

ガウシアングラフィカルモデルに対するロバストなベイズ推定法について

広島大学・先進理工系科学研究科 鬼塚 貴広 (Takahiro Onizuka)

Department of Mathematics, Hiroshima University

広島大学・先進理工系科学研究科 橋本真太郎 (Shintaro Hashimoto)

Department of Mathematics, Hiroshima University

1 はじめに

Graphical model は、変数間のグラフ構造を推定するモデルである．特に、多変量正規分布を仮定すると Gaussian graphical model と呼ばれ、精度行列の構造によって変数間の条件付き独立性を推定することが目的となる．グラフ構造 (変数間の関係) にスパース性を仮定することは、精度行列にスパース性を仮定することに対応し、精度行列の非対角成分のゼロへの縮小を考えることが必要である．Friedman et al. (2008) では、非負定値行列 $\Omega = (\omega_{ij}) \in \mathbb{R}^{p \times p}$ に対して、 $\log |\Omega| - \text{tr}(\mathbf{S}\Omega) - \rho \|\Omega\|_1$ の最大化で定義される graphical Lasso の高速なアルゴリズムを提案した．ただし、 $\mathbf{S}$ は標本分散共分散行列、 $\rho \geq 0$ は調整パラメータ、 $\|\Omega\|_1 = \sum_{i \leq j} |\omega_{ij}|$ である．これは、多変量正規分布の対数尤度に基づく目的関数に対応しており、推定値の正定値性を保証したアルゴリズムが与えられている．同様に、ベイズの枠組みにおいても、多変量正規尤度を仮定し、精度行列の各成分のパラメータに縮小事前分布を仮定した Bayesian graphical Lasso が考えられている (Wang, 2012)：

$$\mathbf{y}_i | \Omega \sim \mathcal{N}_p(\mathbf{0}, \Omega^{-1}), \quad \Omega | \lambda \sim C_\lambda^{-1} \prod_{i < j} \text{Lap}(\omega_{ij} | \lambda) \prod_{i=1}^p \text{Exp}(\omega_{ii} | \lambda/2) 1_{\Omega \in M^+}.$$

ただし、 C_λ は正規化定数で、 $\text{Lap}(\cdot)$, $\text{Exp}(\cdot)$ はそれぞれラプラス分布と指数分布を表す．しかしながら、一般に、正規尤度に基づくパラメータ推定は外れ値の影響を受ける．したがって、外れ値に頑健な精度行列の推定を実現するために、Hirose et al. (2017) は γ -divergence に基づく尤度を用いた Gaussian graphical model を提案した．一般に、 γ -divergence に基づくパラメータ θ の罰則付き推定は、

$$L_\gamma(\theta) = \log \left[\frac{\{n^{-1} \sum_{i=1}^n f(\mathbf{y}_i | \theta)^\gamma\}^{-1/\gamma}}{\{\int f(\mathbf{y} | \theta)^{1+\gamma} d\mathbf{y}\}^{1/(1+\gamma)}} \right] + \lambda \phi(\theta) = R_\gamma(\theta | \mathbf{Y}) + \lambda \phi(\theta) \quad (1)$$

の最小化で与えられる．ただし、第1項 $R_\gamma(\theta | \mathbf{Y})$ は γ -尤度であり、第2項が罰則項である．ここで、 $\gamma > 0$ は推定量のロバスト性と効率性を左右するパラメータであり、 γ が大きくなるにつれ、ロバスト性が強くなり、 $\gamma \rightarrow 0$ のとき通常の負の対数尤度に対応する．ベイズ統計学の文脈においても、外れ値に頑健な推定を実現するために、尤度に γ -divergence を用いる研究は行われてきたが、本研究では、外れ値に頑健な多変量正規

分布の精度行列のベイズ推定による推定法を提案する．特に， γ -divergence に基づく新たな事後分布 (MAP γ -posterior) を定義し，それに基づく Gaussian graphical model を与え，構成した事後分布の性質として，posterior propriety と posterior robustness を示した．また，事後分布の近似計算を行うためのアルゴリズムを与え数値比較と実データへの適用を通して有用性を確認した．

2 γ -posterior

2.1 γ -尤度に基づく事後分布

γ -divergence に基づく事後分布は Hashimoto and Sugasawa (2020) や Nakagawa and Hashimoto (2020) でも考えられている．Hashimoto and Sugasawa (2020) では，パラメータ $\boldsymbol{\theta}$ に対する γ -posterior を以下で定義した：

$$\pi_{\text{HS}}(\boldsymbol{\theta} \mid \mathbf{Y}) \propto \exp(-nR_\gamma(\boldsymbol{\theta} \mid \mathbf{Y})) \pi(\boldsymbol{\theta}). \quad (2)$$

ただし， $\pi(\boldsymbol{\theta})$ は $\boldsymbol{\theta}$ に対する事前分布である．これは，density-power divergence に基づく事後分布 (e.g. Ghosh and Basu, 2016) と同様の定式化による事後分布であり，Kullback–Leibler divergence に基づく通常の事後分布との関係が与えられる．具体的な応用として，線形回帰における回帰係数のスパース推定への応用を与えている．Nakagawa and Hashimoto (2020) では monotone transformed γ -posterior を提案した：

$$\pi_{\text{NH}}(\boldsymbol{\theta} \mid \mathbf{Y}) \propto \exp\left(-n\tilde{R}_\gamma(\boldsymbol{\theta} \mid \mathbf{Y})\right) \pi(\boldsymbol{\theta}). \quad (3)$$

ただし， $\tilde{R}_\gamma(\boldsymbol{\theta} \mid \mathbf{Y}) = -\gamma^{-1}\{\exp(-\gamma R_\gamma(\boldsymbol{\theta} \mid \mathbf{Y})) - 1\}$ である．このような γ -posterior は，事後分布に関する一貫性を満たすため，一般化ベイズとの観点から有用であることが述べられている．これら2つの γ -posterior は Momozaki and Nakagawa (2023) でも ordinal response model に応用されている．

2.2 MAP γ -posterior

本研究では，新たに γ -尤度に基づく事後分布である MAP γ -posterior を以下で定義する：

$$\pi_\gamma(\boldsymbol{\theta} \mid \mathbf{Y}) \propto \exp(-R_\gamma(\boldsymbol{\theta} \mid \mathbf{Y})) \pi(\boldsymbol{\theta}). \quad (4)$$

ただし， $\pi(\boldsymbol{\theta})$ は $\boldsymbol{\theta}$ の事前分布である．一般に，このような事後分布を考えることで次のような利点が考えられる：1) $\pi(\boldsymbol{\theta}) \propto \exp(-\lambda\phi(\boldsymbol{\theta}))$ を仮定すると，事後分布 (4) に基づく MAP 推定量は (1) の最小化で与えられる推定量と等しくなる，2) 頻度論による手法との関係から，最適化に基づく事後分布の近似が行いやすい (e.g. Section 3.3)，3) 事後分布の外れ値に対する頑健性に関する理論的性質が得られやすい (e.g. Section 3.2)．なお，Hashimoto and Sugasawa (2020) や Nakagawa and Hashimoto (2020) による γ -posterior では，これらの利点が得られないため，本研究では，MAP γ -posterior を扱う．

3 提案モデル

3.1 Gaussian graphical model に対する MAP γ -posterior

Gaussian graphical model に対して, MAP γ -posterior は以下で与えられる:

$$\pi_\gamma(\boldsymbol{\Omega} \mid \mathbf{Y}) = \frac{|\boldsymbol{\Omega}|^{1/2(1+\gamma)} \left\{ \sum_{i=1}^n \exp \left(-\frac{\gamma}{2} \mathbf{y}_i^\top \boldsymbol{\Omega} \mathbf{y}_i \right) \right\}^{1/\gamma} \pi(\boldsymbol{\Omega})}{\int |\boldsymbol{\Omega}|^{1/2(1+\gamma)} \left(\sum_{i=1}^n \exp \left(-\frac{\gamma}{2} \mathbf{y}_i^\top \boldsymbol{\Omega} \mathbf{y}_i \right) \right)^{1/\gamma} \pi(\boldsymbol{\Omega}) d\boldsymbol{\Omega}}. \quad (5)$$

以降, $\boldsymbol{\Omega}$ の事前分布については, Bayesian graphical lasso で用いられる事前分布を一般化し $\pi(\boldsymbol{\Omega}) \propto \prod_{i \leq j} \pi(\omega_{ij}) 1_{\boldsymbol{\Omega} \in M^+}$ と仮定する. 事後分布 (5) は, 通常確率分布を尤度に仮定していないため事後分布の propriety は自明ではない. ベイズ法では事後分布が proper であるという性質が必要があるが, 実際, 事後分布 (5) に対しては次の性質が成り立つ.

定理 1 (Posterior propriety). 事前分布 $\pi(\boldsymbol{\Omega}) \propto \prod_{i \leq j} \pi(\omega_{ij}) 1_{\boldsymbol{\Omega} \in M^+}$ について $\prod_{i < j} \pi(\omega_{ij})$ は proper であるとする. このとき, 可積分関数 $g(\omega_{ii})$ が存在して

$$\omega_{ii}^{1/(2+2\gamma)} \pi(\omega_{ii}) \leq g(\omega_{ii}) \quad (i = 1, \dots, p),$$

を満たすならば, (5) で与えられる事後分布 $\pi_\gamma(\boldsymbol{\Omega} \mid \mathbf{Y})$ は任意の $\mathbf{Y} = (\mathbf{y}_1, \dots, \mathbf{y}_n)^\top \in \mathbb{R}^{n \times p}$ に対して proper である.

3.2 Posterior robustness

提案事後分布 (5) の外れ値に対する頑健性について議論する. 観測値 $\mathbf{Y} = (\mathbf{y}_1, \dots, \mathbf{y}_n)^\top \in \mathbb{R}^{n \times p}$ が以下のように表せるとする:

$$y_{ij} = \begin{cases} a_{ij} & (i \in \mathcal{K}), \\ a_{ij} + b_{ij}z & (i \in \mathcal{L}). \end{cases}$$

ただし, $a_{ij} \in \mathbb{R}$, $b_{ij} \in \mathbb{R}$, $z > 0$ ($i = 1, \dots, n$, $j = 1, \dots, p$) である. ここで, $\mathcal{K}$ と $\mathcal{L}$ はそれぞれ非外れ値, 外れ値の添字集合を表し, $\mathcal{K} \cup \mathcal{L} = \{1, \dots, n\}$, $\mathcal{K} \cap \mathcal{L} = \emptyset$ をみたく. したがって, $b_{ij} \neq 0$ かつ z が十分大きな値を取る時 y_{ij} が大きな値を取ることで $\mathbf{y}_i$ を外れ値とみなす. また, $\mathcal{D} = \{\mathbf{y}_i \mid i \in \mathcal{K} \cup \mathcal{L}\}$ をすべての観測値の集合, $\mathcal{D}^* = \{\mathbf{y}_i \mid i \in \mathcal{K}\}$ を外れ値でない観測値の集合とする. このような外れ値の下で, 事後分布 (5) に対して事後分布のロバスト性の一つである posterior robustness を与える.

定理 2 (Posterior robustness). $\pi_\gamma(\boldsymbol{\Omega} \mid \mathbf{Y})$ が任意の観測値に対して proper であるとする. このとき, 事後分布 (5) は L_1 収束に関して以下の posterior robustness を満たす:

$$\lim_{z \rightarrow \infty} \pi_\gamma(\boldsymbol{\Omega} \mid \mathcal{D}) = \pi_\gamma(\boldsymbol{\Omega} \mid \mathcal{D}^*).$$

posterior robustness は, 外れ値の値が十分大きくなると事後分布における外れ値の影響が自動的に除外されることを意味する. 定理 1 と定理 2 により, Bayesian graphical lasso 型の前分布を仮定した MAP γ -posterior (5) は proper であり posterior robustness

を満たすことがわかる．なお，正規尤度に基づく Bayesian graphical lasso や t -分布を仮定した graphical lasso では posterior robustness を満たさないことも証明できる．詳しくは，Onizuka and Hashimoto (2023) を参照されたい．以降，本研究では，Bayesian graphical lasso と同様に L_1 罰則型の事前分布を仮定する：

$$\pi(\mathbf{\Omega}) \propto \exp(-\lambda \|\mathbf{\Omega}\|_1) 1_{\{\mathbf{\Omega} \in M^+\}} = \prod_{i=1}^p \text{Exp}(w_{ii} \mid 1/\lambda) \prod_{i < j}^p \text{Lap}(w_{ij} \mid 1/\lambda) 1_{\{\mathbf{\Omega} \in M^+\}}. \quad (6)$$

3.3 Computation

一般に γ -尤度に基づく事後分布の計算は困難である．特に，本研究では，得られた $\mathbf{\Omega}$ の事後分布（事後サンプル）が正定値行列であることが求められるが，棄却ステップを用いる MCMC ではスパースな正定値行列を生成し採択することが困難であることが考えられる．そのため，本研究では事前分布 (6) の下での MAP γ -posterior からの近似的なサンプリングを weighted Bayesian bootstrap (WBB) を用いて行う．具体的なアルゴリズムは以下の通り：

0. 調整パラメータ λ と初期値 $\mathbf{\Omega}_0$ を与える．
1. 重み $(w_0, w_1, \dots, w_n) = (n+1)\text{Dirichlet}(1, \dots, 1)$ を生成する．
2. Majorize-Minimization (MM) algorithm を用いて，重み付き目的関数

$$L_w(\mathbf{\Omega}) = -\frac{1}{\gamma} \log \left\{ \frac{1}{n} \sum_{i=1}^n w_i f(\mathbf{y}_i \mid \mathbf{\Omega})^\gamma \right\} + \frac{\gamma}{2(1+\gamma)} \log |\mathbf{\Omega}| + w_0 \lambda \|\mathbf{\Omega}\|_1,$$

の最小化を行い，最小化解を（近似的な）事後サンプルとする．

3. Step 1, 2 を m 個の（近似的な）事後サンプルが得られるまで繰り返す．

ただし， $f(\cdot \mid \mathbf{\Omega})$ は $\mathcal{N}_p(\mathbf{0}, \mathbf{\Omega}^{-1})$ の確率密度関数である．なお，各重みに対する最小化 (Step 1, 2) は並列計算が可能である．頻度論に基づく graphical lasso では，推定値が 0 であるかによって構造推定を行うが，ベイズ法では事後分布を用いて非対角成分が 0 であるかを推測する必要がある．本研究では， $\{\omega_{ij} \neq 0\}$ であるか否かを median probability criterion (e.g. Barbieri and Berger, 2004) と呼ばれる $P(|\omega_{ij}| < \varepsilon \mid \mathbf{Y}) \geq 0.5$ で求める．ただし， $\varepsilon > 0$ はある閾値である．

4 数値実験

外れ値を含むデータ生成過程として， $\mathbf{y}_i \sim (1-\delta)\mathcal{N}_p(\mathbf{0}, \mathbf{\Omega}^{-1}) + \delta\mathcal{N}_p(0, 30\mathbf{I}_p)$ ($i = 1, \dots, n$) を考える．ただし， $\mathbf{I}_p$ は単位行列を表し， $p = 12$, $n = 200$ とする． $\mathbf{\Omega}$ として AR(2) の構造 $\omega_{ii} = 1$, $\omega_{i,i-1} = \omega_{i-1,1} = 0.5$, $\omega_{i,i-2} = \omega_{i-2,i} = 0.25$ ($i = 1, \dots, p$) を持つ精度行列を考える． δ は外れ値の割合であり，ここでは， $\delta = 0, 0.1$ の 2 つのデータ生成を考える．比較手法は，1) BR: 提案手法 ($\gamma = 0.1$)，2) BT: 自由度 3 の t -尤度を仮定した Bayesian

graphical lasso, 3) BG: Bayesian graphical lasso (R パッケージ BayesianGLasso), 4) FR: γ -尤度に基づく頻度論の graphical lasso ($\gamma = 0.1$), 5) FG: 頻度論による graphical lasso (R パッケージ glasso). BR, FR に対しては $\lambda = 0.02, 0.06, 0.1$, FG に対しては $\lambda = 0.04, 0.12, 0.2$ の 3 候補ずつ考え, 例えば BR1, BR2, BR3 と表記した.

まず, $\delta = 0.1$ とした外れ値を含むデータに対して, 外れ値を含むデータと外れ値を除いたデータのそれぞれを用いて BR と BG の事後分布を計算し, Figure 1 に示した. 提案手法 (左図) では外れ値の有無で事後分布がほとんど変わっていないことがわかり, posterior robustness が成り立っていることがわかる. 一方で, 外れ値に頑健でない手法である BG では外れ値の有無で事後分布が大きく変化していることがわかり, posterior robustness を満たしていないことが視覚的にわかる.

数値的にこれらの手法を比較するために, root of mean squared error (RMSE), true/false positive rate (TPR/FPR) を計算し, BR, BG, BT に対しては 95% coverage probability (CP) と 95%信用区間の average length (AL) を計算し, 100 回の繰り返しの平均の結果を示した (Table 1). Table 1 を見ると, 外れ値のない $\delta = 0$ の場合, 通常の graphical lasso による方法は最も良い推定を与えているが, 提案手法 BR もそれに近い推定を与えていることがわかる. $\delta = 0.1$ の結果を見ると, 提案手法 BR は外れ値に頑健でないモデルよりも RMSE, TPR/FPR のそれぞれの指標においてよい推定を与えていることがわかる. また, BR の MAP 推定に対応する FR と同等の性能を示しており, 点推定・構造推定の観点で提案手法の有用性が確認できる. また, AL と CP を通した信用区間の精度から, BG, BT に比べて提案手法の事後分布が外れ値に頑健であることもわかる.

Table 1: RMSE, TPR, FPR, AL, CP の 100 回繰り返しの平均.

$\delta = 0$	BR1	BR2	BR3	BG	BT	FR1	FR2	FR3	FG1	FG2	FG3
RMSE	0.07	0.11	0.14	0.06	0.11	0.08	0.14	0.18	0.08	0.13	0.16
TPR	0.03	0.74	0.85	0.10	0.10	0.39	0.68	0.80	0.35	0.64	0.76
FPR	0.00	0.02	0.10	0.14	0.16	0.00	0.03	0.14	0.00	0.02	0.09
AL	0.28	0.25	0.23	0.24	0.36	-	-	-	-	-	-
CP	0.91	0.80	0.72	0.94	0.91	-	-	-	-	-	-
$\delta = 0.1$	BR1	BR2	BR3	BG	BT	FR1	FR2	FR3	FG1	FG2	FG3
RMSE	0.07	0.11	0.14	0.21	0.21	0.08	0.14	0.17	0.21	0.21	0.22
TPR	0.01	0.72	0.85	0.06	0.06	0.37	0.67	0.80	0.06	0.18	0.28
FPR	0.00	0.02	0.10	0.06	0.05	0.00	0.03	0.15	0.04	0.13	0.21
AL	0.30	0.27	0.24	0.19	0.07	-	-	-	-	-	-
CP	0.92	0.82	0.73	0.47	0.52	-	-	-	-	-	-

5 実データ解析

提案手法を yeast gene expression data (Gasch et al., 2000) に適用した結果を示す. このデータは $p = 8$ の遺伝子関係に着目したものであり, 本研究では MAD で基準化した $n = 136$ のデータを用いる. Hirose et al. (2017) によって特定されている 13 個の外れ値を

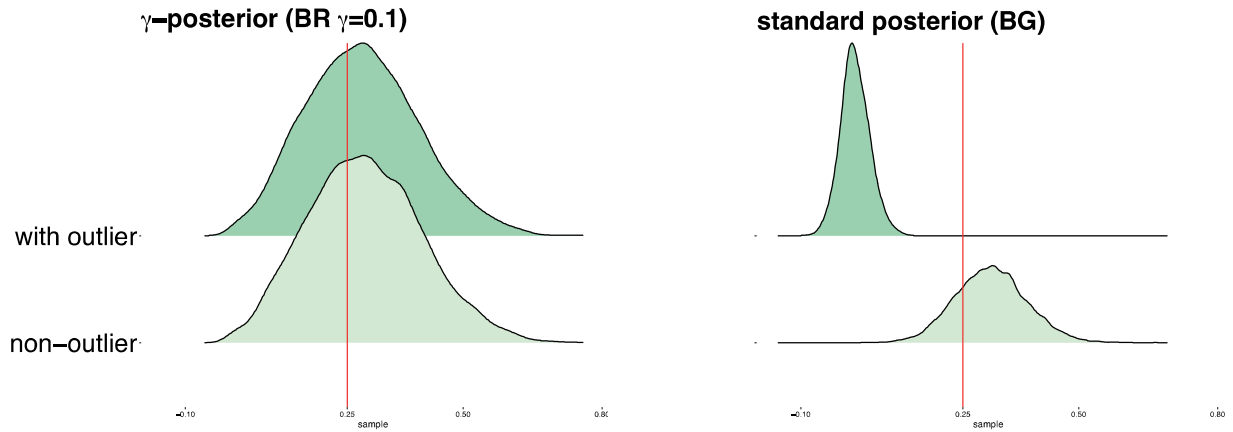


Figure 1: $\omega_{2,5}$ に対する事後分布（上段：外れ値含むデータによる事後分布，下段：外れ値を除いたデータによる事後分布）．左：BR ($\gamma = 0.1$), 右：BG.

含むデータ (136 個) と除いたデータ (123 個) のそれぞれで推定を行い結果を比較する． λ は辺の数が 9 つになるように選択した．推定結果は Figure 2 に示した．BR と FR は，外れ値の有無で推定結果は変わらず，それぞれ (1), (2) のような構造が推定された．一方で，外れ値に頑健でない FG では (3), (4) のように外れ値の有無で結果が大きく異なっていることがわかる．実データ解析の結果の詳細については，Onizuka and Hashimoto (2023) を参照されたい．

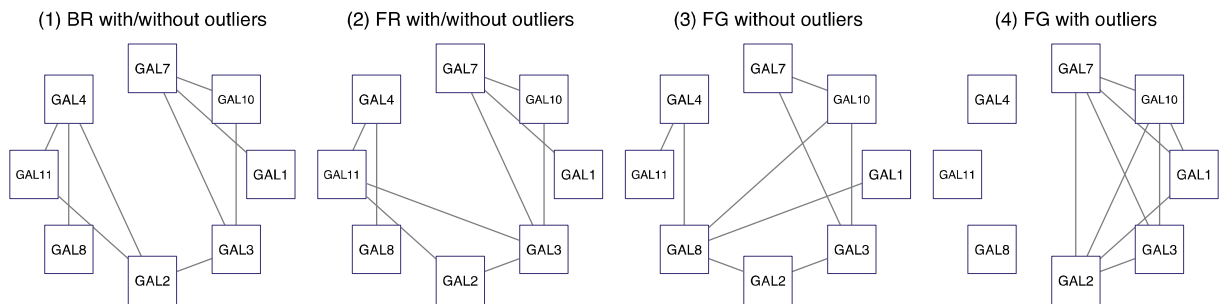


Figure 2: 上段: (1) 外れ値 (有・無) による BR の結果, (2) 外れ値 (有・無) による FR の結果, (3) 外れ値を除いたデータによる FG の結果, (4) 外れ値を含むデータによる FG の結果.

6 おわりに

本稿では，外れ値が存在する場合の共分散行列の逆行列のベイズ推定について論じた．特に，ダイバージェンスに基づくロバスト推定に焦点を当て， γ -divergence に基づく方法が事後分布のロバスト性を導くことがわかった．ダイバージェンスに基づく方法は様々なモデルに対して汎用的に適用できる利点がある一方で，事後分布の計算が複雑になる点が問題点として挙げられる．本稿では，ラプラス型の事前分布に基づくロバストなベイズ

アングラフィカルラッソの最大事後確率 (MAP) 推定値が Hirose et al. (2017) で提案された手法の最適解と一致するような事後分布を考えているため、最適化に基づく事後分布からのサンプリング法を用いて計算を行った。しかし、これはあくまでも近似的な計算方法であるため特に事後分布の不確実性の定量化に関する妥当性については課題が残る。モデルベースの外れ値に対して頑健な方法としては、裾の重い分布を用いた方法が古くから知られており、近年 Gagnon et al. (2020) などは超裾厚な分布 (super-heavy tailed distribution) を用いれば事後分布のロバスト性を達成することを示している (Hamura et al. (2024) も参照されたい)。本研究についても、同様にロバスト性を保持しつつ計算が簡便な多変量の裾が重い分布を用いた方法が提案できるかもしれない。

References

- [1] Barbieri, M. M. and Berger, J. O. (2004). Optimal predictive model selection. *The Annals of Statistics*, **32**(3), 870–897.
- [2] Ghosh, A. and Basu, A. (2016). Robust Bayes estimation using the density power divergence. *Annals of the Institute of Statistical Mathematics*, **68**, 413–437.
- [3] Friedman, J., Hastie, T. and Tibshirani, R. (2008). Sparse inverse covariance estimation with the graphical lasso. *Biostatistics*, **9**(3), 432–441.
- [4] Gagnon, P., Desgagné, A. and Bédard, M. (2020). A new Bayesian approach to robustness against outliers in linear regression. *Bayesian Analysis*, **15**, 389–414.
- [5] Hamura, Y., Irie, K. and Sugawara, S. (2024). Posterior robustness with milder conditions: Contamination models revisited. To appear in *Statistics and Probability Letters*.
- [6] Hashimoto, S. and Sugawara, S. (2020). Robust Bayesian regression with synthetic posterior distributions. *Entropy*, **22**(6), 661.
- [7] Hirose, K., Fujisawa, H. and Sese, J. (2017). Robust sparse Gaussian graphical modeling. *Journal of Multivariate Analysis*, **161**, 172–190.
- [8] Momozaki, T. and Nakagawa, T. (2023). Robustness of bayesian ordinal response model against outliers via divergence approach. arXiv preprint arXiv:2305.07553.
- [9] Nakagawa, T. and Hashimoto, S. (2020). Robust Bayesian inference via γ -divergence. *Communications in Statistics-Theory and Methods*, **49** (2), 343–360.
- [10] Onizuka, T., Hashimoto, S. (2023). Robust Bayesian graphical modeling using γ -divergence. *arXiv*: 2312.07262.
- [11] Wang, H. (2012). The Bayesian graphical lasso and efficient posterior computation. *Bayesian Analysis*, **7**(2), 771–790.