京都産業大学経済学レビュー　No.10　（令和５年３月）

# モデル不確実性下の線形回帰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

本論文はモデル不確実性下の線形回帰における回帰係数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) により提案された局所特定化過誤の枠組みを一般化した回帰モデルにおいて、標準的な事前分布の仮定の下での Bernstein-von Mises 定理のバージョンを示す。さらに漸近的な情報行列等式を示し、最小二乗推定量のバイアスが標準誤差より大きい場合でさえ、回帰係数の信用集合と信頼集合は互いに近づくことを示す。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状況でさえ、集合は互いに近づく可能性が示唆される。この結果は信用集合に信頼集合と同程度の長期的な被覆性質を持つとする解釈に正当化を与える。さらにこの結果は信用/信頼集合の計算に互いの計算手法を代替することの根拠を与え、特にブートストラップ信頼集合による信用集合の代用の妥当性を示唆する。


キーワード：Bernstein-von Mises 定理、局所特定化過誤の枠組み、信用集合、ブートストラップ推測、情報行列等式

## 1. 序論

近年の計量経済学や統計学において Bernstein-von Mises 定理（以後、BvM 定理）の拡張研究に興味が持たれている。Bernstein (1927) と von Mises (1919) らによる古典的な定理によれば、有限次元パラメータを伴う滑らかなモデルにおいて、パラメータの結合事後分布は標本サイズが大きくなるにつれて多変量正規分布に近づき、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（van der Vaart (1998) など）。定理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を意味する。このとき正則条件を満たす事前分布の存在は二つの立場の推論の違いに本質的な影響をもたらさない。定理により、信用集合に

は信頼集合と同等の長期的な被覆性質を持つとする解釈が与えられるほか、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が与えられる。これにより、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）による信用集合が信頼集合を代用すること（Chernozhukov and Hong (2003)）、さらにブートストラップによる信頼集合が信用集合を代用することが正当化される（Kitagawa, Montiel Olea, Payne and Velez (2020)）。ベイジアンと頻度論の立場には基本的な哲学の違いが存在しながらも、集合間の一致の可能性はいずれの立場にとっても有益であり、定理の拡張研究に関心が持たれている。

近年の BvM 定理の進展においては、無限次元モデルへの拡張とそこでの集合の一致の成立条件に大きな関心が持たれてきた。特に、興味あるパラメータが無限次元の場合、信用/信頼集合の一致が一般には成立しないことが Cox (1993)、Freedman (1999) 等によって指摘された。このネガティブな結果は定理の拡張と、ベイジアンと頻度論の関連に関する基本的な限界を示唆するかのように思われる。一方で、適切な事前分布の想定の下では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であるセミパラメトリックモデルを中心に、いくつものポジティブな結果が報告されている（Leahu (2011)、Knapik, van der Vaart and van Zanten (2011)、Bontemps (2011)、Bickel and Kleijn (2012)、Castillo (2012)、Rivoirard and Rousseau (2012)、Kato (2013)、Castillo and Nickl (2013, 2014)、Castillo and Rousseau (2015)、Szabó, van der Vaart and van Zanten (2015)、Norets (2015)、Chae, Kim and Kleijn (2019) など）。定理の無限次元モデルへの拡張に関する近年の展開は Ghosal and van der Vaart (2017) に整理されている。今なおこの分野においては活発な研究が続けられている。

パラメータの次元と関連して、信用集合と信頼集合の不一致をもたらす重要な要因が、モデルの特定化過誤である。事実、モデルの特定化過誤の下では、有限次元モデルにおいてさえ、パラメータの事後分布の分散共分散行列は、最尤推定量の“サンドイッチ形式”の分散共分散行列（Huber (1967)、White (1982)）と一致しない。この事実はテキストレベルにおいて古くから認識されていたが（Hartigan (1983)、Geweke (2005) など）、近年になり Kleijn and van der Vaart (2012)、Müller (2013) が再び取りあげた。特に前者は、ベイジアンと頻度論にとって共通の基礎的枠組みにおいて、特定化過誤 BvM 定理を示し、一般に信用集合は信頼集合に一致しないことを指摘した。このような集合の不一致は事前分布の影響が漸近的に消えていく状況でさえ生じることに注意したい。彼らの結果は、ひとたびモデル特定化を誤れば、信用集合と信頼集合の関連性は大きく失われうることを意味する。複雑現象のモデリングでは分布や関数形の特定化を誤る脅威は普遍的に存在するため、この点は非実験データを扱う計量経済学において特に重要であるように思われる。ノン-セミパラメトリックモデルがしばしば特定化過誤バイアスの回避のために適用されることを考慮すれば、特定化過誤はパラメータの次元とも密接に関連する点である。

信用集合と信頼集合の一致の可能性に関して明らかでないのは、モデル不確実性

（model uncertainty）との関連である。例えば、回帰モデルのデータ生成過程が分析者にとって未知なる部分を含む場合、最終的にいかなる回帰モデルを用いるべきかについては確実的ではない（Draper (1995)、Chatfield (1995) など）。モデル不確実性に対する典型的なアプローチの一つは、深刻なバイアスの回避のためにコントロール変数や相互作用項あるいは非線形項等を追加することで、大きなモデルを想定することである。ところが一方で、過度に大きすぎるモデルを想定すると解釈が複雑になることに加えて、分散の著しい増加をもたらす。このとき、大きすぎるモデルは回帰係数の推定のために必ずしも最適なモデルとは限らない（Claeskens and Hjort (2008))。実際、バイアス-分散のトレードオフにより、極めてゼロに近い係数の変数をあえて除外した単純なモデルは、それらを含めるような大きなモデルに比べて、より正確な推定をもたらしうる。この事実は、モデル不確実性下の回帰において、小さいバイアスあるいは“弱い”特定化過誤をあえて許容するモデリングを正当化する。ところが、この状況下における信用/信頼集合の一致の可能性については、既存研究では十分に明らかにされていないようである。Kleijn and van der Vaart (2012) は特定化過誤下の BvM 定理に関する基礎を確立したものの、特定化の誤りが深刻な場合のみを検討している。Müller (2013) は頻度論のサンドイッチ形式を事後分散に用いる方法を提示したが、やはり特定化の誤りが深刻な場合を検討している。Norets (2015)、Chae, Kim and Kleijn (2019) は部分的な特定化過誤を許容した回帰モデルにおいてセミパラメトリック BvM 定理を示したが、モデル不確実性や弱い特定化過誤を検討してはいない。もしバイアスが深刻でないならば、集合間の不一致の度合いは幾らか緩和される可能性があるが、そのための枠組みと条件は既存研究では明確にされていない。

本論文はモデル不確実性下の線形回帰モデルにおける回帰係数の信用/信頼集合の一致の可能性を検討する。このために本論文は Hjort and Claeskens (2003) の提案した局所特定化過誤の枠組みと、そのバイアスの収束レートを一般化した枠組みに基づく回帰モデルに焦点を当てる。特に、局所特定化過誤の枠組みはモデル不確実的状況の近似として集中情報量規準（Claeskens and Hjort (2003)）や頻度論モデル平均（Hjort and Claeskens (2003)）の基礎であり、回帰モデルを含む多くの応用例を持つ（Claeskens and Hjort (2008)、Zhang and Liang (2011)、Hansen (2014)、Liu (2015) など）。実際、この枠組みに基づく回帰モデルは想定する最大のモデルに含まれる幾つかの係数がゼロに近い状況の近似と見なせる。ところがモデリングバイアスは高々 $O(1/\sqrt{n})$ であるため、バイアス-分散のトレードオフにより、バイアスが出ることをあえて許容しつつ、係数の小さい変数を除外した特定化過誤モデルの選択が正当化される。このような状況はモデル不確実性に直面する実際の回帰モデリングにおいて典型的に現れる。さらに、本論文は局所特定化過誤の枠組みを一般化した回帰モデルを検討する。この設定は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が、そこでの信用/信頼集合の一致可能性は自明ではない。特に、この枠組みは真値の被覆可能性を検討するのに十分一般的な枠組みであり、深刻な特定化過誤下の“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。加えて、本論文はモンテカルロ研究により有限標本における信用集合と信頼集合の真値の被覆性質および集合の一致可能性を検証する。とりわけブートストラップ信頼集合による信用集合の代用の可能性は重要であるが、その近似精度については自明ではない。著者の知る限り、特定化過誤下の事後分布のブートストラップ近似の可能性については既存研究で明らかにされていない。

本論文の貢献はモデル不確実性下の回帰における信用/信頼集合の一致の十分条件を明らかにしたことである。一般化された局所特定化過誤に基づく回帰において、本論文は BvM 定理のバージョンと漸近的な情報行列等式を示す。特に、この結果は局所特定化過誤による近似が正当化される多くの線形回帰において回帰係数の集合の一致を示唆するものである。この結論はベイジアンが誤差項の分布に関して特定化を誤る場合でさえ漸近的には影響を受けない。さらに、結果はバイアスの収束レートが標準誤差より遅い場合であっても、標本サイズが大きい限り、いずれの集合も互いに近づくことを示唆する。モンテカルロ研究からは、信用/信頼集合のいずれもが真値の被覆に失敗する状況でさえ、いずれの集合も互いに近づくことが示唆される。このとき信用集合には信頼集合と同程度の信頼度による被覆性質を期待できるとともに、信用/信頼集合の計算に互いの計算手法を代用することが正当化される。とりわけブートストラップ信頼集合は信用集合の高い精度での近似を与えうることが示唆される。これらの発見は既存研究で指摘されていない新たな知見であり、ベイジアンと頻度論のいずれの立場にとっても有益であると期待される。

本論文の構成は以下である。第 2 章でモデル不確実性下の回帰の枠組みを提示し、信用/信頼集合の一致に関する理論を提示する。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。補論で定理の証明を提供する。

## 2. モデル不確実性下の特定化過誤回帰

### 2-1. データ生成過程とモデル

大きさ $n$ の観測値を持つデータを $(Y^n, X^n) = (Y_1, \cdots, Y_n, X_1, \cdots, X_n)$, $Y_i \in \mathbb{R}$, $X_i \in \mathbb{R}^k, i = 1, \cdots, n$ とする。所与の $n$ に対して、観測値のデータ生成過程（data generating process、DGP）は独立同一分布で $(Y_i, X_i) \sim F_{true}$ とする。$(Y_i, X_i)$ の結合分布 $F_{true}$ は通常の意味での条件付き密度関数 $f_{true}(y|x)$ を持つとする。$(Y^n, X^n)$ の結合分布を $F^n_{true}$ と書くとき、$E_n[\cdot], E_n[\cdot \mid \cdot]$ を $F^n_{true}$ に関する期待値とする。

所与の $n$ に対して、本論文が焦点を当てる $(Y_i, X_i)$ の DGP には、以下を仮定する。

$$Y_i = X_i^T \beta_0 + g(X_i) n^{-\alpha_1} + \epsilon_i,\ E_n[\epsilon_i | X_i] = 0,\ E_n[\epsilon_i^2 \big| X_i] = \tau_0^{-1} + h(X_i) n^{-2\alpha_2}. \quad (1)$$

パラメータ $(\beta_0, \tau_0)$ は $f_{true}$ を決定する真値とする。$\tau_0$ は誤差項の精度（分散の逆数）を決定するパラメータで、ある $0 < \tau_l < \tau_u < \infty$ に対して $\tau_0 \in (\tau_l, \tau_u)$ と仮定する。一般に $g, h$ には未知関数を許容し、特に条件付き不均一分散 $h \geq 0$ を許容する。$\alpha_1, \alpha_2 \geq \alpha_0 > 0$ を所与のバイアスコントロールとして、$\alpha_0 < 1/2$ を許容する。(1) は triangular array ゆえに、

$Y,\epsilon$ の $E_n[\cdot],E_n[\cdot\,|\,\cdot]$ は $n$ に依存することに注意したい。$X$ の分布は $n$ に依存しないものと仮定するが、表記の一貫性のため、その期待値を他の変数と同様に $E_n[X]$ と書く。ここで $E_n[XX^T]$ を有限で可逆、$\lambda_{min}$ を行列の最小固有値として $\lambda_{min}(E_n[XX^T])>0$ と仮定するが、$X$の分布は特定化しない。

未知部分を含むDGPである (1) に対して、ベイジアンと頻度論者はそれぞれの立場から $(\beta_0,\tau_0)$ の推論を考える。モデル不確実性下において、いずれの立場も、共通のモデルとして以下の均一分散の線形回帰を想定する。

$$Y_i=X_i^T\beta+\epsilon_i,\ \ E[\epsilon_i|X_i]=0,\ \ E[\epsilon_i^2\big|X_i]=\tau^{-1}. \tag{2}$$

ただし、モデルは条件付き密度 $f_{\beta,\tau}(y|x)$ を持つとする。$(\beta,\tau)$ のパラメータ空間 $\Theta\subset\mathbb{R}^k\times(\tau_l,\tau_u)$ を凸集合として、$(\beta_0,\tau_0)$ を $\Theta$ の内点と仮定する。所与の $\alpha_1,\alpha_2$ と有限の $n$ の下では $f_{true}\notin\{f_{\beta,\tau}:(\beta,\tau)\in\Theta\}$ であり、DGP (1) の分布はモデル (2) に含まれないという意味で、モデルは特定化過誤である。ただし、$n$ が大きくなるにつれて特定化の誤りは小さくなるため、この意味でモデルは“弱い”特定化過誤である。

本論文が焦点を当てる (1) と (2) に基づく設定は、モデル不確実的状況下の回帰モデリングの近似として十分に一般的である。$X$ に基づく線形部分はモデルに必ず含める一方で、モデル (2) で除外される $g,h$ はモデル不確実性の存在を示唆する。$\alpha_1,\alpha_2$ は $g,h$ の除外に伴う局所バイアスを決定する局外パラメータであり、設定は $\alpha_1=\alpha_2=1/2$ として局所特定化過誤の枠組みに基づく回帰を含む（Liu (2015) など）。もし $g,h$ が $X$ の既知関数ならこれらをモデルに含めることも可能であるが、バイアスの局所性より、これらを除外した特定化過誤モデル (2) は正当化される。例えば $g,h$ には $X_j\,(j=1,\cdots,k)$ の相互作用項や非線形変換を許容する。$(\beta_0)_k=0,\ g(X)=g(X_k),\ h(X)=h(X_k)$ とする場合、$g,h$ は線形部分の $X_j\,(j=1,\cdots,k-1)$ に含まれない $X_k$ を含んでよく、それには（観測されない）欠落変数が許容される。さらに、後述の仮定を満たす限りにおいて $g,h$ には $X$ の未知関数が許容され、(1) はノンパラメトリック部分線形モデルやランダム係数に伴う不均一分散回帰を含む。このとき $g,h$ のノンパラメトリック近似に基づく推定は原理的に可能であるが、再びバイアスの局所性より、これらを除外した (2) は正当化される。

ここでベイジアンは事後分布の構成のため、(2) に加えて $\epsilon$ の正規性を想定する。

$$\epsilon_i|X_i,\beta,\tau\sim N_1(0,\tau^{-1}). \tag{3}$$

ただし $N_d(\mu,\Sigma)$ は平均 $\mu$、分散共分散行列 $\Sigma$ の $d$ 変量正規分布とする。(1) に正規性は仮定されないため、(3) は分布の想定に関してさえも潜在的に特定化過誤である。ここで $(\beta,\tau)$ の事前分布は $\Theta$ 上で事前密度 $\pi(\beta,\tau)$ を持つとする。$(Y^n,X^n)$ が与えられた下で、$(\beta,\tau)$ の結合事後密度は次で与えられる。

$$\pi_n(\beta,\tau|Y^n,X^n)=\frac{\pi(\beta,\tau)\prod_{i=1}^n f_{\beta,\tau}(Y_i|X_i)}{\int_\Theta\pi(\beta,\tau)\prod_{i=1}^n f_{\beta,\tau}(Y_i|X_i)d(\beta,\tau)}.$$

このとき、$\prod_{i=1}^n f_{\beta,\tau}(Y_i|X_i)$ の $(\beta,\tau)\in\Theta$ に関する最大価を $(\hat\beta_n,\hat\tau_n)$ で表すと $\hat\beta_n=$

$\left(\sum_{i=1}^{n} X_i X_i^T\right)^{-1} \sum_{i=1}^{n} X_i Y_i$, $(\hat{\tau}_n)^{-1} = \frac{1}{n}\sum_{i=1}^{n}\left(Y_i - X_i^T \hat{\beta}_n\right)^2$ であり、特に $\hat{\beta}_n$ は (2) における最小二乗推定値と同一である。ここで $f_{true}$ と $f_{\beta,\tau}$ の Kullback-Leibler ダイバージェンス $E_n\left[\log\left(f_{true}/f_{\beta,\tau}\right)\right]$ の $(\beta,\tau) \in \Theta$ に関する最小価を $(\beta_n^*, \tau_n^*)$ で表すとして、これを疑似真値と呼ぶ。このとき疑似真値は $E_n\left[\log f_{\beta,\tau}\right]$ の最大価と等しいため、$\beta_n^* = E_n[XX^T]^{-1}E_n[XY]$, $(\tau_n^*)^{-1} = E_n(Y - X^T\beta_n^*)^2$ として陽の形式で書ける。特に $\beta_n^*$ は $Y$ のモデル変数 $X$ への線形射影係数に相当するが、これは (1) の下で $n$ に依存することに注意したい。

## 2-2. Bernstein-von Mises 定理

BvM 定理のバージョンに相当する次の定理は、尤度関数のモード $\left(\hat{\beta}_n, \hat{\tau}_n\right)$ で中心化された $(\beta,\tau)$ の結合事後分布の漸近正規性である。定理のための仮定は以下である。

**A1**. $\pi(\beta,\tau)$ は任意の $(\beta,\tau) \in \Theta$ に対して連続で、$(\beta_0,\tau_0) \in \Theta$ で正値とする。

**A2**. $E_n[Xg(X)] < \infty$, $E_n[g(X)^2] < \infty$, $E_n[h(X)] < \infty$.

A1 は真値の近傍での $\pi$ の正値性を要請するものであり、有限次元 BvM 定理に必要な仮定である。もし $\pi$ が $\Theta$ の広い範囲に正値を与える確率密度なら、この仮定は正当化される。ただし $\pi$ は既知のパラメトリックな確率密度である必要はない。A2 は $g, h$ に関するモーメントの条件である。次の定理のためには $\left(\hat{\beta}_n, \hat{\tau}_n\right) \to_{P_{true}} (\beta_0, \tau_0)$ が必要であるが、それは A2 の下で保証される（補論の補題 1）。ただし $\to_{P_{true}}$ は $F_{true}^n$ の下での確率収束を表す。

ここで $\ell_{\beta,\tau} = \log f_{\beta,\tau}$ と書くと、(3) の下での $\ell_{\beta,\tau}$ のヘシアン行列の $(\beta_0,\tau_0)$ における期待値は

$$E_n\left[\ddot{\ell}_{\beta_0,\tau_0}\right] = \begin{bmatrix} -\tau_0 E_n[XX^T] & E_n[X^T(Y - X^T\beta_0)] \\ E_n[X(Y - X^T\beta_0)] & -(2\tau_0^2)^{-1} \end{bmatrix}$$

であるが、A2 の下で、これにマイナスを掛けた行列の極限が存在する：

$$J_{\beta_0,\tau_0} = -\lim_{n\to\infty} E_n\left[\ddot{\ell}_{\beta_0,\tau_0}\right] = \begin{bmatrix} \tau_0 E_n[XX^T] & 0 \\ 0 & (2\tau_0^2)^{-1} \end{bmatrix}.$$

ここで密度関数 $f$ の全変動ノルムを $\|f\|_{TV} = \int |f(t)| dt$ とする。特に、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ことに注意したい（van der Vaart (1998) の 2.9 節など）。

**定理 1**. $(Y^n, X^n)$ は $\alpha_1, \alpha_2$ に対して (1) に従うとする。A1 と A2 を仮定する。このとき、(2) と (3) に基づく結合事後密度について、以下が成り立つ

$$\left\|\pi_n\left(\sqrt{n}\left((\beta,\tau) - \left(\hat{\beta}_n, \hat{\tau}_n\right)\right) \middle| Y^n, X^n\right) - N_{k+1}\left(0, J_{\beta_0,\tau_0}^{-1}\right)\right\|_{TV} \to_{P_{true}} 0.$$

定理 1 の証明は補論で提供される。証明は Bickel and Doksum (2001) の定理 5.2.2 の証明ステップを辿ることで殆ど同様に示される。

全変動ノルムは位置とスケールの変更に関して不変であるため、定理 1 は次を意味する：$\left\| \pi_n(\beta, \tau | Y^n, X^n) - N_{k+1}\left( (\hat{\beta}_n, \hat{\tau}_n),\ J^{-1}_{\beta_0, \tau_0}/n \right) \right\|_{TV} \rightarrow_{P_{true}} 0.$ 特に周辺分布について、これは次の近似を意味する。

$$\beta | Y^n, X^n \sim N_k\left(\hat{\beta}_n, \tau_0^{-1} E_n[XX^T]^{-1}/n\right).$$

よって、$\beta$ の周辺事後分布は多変量正規分布で近似され、その平均は最小二乗推定値、分散共分散行列は $\tau_0^{-1} E_n[XX^T]^{-1}/n$ である。(1) に誤差項の正規性は仮定されないため、(3) は潜在的に特定化過誤であるにも関わらず、$\beta$ の事後分布は再び正規分布に従う。本論文の枠組みでは $(\hat{\beta}_n, \hat{\tau}_n)$ に一致性が成り立つために、定理 1 の漸近的な事後分散共分散行列は真値によって評価されている。漸近分布は $\alpha_1, \alpha_2$ に依存しないため、バイアスの収束レートは結論に影響を及ぼさない。

## 2-3. 漸近的な情報行列等式

モデル (2) における $\beta$ の頻度論推測には、効率性の観点に基づき、典型的に最小二乗推定量 $\hat{\beta}_n$ の標本分布が考えられる (Hayashi (2000)、Chamberlain (1987))。ここで、(1) と (2) に限定されない一般の特定化過誤的状況において、$\hat{\beta}_n$ の標本分布は次で近似されることを思い出したい（van der Vaart (1998) の定理 5.23 など）。

$$\hat{\beta}_n \sim N_k(\beta^*, E[XX^T]^{-1} E[(Y - X^T\beta^*)^2 XX^T] E[XX^T]^{-1}/n).$$

ただし、$\beta^*$ を $E(Y - X^T\beta)^2$ の最小価、$E[\cdot]$ をデータ生成過程の分布に関する期待値とする。このとき、$\hat{\beta}_n$ の“サンドイッチ形式”の分散共分散行列は、本論文の定理 1 で示唆される $\beta$ の事後分散共分散行列 $\tau_0^{-1} E_n[XX^T]^{-1}/n$ とは一般に異なる。よって、Kleijn and van der Vaart (2012) における示唆と同様に、一般の特定化過誤回帰において回帰係数の信用/信頼集合は一致しない。それにも関わらず、本論文が焦点を当てる (1) と (2) においてこの不一致は緩和される。実際、次の定理は $(\beta_n^*, \tau_n^*)$ に基づく漸近的な情報行列等式である。この定理のために $n \rightarrow \infty$ につれて $(\beta_n^*, \tau_n^*) \rightarrow (\beta_0, \tau_0)$ が必要であるが、それは A2 の下で保証される（補論の補題 2）。定理のために追加的に以下を仮定する。

A3．$E_n[g(X)^2 XX^T] < \infty,\ E_n[h(X)XX^T] < \infty$．すべての $j, l = 1, \cdots, k$ に対して $E_n\left[(X_j X_l) XX^T\right] < \infty$. すべての $j = 1, \cdots, k$ に対して $E_n\left[(g(X)X_j) XX^T\right] < \infty$.

A3 は $g, h$ と $X$ の交差モーメントに関する条件であるが、特に $X$ の 4 次モーメントは誤差項の分散推定を含む回帰において典型的に仮定される（Hayashi (2000) の 2 章など）。

**定理 2**．$(Y^n, X^n)$ は $\alpha_1, \alpha_2$ に対して (1) に従うとする。A2 と A3 を仮定する。このとき

$$(\tau_n^*)^{-1}E_n[XX^T]^{-1} = E_n[XX^T]^{-1}E_n[(Y - X^T\beta_n^*)^2XX^T]E_n[XX^T]^{-1} + o(1).$$

定理 2 の証明は補論で提供される。

定理 2 は、$n$ が大きい限り、$\hat{\beta}_n$ の分散共分散行列を $(\tau_n^*)^{-1}E_n[XX^T]^{-1}/n$ で置き換えること（あるいはその逆）を正当化する。したがって、定理 1 と 2 により、$\beta$ の事後分布と $\hat{\beta}_n$ の標本分布は、有限の $n$ において、それぞれ次で近似される。

$$\beta|Y^n, X^n \sim N_k\left(\hat{\beta}_n, (\tau_n^*)^{-1}E_n[XX^T]^{-1}/n\right), \qquad \hat{\beta}_n \sim N_k(\beta_n^*, (\tau_n^*)^{-1}E_n[XX^T]^{-1}/n).$$

二つの分布は平均が互いに“逆転”した形式を持ち、いずれの分散共分散行列も共通である。したがって、結果は $\beta_n^*$ の信用集合と信頼集合が任意の信頼水準に対して互いに近づくことを意味する。さらに、$n$ が大きい限り $\beta_n^*$ は $\beta_0$ の任意の近傍に近づくため、十分に大きな $n$ では、この信用/信頼集合は $\beta_0$ の推論のための集合として解釈可能である。また、定理 2 の証明より、情報行列等式の乖離の収束レートは $O\left(n^{-2(\alpha_1 \wedge \alpha_2)}\right)$ である。すなわち、$\alpha_1, \alpha_2 < 1/2$ の場合でさえ、$n$ が大きい限り、定理は漸近的な等式の成立を示唆する。ただし $\alpha_1 \neq \alpha_2$ なら、収束レートは $n^{-2\alpha_1}$ と $n^{-2\alpha_2}$ のより遅い方で決定される。特定化の深刻な誤りが不均一分散にのみ存在する場合でさえ、等式の乖離は大きくなりうる。

有限標本における信用/信頼集合の $\beta_0$ の被覆性質については、本論文の理論からは自明ではない。実際、$\alpha_1, \alpha_2 < 1/2$ ならバイアスの収束レートは標準誤差より遅いため、有限の $n$ における $\beta_n^*$ と $\beta_0$ の乖離は小さくない可能性がある。このとき、名目水準を達成するような集合の被覆性質はおそらく期待できない。この理由は、$n$ が大きくなるにつれてバイアスよりも先に標準誤差が小さくなることで、真値とやや離れた領域に分布が集中するためである。この傾向は事後分布と標本分布のいずれにおいても予想される。回帰係数の真値の被覆性質は応用上の観点から重要であるため、次章のモンテカルロ研究において更なる検証を行う。

本論文が焦点を当てた枠組みの重要な例は、$\alpha_1 = \alpha_2 = 1/2$ の場合に相当する局所特定化過誤回帰である。この枠組みはモデル不確実性下の回帰モデルの近似として、変数選択のための集中情報量規準や頻度論モデル平均の基礎である。実際、このとき生ずるバイアスは標準誤差と同じ高々 $O\left(1/\sqrt{n}\right)$ ゆえに、真値で中心化された推定量の漸近分布と平均二乗誤差の導出を可能とする。このとき、真値の推論のための最適な変数選択規準やモデル平均推定量を構成できる。定理 1 と 2 を $\alpha_1 = \alpha_2 = 1/2$ に特殊化した結果は、このようなモデル不確実性下の回帰における回帰係数の事後正規性と情報行列等式の成立を意味する。これは例えば集中情報量規準に基づく変数選択を行う状況における信用/信頼集合の接近を意味する。このような状況下で信用集合に信頼集合と同等の頻度論的解釈が与えられること、そして信用/信頼集合の計算に互いの計算手法を代用することの妥当性が与えられることは、応用上重要な知見であると指摘できる。

## 3．モンテカルロ研究

信用集合と信頼集合の真値の被覆確率とこれらの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。データ生成過程は、

$$Y_i = \beta_{0,0} + \beta_{0,1}X_{1i} + \beta_{0,2}X_{2i} + n^{-\alpha_1}X_{3i} + \epsilon_i, \quad i = 1, \cdots, n$$

ただし $\beta_0 = \left(\beta_{0,0}, \beta_{0,1}, \beta_{0,2}\right)^T = (1,1,1)^T$, $X_{j\,i} \sim N_1(1,1), (j = 1,2,3)$, $X_j$ の間の共分散を 1/2 とする。条件付き不均一分散を許容する $\epsilon$ には、次の二つを考える。

$$\text{(I)}\ \epsilon_i \sim (\tau_0^{-1} + n^{-2\alpha_2}X_{3i}^4)^{1/2}(\chi_3^2 - 3)/\sqrt{6}, \qquad \text{(II)}\ \epsilon_i \sim N_1(0, \tau_0^{-1} + n^{-2\alpha_2}X_{1i}^4).$$

ただし $\tau_0 = 1$、$\chi_d^2$ は自由度 $d$ のカイ二乗分布に従う変数として、$X_j$ と独立とする。これらの DGP に対して、変数に$X_i = (1, X_{1i}, X_{2i})^T$を用いた均一分散モデル$Y_i = X_i^T\beta + u_i$, $E[u_i|X_i] = 0$, $E[u_i^2|X_i] = \tau^{-1}$ を想定する。ベイジアンは (3) と同様に $u_i \sim N_1(0, \tau^{-1})$ を想定する。DGP (I) と (II) のいずれの下でもモデルは回帰関数と不均一分散に関して特定化過誤である。正規尤度を想定するベイジアンは DGP (I) において $\epsilon$ の分布に関してさえも特定化過誤である。バイアスコントロールの設定を $\alpha_1, \alpha_2 \in \{1/2, 1/4\}$ とする。$\alpha_1 = \alpha_2 = 1/2$ は局所特定化過誤、$\alpha_1 = 1/4$（$\alpha_2 = 1/4$）の場合は回帰関数（不均一分散）の特定化の誤りがより深刻な場合に相当する。

興味あるパラメータを $\beta_{0,1}$ とする。正規尤度の想定に基づく $\beta_{0,1}$ の 95%信用区間（CrI）には、$(\beta, \tau)$ の正規-ガンマ事前分布 $\pi(\beta, \tau) = N_k(\beta|\, \underline{\beta}, \tau^{-1}\underline{\Omega}^{-1}) Gamma\,(\tau|\underline{a}, \underline{b})$ に基づく以下とする（Bernardo and Smith (1994) など）。

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

ただし、$\beta_{post} = \left(\underline{\Omega} + \sum_{i=1}^n X_i X_i^T\right)^{-1}\left(\underline{\Omega}\,\underline{\beta} + \sum_{i=1}^n X_i Y_i\right)$, $a_n = \underline{a} + n/2$, $b_n = \underline{b} + \sum_{i=1}^n \big(Y_i - X_i^T\beta_{post}\big)Y_i/2 + \left(\underline{\beta} - \beta_{post}\right)^T \underline{\Omega}\,\underline{\beta}/2$, $t_a(k)$ は自由度 $k$ の標準 $t$ 分布の上側 100a%分位点とする。ここで $\beta_{post}$ は解析に基づく $\beta$ の事後平均である。ハイパーパラメータは $\underline{\beta} = (0,0,0)^T$, $\underline{\Omega} = \text{diag}(1,1,1)/100$, $\underline{a} = 1$, $\underline{b} = 1$とする。異なるハイパーパラメータの場合も検討したものの、結果の傾向に大きな違いは見られなかったため省略する。また信用区間の構成には典型的に最高事後密度区間が用いられるが、この設定では周辺事後分布に対称性が期待されるため対称信用区間に焦点を当てる。

頻度論推測として、Kleijn and van der Vaart (2012) に基づき、特定化過誤を想定した下での $\beta_{0,1}$ の 95%信頼区間（CI）を考える。$\hat{\beta}_n = \left(\sum_{i=1}^n X_i X_i^T\right)^{-1}\sum_{i=1}^n X_i Y_i$ とそのサンドイッチ形式の漸近分散に基づく信頼区間を以下とする。

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

ただし、$\widehat{Var}(\hat{\beta}_n)=\left(\frac{1}{n}\sum_{i=1}^{n}X_iX_i^T\right)^{-1}\left(\frac{1}{n}\sum_{i=1}^{n}\left(Y_i-X_i^T\hat{\beta}_n\right)^2X_iX_i^T\right)\left(\frac{1}{n}\sum_{i=1}^{n}X_iX_i^T\right)^{-1}$、$z_a$を標準正規分布の上側 100a%分位点とする。$\widehat{Var}(\hat{\beta}_n)$ は $\hat{\beta}_n$ の漸近分散の不均一分散ロバスト推定量（White (1980)）と関連するが、不均一分散に対してロバストであることを根拠にこの分散推定量を用いるわけではない。あくまで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る。さらに、CI に加えてブートストラップ信頼区間（B-CI）を検討する。今回はノンパラメトリックブートストラップとして、$(Y,X)$ を組ごとにリサンプリングするブートストラップ（pairs bootstrap）に焦点を当てる。Efron の分位点法（van der Vaart (1998)、p.327）に基づき、信頼限界には $\left(\hat{\beta}_n\right)_1$ のブートストラップ分布の上側 2.5%点と 97.5%点を用いる。

このモンテカルロ実験では仮想的な繰り返し標本における信用/信頼区間の $\beta_{0,1}=1$ の被覆確率を調査する。加えて、CrI と CI それぞれの区間長の中央値の差（LD）、CrI と B-CI における同様の区間長の差（LDB）を調査する。実験に用いる標本サイズは $n\in\{20,50,100,1000\}$ とする。ブートストラップの繰り返し標本の数は 400、モンテカルロ実験の回数は 2000 とする。

表 1 は DGP (I) に対する各区間の被覆確率と区間長の差を表したものである。まず $\alpha_2=1/2$ の場合について、$\alpha_1=1/2$ では、$n$ が大きいほど信用区間の被覆確率は 0.95 に近づいている。このとき、$n$ が大きいほど CrI と CI の被覆確率は近い値をとり、区間長の差も小さいため、結果は区間の境界そのものが互いに近づくことを示唆する。ここに真値を中心とした信用/信頼区間の一致の可能性を確認できる。ただし、この結果はベイジアンが誤差項の分布の特定化を誤った状況であるにも関わらず得られたことに注意したい。さらに、$n$ が大きいとき B-CI は CI と CrI を良く近似していることが示唆される。この結果はブートストラップ信頼区間による信用区間の代用を支持する根拠になる。特に $n$ が小さいとき、B-CI は CI の被覆確率を改良することも示唆される。これはブートストラップ信頼区間の使用がベイジアンと頻度論者のいずれにとっても有益であることを示唆する。一方、バイアスがより深刻な $\alpha_1=1/4$ の場合、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見てとれる。このとき、$n$ 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の方がバイアスより小さくなることで真値からやや離れた領域に分布が集中するためであると思われる。それにも関わらず、$n$ が大きくなれば二つの区間の被覆確率は同じ値に近づき（およそ 0.63）、区間長の差も小さくなる。この傾向は $n=100$ でさえ生ずる。この結果は本論文の理論と整合的である。$\alpha_2=1/4$ でも同様の傾向を見てとれる。DGP (I) では不均一分散の特定化の誤りは区間の一致の結論に大きな影響をもたらすわけではないように見える。

表 2 は DGP (II) に対する同様の実験の結果である。$\alpha_2=1/2$ の場合、DGP (I) とほぼ同様の傾向を見てとれる。特に $\alpha_1=1/2$ の場合は BvM 定理と整合的な結果であり、これは局所特定化過誤における信用/信頼区間の一致を再び支持するものである。ところが、不均

一分散の特定化の誤りがより深刻な $\alpha_2 = 1/4$ の場合、$n = 100, 1000$ において CrI と CI の被覆確率に無視できない乖離を見てとれる。このとき、DGP (I) とは対照的に、区間長についても小さくない差が存在するように見える。おそらくこれは CI の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ためであると思われる。実際、CrI は潜在的な不均一分散を捉えるよう設計されてはいないが、一方の CI は不均一分散を考慮に入れた標準誤差と関連する。おそらくこの理由で CI には $n$ が大きくなるに従って、被覆確率の改善がもたらされることで、$n = 100$ における CrI と CI の被覆確率の差が生ずるものと思われる。それにも関わらず、$n = 1000$ では CrI と CI の被覆確率は再び近づいており、区間長の差も小さくなる。このとき B-CI は CI をよく近似しており、さらに、$n$ が大きいほど B-CI と CrI の区間長の差も小さくなる。したがって、$n$ が大きくなるほど乖離が小さくなる点で、結果は再び漸近理論と整合的であると解釈できる。ただし、あくまで有限標本では情報行列等式に乖離が残りうることがここに改めて明確になる。$\alpha_1 = 1/4$ でも同様の傾向があり、CrI と CI の区間長の乖離は $n$ が大きくなるほど小さくなる。

図 1 は不均一分散の特定化過誤が深刻な $\alpha_2 = 1/4$ に焦点を当て、$[1/10, 1]$ を 10 等分する $\alpha_1$ に対する各区間の被覆確率をプロットしたものである。上段は DGP (I)、下段は DGP (II) であり、それぞれ左から $n = 20, 100, 1000$ である。この実験のためのブートストラップ標本の数は 400、モンテカルロ実験の回数は 1000 とする。結果として、DGP (I) のいずれの $n$ でも、$\alpha_1$ の値に応じて連続的に被覆確率が変化する傾向を見てとれる。$n = 20$ なら CrI と CI の被覆確率に乖離が見られるが、$n = 100, 1000$ ならこれらの被覆確率は $\alpha_1$ に一様に近づいている。このとき B-CI は CrI と CI のいずれの被覆確率とも極めて近く、これらの良い近似を与えることが示唆される。一方 DGP (II) では、$n = 20, 100$ で生じた被覆確率の乖離は、$n = 1000$ において $\alpha_1$ に一様に小さくなる傾向を見てとれる。このとき、$\alpha_1 \leq 1/4$ では、いずれの区間も真値の被覆に頻繁に失敗しつつも、$n = 100$ の場合に比べて互いに近づくことが示唆される。再びこれは理論と整合的な結果である。

| | | $\alpha_1 = 1/2$ | | | | | $\alpha_1 = 1/4$ | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| $\alpha_2$ | $n$ | CrI | CI | B-CI | LD | LDB | CrI | CI | B-CI | LD | LDB |
| 1/2 | 20 | 0.925 | 0.884 | 0.923 | 0.173 | 0.028 | 0.890 | 0.845 | 0.894 | 0.161 | 0.071 |
| | 50 | 0.940 | 0.927 | 0.930 | 0.054 | 0.016 | 0.885 | 0.862 | 0.865 | 0.054 | 0.017 |
| | 100 | 0.927 | 0.924 | 0.916 | 0.025 | 0.016 | 0.861 | 0.845 | 0.845 | 0.021 | 0.013 |
| | 1000 | 0.940 | 0.942 | 0.935 | 0.001 | 0.002 | 0.632 | 0.626 | 0.622 | 0.001 | 0.003 |
| 1/4 | 20 | 0.911 | 0.882 | 0.921 | 0.186 | 0.125 | 0.898 | 0.875 | 0.907 | 0.187 | 0.113 |
| | 50 | 0.925 | 0.930 | 0.922 | 0.059 | 0.011 | 0.910 | 0.906 | 0.891 | 0.058 | 0.016 |
| | 100 | 0.937 | 0.944 | 0.938 | 0.016 | 0.009 | 0.896 | 0.910 | 0.902 | 0.016 | 0.004 |
| | 1000 | 0.932 | 0.941 | 0.934 | 0.002 | 0.001 | 0.719 | 0.730 | 0.715 | 0.002 | 0.000 |

表１：DGP (I)における信用区間（CrI）と信頼区間（CI, B-CI）の被覆確率と区間長の差

| | | $\alpha_1 = 1/2$ | | | | | $\alpha_1 = 1/4$ | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| $\alpha_2$ | $n$ | CrI | CI | B-CI | LD | LDB | CrI | CI | B-CI | LD | LDB |
| 1/2 | 20 | 0.863 | 0.850 | 0.901 | 0.049 | 0.161 | 0.849 | 0.831 | 0.897 | 0.059 | 0.161 |
| | 50 | 0.902 | 0.909 | 0.922 | 0.007 | 0.035 | 0.853 | 0.859 | 0.877 | 0.008 | 0.038 |
| | 100 | 0.928 | 0.931 | 0.930 | 0.013 | 0.018 | 0.833 | 0.842 | 0.851 | 0.012 | 0.017 |
| | 1000 | 0.934 | 0.934 | 0.929 | 0.001 | 0.001 | 0.642 | 0.643 | 0.639 | 0.001 | 0.001 |
| 1/4 | 20 | 0.809 | 0.839 | 0.895 | 0.074 | 0.368 | 0.803 | 0.825 | 0.899 | 0.063 | 0.347 |
| | 50 | 0.847 | 0.915 | 0.927 | 0.164 | 0.189 | 0.805 | 0.878 | 0.894 | 0.153 | 0.180 |
| | 100 | 0.861 | 0.929 | 0.933 | 0.140 | 0.139 | 0.801 | 0.887 | 0.888 | 0.127 | 0.127 |
| | 1000 | 0.894 | 0.951 | 0.947 | 0.029 | 0.027 | 0.674 | 0.783 | 0.772 | 0.029 | 0.027 |

表 2：DGP (II)における信用区間（CrI）と信頼区間（CI, B-CI）の被覆確率と区間長の差

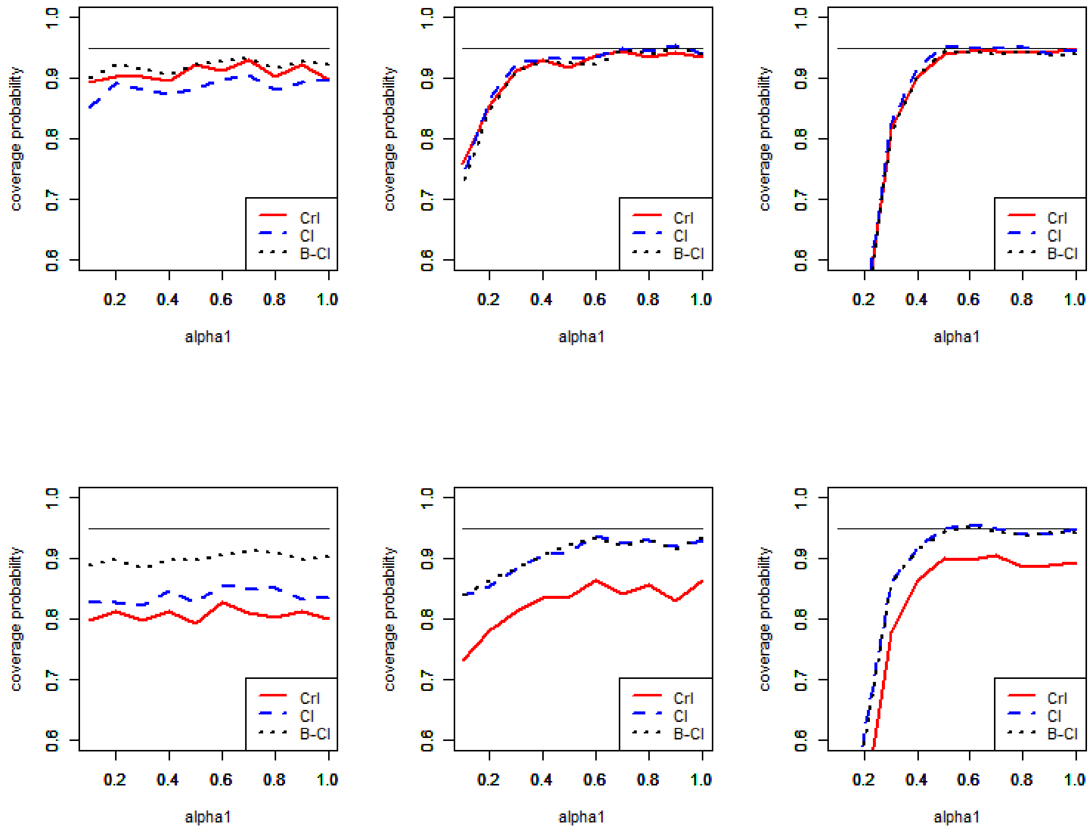


図 1 ：信用区間（CrI）と信頼区間（CI, B-CI）の被覆確率（$\alpha_2 = \frac{1}{4}$）。横軸は$\alpha_1$、縦軸は被覆確率。上段はDGP (I)、下段は(II)、左から $n = 20,100,1000$。黒線は名目水準95%。

## 4. 結論

本論文はモデル不確実性が存在する線形回帰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。回帰において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さえ信用/信頼集合は近づくことが示唆された。このとき、誤差項の分布の特定化過誤や事前分布の存在、そしてバイアスの収束レートの遅さは集合の一致に強い影響を及ぼさない。モンテカルロ研究からは、無視できないバイアスにより信用/信頼集合のいずれもが真値の被覆に失敗する場合でさえ、標本サイズが大きい限り、集合は互いに近づくことが示唆された。これにより、変数選択を必要とするような状況において信用集合に信頼集合と同程度の長期的な被覆性質を期待できる。さらに結果はブートストラップ信頼集合により信用集合を代用することを正当化する。これらの新たな知見は、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と期待される。

## 引用文献

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.

Chamberlain, G. (1987). “Asymptotic efficiency in estimation with conditional moment restrictions,” *Journal of Econometrics*, 34(3), 305 - 334.

Chatfield, C. (1995). “Model uncertainty, data mining and statistical inference,” *Journal of the Royal Statistical Society: Series A*, 158(3), 419 - 444.

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.

Claeskens, G. and N. L. Hjort. (2008). *Model Selection and Model Averaging*, Cambridge University Press.

Cox, D.D. (1993). “An analysis of Bayesian inference for nonparametric regression,” *The Annals of Statistics*, 21(2), 903 - 923.

Draper, D. (1995). “Assessment and propagation of model uncertainty,” *Journal of the Royal Statistical Society: Series B*, 57(1), 45 - 70.

Freedman, D. A. (1999). “On the Bernstein-Von Mises theorem with infinite-dimensional parameters,” *The Annals of Statistics*, 27(4) , 1119-1140.

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.

Hansen, B. E. (2014). “Model averaging, asymptotic risk, and regressor groups,” *Quantitative Economics*, 5(3), 495 - 530.

Hartigan, J. A. (1983). *Bayes Theory*, Springer.

Hayashi, F. (2000). *Econometirics*, 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, 221 - 233.

Kato, K. (2013). “Quasi-Bayesian analysis of nonparametric instrumental variables models,” *The Annals of Statistics*, 41(5), 2359 - 2390.

Kitagawa. T, J. L. Montiel Olea, J. Payne, and 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, and J.H.van Zanten. (2011). "Bayesian inverse problems with Gaussian priors," *The Annals of Statistics*, 39(5), 2626 - 2657.

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.

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.

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.

## 補論

補論では一致性に関する二つの補題、定理1および定理2の証明を与える。ベクトル $a$ のユークリッドノルムを $\|a\|$、行列 $A=(a_{jl})$のノルムを $\|A\|=tr(A^TA)^{1/2}=\left(\sum_j\sum_l a_{jl}^2\right)^{1/2}$ とする。以後の確率収束 $\to_P$ および $O_P, o_P$ は $F_{true}^n$の下での収束とする。

**補題1．** $(Y^n, X^n)$ は (1) に従うとする。A2を仮定する。このとき $(\hat{\beta}_n, \hat{\tau}_n) \to_P (\beta_0, \tau_0)$.

**証明．** $\hat{\beta}_n=\beta_0+n^{-\alpha_1}\left(\frac{1}{n}\sum_{i=1}^n X_iX_i^T\right)^{-1}\frac{1}{n}\sum_{i=1}^n X_i g(X_i)+\left(\frac{1}{n}\sum_{i=1}^n X_iX_i^T\right)^{-1}\frac{1}{n}\sum_{i=1}^n X_i\epsilon_i$ であるが、（triangular array の）大数の法則と A2 より $\hat{\beta}_n-\beta_0=O_P(n^{-\alpha_1})+o_P(1)=o_P(1)$. 同様に

$$
\begin{aligned}
(\hat{\tau}_n)^{-1} &= \left(\beta_0 - \hat{\beta}_n\right)^T \left(\frac{1}{n}\sum_{i=1}^{n} X_i X_i^T\right)\left(\beta_0 - \hat{\beta}_n\right) + n^{-2\alpha_1}\frac{1}{n}\sum_{i=1}^{n} g(X_i)^2 + \frac{1}{n}\sum_{i=1}^{n}\epsilon_i^2 \\
&\quad + 2n^{-\alpha_1}\left(\beta_0 - \hat{\beta}_n\right)^T \frac{1}{n}\sum_{i=1}^{n} X_i g(X_i) + 2n^{-\alpha_1}\frac{1}{n}\sum_{i=1}^{n} g(X_i)\epsilon_i + 2\left(\beta_0 - \hat{\beta}_n\right)^T \frac{1}{n}\sum_{i=1}^{n} X_i\epsilon_i \\
&= \tau_0^{-1} + O_P\left(\left\|\hat{\beta}_n - \beta_0\right\|^2\right) + O_P\left(n^{-\alpha_1}\left\|\hat{\beta}_n - \beta_0\right\|\right) + O_P(n^{-2\alpha_1}) + O_P(n^{-2\alpha_2})
\end{aligned}
$$

であるが、直前に得られた $\hat{\beta}_n - \beta_0 = O_P(n^{-\alpha_1})$より $(\hat{\tau}_n)^{-1} - \tau_0^{-1} = O_P\left(n^{-2(\alpha_1 \wedge \alpha_2)}\right) = o_P(1)$ が従う。連続写像定理より、これは $\hat{\tau}_n \to_P \tau_0$ を意味する。このとき van der Vaart (1998) の定理 2.7 より $\left(\hat{\beta}_n, \hat{\tau}_n\right)$ は結合で確率収束するため、補題の結果を得る。 □

**補題 2.** $(Y^n, X^n)$ は (1) に従うとする。A2 を仮定する。このとき $\beta_n^* = \beta_0 + O(n^{-\alpha_1})$, $\tau_n^* = \tau_0 + O\left(n^{-2(\alpha_1 \wedge \alpha_2)}\right)$. 特に $(\beta_n^*, \tau_n^*) \to (\beta_0, \tau_0)$.

**証明.** 証明は補題1と同様である。(1)の下で $\beta_n^* = \beta_0 + n^{-\alpha_1} E_n[XX^T]^{-1} E_n[Xg]$ であるが、第 2 項は仮定より $O(n^{-\alpha_1})$. 同様に $(\tau_n^*)^{-1} = \tau_0^{-1} + n^{-2\alpha_2}E_n[h] + n^{-2\alpha_1} E_n[g^2] + (\beta_0 - \beta_n^*)^T E_n[XX^T](\beta_0 - \beta_n^*) + 2n^{-\alpha_1}(\beta_0 - \beta_n^*)^T E_n[Xg]$ であるが、$\beta_n^* - \beta_0 = O(n^{-\alpha_1})$ より $(\tau_n^*)^{-1} - \tau_0^{-1} = O\left(n^{-2(\alpha_1 \wedge \alpha_2)}\right)$. 結果は $n \to \infty$ につれて $(\beta_n^*, \tau_n^*) \to (\beta_0, \tau_0)$ を意味する。 □

**注意.** Kullback-Leibler ダイバージェンスの最小価 $(\beta_n^*, \tau_n^*)$ は (2) と (3) の下で陽の形式が導かれたが、補題 2 の証明において (2) と (3) は必要ではない。特に、正規性の仮定がなくとも補題 2 は成立する。

**定理 1 の証明.** Bickel and Doksum (2001) の定理 5.2.2 の証明と本質的に同様であるためスケッチのみを示す。定理の結果のためには、$\ell_{\beta,\tau}$ に対して次の二つの条件：任意の $\epsilon_n \to_P 0$ に対して

$$
\sup_{\|(\beta,\tau)-(\beta_0,\tau_0)\| \le \epsilon_n} \left\| \frac{1}{n}\sum_{i=1}^{n} \ddot{\ell}_{\beta,\tau}(Y_i) - \left(-J_{\beta_0,\tau_0}\right) \right\| \to_P 0, \tag{4}
$$

そして任意の $\gamma > 0$ に対して $\delta(\gamma) > 0$ が存在して、確率 1 に近づく事象として（w.p.a.1）

$$
\sup_{\|(\beta,\tau)-(\beta_0,\tau_0)\| > \gamma} \frac{1}{n}\sum_{i=1}^{n}\{\ell_{\beta,\tau}(Y_i) - \ell_{\beta_0,\tau_0}(Y_i)\} \le -\delta(\gamma) \tag{5}
$$

を示せば十分である。実際、$\left(\hat{\beta}_n, \hat{\tau}_n\right)$ において中心化とリスケールしたパラメータ $t = \sqrt{n}\left((\beta,\tau) - \left(\hat{\beta}_n, \hat{\tau}_n\right)\right)$ の事後密度の正規密度への収束を示すにあたり、$\ell_{(\hat{\beta}_n, \hat{\tau}_n)+t/\sqrt{n}}$ を $\left(\hat{\beta}_n, \hat{\tau}_n\right)$ において Taylor 展開するとき、$\left\|\left(\bar{\beta}(t), \bar{\tau}(t)\right) - \left(\hat{\beta}_n, \hat{\tau}_n\right)\right\| \le \|t\|/\sqrt{n}$ を満たすある $\bar{\beta}(t), \bar{\tau}(t)$ に対して、$t$ に一様に、次式を示す必要がある：

$$\frac{1}{n}\sum_{i=1}^{n}\ddot{\ell}_{\bar{\beta}(t),\bar{\tau}(t)}(Y_i)\to_P -J_{\beta_0,\tau_0}. \tag{6}$$

ところが、任意の $M>0$ に対して，$\epsilon_n = M/\sqrt{n}+\left\|\left(\hat{\beta}_n,\hat{\tau}_n\right)-(\beta_0,\tau_0)\right\|$ とおけば、$\|t\|\le M$ となる任意の $t$ に対して、補題 1 より $\left\|\left(\bar{\beta}(t),\bar{\tau}(t)\right)-(\beta_0,\tau_0)\right\|\le\epsilon_n$ かつ $\epsilon_n\to_P 0$ である。

よって、 (6) 式のためには (4) 式を示せば十分である。一方の (5) 式は事後密度の $(\beta,\tau)$ に関する積分の収束のための十分条件である。証明の残る部分は A1 と $\left(\hat{\beta}_n,\hat{\tau}_n\right)$ の一致性（補題 1）の下で同様に示すことができる。

いま (4) 式を示すため、任意の $\epsilon_n\to_P 0$ を固定する。まず、(4) 式の左辺は

$$\sup_{\|(\beta,\tau)-(\beta_0,\tau_0)\|\le\epsilon_n}\left\|\frac{1}{n}\sum_{i=1}^{n}\{\ddot{\ell}_{\beta,\tau}(Y_i)-\ddot{\ell}_{\beta_0,\tau_0}(Y_i)\}\right\|+\left\|\frac{1}{n}\sum_{i=1}^{n}\ddot{\ell}_{\beta_0,\tau_0}(Y_i)-\left(-J_{\beta_0,\tau_0}\right)\right\|$$

という分解により、上から押さえられる。第 1 項のノルムの内側は

$$\frac{1}{n}\sum_{i=1}^{n}\{\ddot{\ell}_{\beta,\tau}(Y_i)-\ddot{\ell}_{\beta_0,\tau_0}(Y_i)\}=\begin{bmatrix}(\tau_0-\tau)\frac{1}{n}\sum_{i=1}^{n}X_iX_i^T & (\beta_0-\beta)^T\frac{1}{n}\sum_{i=1}^{n}X_iX_i^T\\ \frac{1}{n}\sum_{i=1}^{n}X_iX_i^T(\beta_0-\beta) & (\tau_0^{-2}-\tau^{-2})/2\end{bmatrix}$$

であるが、$\|(\beta,\tau)-(\beta_0,\tau_0)\|\le\epsilon_n$ を満たす任意の $(\beta,\tau)$ に対して、大数の法則と A2 より、行列の全てのブロック成分のノルムが $O_P(\epsilon_n)$ であることを示せる。実際

$$\sup_{\|(\beta,\tau)-(\beta_0,\tau_0)\|\le\epsilon_n}\left\|(\tau_0-\tau)\frac{1}{n}\sum_{i=1}^{n}X_iX_i^T\right\|^2\le\epsilon_n^2\sum_{j=1}^{k}\sum_{l=1}^{k}\left(\frac{1}{n}\sum_{i=1}^{n}X_{j\,i}X_{l\,i}\right)^2=O_P(\epsilon_n^2).$$

同様に、行列ノルムの劣乗法性より $\sup\|(\beta_0-\beta)^T\frac{1}{n}\sum_{i=1}^{n}X_iX_i^T\|^2\le\sup\|(\beta_0-\beta)\|^2\left\|\frac{1}{n}\sum_{i=1}^{n}X_iX_i^T\right\|^2=O_P(\epsilon_n^2)$. さらに仮定よりある定数 $C_0>0$ に対して $\sup|\tau_0^{-2}-\tau^{-2}|^2\le C_0\sup|\tau_0-\tau|^2=O(\epsilon_n^2)$. ここで、行列 $A$ のノルムはブロック行列 $A_{j\,l}(j,l=1,2)$ について $\|A\|^2=\|A_{1\,1}\|^2+\|A_{1\,2}\|^2+\|A_{2\,1}\|^2+\|A_{2\,2}\|^2$ である事実に注意すれば、分解の第 1 項のノルムは $O_P(\epsilon_n)$ が従い、特に $o_P(1)$ である。一方、分解の第 2 項が $o_P(1)$ を示すためには

$$\tau_0\frac{1}{n}\sum_{i=1}^{n}X_iX_i^T\to_P\tau_0E_n[XX^T],\qquad\frac{1}{n}\sum_{i=1}^{n}X_i\left(Y_i-X_i^T\beta_0\right)\to_P 0$$

で十分であるが、左式は大数の法則より明らかである。右式は $\frac{1}{n}\sum_{i=1}^{n}X_ig(X_i)n^{-\alpha_1}+\frac{1}{n}\sum_{i=1}^{n}X_i\epsilon_i$ と書けるが、第 1 項は大数の法則と A2 より $n^{-\alpha_1}E_n[Xg]+n^{-\alpha_1}\left(\frac{1}{n}\sum_{i=1}^{n}X_ig(X_i)-E_n[Xg]\right)=O(n^{-\alpha_1})+o_P(n^{-\alpha_1})=o_P(1)$、第 2 項も同様に $o_P(1)$. 以上より、(4) 式が従う。

次に (5) 式のため、任意の $\gamma>0$ を固定する。まず簡単な計算より

$$2E_n\left[\ell_{\beta_0,\tau_0}-\ell_{\beta,\tau}\right] = \left(\log\left(\frac{\tau_0}{\tau}\right)-\left(1-\frac{\tau}{\tau_0}\right)\right)+\tau(\beta_0-\beta)^T E_n[XX^T](\beta_0-\beta)$$
$$+(\tau-\tau_0)(E_n[g^2]n^{-2\alpha_1}+E_n[h]n^{-2\alpha_2})+2\tau(\beta_0-\beta)^T E_n[Xg]n^{-\alpha_1}.$$

ここで Norets (2015) の補題 10 より、$(\beta,\tau)$ に依存しないある定数 $C_1, C_2>0$ に対して、第 1 項は $\log\left(\frac{\tau_0}{\tau}\right)-\left(1-\frac{\tau}{\tau_0}\right)>C_1|\tau_0-\tau|^2$、第 2 項は $\tau(\beta_0-\beta)^T E_n[XX^T](\beta_0-\beta)>C_2\|\beta_0-\beta\|^2$ と下から押さえられる。一方、分解の第 3 項と第 4 項は負になりうるが、A2 の下で、この $C_1, C_2$ に対して十分大きな $n$ が存在して、任意の $(\beta,\tau)$ に対して $|(\tau-\tau_0)(E_n[g^2]n^{-2\alpha_1}+E_n[h]n^{-2\alpha_2})|<C_1|\tau_0-\tau|^2$ そして $|2\tau(\beta_0-\beta)^T E_n[Xg]n^{-\alpha_1}|<C_2\|\beta_0-\beta\|^2$ が成り立つ。したがって、普遍的な定数 $C>0$ が存在して、十分大きな $n$ で、$\|(\beta,\tau)-(\beta_0,\tau_0)\|>\gamma$ となる任意の $(\beta,\tau)$ に対して $E_n\left[\ell_{\beta_0,\tau_0}-\ell_{\beta,\tau}\right]\geq C\,\|(\beta,\tau)-(\beta_0,\tau_0)\|^2>C\gamma^2$ と下から押さえられる。ここで $\delta(\gamma)=C\gamma^2$ とおけば、$\|(\beta,\tau)-(\beta_0,\tau_0)\|>\gamma$ を満たす任意の $(\beta,\tau)$ に対して、w.p.a.1 で $\frac{1}{n}\sum_{i=1}^{n}\{\ell_{\beta,\tau}-\ell_{\beta_0,\tau_0}\}\leq-\delta(\gamma)$ が従う。この $\delta(\gamma)$ は $(\beta,\tau)$ に一様にとれるため、これは (5) 式を意味する。(4) 式と (5) 式が示されたので、定理の結果が従う。 □

**定理 2 の証明.** 結果は次を示すことに等しい。

$$E_n[(Y-X^T\beta_n^*)^2XX^T]-E_n(Y-X^T\beta_n^*)^2E_n[XX^T]=o(1). \tag{7}$$

このために、まず (1) より、次の二つの式が成り立つ：

$$E_n(Y-X^T\beta_n^*)^2=E_n[(\beta_0-\beta_n^*)^TXX^T(\beta_0-\beta_n^*)]+n^{-2\alpha_1}E_n[g^2]$$
$$+n^{-2\alpha_2}E_n[h]+2n^{-\alpha_1}E_n[(\beta_0-\beta_n^*)^TXg]+\sigma_0^2,$$
$$E_n[(Y-X^T\beta_n^*)^2XX^T]=E_n[(\beta_0-\beta_n^*)^TXX^T(\beta_0-\beta_n^*)XX^T]+n^{-2\alpha_1}E_n[g^2XX^T]$$
$$+n^{-2\alpha_2}E_n[hXX^T]+2n^{-\alpha_1}E_n[((\beta_0-\beta_n^*)^TXg)XX^T]+\sigma_0^2E_n[XX^T].$$

ここで $\widetilde{XX^T}=XX^T-E_n[XX^T]$ とおき、次を定義する。

$$D_1=E_n\left[(\beta_0-\beta_n^*)^TXX^T(\beta_0-\beta_n^*)\widetilde{XX^T}\right],\quad D_2=2n^{-\alpha_1}E_n\left[((\beta_0-\beta_n^*)^TXg)\widetilde{XX^T}\right],$$
$$D_3=n^{-2\alpha_1}E_n\left[g^2\widetilde{XX^T}\right],\quad D_4=n^{-2\alpha_2}E_n\left[h\widetilde{XX^T}\right].$$

このとき、(7) の左辺は $D_1+D_2+D_3+D_4$ と分解できる。ところが、行列ノルムの三角不等式、A3、そして Cauchy-Schwarz の不等式より、

$$\|D_1\|\leq\sum_{j=1}^{k}\sum_{l=1}^{k}\left|(\beta_0-\beta_n^*)_j\right|\,|(\beta_0-\beta_n^*)_l|\left\|E_n\left[(XX^T)_{j\,l}\widetilde{XX^T}\right]\right\|$$

$$\leq\max_{1\leq j,l\leq k}\left\|E_n\left[(X_jX_l)\widetilde{XX^T}\right]\right\|\sum_{j=1}^{k}\left|(\beta_0-\beta_n^*)_j\right|\sum_{l=1}^{k}|(\beta_0-\beta_n^*)_l|=O(\|\beta_n^*-\beta_0\|^2).$$

同様に、A2 と A3 より $\|D_2\|\leq 2n^{-\alpha_1}\sum_{j=1}^{k}\left|(\beta_0-\beta_n^*)_j\right|\left\|E_n\left[\left(X_jg(X)\right)\widetilde{XX^T}\right]\right\|=O(n^{-\alpha_1}\|\beta_n^*-\beta_0\|)$, $\|D_3\|=O(n^{-2\alpha_1})$, $\|D_4\|=O(n^{-2\alpha_2})$ が従う。したがって、補題 2 より (7) の左辺は

$O\left(n^{-2(\alpha_1 \wedge \alpha_2)}\right)$ であるが、これは $o(1)$ を意味する。 □