 一般の固定された  $k < \infty$  で同様に示すことができる. 証明はBickel and Doksum (2001) の定理5.5.2の証明と同ステップを辿るものである. 中心化とスケールングを適用した  $t = \sqrt{n}(\vartheta - \widehat{\theta}_n)$  の事後密度は, 密度の変数変換公式より

$$q_n(t) = c_n^{-1} \pi(\widehat{\theta}_n + t/\sqrt{n}) \prod_{i=1}^n f_{\widehat{\theta}_n + t/\sqrt{n}}(Y_i),$$

ただし、 $c_n = \int \pi(\widehat{\theta}_n + s/\sqrt{n}) \prod_{i=1}^n f_{\widehat{\theta}_n + s/\sqrt{n}}(Y_i) ds$  である。この分母と分子を  $t$  に依らない  $\prod_{i=1}^n f_{\widehat{\theta}_n}(Y_i)$  で割ることで、 $q_n$  は以下で書き直せる：

$$q_n(t) = d_n^{-1} \pi(\widehat{\theta}_n + t/\sqrt{n}) \exp\left(\sum_{i=1}^n (\ell_{\widehat{\theta}_n + t/\sqrt{n}}(Y_i) - \ell_{\widehat{\theta}_n}(Y_i))\right),$$

ただし、 $d_n = \int \pi(\widehat{\theta}_n + s/\sqrt{n}) \exp\left(\sum_{i=1}^n (\ell_{\widehat{\theta}_n + s/\sqrt{n}}(Y_i) - \ell_{\widehat{\theta}_n}(Y_i))\right) ds$  である。全変動収束を示すために、まず密度関数の各点収束として

$$q_n(t) \rightarrow_p q(t) = \frac{\sqrt{J_0}}{\sqrt{2\pi}} \exp\left(-\frac{t^2 J_0}{2}\right), \forall t. \tag{4}$$

を示す。このためには以下を示せば十分である

$$d_n q_n(t) \rightarrow_p \pi(\theta_0) \exp\left(-\frac{t^2 J_0}{2}\right), \forall t, \tag{5}$$

$$d_n \rightarrow_p \pi(\theta_0) \int \exp\left(-\frac{s^2 J_0}{2}\right) ds. \tag{6}$$

(5) 式のため、まず  $\ell_\theta$  の展開より、 $|\bar{\theta}(t) - \widehat{\theta}_n| \leq t/\sqrt{n}$  となる  $\bar{\theta}(t) \in \Theta$  が存在して、次が成り立つ

$$\sum_{i=1}^n \{\ell_{\widehat{\theta}_n + t/\sqrt{n}}(Y_i) - \ell_{\widehat{\theta}_n}(Y_i)\} = \frac{t^2}{2} \frac{1}{n} \sum_{i=1}^n \ddot{\ell}_{\bar{\theta}(t)}(Y_i).$$

ここで  $\widehat{\theta}_n$  の定義より  $\sum_{i=1}^n \dot{\ell}_{\widehat{\theta}_n}(Y_i) = 0$  を用いた。このとき、 $t$  に対して一様に

$$\frac{1}{n} \sum_{i=1}^n \ddot{\ell}_{\bar{\theta}(t)}(Y_i) \rightarrow_p E_0[\ddot{\ell}_0] = -J_0 \tag{7}$$

が成り立つ。実際

$$\begin{aligned} & \frac{1}{n} \sum_{i=1}^n \ddot{\ell}_{\bar{\theta}(t)}(Y_i) - E_0[\ddot{\ell}_0] \\ &= \left\{ \frac{1}{n} \sum_{i=1}^n \ddot{\ell}_{\bar{\theta}(t)}(Y_i) - \frac{1}{n} \sum_{i=1}^n \ddot{\ell}_0(Y_i) \right\} + \left\{ \frac{1}{n} \sum_{i=1}^n \ddot{\ell}_0(Y_i) - E_{true}[\ddot{\ell}_0] \right\} + \{E_{true}[\ddot{\ell}_0] - E_0[\ddot{\ell}_0]\}. \end{aligned}$$

と分解すると、まず第2項は (triangular array に対する) 大数の法則より  $o_p(1)$ 、そして第3項は (2) と A3 より  $E_0[\ddot{\ell}_0 r] n^{-\alpha} + E_0[\ddot{\ell}_0 R(\cdot, n^{-\alpha})] = O(n^{-\alpha}) + o(n^{-\alpha}) = o(1)$  である。よって、分解の第1項について  $t$  に一様に以下を示せば、(7) 式に十分である：

$$\sup_{|t| \leq M} \left| \frac{1}{n} \sum_{i=1}^n \ddot{\ell}_{\bar{\theta}(t)}(Y_i) - \frac{1}{n} \sum_{i=1}^n \ddot{\ell}_0(Y_i) \right| \rightarrow_p 0, \forall M.$$

このために、任意の  $M > 0$  に対して  $\epsilon_n = M/\sqrt{n} + |\widehat{\theta}_n - \theta_0|$  とおけば、A7 より  $\epsilon_n \rightarrow_p 0$ 、 $t$  に一様に  $|\bar{\theta}(t) - \theta_0| \leq \epsilon_n$  である。このとき明らかに  $\{t: |t| \leq M\} \subset \{t: |\bar{\theta}(t) - \theta_0| \leq \epsilon_n\}$  であるため、任意の  $M > 0$  に対して、前式の左辺は  $\sup_{|\bar{\theta}(t) - \theta_0| \leq \epsilon_n} \left| \frac{1}{n} \sum_{i=1}^n \ddot{\ell}_{\bar{\theta}(t)}(Y_i) - \frac{1}{n} \sum_{i=1}^n \ddot{\ell}_0(Y_i) \right|$  で優越される。ところが

A5よりこれはゼロに確率収束する。したがって (7) 式が従う。さらに, A6, A7より, 任意の  $M > 0$  に対して  $\sup_{|t| \leq M} |\pi(\widehat{\theta}_n + t/\sqrt{n}) - \pi(\theta_0)| \rightarrow_p 0$  ゆえに, 特に任意の  $t$  に対して  $\pi(\widehat{\theta}_n + t/\sqrt{n}) \rightarrow_p \pi(\theta_0)$  である。以上の結果より, (5) 式を得る。

次に (6) 式のため,  $\gamma > 0$  に対して,  $d_n = \int_{|s| \leq \gamma\sqrt{n}} d_n q_n(s) ds + \int_{|s| > \gamma\sqrt{n}} d_n q_n(s) ds$  と分解する。第2項の積分について, まず  $s = \sqrt{n}(t - \widehat{\theta}_n)$  で変数変換しつつ, A4 と  $\widehat{\theta}_n \rightarrow_p \theta_0$  より,  $\epsilon(\gamma) > 0$  が存在して, 確率が1に近づく事象として (w.p.a.1), 次の不等式が従うことに注意する

$$\sup_{|t - \widehat{\theta}_n| > \gamma} \exp\left(\sum_{i=1}^n (\ell_t(Y_i) - \ell_{\widehat{\theta}_n}(Y_i))\right) \leq \exp(-n\epsilon(\gamma)).$$

これより, 分解の第2項は任意の  $\gamma > 0$  に対して, w.p.a.1 で

$$\sqrt{n} \int \pi(t) \exp\left(\sum_{i=1}^n (\ell_t(Y_i) - \ell_{\widehat{\theta}_n}(Y_i))\right) 1[|t - \widehat{\theta}_n| > \gamma] dt \leq \sqrt{n} \exp(-n\epsilon(\gamma))$$

で優越される。この右辺は明らかにゼロに収束するため, 分解の第2項はゼロに確率収束する。分解の第1項については, まず Bickel and Doksum (2001) の定理 5.5.2 の証明と同様にすることで, 十分小さい  $\bar{\gamma} > 0$  が存在して,  $|t| \leq \bar{\gamma}\sqrt{n}$  となる任意の  $t$  に対して, w.p.a.1 で

$$\pi(\widehat{\theta}_n + t/\sqrt{n}) \exp\left(\frac{t^2}{2} \frac{1}{n} \sum_{i=1}^n \ddot{\ell}_{\widehat{\theta}_n(t)}(Y_i)\right) 1[|t| \leq \bar{\gamma}\sqrt{n}] \leq 2\pi(\theta_0) \exp\left(-\frac{t^2 J_0}{8}\right)$$

で優越される。ただし, 右辺は明らかに  $t$  に関して積分可能である。このとき, 分解の第1項について, (5) 式の結果から

$$\int d_n q_n(s) 1[|s| \leq \bar{\gamma}\sqrt{n}] ds \rightarrow_p \int \pi(\theta_0) \exp\left(-\frac{s^2 J_0}{2}\right) ds, \quad (8)$$

を示せる。実際, Markov の不等式より

$$\begin{aligned} P_{true} \left( \left| \int d_n q_n(s) 1[|s| \leq \bar{\gamma}\sqrt{n}] ds - \int \pi(\theta_0) \exp\left(-\frac{s^2 J_0}{2}\right) ds \right| > \epsilon \right) \\ \leq \frac{1}{\epsilon} \int \left| \int d_n q_n(s) 1[|s| \leq \bar{\gamma}\sqrt{n}] - \pi(\theta_0) \exp\left(-\frac{s^2 J_0}{2}\right) \right| dP_{true} ds \end{aligned}$$

であるが, 右辺の重積分がゼロに収束することを言えば, (8) 式に十分である。このために, まず直前の結果から, 重積分の被積分関数は任意の  $n$  に対して非確率的な  $b(s) = 4\pi(\theta_0) \exp(-s^2 J_0/8)$  で優越されることに注意する。ここで不等式  $\exp(-x^2) \leq \exp(-x^2/4)$  を用いた。よって, (5) 式に優収束定理を適用することで, 任意の  $s$  に対して

$$\int \left| d_n q_n(s) 1[|s| \leq \bar{\gamma}\sqrt{n}] - \pi(\theta_0) \exp\left(-\frac{s^2 J_0}{2}\right) \right| dP_{true} \rightarrow 0.$$

さらに, 外側の積分の被積分関数も  $b(s)$  で優越されることから, 再び優収束定理により, Markov 不等式の右辺はゼロに収束する。これで (8) 式が従うため, 以上の結果より (6) 式が従う。



最後に、密度関数の各点収束 (4) は確率収束の下での全変動収束を意味することを示す。まず全変動距離について  $\int |q_n(t) - q(t)| dt = 2 \int [q(t) - q_n(t)]^+ dt$  が成り立つことに注意する。ただし  $[f]^+ = \max\{0, f\}$  は  $f$  の正部分を表わす。(4) 式は任意の  $t$  について  $[q(t) - q_n(t)]^+ \rightarrow_p 0$ , a.s. で任意の  $n$  について  $[q(t) - q_n(t)]^+ \leq q(t)$  を意味する。したがって、優収束定理を適用して、任意の  $t$  に対して  $\int [q(t) - q_n(t)]^+ dP_{true} \rightarrow 0$ 。また、明らかに  $\int [q(t) - q_n(t)]^+ dP_{true} \leq q(t)$  より、再び優収束定理より、 $\iint [q(t) - q_n(t)]^+ dP_{true} dt \rightarrow 0$ 。したがって、任意の  $\epsilon > 0$  に対して、Markov の不等式とこれまでの結果から

$$P_{true} \left( \int [q(t) - q_n(t)]^+ dt > \epsilon \right) \leq \frac{1}{\epsilon} \int \int [q(t) - q_n(t)]^+ dP_{true} dt \rightarrow 0.$$

これは  $\int [q(t) - q_n(t)]^+ dt \rightarrow_p 0$  を意味するため、確率収束の下での全変動収束が従う。□

(2) 定理 2 の仮定と証明

A8. Kullback-Leibler ダイバージェンス  $KL_{\theta,n} = E_{true}[\log(f_{true}/f_\theta)]$  は任意の  $\theta \in \Theta$  で有限で、 $\theta \in \Theta$  に関する最小値  $\theta_n^* \in \Theta$  が一意に存在する。

A9.  $\ell_\theta(y) = \log f_\theta(y)$  は  $\theta_0$  の近傍で 3 回連続微分可能とする。 $f_0(y)$  の下で積分可能な関数  $\varphi_1(y), \varphi_2(y), \varphi_3(y)$  が存在し、任意の  $\theta \in \Theta$  で  $|\ell_\theta^2(y)| \leq \varphi_1(y)$ ,  $|\ddot{\ell}_\theta(y)| \leq \varphi_2(y)$ ,  $|\ell_\theta^{(3)}(y)| \leq \varphi_3(y)$ 。

A10.  $f_0(y)$  の下で積分可能な関数  $\bar{\varphi}_1(y), \bar{\varphi}_2(y), \bar{\varphi}_3(y)$  が存在して、任意の  $\theta \in \Theta$  で  $|\ell_\theta^2(y)r(y)| \leq \bar{\varphi}_1(y)$ ,  $|\ddot{\ell}_\theta(y)r(y)| \leq \bar{\varphi}_2(y)$ ,  $|\ell_\theta^{(3)}(y)r(y)| \leq \bar{\varphi}_3(y)$ 。

A11.  $E_0[\dot{\ell}_0 r] < \infty$ ,  $E_0[\dot{\ell}_0 R(\cdot, t)] = o(|t|)$ ,  $E_0[\dot{\ell}_0^2 R(\cdot, t)] = o(|t|)$ ,  $E_0[\ell_0^{(3)} R(\cdot, t)] = o(|t|)$ 。

A12. 任意の  $\epsilon_n \rightarrow 0$  に対して

$$E_0 \left[ \sup_{|\theta - \theta_0| < \epsilon_n} |\dot{\ell}_\theta^2 - \dot{\ell}_0^2| |R(\cdot, t)| \right] = o(|t|), \quad E_0 \left[ \sup_{|\theta - \theta_0| < \epsilon_n} |\ddot{\ell}_\theta - \ddot{\ell}_0| |R(\cdot, t)| \right] = o(|t|),$$

$$E_0 \left[ \sup_{|\theta - \theta_0| < \epsilon_n} |\ell_\theta^{(3)} - \ell_0^{(3)}| |R(\cdot, t)| \right] = o(|t|).$$

A13. 任意の  $\theta \in \Theta$  で  $\int \frac{\partial}{\partial \theta} f_\theta(y) dy = \frac{\partial}{\partial \theta} \int f_\theta(y) dy$ ,  $\int \frac{\partial^2}{\partial \theta^2} f_\theta(y) dy = \frac{\partial^2}{\partial \theta^2} \int f_\theta(y) dy$ 。

A14.  $n \rightarrow \infty$  につれて、 $\theta_n^* \rightarrow \theta_0$ 。

A9, A13 は最尤推定量の漸近正規性において典型的に仮定される。A10 は Hjort and Claeskens (2003) の C4 をやや強めた条件であるが、優収束定理を適用するために十分である。A12 は  $\ell_\theta, \dot{\ell}_\theta, \ell_\theta^{(3)}$  の局所的な一様連続性に関連するが、 $R(\cdot, t) = o(|t|)$  を想定する下では制約的な条件ではない。A14 はハイレベルであるが、A7 と同様に対数尤度の凹性の下で正当化される (命題 2 と命題 3)。

**定理 2 の証明.** 任意の  $\theta \in \Theta$  に対して、次の等式が成り立つことに注意する

$$\ddot{\ell}_\theta(y) = f_\theta^{(2)}(y) \frac{1}{f_\theta(y)} - \left(\dot{\ell}_\theta(y)\right)^2. \quad (9)$$

この両辺を  $f_{true}$  により期待値をとることで、任意の  $\theta \in \Theta$  に対して以下が成り立つ

$$E_{true}[\ddot{\ell}_\theta] = E_{true}\left[f_\theta^{(2)} \frac{1}{f_\theta}\right] - E_{true}\left[\dot{\ell}_\theta^2\right].$$

したがって、 $\theta = \theta_n^*$  で評価した右辺第1項が  $o(1)$  を示せば結果に十分である。このために、まず

$$\int f_{\theta_n^*}^{(2)} \frac{f_{true}}{f_{\theta_n^*}} dy = \int f_{\theta_n^*}^{(2)} \frac{f_0}{f_{\theta_n^*}} dy + n^{-\alpha} \int f_{\theta_n^*}^{(2)} \frac{f_0}{f_{\theta_n^*}} r dy + \int f_{\theta_n^*}^{(2)} \frac{f_0}{f_{\theta_n^*}} R(\cdot, n^{-\alpha}) dy. \quad (10)$$

であるが、右辺の全ての項が  $o(1)$  であることを示す。(10) 式の第1項は、A14と  $f_\theta(y)$ ,  $f_\theta^{(2)}(y)$  の  $\theta$  に関する連続性より、 $y$  の関数列として  $f_{\theta_n^*}^{(2)}(y)/f_{\theta_n^*}(y)$  は  $f_0^{(2)}(y)/f_0(y)$  に各点収束する。さらに、再び等式 (9) より、任意の  $\theta$  に対して

$$\left| f_\theta^{(2)}(y) \frac{1}{f_\theta(y)} \right| \leq \dot{\ell}_\theta^2(y) + |\ddot{\ell}_\theta(y)| \leq \varphi_1(y) + \varphi_2(y),$$

であるが、A9より  $\varphi_1, \varphi_2$  は  $f_0$  の下で積分可能である。したがって、優収束定理より

$$\int f_{\theta_n^*}^{(2)} \frac{f_0}{f_{\theta_n^*}} dy \rightarrow \int f_0^{(2)} dy = 0.$$

この最後の等号で A13 を用いた。したがって、(10) 式の第1項は  $o(1)$ 。(10) 式の第2項の積分は、等式 (9) と A10 より、優収束定理を2度適用することで

$$E_0 \left[ f_{\theta_n^*}^{(2)} \frac{r}{f_{\theta_n^*}} \right] = E_0[\ddot{\ell}_{\theta_n^*} r] + E_0[\dot{\ell}_{\theta_n^*}^2 r] \rightarrow E_0[\ddot{\ell}_0 r] + E_0[\dot{\ell}_0^2 r] < \infty.$$

よって、(10) 式の第2項は  $O(n^{-\alpha}) = o(1)$ 。(10) 式の第3項は、A14より、任意の  $\delta > 0$  に対して十分大きな  $n$  で  $\theta_n^* \in \{\theta \in \Theta: |\theta - \theta_0| < \delta\}$  より、等式 (9) より

$$\begin{aligned} & \left| E_0 \left[ f_{\theta_n^*}^{(2)} \frac{R(\cdot, n^{-\alpha})}{f_{\theta_n^*}} \right] - E_0 \left[ f_0^{(2)} \frac{R(\cdot, n^{-\alpha})}{f_0} \right] \right| \\ & \leq E_0 \left[ \sup_{|\theta - \theta_0| < \delta} |\ddot{\ell}_\theta - \ddot{\ell}_0| |R(\cdot, n^{-\alpha})| \right] + E_0 \left[ \sup_{|\theta - \theta_0| < \delta} |\dot{\ell}_\theta^2 - \dot{\ell}_0^2| |R(\cdot, n^{-\alpha})| \right] \end{aligned}$$

であるが、この右辺は A12 より  $o(1)$ 。さらに、(9)、A3、A11より、 $E_0 \left[ f_0^{(2)} \frac{R(\cdot, n^{-\alpha})}{f_0} \right] = E_0[\ddot{\ell}_0 R(\cdot, n^{-\alpha})] + E_0[\dot{\ell}_0^2 R(\cdot, n^{-\alpha})] = o(n^{-\alpha}) = o(1)$ 。よって (10) 式の第3項は  $o(1)$ 。したがって結果を得る。□

### (3) 一貫性と収束レートに関する命題

ハイレベルの仮定 A7 と A14 の十分条件として、対数尤度関数の凹性の下での最尤推定量の一貫性 (命題1) と疑似真値の真値への収束 (命題2) を示す。また疑似真値の真値への収束レートが明示的に  $O(n^{-\alpha})$  となるための十分条件を示す (命題3)。 $\widehat{\theta}_n$  の定義と A1, A8 の下で、 $\widehat{\theta}_n$ ,

$\theta_n^*, \theta_0$  は  $\widehat{\theta}_n = \operatorname{argmax}_{\theta \in \Theta} \frac{1}{n} \sum_{i=1}^n \ell_\theta(Y_i)$ ,  $\theta_n^* = \operatorname{argmax}_{\theta \in \Theta} E_{true}[\ell_\theta]$ ,  $\theta_0 = \operatorname{argmax}_{\theta \in \Theta} E_0[\ell_\theta]$  を満たす。ただし所与の仮定の下でそれぞれ一意の最大値である。ここで、A15として  $\ell_\theta$  の凹性とそれに伴う正則条件を仮定する。この仮定は線形回帰やプロビットモデルなどの計量経済学の多くの例で成り立つ (Hayashi (2000), Newey and McFadden (1994) など)。

**A15.**  $\ell_\theta(y) = \log f_\theta(y)$  は  $\theta$  について凹関数とする。任意の  $\theta \in \Theta$  に対して  $E_0[\ell_\theta r] < \infty$ ,  $E_0[\ell_\theta R(\cdot, t)] = o(|t|)$  とする。

**命題 1.**  $Y^n$  が (2) に従うとする。A1, A15 を仮定する。このとき、 $\widehat{\theta}_n \rightarrow_{P_{true}} \theta_0$ 。

**証明.** Newey and McFadden (1994) の定理 2.7 の仮定を満たすことを示す。A1 より  $\Theta$  は凸集合、A15 より  $\frac{1}{n} \sum_{i=1}^n \ell_\theta(Y_i), E_0[\ell_\theta]$  はいずれも  $\theta$  に関して凹関数である。残す条件である目的関数の各点収束については、まず  $\frac{1}{n} \sum_{i=1}^n \ell_\theta(Y_i) - E_0[\ell_\theta] = \left( \frac{1}{n} \sum_{i=1}^n \ell_\theta(Y_i) - E_{true}[\ell_\theta] \right) + (E_{true}[\ell_\theta] - E_0[\ell_\theta])$  と分解すると、第 1 項は (triangular array に対する) 大数の法則より  $o_p(1)$ 。第 2 項は、(2) と A15 より  $E_0[\ell_\theta r] n^{-\alpha} + E_0[\ell_\theta R(\cdot, n^{-\alpha})] = O(n^{-\alpha}) + o(n^{-\alpha}) = o(1)$ 。したがって、任意の  $\theta \in \Theta$  に対して  $\frac{1}{n} \sum_{i=1}^n \ell_\theta(Y_i) - E_0[\ell_\theta] = o_p(1)$ 。これで定理の仮定が満たされるため、一致性が従う。□

**命題 2.**  $Y^n$  が (2) に従うとする。A1, A8, A15 を仮定する。このとき、 $\theta_n^* \rightarrow \theta_0$ 。

**証明.** 命題 1 と全く同様に示せるため省略する。□

**命題 3.**  $Y^n$  が (2) に従うとする。命題 2 の仮定に加え、A3, A9, A10, A11, A12 を仮定する。このとき、 $\theta_n^* = \theta_0 + (-E_0[\ddot{\ell}_0])^{-1} E_0[\dot{\ell}_0 r] n^{-\alpha} + o(n^{-\alpha})$ 。

**証明.**  $\theta_n^*$  の定義と  $\dot{\ell}_\theta$  の  $\theta = \theta_0$  での展開より、 $\bar{\theta}_n \in (\theta_n^*, \theta_0)$  が存在して、 $\theta_n^*$  は次を満たす

$$0 = \int \dot{\ell}_{\theta_n^*} f_{true} dy = \int \left( \dot{\ell}_0 + (\theta_n^* - \theta_0) \ddot{\ell}_0 + \frac{(\theta_n^* - \theta_0)^2}{2} \ell_{\bar{\theta}_n}^{(3)} \right) f_{true} dy.$$

これより、次の等式を得る

$$-(\theta_n^* - \theta_0) \left\{ \int \ddot{\ell}_0 f_{true} dy + \frac{\theta_n^* - \theta_0}{2} \int \ell_{\bar{\theta}_n}^{(3)} f_{true} dy \right\} = \int \dot{\ell}_0 f_{true} dy. \quad (11)$$

ここで、左辺中括弧の第 1 項は、A3 より  $E_0[\ddot{\ell}_0] + E_0[\ddot{\ell}_0 r] n^{-\alpha} + E_0[\ddot{\ell}_0 R(\cdot, n^{-\alpha})] = E_0[\ddot{\ell}_0] + o(1)$ 。中括弧第 2 項の積分については、次を示せる：

$$E_0 \left[ \ell_{\bar{\theta}_n}^{(3)} \right] + E_0 \left[ \ell_{\bar{\theta}_n}^{(3)} r \right] n^{-\alpha} + E_0 \left[ \ell_{\bar{\theta}_n}^{(3)} R(\cdot, n^{-\alpha}) \right] = E_0 \left[ \ell_0^{(3)} \right] + O(n^{-\alpha}) + o(1).$$

実際、命題 2 と優収束定理を適用することで、A9 より第 1 項は  $E_0 \left[ \ell_{\bar{\theta}_n}^{(3)} \right] \rightarrow E_0 \left[ \ell_0^{(3)} \right]$ 、また、A10 より第 2 項は  $E_0 \left[ \ell_{\bar{\theta}_n}^{(3)} r \right] \rightarrow E_0 \left[ \ell_0^{(3)} r \right] < \infty$  である。第 3 項は、命題 2 より、任意の  $\delta > 0$  に対して、十分大きな  $n$  に対して  $\theta_n^* \in \{\theta \in \Theta : |\theta - \theta_0| < \delta\}$  ゆえ

$$\left| E_0 \left[ \ell_{\hat{\theta}_n}^{(3)} R(\cdot, n^{-\alpha}) \right] - E_0 \left[ \ell_0^{(3)} R(\cdot, n^{-\alpha}) \right] \right| \leq E_0 \left[ \sup_{|\theta - \theta_0| < \delta} \left| \ell_\theta^{(3)} - \ell_0^{(3)} \right| |R(\cdot, n^{-\alpha})| \right]$$

であるが, A12 より右辺は  $o(1)$ . さらに, A11 より  $E_0 \left[ \ell_0^{(3)} R(\cdot, n^{-\alpha}) \right] = o(1)$ . したがって, 命題 2 より中括弧第 2 項は  $o(1)$ . 一方, (11) 式の右辺は A11 より  $E_0[\dot{\ell}_0 r]n^{-\alpha} + E_0[\dot{\ell}_0 R(\cdot, n^{-\alpha})] = E_0[\dot{\ell}_0 r]n^{-\alpha} + o(n^{-\alpha})$ . したがって, 結果を得る.  $\square$

#### 参考文献

- Bernardo, J. M. and A. F. M. Smith. (1994). *Bayesian Theory*, John Wiley and Sons Ltd.
- Bernstein, S. N. (1927). *The Theory of Probabilities* (First edition), State Publishing House. Moscow, Leningrad. (Russian).
- Bickel, P. J. and B. J. K. Kleijn. (2012). "The semiparametric Bernstein-von Mises theorem," *The Annals of Statistics*, 40(1), 206–237.
- Bickel, P. J. and K. A. Doksum. (2001). *Mathematical Statistics: Basic Concepts and Selected Ideas*, Vol. I, Prentice-Hall.
- Bontemps, D. (2011). "Bernstein-von Mises theorems for Gaussian regression with increasing number of regressors," *The Annals of Statistics*, 39(5), 2557–2584.
- Castillo, I. (2012). "A semiparametric Bernstein-von Mises theorem for Gaussian process priors," *Probability Theory and Related Fields*, 152, 53–99.
- Castillo, I. and J. Rousseau. (2015). "A Bernstein-von Mises theorem for smooth functionals in semiparametric models," *The Annals of Statistics*, 43(6), 2353–2383.
- Castillo, I. and R. Nickl. (2013). "Nonparametric Bernstein-von Mises theorems in Gaussian white noise," *The Annals of Statistics*, 41 (4), 1999–2028.
- Castillo, I. and R. Nickl. (2014). "On the Bernstein-von Mises phenomenon for nonparametric Bayes procedures," *The Annals of Statistics*, 42 (5), 1941–1969.
- Chae, M., Y. Kim and B. J. K. Kleijn. (2019). "The semi-parametric Bernstein-von Mises theorem for regression models with symmetric errors," *Statistica Sinica*, 29, 1465–1487.
- Chernozhukov, V. and H. Hong. (2003). "An MCMC approach to classical estimation," *Journal of Econometrics*, 115(2), 293–346.
- Claeskens, G. and N. L. Hjort. (2003). "The Focused Information Criterion," *Journal of the American Statistical Association*, 98, 900–916.
- Cox, D. D. (1993). "An analysis of Bayesian inference for nonparametric regression," *The Annals of Statistics*, 21(2), 903–923.
- Freedman, D. A. (1999). "On the Bernstein-Von Mises Theorem with Infinite-Dimensional Parameters," *The Annals of Statistics*, 27(4), 1119–1140.
- Freedman, D. A. (2006). "On The So-Called "Huber Sandwich Estimator" and "Robust Standard Errors"," *The American Statistician*, 60(4), 299–302.
- Geweke, J. (2005). *Contemporary Bayesian Econometrics and Statistics*, John Wiley and Sons, Ltd.
- Ghosal, S. and A. W. van der Vaart. (2017). *Fundamentals of Nonparametric Bayesian Inference*, Cambridge University Press.
- Hartigan, J. A. (1983). *Bayes Theory*, Springer, New York.
- Hayashi, F. (2000). *Econometrics*, Princeton: Princeton University Press.
- Hjort, N. L. and G. Claeskens. (2003). "Frequentist model average estimators," *Journal of the American Statistical Association*, 98(464), 879–899.

- Huber, P. J. (1967). "The behavior of maximum likelihood estimates under nonstandard conditions," in: *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability*, Vol. 1: Statistics, University of California Press, Berkeley, 221-233.
- Kato, K. (2013). "Quasi-Bayesian analysis of nonparametric instrumental variables models," *The Annals of Statistics*, 41(5), 2359-2390.
- Kim, Y. (2006). "The Bernstein-von Mises theorem for the proportional hazard model," *The Annals of Statistics*, 34(4), 1678-1700.
- Kim, Y. and J. Lee. (2004). "A Bernstein-von Mises theorem in the nonparametric right censoring model," *The Annals of Statistics*, 32(4), 1492-1512.
- Kitagawa, T. J. L. Montiel Olea, J. Payne, A. Velez. (2020). "Posterior distribution of nondifferentiable functions," *Journal of Econometrics*, 217(1), 161-175.
- Kleijn, B. J. K. and A. W. van der Vaart. (2012). "The Bernstein-Von Mises theorem under misspecification," *Electronic Journal of Statistics*, 6, 354-381.
- Knapik, B. T., A. W. van der Vaart, J. H. van Zanten. (2011). "Bayesian inverse problems with Gaussian priors," *The Annals of Statistics*, 39(5), 2626-2657.
- Le Cam, L. and G. L. Yang. (1990). *Asymptotics in Statistics: Some Basic Concepts*, 2nd edition (2000). Springer, New York.
- Leahu, H. (2011). "On the Bernstein-von Mises phenomenon in the Gaussian white noise model," *Electronic Journal of Statistics*, 5, 373-404.
- Liu, C. (2015). "Distribution theory of the least squares averaging estimator," *Journal of Econometrics*, 186(1), 142-159.
- Müller, U. K. (2013). "Risk of Bayesian inference in misspecified models, and the sandwich covariance matrix," *Econometrica*, 81(5), 1805-1849.
- Newey, W. K. and D. McFadden. (1994). "Large sample estimation and hypothesis testing," *Handbook of Econometrics*, Vol. 4, Chapter 36, 2111-2245.
- Norets, A. (2015). "Bayesian regression with nonparametric heteroskedasticity," *Journal of Econometrics*, 185(2), 409-419.
- Rivoirard, V. and J. Rousseau. (2012). "Bernstein-von Mises theorem for linear functionals of the density," *The Annals of Statistics*, 40(3), 1489-1523.
- Shen, X. (2002). "Asymptotic normality of semiparametric and nonparametric posterior distributions," *Journal of the American Statistical Association*, 97(457), 222-235.
- Szabó, B., A. W. van der Vaart and J. H. van Zanten. (2015). "Frequentist coverage of adaptive nonparametric Bayesian credible sets," *The Annals of Statistics*, 43(4), 1391-1428.
- van der Vaart, A. W. (1998). *Asymptotic Statistics*, Cambridge University Press.
- von Mises, R. (1919). "Fundamentalsätze der Wahrscheinlichkeitsrechnung," *Mathematische Zeitschrift*, 4, 1-97.
- White, H. (1980). "A heteroskedasticity-consistent covariance matrix estimator and a direct test for heteroskedasticity," *Econometrica*, 48(4), 817-838.
- White, H. (1982). "Maximum likelihood estimation of misspecified models," *Econometrica*, 50(1), 1-25.
- Zhang, X. and H. Liang. (2011). "Focused information criterion and model averaging for generalized additive partial linear models," *The Annals of Statistics*, 39(1), 174-200.

# Equivalence of credible and confidence sets in generalized local misspecification models

Arihiro YOSHIMURA

## Abstract

This paper considers the possibility of equivalence between credible and confidence sets of parameters in smooth finite-dimensional models under weak misspecification. Using the generalized version of the local misspecification framework proposed by Hjort and Claeskens (2003), we present a version of the Bernstein-von Mises theorem under standard prior assumptions. We show that the asymptotic information matrix equality hold and the credible and confidence sets approach each other asymptotically even when the bias of the maximum likelihood estimator is larger than the standard error. Monte Carlo studies suggest that the credible and confidence sets approach each other even when both sets fail to cover the true value because of the bias. The results justify the interpretation that credible sets have the same coverage properties as confidence sets when the misspecification is weak. Furthermore, the results justify substituting the numerical method of confidence sets to compute credible sets, and vice versa. This conclusion is beneficial from Bayesian and frequentist points of view. This study's novel findings are not included in the existing literature and are expected to contribute to the development of statistical inference under model uncertainty.

**Keywords:** Bernstein-von Mises theorem, generalized local misspecification framework, Bayesian inference, credible/confidence set, information matrix equality