Posterior contraction in sparse non-linear marginal mixed model under spike-and-slab prior

Marion Naveau[1,2]

Supervisors: Maud Delattre[2], Laure Sansonnet[1]

[1]Université Paris-Saclay, AgroParisTech, INRAE, UMR MIA Paris-Saclay

[2]Université Paris-Saclay, INRAE, MaIAGE

Statistiques au sommet de Rochebrune

27 Mars 2024

INRAE

université PARIS-SACLAY | ÉCOLE DOCTORALE de mathématiques Hadamard (EDMH)

Stat4Plant

Table of contents

High-dimensional linear regression

$$Y = X\beta + \varepsilon, \quad \varepsilon \sim \mathcal{N}_n(0, \sigma^2 I_n)$$

where $Y \in \mathbb{R}^n$, $X \in \mathcal{M}_{n\times p}(\mathbb{R})$, $\beta \in \mathbb{R}^p$.

- $p = dim(\beta) \to \infty$ as the sample size $n \to \infty$.

High-dimensional linear regression

$$Y = X\beta + \varepsilon, \quad \varepsilon \sim \mathcal{N}_n(0, \sigma^2 I_n)$$

where $Y \in \mathbb{R}^n$, $X \in \mathcal{M}_{n\times p}(\mathbb{R})$, $\beta \in \mathbb{R}^p$.

- $p = dim(\beta) \to \infty$ as the sample size $n \to \infty$.
- **High-dimensional setting:** inference possible only if data are concentrated around some low-dimensional structure.

High-dimensional linear regression

$$Y = X\beta + \varepsilon, \quad \varepsilon \sim \mathcal{N}_n(0, \sigma^2 I_n)$$

where $Y \in \mathbb{R}^n$, $X \in \mathcal{M}_{n\times p}(\mathbb{R})$, $\beta \in \mathbb{R}^p$.

- $p = dim(\beta) \to \infty$ as the sample size $n \to \infty$.
- **High-dimensional setting:** inference possible only if data are concentrated around some low-dimensional structure.
- **Sparsity:** only a few coordinates of the regression vector β are nonzero.

Penalised approach

- Most non-Bayesian approaches use **penalty functions** to encourage sparsity.
  Example: ℓ_1-penalty

$$\hat{\beta}_{\lambda}^{LASSO} = \underset{\beta \in \mathbb{R}^p}{\operatorname{argmin}} \left\{ ||Y - X\beta||^2 + \lambda ||\beta||_1 \right\}$$

Penalised approach

- Most non-Bayesian approaches use **penalty functions** to encourage sparsity.
  Example: ℓ_1-penalty

$$\hat{\beta}_{\lambda}^{LASSO} = \underset{\beta \in \mathbb{R}^p}{\text{argmin}} \left\{ ||Y - X\beta||^2 + \lambda ||\beta||_1 \right\}$$

- **Bayesian framework:** ℓ_1 and ℓ_2 regularisation methods are equivalent to assigning Laplace or Gaussian priors respectively on the regression vector.

Penalised approach

- Most non-Bayesian approaches use **penalty functions** to encourage sparsity.
  Example: ℓ_1-penalty

$$\hat{\beta}_{\lambda}^{LASSO} = \underset{\beta \in \mathbb{R}^p}{\text{argmin}} \left\{ ||Y - X\beta||^2 + \lambda ||\beta||_1 \right\}$$

- **Bayesian framework:** ℓ_1 and ℓ_2 regularisation methods are equivalent to assigning Laplace or Gaussian priors respectively on the regression vector.
- The solution to the corresponding optimisation problem precisely represents the mode of the posterior distribution.

Spike-and-slab priors

- **Mixture spike-and-slab** priors offer a separate control over signal and noise coefficients.

Spike-and-slab priors

- **Mixture spike-and-slab** priors offer a separate control over signal and noise coefficients.
- For each component of regression parameter:

$$\pi(\beta_j) = (1 - r)\phi_0(\beta_j) - r\phi_1(\beta_j)$$

where ϕ_0 is a density highly concentrated at 0, ϕ_1 is a density allowing intermediate and large values of β_j, and r is a small parameter inducing sparsity in the mixture.

Spike-and-slab priors

- **Mixture spike-and-slab** priors offer a separate control over signal and noise coefficients.
- For each component of regression parameter:

$$\pi(\beta_j) = (1-r)\phi_0(\beta_j) - r\phi_1(\beta_j)$$

  where ϕ_0 is a density highly concentrated at 0, ϕ_1 is a density allowing intermediate and large values of β_j, and r is a small parameter inducing sparsity in the mixture.
- **Examples of mixture spike-and-slab priors:** Dirac-Laplace, Laplace-Laplace, Gaussian-Gaussian, etc...

Selection consistency and Posterior contraction

Selection property

Selection consistency: the posterior probability of the true model converges to 1

$$\inf_{\beta_0} \mathbb{E}_0 \left[\Pi(\beta : S_\beta = S_0 | Y^{(n)}) \right] \xrightarrow[n\to\infty]{} 1$$

Estimation property

Posterior contraction: ability of the posterior distribution to recover the true model from the data

$$\sup_{\theta_0} \mathbb{E}_0 \left[\Pi \left(\theta : d_n(\theta, \theta_0) > C\epsilon_n \middle| Y^{(n)} \right) \right] \xrightarrow[n\to\infty]{} 0$$

with $\epsilon_n \xrightarrow[n\to\infty]{} 0$.

State of the art

✤ **With known variance:**

Reference	Model	Spike	Slab	Result
Castillo et al. (2015)	LR	Dirac	Laplace	Consistency
Ročková and George (2018)	LR	Laplace	Laplace	Contraction

State of the art

✤ **With known variance:**

Reference	Model	Spike	Slab	Result
Castillo et al. (2015)	LR	Dirac	Laplace	Consistency
Ročková and George (2018)	LR	Laplace	Laplace	Contraction

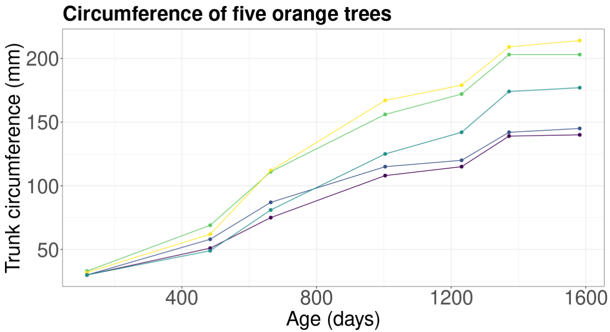
✤ **With unknown variance:**

Reference	Model	Spike	Slab	Result
Narisetty and He (2014)	LR	Gaussian	Gaussian	Consistency
Jiang and Sun (2019)	LR	Generic	Generic	Consistency
Ning et al. (2020)	Multivariate LR	Dirac	Laplace	Consistency
Jeong and Ghosal (2021a)	GLMs	Dirac	Laplace	Contraction
Jeong and Ghosal (2021b)	LR with nuisance	Dirac	Laplace	Consistency
Shen and Deshpande (2022)	Multivariate LR	Laplace	Laplace	Contraction

where LR = Linear Regression.

Framework: repeated measurement data

- **Mixed-effects models:** analyse observations collected repeatedly on several individuals.



- Same overall behaviour but with individual variations.
- Non-linear growth.
- Are these variations due to known characteristics?
  - E.g.: growing conditions, genetic markers, ...

Non-linear marginal mixed model

For $1 \leq i \leq n$,

$$Y_i = f_i(X_i\beta) + Z_i\xi_i + \varepsilon_i^*,\ \varepsilon_i^* \sim \mathcal{N}_{n_i}(0, \sigma^2 I_{n_i}),\ \xi_i \sim \mathcal{N}_q(0, \Gamma),$$

- $Y_i \in \mathbb{R}^{n_i}$, $n_i \in \{1, \dots, J_n\}$, where n_i and J_n can grow with n.

Non-linear marginal mixed model

For $1 \leq i \leq n$,

$$Y_i = f_i(X_i\beta) + Z_i\xi_i + \varepsilon_i^*,\ \varepsilon_i^* \sim \mathcal{N}_{n_i}(0, \sigma^2 I_{n_i}),\ \xi_i \sim \mathcal{N}_q(0, \Gamma),$$

- $Y_i \in \mathbb{R}^{n_i}$, $n_i \in \{1, \dots, J_n\}$, where n_i and J_n can grow with n.
- $f_i(x) = (f(x; t_{i,1}), \dots, f(x; t_{i,n_i}))^\top$, for f non-linear regression function.

Non-linear marginal mixed model

For $1 \leq i \leq n$,

$$Y_i = f_i(X_i\beta) + Z_i\xi_i + \varepsilon_i^*,\ \varepsilon_i^* \sim \mathcal{N}_{n_i}(0, \sigma^2 I_{n_i}),\ \xi_i \sim \mathcal{N}_q(0, \Gamma),$$

- $Y_i \in \mathbb{R}^{n_i}$, $n_i \in \{1, \dots, J_n\}$, where n_i and J_n can grow with n.
- $f_i(x) = (f(x; t_{i,1}), \dots, f(x; t_{i,n_i}))^\top$, for f non-linear regression function.
- $X_i \in \mathcal{M}_{q\times p}$, $Z_i \in \mathcal{M}_{n_i\times q}$, q fixed small and $p >> n$.

Non-linear marginal mixed model

For $1 \leq i \leq n$,

$$Y_i = f_i(X_i\beta) + Z_i\xi_i + \varepsilon_i^*,\ \varepsilon_i^* \sim \mathcal{N}_{n_i}(0, \sigma^2 I_{n_i}),\ \xi_i \sim \mathcal{N}_q(0, \Gamma),$$

- $Y_i \in \mathbb{R}^{n_i}$, $n_i \in \{1, \ldots, J_n\}$, where n_i and J_n can grow with n.
- $f_i(x) = (f(x; t_{i,1}), \ldots, f(x; t_{i,n_i}))^\top$, for f non-linear regression function.
- $X_i \in \mathcal{M}_{q\times p}$, $Z_i \in \mathcal{M}_{n_i\times q}$, q fixed small and $p >> n$.
- We assumed that σ^2 is known.

Non-linear marginal mixed model

For $1 \leq i \leq n$,

$$Y_i = f_i(X_i\beta) + Z_i\xi_i + \varepsilon_i^*,\ \varepsilon_i^* \sim \mathcal{N}_{n_i}(0, \sigma^2 I_{n_i}),\ \xi_i \sim \mathcal{N}_q(0, \Gamma),$$

- $Y_i \in \mathbb{R}^{n_i}$, $n_i \in \{1, \dots, J_n\}$, where n_i and J_n can grow with n.
- $f_i(x) = (f(x; t_{i,1}), \dots, f(x; t_{i,n_i}))^\top$, for f non-linear regression function.
- $X_i \in \mathcal{M}_{q\times p}$, $Z_i \in \mathcal{M}_{n_i\times q}$, q fixed small and $p >> n$.
- We assumed that σ^2 is known.
- **Population parameter:** $\theta = (\beta, \Gamma)$, $\beta \in \mathbb{R}^p$, $\Gamma \in \mathcal{M}_{q\times q}$.

Non-linear marginal mixed model

For $1 \leq i \leq n$,

$$Y_i = f_i(X_i\beta) + Z_i\xi_i + \varepsilon_i^*,\ \varepsilon_i^* \sim \mathcal{N}_{n_i}(0, \sigma^2 I_{n_i}),\ \xi_i \sim \mathcal{N}_q(0, \Gamma),$$

- $Y_i \in \mathbb{R}^{n_i}$, $n_i \in \{1, \ldots, J_n\}$, where n_i and J_n can grow with n.
- $f_i(x) = (f(x; t_{i,1}), \ldots, f(x; t_{i,n_i}))^\top$, for f non-linear regression function.
- $X_i \in \mathcal{M}_{q\times p}$, $Z_i \in \mathcal{M}_{n_i\times q}$, q fixed small and $p >> n$.
- We assumed that σ^2 is known.
- **Population parameter:** $\theta = (\beta, \Gamma)$, $\beta \in \mathbb{R}^p$, $\Gamma \in \mathcal{M}_{q\times q}$.

We assume that n independent observations $Y^{(n)} = (Y_i)_{1\leq i\leq n} \in \mathbb{R}^N$, where $N = \sum_{i=1}^n n_i$, has been generated from this model for a given sparse β_0 and a given Γ_0. The expectation under these true parameters is denoted $\mathbb{E}_0$.

Sparse non-linear marginal mixed model

- This model is called *"marginal"* because the marginal expected value and the covariance matrix of the response variable Y_i are given explicitly through the population parameter vector: $\mathbb{E}[Y_i] = f_i(X_i\beta)$, $Cov(Y_i) = Z_i\Gamma Z_i^\top + \sigma^2 Id_{n_i}$.

Sparse non-linear marginal mixed model

- This model is called *"marginal"* because the marginal expected value and the covariance matrix of the response variable Y_i are given explicitly through the population parameter vector: $\mathbb{E}[Y_i] = f_i(X_i\beta)$, $Cov(Y_i) = Z_i \Gamma Z_i^\top + \sigma^2 Id_{n_i}$.
- The model can be written compactly as:

$$Y_i \sim \mathcal{N}(f_i(X_i\beta), \Delta_{\Gamma,i}), \text{ where } \Delta_{\Gamma,i} = Z_i \Gamma Z_i^\top + \sigma^2 Id_{n_i}.$$

Sparse non-linear marginal mixed model

- This model is called *"marginal"* because the marginal expected value and the covariance matrix of the response variable Y_i are given explicitly through the population parameter vector: $\mathbb{E}[Y_i] = f_i(X_i\beta)$, $Cov(Y_i) = Z_i\Gamma Z_i^\top + \sigma^2 Id_{n_i}$.
- The model can be written compactly as:

$$Y_i \sim \mathcal{N}(f_i(X_i\beta), \Delta_{\Gamma,i}), \text{ where } \Delta_{\Gamma,i} = Z_i\Gamma Z_i^\top + \sigma^2 Id_{n_i}.$$

- **Priors:**

Sparse non-linear marginal mixed model

- This model is called *"marginal"* because the marginal expected value and the covariance matrix of the response variable Y_i are given explicitly through the population parameter vector: $\mathbb{E}[Y_i] = f_i(X_i\beta)$, $Cov(Y_i) = Z_i\Gamma Z_i^\top + \sigma^2 Id_{n_i}$.
- The model can be written compactly as:

$$Y_i \sim \mathcal{N}(f_i(X_i\beta), \Delta_{\Gamma,i}), \text{ where } \Delta_{\Gamma,i} = Z_i\Gamma Z_i^\top + \sigma^2 Id_{n_i}.$$

- **Priors:**
  - Spike-and-slab Dirac-Laplace on (S,β): $(S,\beta) \mapsto \frac{\pi_p(s)}{\binom{p}{s}} g_S(\beta_S)\delta_0(\beta_{S^c})$,

Sparse non-linear marginal mixed model

- This model is called *"marginal"* because the marginal expected value and the covariance matrix of the response variable Y_i are given explicitly through the population parameter vector: $\mathbb{E}[Y_i] = f_i(X_i\beta)$, $Cov(Y_i) = Z_i\Gamma Z_i^\top + \sigma^2 Id_{n_i}$.
- The model can be written compactly as:

$$Y_i \sim \mathcal{N}(f_i(X_i\beta), \Delta_{\Gamma,i}), \text{ where } \Delta_{\Gamma,i} = Z_i\Gamma Z_i^\top + \sigma^2 Id_{n_i}.$$

- **Priors:**
  - Spike-and-slab Dirac-Laplace on (S,β): $(S,\beta) \mapsto \frac{\pi_p(s)}{\binom{p}{s}} g_S(\beta_S)\delta_0(\beta_{S^c})$,
  - Inverse-Wishart(Σ, d) prior on Γ: $\pi(\Gamma) \propto |\Gamma|^{-(d+q+1)/2} \exp\left(-\frac{1}{2} Tr(\Sigma\Gamma^{-1})\right)$.

Sparse non-linear marginal mixed model

- This model is called *"marginal"* because the marginal expected value and the covariance matrix of the response variable Y_i are given explicitly through the population parameter vector: $\mathbb{E}[Y_i] = f_i(X_i\beta)$, $Cov(Y_i) = Z_i\Gamma Z_i^\top + \sigma^2 Id_{n_i}$.
- The model can be written compactly as:

$$Y_i \sim \mathcal{N}(f_i(X_i\beta), \Delta_{\Gamma,i}), \text{ where } \Delta_{\Gamma,i} = Z_i\Gamma Z_i^\top + \sigma^2 Id_{n_i}.$$

- **Priors:**
  - Spike-and-slab Dirac-Laplace on (S,β): $(S,\beta) \mapsto \frac{\pi_p(s)}{\binom{p}{s}} g_S(\beta_S)\delta_0(\beta_{S^c})$,
  - Inverse-Wishart(Σ, d) prior on Γ: $\pi(\Gamma) \propto |\Gamma|^{-(d+q+1)/2} \exp\left(-\frac{1}{2} Tr(\Sigma\Gamma^{-1})\right)$.

Goal

Obtain posterior contraction result in such model for the parameters β and Γ under spike-and-slab Dirac-Laplace prior.

Assumptions

- For some constants $A_1,\ A_2,\ A_3,\ A_4 > 0$,

$$A_1 p^{-A_3} \pi_p(s-1) \leq \pi_p(s) \leq A_2 p^{-A_4} \pi_p(s-1),\ s = 1, \dots p.$$

  - **Example:** $\beta_1, \dots \beta_p \sim (1-r)\delta_0 + r\mathcal{L},\ \pi_p = Bin(p, r)$ where $r \sim Beta(1, p^u),\ u > 1.$

Assumptions

- For some constants $A_1, A_2, A_3, A_4 > 0$,

$$A_1 p^{-A_3} \pi_p(s-1) \leq \pi_p(s) \leq A_2 p^{-A_4} \pi_p(s-1),\ s = 1, \dots p.$$

  - **Example:** $\beta_1, \dots \beta_p \sim (1-r)\delta_0 + r\mathcal{L}$, $\pi_p = Bin(p, r)$ where $r \sim Beta(1, p^u)$, $u > 1$.

- f is assumed to be Lipschitzienne:

$$\forall x, y \in \mathbb{R}^q, \forall t \in \mathbb{R},\ ||f(x,t) - f(y,t)||_2 \leq K||x-y||_2.$$

We denote by $K_n = \sqrt{K^2 J_n}$.

  - **Example:** Log-Gompertz model $y_{ij} = \beta_1 + b_i - Ce^{-\beta_2 t_{ij}} + \varepsilon_{ij}$

Assumptions

- For some constants $A_1,\ A_2,\ A_3,\ A_4 > 0$,

$$A_1 p^{-A_3}\pi_p(s-1) \le \pi_p(s) \le A_2 p^{-A_4}\pi_p(s-1),\ s = 1,\dots p.$$

  - **Example:** $\beta_1,\dots\beta_p \sim (1-r)\delta_0 + r\mathcal{L},\ \pi_p = Bin(p,r)$ where $r \sim Beta(1,p^u),\ u > 1.$

- f is assumed to be Lipschitzienne:

$$\forall x, y \in \mathbb{R}^q, \forall t \in \mathbb{R},\ ||f(x,t) - f(y,t)||_2 \le K||x-y||_2.$$

  We denote by $K_n = \sqrt{K^2 J_n}$.
  - **Example:** Log-Gompertz model $y_{ij} = \beta_1 + b_i - Ce^{-\beta_2 t_{ij}} + \varepsilon_{ij}$

- $g_S(\beta_S) = \prod_{j\in S} \frac{\lambda}{2}\exp(-\lambda|\beta_j|)$, with $\frac{||X||_* K_n}{L_1 p^{L_2}} \le \lambda \le \frac{L_3||X||_* K_n}{\sqrt{n}}$, for some constants $L_1,\ L_2,\ L_3 > 0$, where $||X||_* = \max_j ||X_{.j}||_2$.

Assumptions on true parameters

- $s_0 > 0$, $s_0 \log(p) = o(n)$,

Assumptions on true parameters

- $s_0 > 0$, $s_0 \log(p) = o(n)$,
- $\Gamma_0 \in \mathcal{H}_0$, with $\mathcal{H}_0 = \{\Gamma : 1 \lesssim \rho_{min}(\Gamma) \leq \rho_{max}(\Gamma) \lesssim 1\}$

Assumptions on true parameters

- $s_0 > 0$, $s_0 \log(p) = o(n)$,
- $\Gamma_0 \in \mathcal{H}_0$, with $\mathcal{H}_0 = \{\Gamma : 1 \lesssim \rho_{min}(\Gamma) \leq \rho_{max}(\Gamma) \lesssim 1\}$
- $\beta_0 \in \mathcal{B}_0$, with $\mathcal{B}_0 = \left\{\beta : ||\beta||_\infty \lesssim \lambda^{-1} \log(p)\right\}$,

Assumptions on true parameters

- $s_0 > 0$, $s_0 \log(p) = o(n)$,
- $\Gamma_0 \in \mathcal{H}_0$, with $\mathcal{H}_0 = \{\Gamma : 1 \lesssim \rho_{min}(\Gamma) \leq \rho_{max}(\Gamma) \lesssim 1\}$
- $\beta_0 \in \mathcal{B}_0$, with $\mathcal{B}_0 = \left\{\beta : ||\beta||_\infty \lesssim \lambda^{-1} \log(p)\right\}$,
- $\frac{1}{n} \sum_{i=1}^n \mathbb{1}_{n_i \geq q}$ is bounded,

Assumptions on true parameters

- $s_0 > 0$, $s_0 \log(p) = o(n)$,
- $\Gamma_0 \in \mathcal{H}_0$, with $\mathcal{H}_0 = \{\Gamma : 1 \lesssim \rho_{min}(\Gamma) \leq \rho_{max}(\Gamma) \lesssim 1\}$
- $\beta_0 \in \mathcal{B}_0$, with $\mathcal{B}_0 = \left\{\beta : ||\beta||_\infty \lesssim \lambda^{-1} \log(p)\right\}$,
- $\frac{1}{n} \sum_{i=1}^n \mathbb{1}_{n_i \geq q}$ is bounded,
- $\min_i \{\rho_{\min}^{1/2}(Z_i^\top Z_i) : n_i \geq q\} \gtrsim 1$, *i.e.* Z_i is a full rank,

Assumptions on true parameters

- $s_0 > 0$, $s_0 \log(p) = o(n)$,
- $\Gamma_0 \in \mathcal{H}_0$, with $\mathcal{H}_0 = \{\Gamma : 1 \lesssim \rho_{min}(\Gamma) \leq \rho_{max}(\Gamma) \lesssim 1\}$
- $\beta_0 \in \mathcal{B}_0$, with $\mathcal{B}_0 = \{\beta : ||\beta||_\infty \lesssim \lambda^{-1} \log(p)\}$,
- $\frac{1}{n} \sum_{i=1}^n \mathbb{1}_{n_i \geq q}$ is bounded,
- $\min_i\{\rho_{\min}^{1/2}(Z_i^\top Z_i) : n_i \geq q\} \gtrsim 1$, *i.e.* Z_i is a full rank,
- $\max_i\{\rho_{\max}^{1/2}(Z_i^\top Z_i)\} \lesssim 1$.

Support size theorem

Theorem

Assume that the previous assumptions are satisfied. Then, there exists a constant $C_1 > 0$ *such that:*

$$\sup_{\beta_0 \in \mathcal{B}_0, \Gamma_0 \in \mathcal{H}_0} \mathbb{E}_0 \left[\Pi \left(\beta : |S_\beta| > C_1 s_0 \middle| Y^{(n)} \right) \right] \underset{n \to \infty}{\longrightarrow} 0.$$

Posterior contraction Rényi theorem

Definition

For two n-variates densities $f = \prod_{i=1}^n f_i$ and $g = \prod_{i=1}^n g_i$ of independent variables, the average Rényi divergence (of order $1/2$) is defined by:

$$R_n(f,g) = -\frac{1}{n}\sum_{i=1}^n \log\left(\int \sqrt{f_i g_i}\right)$$

Theorem

Assume that the previous assumptions are satisfied, and $\log(J_n) \lesssim \log(p)$. We denote by $p_{\beta,\Gamma} = \prod_{i=1}^n p_{\beta,\Gamma,i}$ the joint density for $p_{\beta,\Gamma,i}$ the density of the ith observation vector y_i, and p_0 the true joint density. Then, there exists a constant $C_2 > 0$ such that:

$$\sup_{\beta_0\in\mathcal{B}_0,\Gamma_0\in\mathcal{H}_0} \mathbb{E}_0\left[\Pi\left((\beta,\Gamma) : R_n(p_{\beta,\Gamma},p_0) > C_2\frac{s_0\log(p)}{n}\middle| Y^{(n)}\right)\right] \underset{n\to\infty}{\longrightarrow} 0.$$

Posterior contraction rates

Theorem

Assume that the previous assumptions are satisfied, and $\log(J_n) \lesssim \log(p)$*. Then, there exists constants* $C_3, C_4, C_5 > 0$ *such that:*

$$\sup_{\beta_0 \in \mathcal{B}_0, \Gamma_0 \in \mathcal{H}_0} \mathbb{E}_0 \left[\Pi \left(\Gamma : ||\Gamma - \Gamma_0||_F > C_3 \sqrt{\frac{s_0 \log(p)}{n}} \middle| Y^{(n)} \right) \right] \xrightarrow[n \to \infty]{} 0,$$

$$\sup_{\beta_0 \in \mathcal{B}_0, \Gamma_0 \in \mathcal{H}_0} \mathbb{E}_0 \left[\Pi \left(\beta : \sqrt{\frac{1}{n} \sum_{i=1}^{n} ||f_i(X_i\beta) - f_i(X_i\beta_0)||_2^2} > C_4 \sqrt{\frac{s_0 \log(p)}{n}} \middle| Y^{(n)} \right) \right] \xrightarrow[n \to \infty]{} 0,$$

and under an assumption of identifiability on f*, with* $\phi_1(s) = \inf_{\beta : 1 \leq s_\beta \leq s} \frac{||X\beta||_2 \sqrt{s_\beta}}{||X||_* ||\beta||_1}$ *:*

$$\sup_{\beta_0 \in \mathcal{B}_0, \Gamma_0 \in \mathcal{H}_0} \mathbb{E}_0 \left[\Pi \left(\beta : ||\beta - \beta_0||_1 > C_5 \frac{s_0 \sqrt{\log(p)}}{\sqrt{||X||_*^2 \phi_1^2((C_1 + 1)s_0) - s_0^2 \log(p)}} \middle| Y^{(n)} \right) \right] \xrightarrow[n \to \infty]{} 0,$$

1. Introduction

2. Non-linear marginal mixed model

3. Theoretical guarantees

4. Perspectives

Perspectives

- In non-linear: Under spike-and-slab Dirac-Laplace, can we get:
  - Distributional approximation of the posterior?
  - Selection consistency?

under what assumptions?

Perspectives

✤ In non-linear: Under spike-and-slab Dirac-Laplace, can we get:

- Distributional approximation of the posterior?
- Selection consistency?

under what assumptions?

Can the same results be obtained by making the model more complex?

$$\begin{cases} y_i = f_i(\varphi_i) + \varepsilon_i & , \varepsilon_i \overset{\text{ind}}{\sim} \mathcal{N}_{n_i}(0, \sigma^2 I_{n_i}), \\ \varphi_i = X_i\beta + \xi_i & , \xi_i \overset{\text{i.i.d.}}{\sim} \mathcal{N}_q(0, \Gamma). \end{cases}$$

where $y_i \in \mathbb{R}^{n_i}$, $f_i(\varphi_i) = (f(\varphi_i; t_{i,1}), \ldots, f(\varphi_i; t_{i,n_i}))$, $\varphi_i \in \mathbb{R}^q$, $X_i \in \mathcal{M}_{q\times p}$, $\beta \in \mathbb{R}^p$.

Perspectives

✤ In non-linear: Under spike-and-slab Dirac-Laplace, can we get:

- Distributional approximation of the posterior?
- Selection consistency?

under what assumptions?

Can the same results be obtained by making the model more complex?

$$\begin{cases} y_i = f_i(\varphi_i) + \varepsilon_i & , \varepsilon_i \overset{\text{ind}}{\sim} \mathcal{N}_{n_i}(0, \sigma^2 I_{n_i}), \\ \varphi_i = X_i\beta + \xi_i & , \xi_i \overset{\text{i.i.d.}}{\sim} \mathcal{N}_q(0, \Gamma). \end{cases}$$

where $y_i \in \mathbb{R}^{n_i}$, $f_i(\varphi_i) = (f(\varphi_i; t_{i,1}), \ldots, f(\varphi_i; t_{i,n_i}))$, $\varphi_i \in \mathbb{R}^q$, $X_i \in \mathcal{M}_{q\times p}$, $\beta \in \mathbb{R}^p$.

✤ In linear: **Can we obtain a selection consistency theorem under spike-and-slab LASSO prior in LMEM with covariance matrix unknown?**

Et pour conclure sur les posteriEUR, en voici un sympathique...

Thank you for your attention!

References I

Castillo, I., Schmidt-Hieber, J., and Van der Vaart, A. (2015). Bayesian linear regression with sparse priors. *The Annals of Statistics*, 43(5):1986–2018. Publisher: Institute of Mathematical Statistics.

Demidenko, E. (2013). *Mixed models: theory and applications with R*. John Wiley & Sons.

Jeong, S. and Ghosal, S. (2021a). Posterior contraction in sparse generalized linear models. *Biometrika*, 108(2):367–379. Publisher: Oxford University Press.

Jeong, S. and Ghosal, S. (2021b). Unified Bayesian theory of sparse linear regression with nuisance parameters. *Electronic Journal of Statistics*, 15(1):3040–3111. Publisher: The Institute of Mathematical Statistics and the Bernoulli Society.

Jiang, B. and Sun, Q. (2019). Bayesian high-dimensional linear regression with generic spike-and-slab priors. *arXiv preprint arXiv:1912.08993*.

Narisetty, N. N. and He, X. (2014). Bayesian variable selection with shrinking and diffusing priors. *The Annals of Statistics*, 42(2):789–817. Publisher: Institute of Mathematical Statistics.

Ning, B., Jeong, S., and Ghosal, S. (2020). Bayesian linear regression for multivariate responses under group sparsity. *Bernoulli*, 26(3):2353–2382. Publisher: International Statistical Institute.

References II

Ročková, V. and George, E. I. (2018). The spike-and-slab lasso. *Journal of the American Statistical Association*, 113(521):431–444. Publisher: Taylor & Francis.

Shen, Y. and Deshpande, S. K. (2022). On the posterior contraction of the multivariate spike-and-slab LASSO. *arXiv preprint arXiv:2209.04389*.

Model approximation

$$\begin{cases} y_i = f_i(\psi, \varphi_i) + \varepsilon_i & , \varepsilon_i \overset{\text{ind}}{\sim} \mathcal{N}_{n_i}(0, \sigma^2 I_{n_i}), \\ \varphi_i = X_i\beta + \xi_i & , \xi_i \overset{\text{i.i.d.}}{\sim} \mathcal{N}_q(0, \Gamma). \end{cases}$$

where $y_i \in \mathbb{R}^{n_i}$, $f_i(\psi, \varphi_i) = (f(\psi, \varphi_i; t_{i,1}), \dots, f(\psi, \varphi_i; t_{i,n_i}))$, $\psi \in \mathbb{R}^r$, $\varphi_i \in \mathbb{R}^q$, $X_i \in \mathcal{M}_{q\times p}$, $\beta \in \mathbb{R}^p$.

First order approximation of $f_i(\psi, X_i\beta + \xi_i)$ around $\mathbb{E}[\varphi_i] = X_i\beta$:

$$y_i = f_i(\psi, X_i\beta) + Z_i(\beta)\xi_i + \varepsilon_i,$$

where $Z_i = \dfrac{\partial f_i}{\partial \varphi_i}$.

$\Longrightarrow$ Non-linear marginal mixed model with varied matrix of random effects (Demidenko, 2013).

Identifiability/injectivity assumption

$\forall 1 \leq i \leq n$, $\forall \delta > 0$, $\forall t \in \mathbb{R}$,

$$|f(X_i\beta, t) - f(X_i\beta_0, t)| \leq \delta \Rightarrow |f(X_i\beta, t) - f(X_i\beta_0, t)| \gtrsim ||X_i(\beta - \beta_0)||_2$$

Stages of proof

In general, the stages of proof (following Castillo et al. (2015)) are as follows:

1. **Support size:** $\sup_{\beta_0} \mathbb{E}_0 \left[\Pi \left(\beta : |S_\beta| > K|S_0| \middle| Y^{(n)} \right) \right] \longrightarrow 0$
2. **Posterior contraction / Recovery:** $\sup_{\theta_0} \mathbb{E}_0 \left[\Pi \left(\theta : d_n(\theta, \theta_0) > C\epsilon_n \middle| Y^{(n)} \right) \right] \longrightarrow 0$, with $\epsilon_n \longrightarrow 0$
3. **Distributional approximation:** $\sup_{\beta_0} \mathbb{E}_0 \left[\left\| \Pi \left(\beta \in \cdot | Y^{(n)} \right) - \Pi^\infty \left(\beta \in \cdot | Y^{(n)} \right) \right\|_{TV} \right] \longrightarrow 0$
4. **Selection, no supersets:** $\sup_{\beta_0} \mathbb{E}_0 \left[\Pi \left(\beta : S_\beta \supset S_0, S_\beta \neq S_0 \middle| Y^{(n)} \right) \right] \longrightarrow 0$
5. **Selection consistency:** $\inf_{\beta_0} \mathbb{E}_0 \left[\Pi(\beta : S_\beta = S_0 | Y^{(n)}) \right] \longrightarrow 1.$

Idea of the proof

Set $B = \{(\beta, \Gamma) : |S_\beta| > \tilde{s}\}$, with any integer $\tilde{s} \geq s_0$.

Yet, by Bayes' formula: $\Pi(B|y) = \dfrac{\int_B \Lambda_n(\beta,\Gamma)d\Pi(\beta,\Gamma)}{\int \Lambda_n(\beta,\Gamma)d\Pi(\beta,\Gamma)}$, where $\Lambda_n(\beta,\Gamma) = \prod_{i=1}^n \dfrac{p_{\beta,\Gamma,i}}{p_{0,i}}$ likelihood ratio.

Thus, the following lemma shows that the denominator of the posterior distribution is bounded below by a factor with probability tending to one:

Lemma

Let's assume that the previous hypotheses are satisfied. Then, there exists a constant M such that:

$$\mathbb{P}_0\left(\int \Lambda_n(\beta,\Gamma)d\Pi(\beta,\Gamma) \geq \pi_p(s_0)e^{-M(s_0\log(p)+\log(n))}\right) \longrightarrow 1.$$

This event is denoted by $\mathcal{A}_n$.

Idea of the proof

Then, $\mathbb{E}_0 \left[\Pi\left(B|y\right)\right] = \mathbb{E}_0 \left[\Pi\left(B|y\right)\mathbb{1}_{\mathcal{A}_n}\right] + \underbrace{\mathbb{E}_0 \left[\Pi\left(B|y\right)\mathbb{1}_{\mathcal{A}_n^c}\right]}_{\longrightarrow 0 \text{ by lemma}}$.

And by the lemma and Fubini-Tonelli's theorem the first term is bounded by a term tending towards 0 with n:

$$\begin{aligned}\mathbb{E}_0 \left[\Pi\left(B|y\right)\mathbb{1}_{\mathcal{A}_n}\right] &= \mathbb{E}_0 \left[\frac{\int_B \Lambda_n(\beta,\Gamma)d\Pi(\beta,\Gamma)}{\int \Lambda_n(\beta,\Gamma)d\Pi(\beta,\Gamma)}\mathbb{1}_{\mathcal{A}_n}\right] \\ &\leq \pi_p(s_0)^{-1}\exp\left\{M(s_0\log(p)+\log(n))\right\}\Pi(B) \longrightarrow 0.\end{aligned}$$

This leads to the theorem: there exist a constant C_1 such that $\mathbb{E}_0\left[\Pi\left(|S_\beta| > C_1 s_0|y\right)\right] \longrightarrow 0$.