

氏 名 書川 侑子

学位(専攻分野) 博士(統計科学)

学位記番号 総研大甲第 2678 号

学位授与の日付 2026 年 3 月 24 日

学位授与の要件 先端大学院先端学術専攻 (統計科学コース)  
学位規則第6条第1項該当

学位論文題目 BAYESIAN MODELING FRAMEWORK FOR DATA WITH  
GROUP STRUCTURE VIA VARIABLE FUSION

論文審査委員 主 査 村上 大輔  
統計科学コース 准教授  
藤澤 洋徳  
統計科学コース 教授  
二宮 嘉行  
統計科学コース 教授  
松井 秀俊  
滋賀大学 データサイエンス学部 教授

# 博士論文の要旨

氏 名：書川 侑子

論文題目：

BAYESIAN MODELING FRAMEWORK FOR DATA WITH GROUP STRUCTURE VIA VARIABLE FUSION

多くの実データにおいて、説明変数は空間的・時間的な依存性を持ち、隣接する変数の回帰係数は目的変数に対して似た効果を示すことが多い。この隣接性に基づく依存関係は、一種のグループ構造と見なすことができる。このようなデータの持つ構造を無視した解析手法は、推定の不安定化や過学習、解釈性の低下を招くおそれがある。この問題に対処するために提案されたのが変数融合 (variable fusion) の枠組みである。変数融合による手法は、回帰係数だけでなく、隣接する係数間の差にも正則化を課すことで、隣接する回帰係数の値を互いに近づけ、グループ化を促す。変数融合の代表的な手法である **fused lasso** およびその一般化形である **generalized fused lasso** は、ゲノム解析や信号処理などの分野で広く活用されている。しかし、これらの頻度論的手法は、不確実性の定量化や正則化の課し方が異なるモデル間の体系的な比較が困難であるといった点で限界を抱えている。

この課題を克服するため、回帰係数およびその差にラプラス分布を導入することで、**fused lasso** をベイズ的に拡張した変数融合のための手法が提案されてきた。さらに近年では、柔軟な推定を可能とする **global-local** 収縮事前分布、例えば馬蹄事前分布を変数融合型の正則化法に導入する試みも行われている。しかし、説明変数に対応する回帰係数間の融合を行う **Bayesian generalized fused lasso** については、線形回帰以外の枠組みへの拡張がほとんど行われていないのが現状である。さらに、ベイズ推定ではハイパーパラメータに事前分布を設定することで値の調整を自動的に行うことができるものの、推論が事前分布の選択に依存することは避けられない。そのため、モデル比較を行うための客観的な尺度が必要となる。広く用いられている **WAIC** (Watanabe, 2010) は、真の分布と予測分布間の **Kullback-Leibler** ダイバージェンスに基づく汎用的な情報量規準であるが、事前分布の複雑度を反映するバイアス補正項を持たない。その結果、複雑な事前分布を選ぶ傾向があり、特に事前分布のクラスが異なるモデル間の比較において問題となる。この問題を解決するために、**prior intensified information criterion** (PIIC; Ninomiya, 2021) が提案されている。PIIC は、事前分布の対数が  $O(n)$  である設定に基づいて導出され、事前分布の複雑さを反映するペナルティ項を導入することで、事前分布の影響を適切に反映したモデル選択を可能にする。しかし、既存の PIIC を **Bayesian generalized fused lasso** に基づくモデルに適用することはできない。

本論文は以上の課題を踏まえ、(i) 変数融合に基づくロジスティック回帰の階層ベイズ推定、(ii) 事前分布の影響を考慮した客観的なモデル選択を可能にする、**Bayesian**

generalized fused lasso に基づくモデルに対する情報量規準の開発，という二つの柱からなるグループ構造を持つデータのための包括的ベイズモデリング枠組みを構築する．

第一に，回帰係数とその差の両方にラプラス分布を課す Bayesian fused lasso 型ロジスティック回帰モデルを提案する．これにより，小さな係数はゼロへ縮小されつつも，類似した係数同士は融合され，回帰係数の持つグループ構造が推定できる．また，ラプラス分布が大きな回帰係数を過度に縮小してしまう問題を解決するため，回帰係数の差に馬蹄事前分布を適用したモデルも構築する．また，Pólya-Gamma データ拡大法および half-Cauchy 分布の階層表現を用いることで，Gibbs サンプルングによるパラメータの効率的な推定を実現した．数値実験および半導体製造時の異常検知に用いられる時系列データへの応用から，提案手法は推定精度・予測性能・グループ同定のいずれにおいても既存手法を上回る事が確認された．

第二に，Bayesian generalized fused lasso のための情報量規準を構築する．まず，正則化項が  $o(\sqrt{n})$  であるという設定のもとでなされた Viallon et al. (2016) の generalized fused lasso 推定量に対する漸近解析を  $O(n)$  の場合に拡張し，アクティブセットに対応する推定量が急速にゼロへ収束すること，および非アクティブセットに対して漸近正規性が成り立つことを示した．これにより，パラメータ変換を用いることなく PIIC を導出できる理論的基盤を整備した．さらに，グループ構造データの代表例である空間データへの応用を念頭に，generalized fused lasso を用いた空間変動係数 (SVC) モデルに対して PIIC を拡張した．数値実験および米国の住宅における塵のデータの解析により，提案した PIIC は WAIC よりも適切なモデル選択を行い，解釈しやすい空間的なクラスター構造を抽出できることが示された．

本論文の主な貢献は以下の 4 点である．

- (i) ラプラス事前分布と馬蹄事前分布を統合した階層ベイズ型ロジスティック回帰モデルの開発．
- (ii) Pólya-Gamma 拡張および half-Cauchy 階層表現に基づく効率的な Gibbs サンプルングアルゴリズムの構築．
- (iii) 事前分布の対数が  $O(n)$  である設定における generalized fused lasso 推定量の漸近的な性質の導出．
- (iv) PIIC を SVC モデルに対して拡張することによる，事前分布の複雑性を考慮したモデル選択手法の構築．

まとめると，本論文は，変数融合に基づく柔軟な推定と客観的なモデル評価を統合した，グループ構造データのための統一的ベイズモデリングおよび評価の枠組みを確立した．提案手法は理論的にも実用的にも有用であり，グループ構造を持つデータの解析のための新たな基盤を提供するものである．

Results of the Doctoral Thesis Defense

博士論文審査結果

Name in Full

氏名 書川 侑子

Title

論文題目 BAYESIAN MODELING FRAMEWORK FOR DATA WITH GROUP STRUCTURE VIA VARIABLE FUSION

2026年2月2日の午前10時から2時間にわたり、書川侑子氏の博士論文審査を、本人および4名の審査委員全員の出席のもとに行った。出願者による1時間15分の公開發表による概要説明と質疑応答、さらに45分の出願者と審査委員のみによる非公開の質疑応答をもとに審査を行った結果、審査委員会は本論文が学位（統計科学）の授与に値すると判断した。

[論文の概要]

回帰分析には、回帰構造が空間的・時間的な依存性を持ち、隣接する説明変数の回帰係数が似た値となる、という設定のものがある。この依存関係は一種のグループ構造と見なすことができ、その構造を無視した解析手法は推定の不安定化や過学習を招く。この問題に対処するために提案されたのが変数融合（variable fusion）であり、隣接する回帰係数間の差に正則化を課すことで回帰係数のグループ化を促す。変数融合の代表的な手法である fused lasso およびその一般化である generalized fused lasso が種々の分野で活用されている一方、その頻度論的手法は不確実性の定量化などの点において限界を抱える。そこで提案されたのが回帰係数の差にラプラス分布などの事前分布を仮定する Bayesian generalized fused lasso である。しかし、線形回帰以外はほとんど扱われておらず、またモデル比較を行うための客観的な尺度が十分に開発されていない。その尺度の最有力候補は widely applicable information criterion (WAIC) であるが、事前分布のクラスの複雑度を反映するバイアス補正項がないため、そのクラスの候補が複数存在するときのモデル比較ができない、などの問題がある。また、その問題を解決するために prior-intensified information criterion (PIIC) と呼ばれるものが提案されているが、それは Bayesian generalized fused lasso を対象としていなかった。

以上を踏まえ、本論文は (i) 変数融合に基づくロジスティック回帰の階層ベイズ推定、(ii) Bayesian generalized fused lasso に対して事前分布のクラスも選択できるような PIIC の開発、という二つの柱からなる、グループ構造を持つデータのための包括的ベイズモデリングの枠組みを構築している。(i) では、まず回帰係数の差にラプラス分布を仮定する Bayesian fused lasso 型ロジスティック回帰モデルを提案している。また、ラプラス分布が大きな差を過度に縮小してしまう問題を解決するため、差に馬蹄事前分布を仮定したモデルも提案している。そして Pólya-Gamma データ拡大法および half-Cauchy 分布の階層表現を用いることで、Gibbs サンプリングによるパラメータの効率的な推定を実現している。(ii) では、まず正則化項が  $o(\sqrt{n})$  である設定のもとでなされた Viallon et al. (2016) の generalized fused lasso 推定量に対する漸近解析を  $O(n)$  の設定に拡張し、非アクティブセットに対応する推定量が急速にゼロへ収束することと、アクティブセットに対応する推定量が漸近正規性をもつことを示している。そして、グループ構造データの代表例である空間データへの応用を念頭に、空間変動係数モデルに対して PIIC を導出している。

1 章は序論, 2 章は背景, 3 章は (i) の内容, 4 章は (ii) の内容, 5 章は結論となっている. 英語で書かれており, 計 93 ページである.

#### [論文の評価]

本論文は, 変数融合に基づく柔軟な推定と客観的なモデル評価を統合した, グループ構造データのための統一的ベイズモデリングおよび評価の枠組みを確立している. 提案手法は方法論的にも実用的にも有用であり, グループ構造を持つデータの解析のための新たな基盤を提供するものであるといえる. また, **generalized fused lasso** 推定量に対して導かれている漸近解析は自明なものではなく, 理論的にも価値がある. 以上から, 本審査委員会は, 本論文が統計科学の博士論文として十分な内容を有していると判断した.

#### [その他]

3 章の内容は査読付国際学術雑誌 *Japanese Journal of Statistics and Data Science* に採択されている.