2025年3月10日最終更新

Priors for Second-Order Unbiased Bayes Estimators

arXiv:2412.19187

酒井 真菜　東大経済・理研AIP

松田 孟留　東大情報理工・理研CBS

久保川 達也　東大経済

アウトライン

1. イントロダクション
2. 主結果1: ベイズ推定量のバイアスの評価
3. 主結果2: 事前分布の構成
4. 例
5. シミュレーション
6. まとめ

アウトライン

1. イントロダクション
2. 主結果1: ベイズ推定量のバイアスの評価
3. 主結果2: 事前分布の構成
4. 例
5. シミュレーション
6. まとめ

背景

- ベイズ推論における様々な無情報事前分布が提案されてきた.
  - Jeffreys' prior, reference prior, etc.
- Hartigan (1965) は，ある事前分布に対応するベイズ推定量のバイアスがサンプルサイズ n に対して $o(n^{-1})$ になるとき，その事前分布は漸近不偏 (asymptotically unbiased) であると呼んだ.
  - 一般に，ベイズ推定量は $O(n^{-1})$ のバイアスを持つ.
  - 特に，サンプルサイズが小さいときに有用.
- Hartigan (1965) の結果はモデルのi.i.d.性を仮定しているため，限定的.
  - E.g. ベイズ回帰分析では，説明変数 x_i を固定する.

    → データはi.i.d.ではないため，Hartigan (1965) の結果は使えない.

本研究の貢献

1. Asymptotically unbiased priors が満たすべき条件を，i.i.d.ではない状況へ拡張した.
   - 事前分布は，偏微分方程式系の解として特徴づけられる.
2. 上の偏微分方程式系の解が存在する必要十分条件を検討し，asymptotically unbiased priors の構成手順をまとめた.
   - これらの結果は，moment matching priors (Ghosh and Liu, 2011) など他の事前分布にもそのまま適用可能.
3. Nested error regression (NER) モデル (変量効果モデル) に応用し，新しい事前分布の提案を行った.

アウトライン

1. イントロダクション
2. 主結果1: ベイズ推定量のバイアスの評価
3. 主結果2: 事前分布の構成
4. 例
5. シミュレーション
6. まとめ

設定

- $(X_1, \dots, X_n)$ は確率密度関数 $f_n(x_1, \dots, x_n|\theta)$ を持つ.
  - i.i.d.性や独立性は仮定しない！
- $\theta = (\theta_1, \dots, \theta_p) \in \Theta$: p次元のパラメータ.
- $\Theta \subset \mathbb{R}^p$: パラメータ空間 (rectangular).
- $\ell_n(\theta) = \log f_n(x_1, \dots, x_n|\theta)$: 対数尤度.
- $\pi(\theta)$: θ の事前分布の密度.
- これらの設定のもとで，パラメータ θ の**事後平均**は

$$\hat{\theta}^B = \frac{\int_\Theta \theta \pi(\theta) \exp\bigl(\ell_n(\theta)\bigr)\, d\theta}{\int_\Theta \pi(\theta) \exp\bigl(\ell_n(\theta)\bigr)\, d\theta}.$$

Asymptotically unbiased priors

定義 (asymptotically unbiased prior)

ある事前分布に対応するベイズ推定量 (= 事後平均) のバイアスが$o(n^{-1})$であるとき，その事前分布は漸近不偏 (asymptotically unbiased) であるという．また，このとき，ベイズ推定量は second-order unbiased（2次不偏）であるという．

- Hartigan (1965) では，一般の損失関数を考えているので，上の定義とは表現が異なる．
- 2乗損失を考える場合には， Hartigan (1965) の定義と上の定義は一致．

ベイズ推定量のバイアスの評価

Corollary 2
正則条件のみから導出される結果

↓ 対数尤度の微分の極限に関する仮定を追加

Corollary 3
追加的な仮定のもとでの結果

↓ i.i.d.性の仮定を追加

Theorem 4
i.i.d.の場合に一般に成り立つ結果

Assumption 1（正則条件）

(i) $\hat{\theta}^{ML} - \theta = O_p\left(n^{-1/2}\right)$. ただし $\hat{\theta}^{ML}$ はモデルの最尤推定量.

(ii) $\ell_n(\theta) = O_p(n)$.

(iii) $\ell_n(\theta)$ は3回連続微分可能.

(iv) $\pi(\theta)$ は微分可能.

(v) $H_n(\theta) = -\frac{\partial^2}{\partial\theta\partial\theta^\top}\ell_n(\theta)$は可逆. $H_n(\hat{\theta}^{ML})$ は正定値.

ノーテーション

- $H_n(\theta) = -\frac{1}{n}\frac{\partial^2}{\partial\theta\partial\theta^\top}\ell_n(\theta).$
- $I_{n,rs}(\theta) = \frac{1}{n}\left(\frac{\partial\ell_n(\theta)}{\partial\theta_r}\right)\left(\frac{\partial\ell_n(\theta)}{\partial\theta_s}\right).$
- $J_{n,rs,t}(\theta) = \frac{1}{n}\left(\frac{\partial^2}{\partial\theta_r\partial\theta_s}\ell_n(\theta)\right)\left(\frac{\partial}{\partial\theta_t}\ell_n(\theta)\right).$
- $K_{n,rst}(\theta) = \frac{1}{n}\frac{\partial^3}{\partial\theta_r\partial\theta_s\partial\theta_t}\ell_n(\theta).$

正則条件の下でのベイズ推定量のバイアスの評価

Corollary 2 (正則条件の下でのバイアスの評価)

Assumption 1 の下で，事後平均 $\hat{\theta}^B$ のバイアスは

$$\mathbb{E}\left(\hat{\theta}^B - \theta\right)$$

$$= \frac{1}{n}\mathbb{E}\left[H_n^{-1}(\theta)\left(\frac{\partial}{\partial\theta}\left(\log\pi(\theta) + 2\ell_n(\theta)\right) + \frac{1}{2}\sum_{r=1}^{p}\sum_{s=1}^{p} H_n^{rs}(\theta)A_{n,rs}(\theta)\right)\right] + o\left(\frac{1}{n}\right).$$

ただし，

$$A_{n,rs}(\theta) = \begin{bmatrix} K_{n,1rs}(\theta) + 2J_{n,1r,s}(\theta) \\ \vdots \\ K_{n,prs}(\theta) + 2J_{n,pr,s}(\theta) \end{bmatrix} + \sum_{t=1}^{p}\sum_{u=1}^{p} H_n^{tu}(\theta)I_{n,su}(\theta)\begin{bmatrix} K_{n,rt1}(\theta) \\ \vdots \\ K_{n,rtp}(\theta) \end{bmatrix}.$$

$$H_n(\theta) = -\frac{1}{n}\frac{\partial^2}{\partial\theta\partial\theta^\top}\ell_n(\theta) \qquad J_{n,rs,t}(\theta) = \frac{1}{n}\left(\frac{\partial^2}{\partial\theta_r\partial\theta_s}\ell_n(\theta)\right)\left(\frac{\partial}{\partial\theta_t}\ell_n(\theta)\right)$$

$$K_{n,rst}(\theta) = \frac{1}{n}\frac{\partial^3}{\partial\theta_r\partial\theta_s\partial\theta_t}\ell_n(\theta) \qquad I_{n,rs}(\theta) = \frac{1}{n}\left(\frac{\partial\ell_n(\theta)}{\partial\theta_r}\right)\left(\frac{\partial\ell_n(\theta)}{\partial\theta_s}\right)$$

Corollary 2 の結果について

- Corollary 2 の結果は，期待値の計算が大変なので，一般にはあまり使い勝手が良くない.

- $H_n(\theta) = -\frac{\partial^2}{\partial\theta\partial\theta^\top}\ell_n(\theta)$ が確率的でないならば，

$$\mathbb{E}\left(\hat{\theta}^B - \theta\right) = H_n^{-1}(\theta)\left(\frac{\partial}{\partial\theta}\log\pi(\theta) + \frac{1}{2}\sum_{r=1}^{p}\sum_{s=1}^{p} H_n^{rs}(\theta)\mathbb{E}\left[A_{n,rs}(\theta)\right]\right) + o(n^{-1}).$$

よって，

$$\frac{\partial}{\partial\theta}\log\pi(\theta) = -\frac{1}{2}\sum_{r=1}^{p}\sum_{s=1}^{p} H_n^{rs}(\theta)\mathbb{E}\left[A_{n,rs}(\theta)\right]$$

$\Longrightarrow \quad \hat{\theta}^B$ は2次不偏.

事前分布は n に依存

Assumption 2（対数尤度の微分の極限の仮定）

(i) 可逆な $p \times p$ 次元定数行列 $H(\theta)$ が存在し，

$$H_n(\theta) = H(\theta) + O_p\left(n^{-1/2}\right).$$

(ii) 各 $r, s, t \in \{1, \dots, p\}$ に対して定数 $K_{rst}(\theta)$ が存在し，

$$K_{n,rst}(\theta) = K_{rst}(\theta) + o_p(1).$$

(iii) $p \times p$ 次元定数行列 $I(\theta)$ が存在し，

$$\mathbb{E}\left[I_{n,rs}(\theta)\right] = I(\theta) + o(1).$$

(iv) 各 $r, s, t \in \{1, \dots, p\}$ に対して定数 $J_{rs,t}(\theta)$ が存在し，

$$\mathbb{E}\left[J_{n,rs,t}(\theta) =\right] = J_{rs,t}(\theta) + o(1).$$

追加的な仮定の下でのベイズ推定量のバイアスの評価

Corollary 3 (追加的な仮定の下でのバイアスの評価)

Assumptions 1, 2の下で，事後平均 $\hat{\theta}^B$ のバイアスは

$$\mathbb{E}\left(\hat{\theta}^B - \theta\right) = \frac{1}{n} H^{-1}(\theta) \left(\frac{\partial}{\partial \theta} \log \pi(\theta) + \frac{1}{2} \sum_{r=1}^{p} \sum_{s=1}^{p} H^{rs}(\theta) A_{rs}(\theta) \right) + o\left(\frac{1}{n}\right).$$

ただし，

$$A_{rs}(\theta) = \begin{bmatrix} K_{1rs}(\theta) + 2J_{1r,s}(\theta) \\ \vdots \\ K_{prs}(\theta) + 2J_{pr,s}(\theta) \end{bmatrix} + \sum_{t=1}^{p} \sum_{u=1}^{p} H^{tu}(\theta) I_{su}(\theta) \begin{bmatrix} K_{rt1}(\theta) \\ \vdots \\ K_{rtp}(\theta) \end{bmatrix}.$$

よって，

$$\frac{\partial}{\partial \theta} \log \pi(\theta) = -\frac{1}{2} \sum_{r=1}^{p} \sum_{s=1}^{p} H^{rs}(\theta) A_{rs}(\theta)$$

$\Longrightarrow$ $\hat{\theta}^B$ は2次不偏.

i.i.d.の場合の仮定2の成立

- $X_1, \dots, X_n$ は i.i.d.で，それぞれ確率密度関数 $f(x|\theta)$ を持つとする.
- 直接計算より,

$$I(\theta) = \mathbb{E}\left[\left(\frac{\partial}{\partial\theta}\log f(X|\theta)\right)\left(\frac{\partial}{\partial\theta}\log f(X|\theta)\right)^{\top}\right]$$

(i.e. Fisher情報行列）かつ

$$J_{rs,t}(\theta) = \mathbb{E}\left[\left(\frac{\partial^2}{\partial\theta_r\partial\theta_s}\log f(X|\theta)\right)\left(\frac{\partial}{\partial\theta_t}\log f(X|\theta)\right)\right].$$

- 中心極限定理より,

$$H(\theta) = I(\theta).$$

- 大数の法則より,

$$K_{rst}(\theta) = \mathbb{E}\left[\frac{\partial^3}{\partial\theta_r\partial\theta_s\partial\theta_t}\log f\left(X|\theta\right)\right].$$

i.i.d.の場合のベイズ推定量のバイアスの評価

Corollary 4 (i.i.d.の場合のバイアスの評価)

i.i.d.のとき，事後平均 $\hat{\theta}^{\mathrm{B}}$ のバイアスは

$$\mathbb{E}\left(\hat{\theta}^B - \theta\right) = \frac{1}{n} I^{-1}(\theta) \left(\frac{\partial}{\partial \theta} \log \pi(\theta) - \sum_{r=1}^{p} \sum_{s=1}^{p} I^{rs}(\theta) \frac{\partial}{\partial \theta_s} \begin{bmatrix} I_{1r}(\theta) \\ \vdots \\ I_{pr}(\theta) \end{bmatrix} \right) + o\left(\frac{1}{n}\right).$$

よって，

$$\frac{\partial}{\partial \theta} \log \pi(\theta) = \sum_{r=1}^{p} \sum_{s=1}^{p} I^{rs}(\theta) \frac{\partial}{\partial \theta_s} \begin{bmatrix} I_{1r}(\theta) \\ \vdots \\ I_{pr}(\theta) \end{bmatrix}$$

Fisher 情報のみから計算可能！

$\Longrightarrow$ $\hat{\theta}^B$ は2次不偏.

$I(\theta) = \mathbb{E}\left[\left(\frac{\partial}{\partial \theta} \log f(X|\theta)\right)\left(\frac{\partial}{\partial \theta} \log f(X|\theta)\right)^{\top}\right]$: Fisher 情報行列

Corollary 4と関連研究の比較

	特徴付け	性質
本研究 Corollary 4	$\frac{\partial}{\partial\theta}\log\pi(\theta) = \sum_{r=1}^{p}\sum_{s=1}^{p} I^{rs}(\theta)(-1)\begin{bmatrix}\frac{\partial}{\partial\theta_s}I_{1r}(\theta)\\ \vdots \\ \frac{\partial}{\partial\theta_s}I_{pr}(\theta)\end{bmatrix} = \underbrace{\sum_{r=1}^{p}\sum_{s=1}^{p} I^{rs}(\theta)\begin{bmatrix}J_{1r,s}(\theta)+K_{1rs}(\theta)\\ \vdots \\ J_{pr,s}(\theta)+K_{prs}(\theta)\end{bmatrix}}_{①}$	**事後平均**が 2次不偏
Firth's method	$\frac{\partial}{\partial\theta}\ell_n(\theta) = \underbrace{\sum_{r=1}^{p}\sum_{s=1}^{p} I^{rs}(\theta)\left(\begin{bmatrix}J_{1r,s}(\theta)\\ \vdots \\ J_{pr,s}(\theta)\end{bmatrix} + \frac{1}{2}\begin{bmatrix}K_{1rs}(\theta)\\ \vdots \\ K_{prs}(\theta)\end{bmatrix}\right)}_{②}$	**MLE**の 2次バイアス補正
Moment matching prior	$\frac{\partial}{\partial\theta}\log\pi(\theta) = \underbrace{\sum_{r=1}^{p}\sum_{s=1}^{p} I^{rs}(\theta)\frac{1}{2}\begin{bmatrix}K_{1rs}(\theta)\\ \vdots \\ K_{prs}(\theta)\end{bmatrix}}_{③}$	**MLE**と**事後平均**が $o_p(1/n)$ で一致

① = ② + ③ という関係が成立している.

→Asymptotically unbiased prior は，moment matching prior の Firth's method によるバイアス補正であると解釈できる.

これまでのまとめ

$\frac{\partial}{\partial\theta}\log\pi(\theta)=\phi(\theta)$ の解 $\pi(\theta)$ に対応するベイズ推定量は2次不偏.

Corollary	$\phi(\theta)$	特徴
2	$-\frac{1}{2}\sum_{r=1}^{p}\sum_{s=1}^{p}H_n^{rs}(\theta)\mathbb{E}[A_{n,rs}(\theta)]$	n に依存する事前分布. $H_n(\theta)$が確率的でないときしか使えない.
3	$-\frac{1}{2}\sum_{r=1}^{p}\sum_{s=1}^{p}H^{rs}(\theta)A_{rs}(\theta)$	対数尤度の微分の極限に関する仮定が必要.
4	$\sum_{r=1}^{p}\sum_{s=1}^{p}I^{rs}(\theta)\frac{\partial}{\partial\theta_s}\begin{bmatrix}I_{1r}(\theta)\\ \vdots\\ I_{pr}(\theta)\end{bmatrix}$	i.i.d.の場合に標準的に使える.

仮定弱い
使いにくい

仮定強い
使いやすい

- $A_{n,rs}(\theta)=\begin{bmatrix}K_{n,1rs}(\theta)+2J_{n,1r,s}(\theta)\\ \vdots\\ K_{n,prs}(\theta)+2J_{n,pr,s}(\theta)\end{bmatrix}+\sum_{t=1}^{p}\sum_{u=1}^{p}H_n^{tu}(\theta)\left(\frac{\partial\ell_n(\theta)}{\partial\theta_s}\right)\left(\frac{\partial\ell_n(\theta)}{\partial\theta_u}\right)\begin{bmatrix}K_{n,rt1}(\theta)\\ \vdots\\ K_{n,rtp}(\theta)\end{bmatrix}$
- $A_{rs}(\theta)=\begin{bmatrix}K_{1rs}(\theta)+2J_{1r,s}(\theta)\\ \vdots\\ K_{prs}(\theta)+2J_{pr,s}(\theta)\end{bmatrix}+\sum_{t=1}^{p}\sum_{u=1}^{p}H^{tu}(\theta)I_{su}(\theta)\begin{bmatrix}K_{rt1}(\theta)\\ \vdots\\ K_{rtp}(\theta)\end{bmatrix}$

アウトライン

1. イントロダクション
2. 主結果1: ベイズ推定量のバイアスの評価
3. 主結果2: 事前分布の構成
4. 例
5. シミュレーション
6. まとめ

Asymptotically unbiased priors の存在条件

一般に $\frac{\partial}{\partial\theta}\log\pi(\theta) = \phi(\theta)$ となる条件を考える.

Theorem 5 (偏微分方程式系の解の存在条件)

$\phi(\theta)$ は微分可能であり，$\phi(\theta)$ の θ に関する微分と積分の順序を交換できるとする．このとき，$\frac{\partial}{\partial\theta}\log\pi(\theta) = \phi(\theta)$ を満たす2回連続微分可能な $\pi(\theta)$ が存在するための必要十分条件は，任意の $t, u \in \{1, \dots, p\}$ について

$$\frac{\partial}{\partial\theta_u}\phi_t(\theta) = \frac{\partial}{\partial\theta_t}\phi_u(\theta)$$

可積分条件

が成り立つことである.

Theorem 5 の証明の概略

(解が存在 $\Longrightarrow$ 可積分条件)

- $\frac{\partial}{\partial \theta} \log \pi(\theta) = \phi(\theta)$ より

$$\frac{\partial^2}{\partial \theta \partial \theta^\top} \log \pi(\theta) = \frac{\partial}{\partial \theta} \phi(\theta).$$

- 左辺は対称行列なので，右辺も対称行列．よって，

$$\frac{\partial}{\partial \theta_u} \phi_t(\theta) = \frac{\partial}{\partial \theta_t} \phi_u(\theta) \qquad (t, u \in \{1, \ldots, p\}).$$

(可積分条件 $\Longrightarrow$ 解が存在)

- 可積分条件が成り立つとき，次の Corollary 7 の手順で解を構成できる．

事前分布の構成手順

Corollary 7 (Asymptotically unbiased prior の構成手順)

以下の手順によって構成された $\pi(\theta)$ は，可積分条件の下で

$$\frac{\partial}{\partial\theta}\log\pi(\theta) = \phi(\theta)$$

を満たす．

(i) 定数ベクトル $(c_1, \ldots, c_p) \in \Theta$ を任意に固定する．

(ii) $\psi_t(\theta_t, \ldots, \theta_p) \coloneqq \phi_t(c_1, \ldots, c_{t-1}, \theta_t, \ldots, \theta_p)$ を用いて，

$$\tilde{\pi}(\theta) \coloneqq \exp\left(\sum_{t=1}^{p}\int_{c_t}^{\theta_t}\psi_t(z, \theta_{t+1}, \ldots, \theta_p)dz\right)$$

を計算する．

$\theta_1, \ldots, \theta_p$ と順番に積分して和を取る

(iii) $\pi(\theta) \propto \tilde{\pi}(\theta)$ とする．

Corollary 7の証明

$s > r$ について，可積分条件より

$$\frac{\partial}{\partial \theta_s}\psi_r(\theta_r, \dots, \theta_p) = \frac{\partial}{\partial \theta_s}\phi_r(\theta)\Big|_{(\theta_1,\dots,\theta_{r-1})=(c_1,\dots,c_{r-1})} = \frac{\partial}{\partial \theta_r}\phi_s(\theta)\Big|_{(\theta_1,\dots,\theta_{r-1})=(c_1,\dots,c_{r-1})}$$

が成り立つことに注意すると，$s = 1, \dots, p$ について

$$\frac{\partial}{\partial \theta_s}\log\tilde{\pi}(\theta) = \sum_{r=1}^{s-1}\int_{c_r}^{\theta_r}\left(\frac{\partial}{\partial \theta_s}\psi_r(z, \theta_{r+1}, \dots, \theta_p)\right)dz + \psi_s(\theta_s, \dots, \theta_p)$$

$$= \sum_{r=1}^{s-1}\int_{c_r}^{\theta_r}\left(\frac{\partial}{\partial z}\phi_s(\theta_1, \dots, \theta_{r-1}, z, \theta_{r+1}, \dots, \theta_p)\Big|_{(\theta_1,\dots,\theta_{r-1})=(c_1,\dots,c_{r-1})}\right)dz + \psi_s(\theta_s, \dots, \theta_p)$$

$$= \sum_{r=1}^{s-1}\left(\phi_s(c_1, \dots, c_{r-1}, \theta_r, \dots, \theta_p) - \phi_s(c_1, \dots, c_r, \theta_{r+1}, \dots, \theta_p)\right) + \phi_s(c_1, \dots, c_{s-1}, \theta_s, \dots, \theta_p)$$

$$= \phi_s(\theta_1, \dots, \theta_p).$$

Asymptotically unbiased prior の導出

$\frac{\partial}{\partial \theta} \log \pi(\theta) = \phi(\theta)$ を満たす $\pi(\theta)$ を導出するために…

可積分条件を確認

↓

Corollary 7 の構成手順を用いて事前分布を構成

事前分布の構成方法の汎用性について

- 前ページの導出手順は，asymptotically unbiased prior 以外にも，

$$\frac{\partial}{\partial\theta}\log\pi(\theta) = \phi(\theta)$$

という形で定義される $\pi(\theta)$ の導出に対して一般に応用可能である.

  - Moment matching prior などでも同様の手順で解の存在の確認と事前分布の導出が行える.

アウトライン

1. イントロダクション
2. 主結果1: ベイズ推定量のバイアスの評価
3. 主結果2: 事前分布の構成
4. 例
5. シミュレーション
6. まとめ

1次元パラメータ i.i.d.モデル

- モデルがi.i.d.かつパラメータが1次元 ($p = 1$) の場合， 可積分条件は常に成り立つ.
- Corollary 7 の手順によって求める事前分布を導出する：
  - $\phi_1(\theta) = I^{-1}(\theta)I'(\theta) = \psi_1(\theta)$.
  - 任意の $c \in \Theta$について,
$$\tilde{\pi}(\theta) = \exp\left(\int_c^\theta \frac{I'(z)}{I(z)} dz\right) = \exp(\log I(\theta) - \log I(c)) = \frac{I(\theta)}{I(c)}.$$
  - よって， $\pi(\theta) \propto I(\theta)$.

- $\phi(\theta) = \sum_{r=1}^p \sum_{s=1}^p I^{rs}(\theta) \frac{\partial}{\partial \theta_s} \begin{bmatrix} I_{1r}(\theta) \\ \vdots \\ I_{pr}(\theta) \end{bmatrix}$
- $\psi_t(\theta_t, \dots, \theta_p) = \phi_t(c_1, \dots, c_{t-1}, \theta_t, \dots, \theta_p)$
- $\tilde{\pi}(\theta) = \exp\left(\sum_{t=1}^p \int_{c_t}^{\theta_t} \psi_t(z, \theta_{t+1}, \dots, \theta_p) dz\right)$

Example 1: 二項分布

- $X_1, \dots, X_n \overset{i.i.d.}{\sim} Ber(\theta)$ とする.
- Fisher 情報量は, $I(\theta) = \theta^{-1}(1-\theta)^{-1}$.
- 事前分布を $\pi(\theta) \propto I(\theta) = \theta^{-1}(1-\theta)^{-1}$ とするとき, 事後分布は
$$\pi(\theta | x_1, \dots, x_n) \propto \theta^{n\bar{X}-1}(1-\theta)^{n(1-\bar{X})-1}.$$
    - すなわち, $\theta | X_1, \dots, X_n \sim \mathrm{Beta}(n\bar{X}, n(1-\bar{X}))$.
- これより, $\hat{\theta}^B = \bar{X}$ であるが, これは θ の不偏推定量.

Example 2: 正規分布（期待値，分散パラメータ）

- $X_1, \dots, X_n \overset{i.i.d.}{\sim} N(\mu, \sigma^2)$， $\theta = (\theta_1, \theta_2) = (\mu, \sigma^2)$ とする.
- $I(\theta) = \begin{bmatrix} \sigma^{-2} & 0 \\ 0 & \sigma^{-4}/2 \end{bmatrix}$ より $\phi_1(\theta) = 0,\ \phi_2(\theta) = -\frac{2}{\sigma^2}$ と計算できる．したがって $\frac{\partial}{\partial \theta_2}\phi_1(\theta) = \frac{\partial}{\partial \theta_1}\phi_2(\theta) = 0$ なので，可積分条件が成立.
- $\psi_1(\mu, \sigma) = 0,\ \psi_2(\sigma) = -\frac{2}{\sigma^2}$ より， $\pi(\theta) \propto \tilde{\pi}(\theta) = \exp\left(0 - \int_c^{\sigma^2} \frac{2}{z} dz\right) \propto \frac{1}{\sigma^4}$.
- $\bar{X} := \frac{1}{n}\sum_{i=1}^n X_i$,$S^2 := \sum_{i=1}^n (X_i - \bar{X})^2$ と書くとき事後平均

$$(\hat{\mu}^B, \hat{\sigma}^{2,B}) = \left(\bar{X}, \frac{1}{n-1} S^2\right)$$

は (μ, σ^2) の不偏推定量.

$$\phi(\theta) = \sum_{r=1}^p \sum_{s=1}^p I^{rs}(\theta) \frac{\partial}{\partial \theta_s} \begin{bmatrix} I_{1r}(\theta) \\ \vdots \\ I_{pr}(\theta) \end{bmatrix}$$

$$\psi_t(\theta_t, \dots, \theta_p) = \phi_t(c_1, \dots, c_{t-1}, \theta_t, \dots, \theta_p)$$

$$\tilde{\pi}(\theta) = \exp\left(\sum_{t=1}^p \int_{c_t}^{\theta_t} \psi_t(z, \theta_{t+1}, \dots, \theta_p) dz\right)$$

Example 5: 線形回帰モデル

- モデルは,

$$y_i = x_i^\top \beta + \epsilon_i \qquad (i = 1, \ldots, n)$$

- $y_i \in \mathbb{R}, x_i \in \mathbb{R}^p$を観測.
- $\epsilon_1, \ldots, \epsilon_n \overset{i.i.d.}{\sim} N(0, \sigma^2)$.
- x_i は非確率的であるとする (conditional inference).

  → モデルは i.i.d.ではない！
- モデルのパラメータは, $\theta = \left(\theta_1, \ldots, \theta_p, \theta_{p+1}\right) = (\beta, \sigma^2) \in \mathbb{R}^{p+1}$.
- 求める事前分布は

$$\pi(\theta) \propto \sigma^{-4}.$$

線形回帰モデルの事後分布

- 事後分布は,

$$\beta|\sigma^2, y_1, \ldots, y_n \sim N\left(\hat{\beta}^{OLS}, \sigma^2 (X^\top X)^{-1}\right),$$

$$\sigma^2|y_1, \ldots, y_n \sim \mathrm{IG}\left(\frac{n-p}{2} + 1, \frac{1}{2} y^\top (I_n - X(X^\top X)^{-1} X^\top) y\right).$$

- したがって，ベイズ推定量は

$$\hat{\beta}^B = \hat{\beta}^{OLS},$$

$$\hat{\sigma}^{2,B} = \frac{1}{n-p} (y^\top (I_n - X(X^\top X)^{-1} X^\top) y) \quad \text{if} \quad n > p.$$

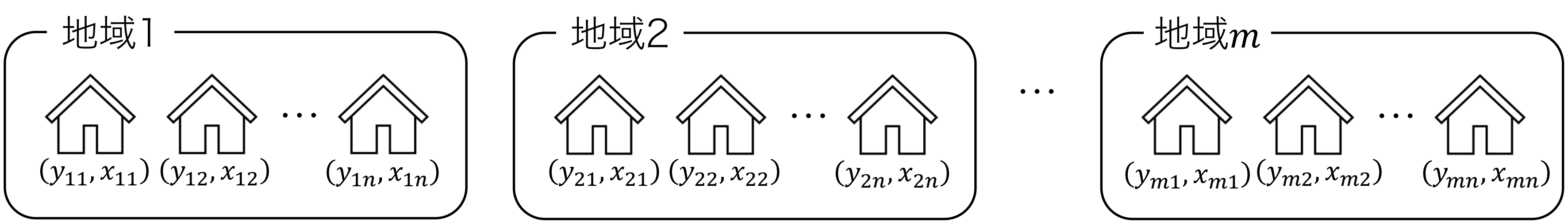
- これらは，$\theta = (\beta, \sigma^2)$ の exact な不偏推定量である！

Examples 6&7: Nested error regression (NER)

- m：地域数，n：地域内のユニット数（固定）とするとき，

$$y_{ij} = x_{ij}^\top \beta + v_i + \epsilon_{ij} \quad (i = 1, \dots, m,\ j = 1, \dots, n)$$

- $y_{ij} \in \mathbb{R},\ x_{ij} \in \mathbb{R}^p$を観測．
- $v_i \sim N(0, \tau^2)$，$\epsilon_{i1}, \dots, \epsilon_{in} \overset{i.i.d.}{\sim} N(0, \sigma^2)$で，$v_i$と$\epsilon_{ij}$は独立．
- モデルのパラメータは，$\theta = (\theta_1, \dots, \theta_p, \theta_{p+1}, \theta_{p+2}) = (\beta, \tau^2, \sigma^2) \in \mathbb{R}^{p+2}$.
- $m \to \infty$という状況を考える．nは固定.
- データは i に関して独立であるとする.
- このとき，asymptotically unbiased prior は $\pi(\theta) \propto [\sigma^2(\sigma^2 + n\tau^2)]^{-2}$.



NERモデルの Gibbs sampler

- $\theta = (\beta, \tau^2, \sigma^2)$ を直接サンプリングする代わりに， $\bar{\theta} = (\beta, \rho, \sigma^2)$ をサンプリングする．ただし，

$$\rho = \frac{\sigma^2}{\sigma^2 + n\tau^2} \in (0,1).$$

- $\bar{\theta}$ のパラメータの full conditional は，

$$\beta_r \mid \beta_{-r}, \rho, \sigma^2, \mathcal{D} \sim N\left(\frac{\sum_{i=1}^m \left(y_i - \sum_{s \neq r} x_{i\cdot,s}\,\beta_s\right)^\top \bar{V}^{-1} x_i}{\sum_{i=1}^m x_{i\cdot,r}^\top \bar{V}^{-1} x_{i\cdot,r}},\ \left(\sum_{i=1}^m x_{i\cdot,r}^\top \bar{V}^{-1} x_{i\cdot,r}\right)^{-1}\right),$$

$$\rho \mid \beta, \sigma^2, \mathcal{D} \sim \text{Truncated Gamma}\left(\frac{m}{2} + 1,\ \frac{1}{2n\sigma^2} \sum_{i=1}^m (y_i - x_i\beta)^\top \iota_n \iota_n^\top (y_i - x_i\beta)\right)$$

truncated on (0,1),

$$\sigma^2 \mid \beta, \rho, \mathcal{D} \sim \text{IG}\left(\frac{nm}{2} + 2,\ \frac{1}{2}\sum_{i=1}^m (y_i - x_i\beta)^\top \left(I_n - \frac{1-\rho}{n}\iota_n \iota_n^\top\right)(y_i - x_i\beta)\right).$$

- ただし $\bar{V}^{-1} = \sigma^{-2}\left(I_n - \frac{(1-\rho)}{n}\iota_n\iota_n^\top\right), x_{i\cdot,r} = \left(x_{i1,r}, \dots, x_{in,r}\right)$.

アウトライン

1. イントロダクション
2. 主結果1: ベイズ推定量のバイアスの評価
3. 主結果2: 事前分布の構成
4. 例
5. シミュレーション
6. まとめ

NERモデルのシミュレーション

設定

- 地域内のユニット数：$n = 5$.
- 地域数 $m \in \{10,\ 32,\ 100,\ 316,\ 1000\}$ の5通りで推定.

目的

- 各 m に対し，事後平均 $\hat{\theta}^B$ のバイアス $\mathbb{E}(\hat{\theta}^B - \theta)$ を評価する.

シミュレーションの手順

各 $m \in \{10,\ 32,\ 100,\ 316,\ 1000\}$ について,

1. 真の分布から $\mathcal{D}^m = \{(y_i, x_i): i \leq \mathrm{m}\}$ を10000回発生させる．それぞれ $\mathcal{D}^{m,1}, \dots, \mathcal{D}^{m,10000}$ と呼ぶ.
2. 各 $\mathcal{D}^{m,k}$ に対して，対応する事後平均 $\hat{\theta}^{m,k}$ を計算する.
   - MCMC (Gibbs sampler) を用いて事後分布からのサンプリングを行う.
3. 平均 $\hat{\mathbb{E}}(\hat{\theta}) := \frac{1}{10000}\sum_{k=1}^{10000} \hat{\theta}^{m,k}$ によって $\mathbb{E}(\hat{\theta})$ を近似する.

比較する事前分布

[AU] Asymptotically unbiased prior: $\pi(\theta) \propto [\sigma^2(\sigma^2 + n\tau^2)]^{-2}$.

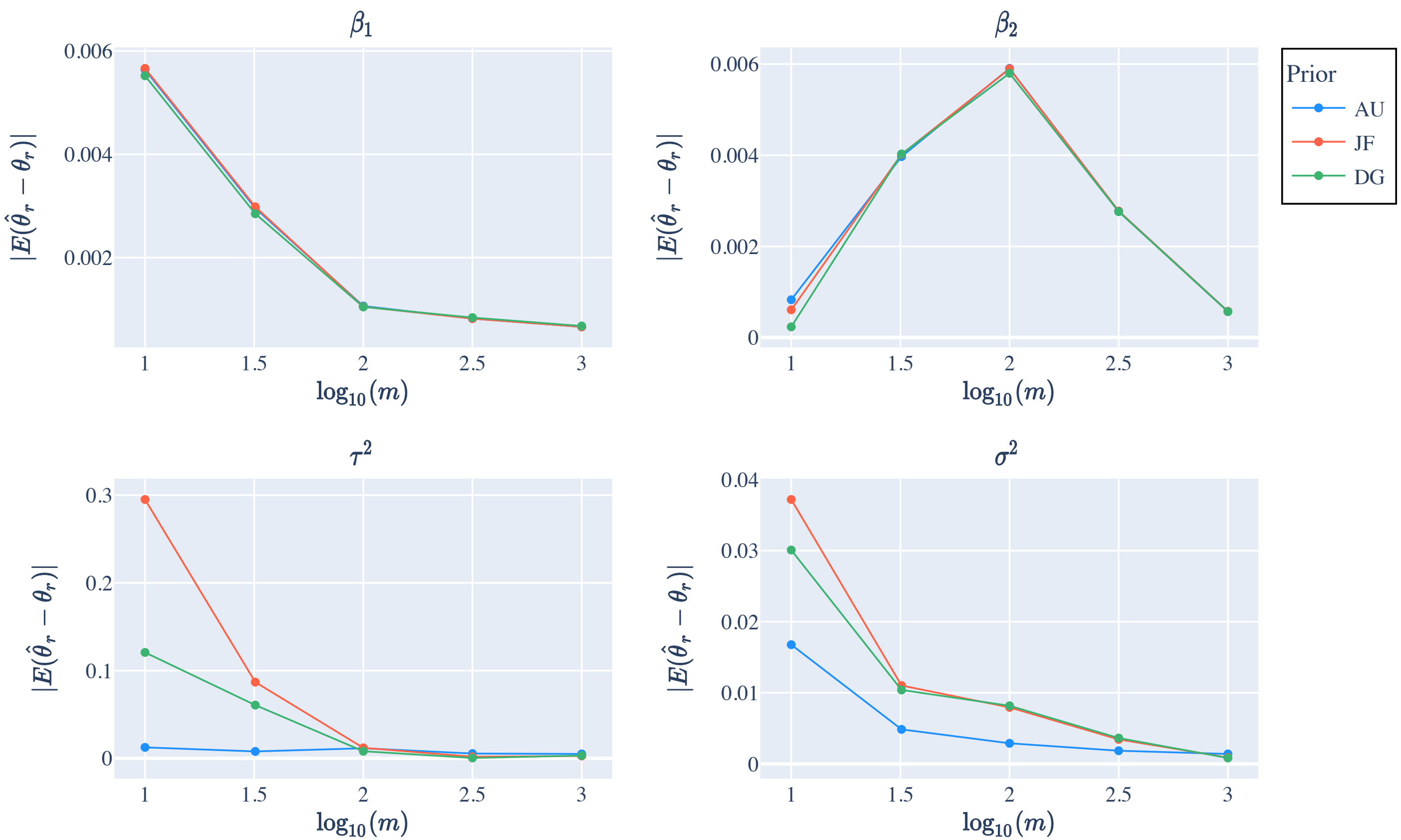
[JF] (部分的) Jeffreys' prior: $\pi(\theta) \propto [\sigma^2(\sigma^2 + n\tau^2)]^{-1}$.

- 回帰係数に関して $\pi(\beta) \propto 1$，分散成分 (σ^2, τ^2) に対してJeffreys' priorを考えたもの．Cf. Tiao and Tan (1965).

[DG] Datta and Ghosh (1991): $\pi(\beta) \propto 1,\ \tau^2 \sim \mathrm{IG}(a_\tau, b_\tau),\ \sigma^2 \sim \mathrm{IG}(a_\sigma, b_\sigma)$.

- 実験では $a_\tau = b_\tau = a_\sigma = b_\sigma = 5$ と設定.

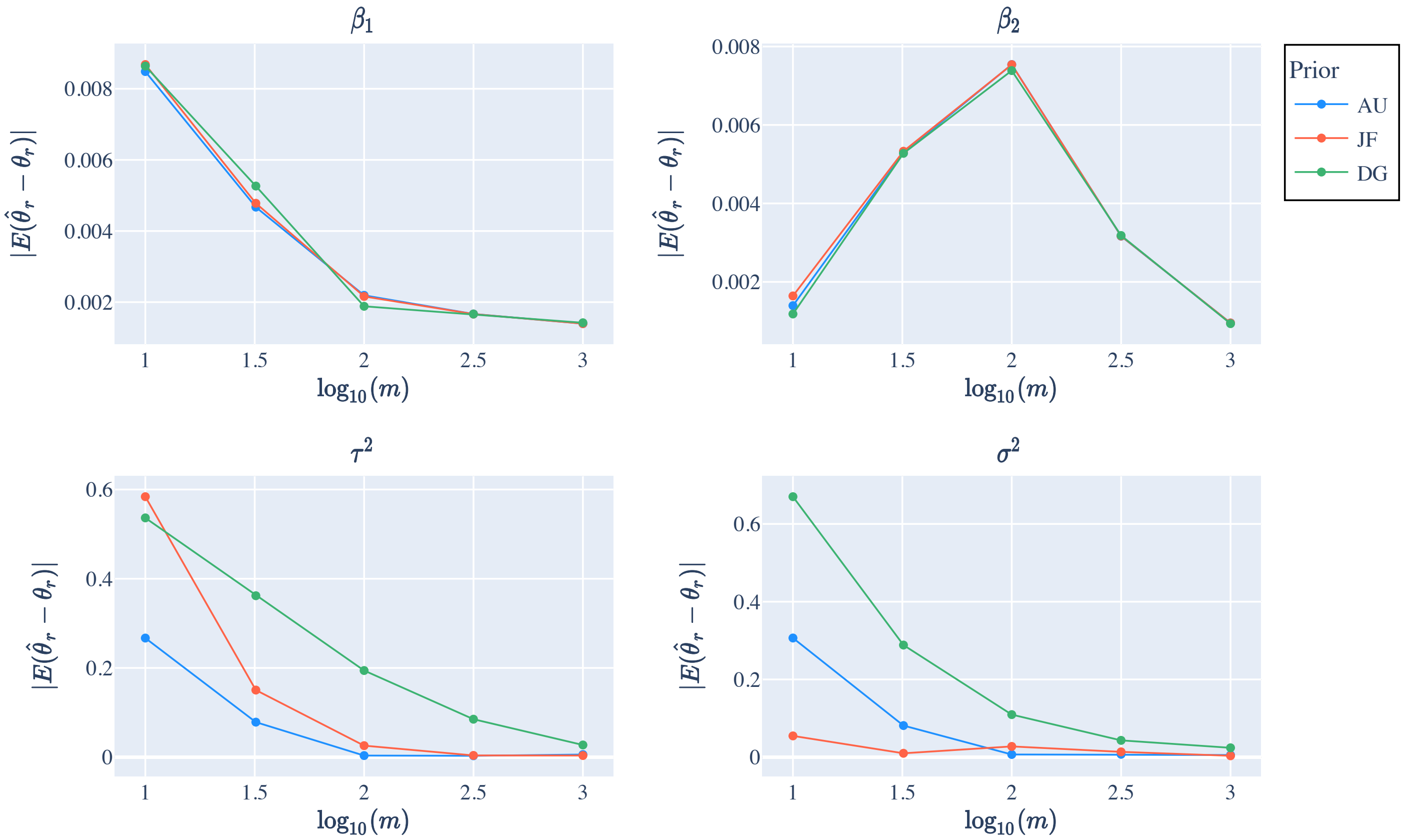
バイアスの比較: 真値 $(\beta_1, \beta_2, \tau^2, \sigma^2) = (1, 1, 1, 1)$



[Figure 1]

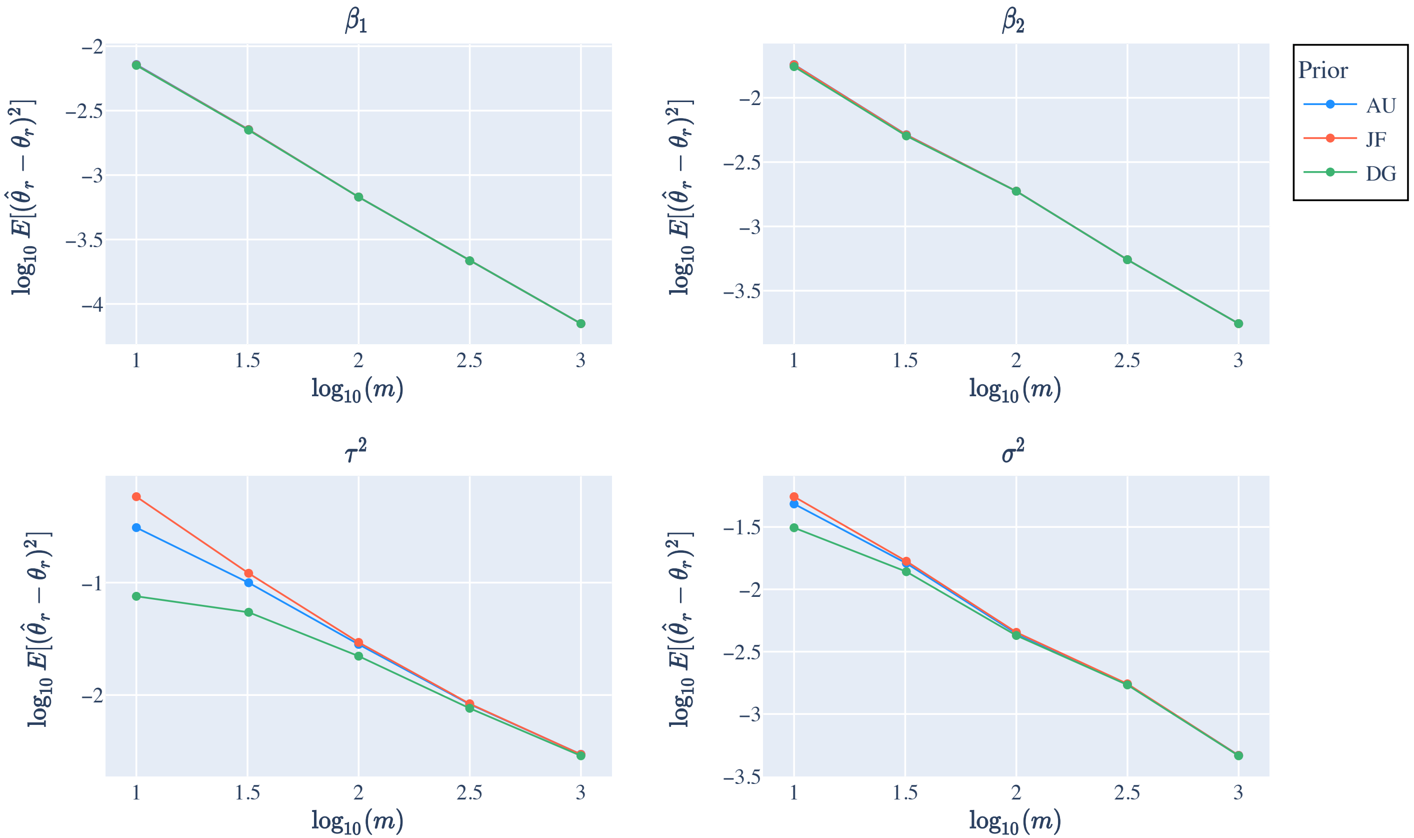
- サンプルサイズが小さいとき，τ^2 と σ^2 のバイアスは事前分布の選択に依存.
- どの事前分布がバイアスの意味で最適かはパラメータの真値による (次スライド) が，提案した事前分布である **[AU]** は良い選択肢である.

バイアスの比較: 真値 $(\beta_1, \beta_2, \tau^2, \sigma^2) = (1, 1, 0.5, 4)$



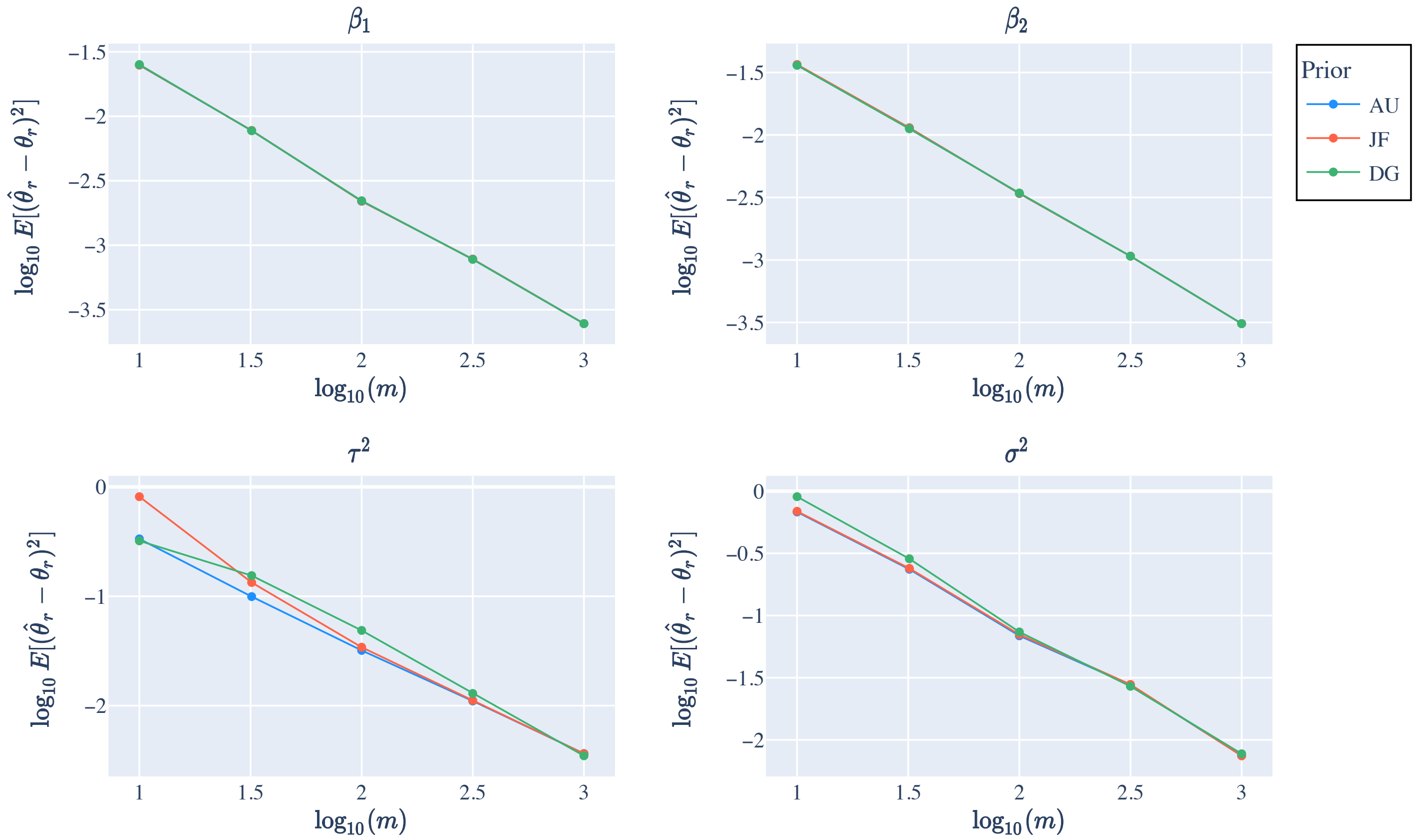
[Figure 2]

MSEの比較: 真値 $(\beta_1, \beta_2, \tau^2, \sigma^2) = (1, 1, 1, 1)$



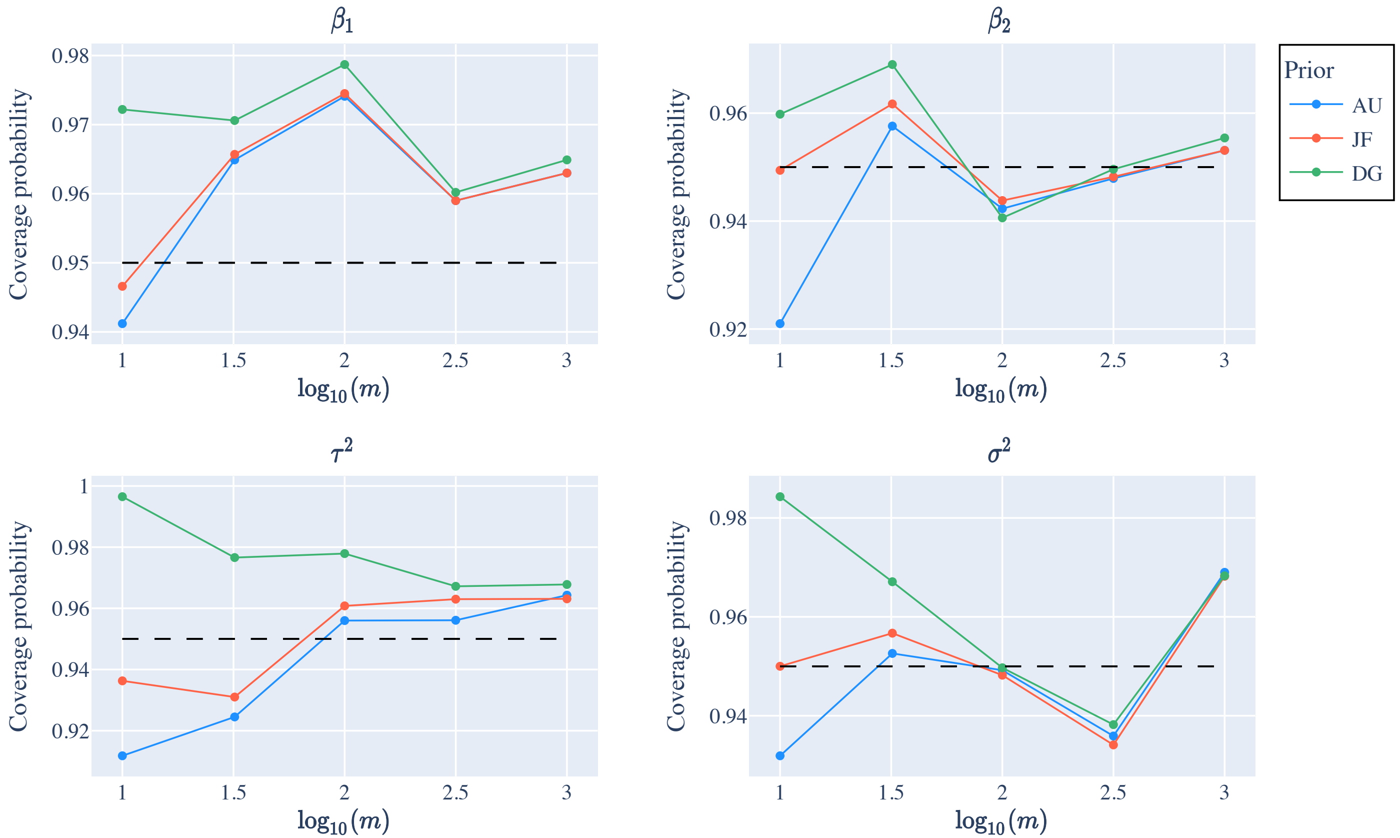
[Figure 3]

MSEの比較: 真値 $(\beta_1, \beta_2, \tau^2, \sigma^2) = (1, 1, 0.5, 4)$



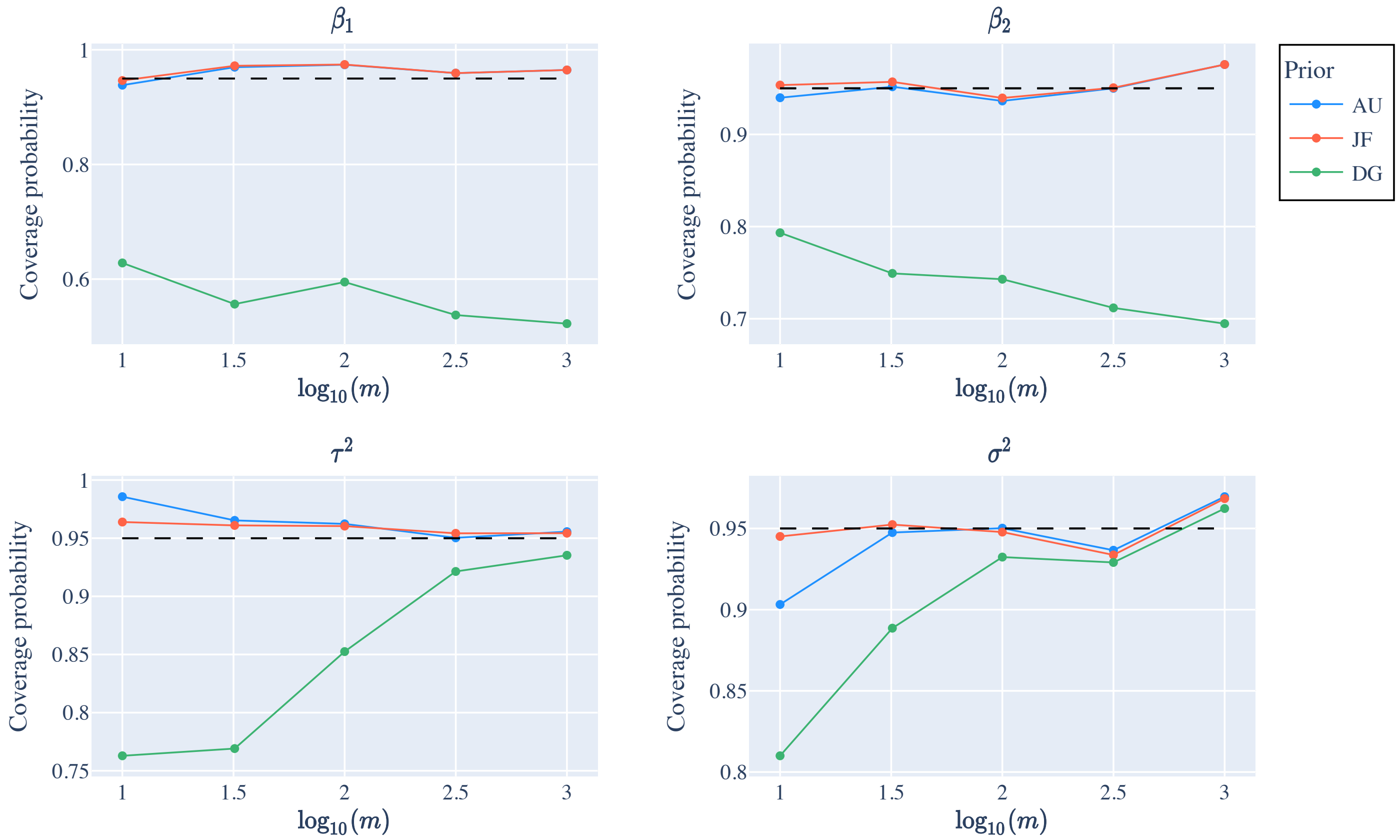
[Figure 4]

95%被覆確率の比較: 真値 $(\beta_1, \beta_2, \tau^2, \sigma^2) = (1, 1, 1, 1)$



[Figure 5]

95%被覆確率の比較: 真値 $(\beta_1, \beta_2, \tau^2, \sigma^2) = (1, 1, 0.5, 4)$



[Figure 6]

シミュレーションのまとめ

- どの事前分布がバイアスの意味で最適かはパラメータの真値によるが，提案した asymptotically unbiased prior は良い選択肢である．
- MSEや coverage probability の結果と併せても，asymptotically unbiased prior は全体的に悪くないパフォーマンスを示す．

アウトライン

1. イントロダクション
2. 主結果1: ベイズ推定量のバイアスの評価
3. 主結果2: 事前分布の構成
4. 例
5. シミュレーション
6. まとめ

まとめ

- i.i.d.とは限らないモデルにおける asymptotically unbiased priors の特徴づけを行った.
- Asymptotically unbiased priors の存在条件と構成手順をまとめた.
- いくつかのモデルへの応用を紹介し，NERモデルのシミュレーション結果を提示した.

参考文献

- Datta, G. S. and Ghosh, M. (1991). Bayesian prediction in linear models: Applications to small area estimation. *The Annals of Statistics 19*(4), 1748–1770.
- Firth, D. (1993). Bias reduction of maximum likelihood estimates. *Biometrika 80*(1), 27-38.
- Ghosh, M. and R. Liu (2011). Moment matching priors. *Sankhya: The Indian Journal of Statistics, Series A (2008-) 73* (2), 185–201.
- Hartigan, J. A. (1965). The asymptotically unbiased prior distribution. *The Annals of Mathematical Statistics 36* (4), 1137–1152.
- Jeffreys, H. (1961). *The Theory of Probability* (Third ed.). Oxford University Press.
- Tiao, G. C. and Tan, W. Y. (1965). Bayesian analysis of random-effect models in the analysis of variance. I. Posterior distribution of variance-components. *Biometrika 52*(1/2), 37–53.