A Bayesian Approach for Envelope Models

Subhadip Pal

Department of Biostatistics and Bioinformatics, Emory University

Collaborator: Kshitij Khare and Zhihua Su

Envelope Approach for Multivariate Regression

Standard multivariate linear regression model: $\mathbf{Y}_i = \boldsymbol{\mu} + \boldsymbol{\beta}\mathbf{X}_i + \boldsymbol{\varepsilon}_i,\ i = 1, 2, \cdots, n$

$$\mathbf{Y} \in \mathbb{R}^r, \mathbf{X} \in \mathbb{R}^p, \boldsymbol{\mu} \in \mathbb{R}^r, \boldsymbol{\beta} \in \mathbb{R}^{r \times p}, \boldsymbol{\varepsilon} \in \mathbb{R}^r$$

Envelopes arise by re-parametrization of the SLM in terms of the smallest subspace $\mathcal{E} \subseteq \mathbb{R}^r$ such that ($\mathbf{P}_{\mathcal{E}}$ is projection onto the space $\mathcal{E}$ and $\mathbf{Q}_{\mathcal{E}} = \mathbf{I} - \mathbf{P}_{\mathcal{E}}$)

$$\mathbf{Q}_{\mathcal{E}}\mathbf{Y} \mid \mathbf{X} \sim \mathbf{Q}_{\mathcal{E}}\mathbf{Y} \text{ and } \mathbf{P}_{\mathcal{E}}\mathbf{Y} \perp \mathbf{Q}_{\mathcal{E}}\mathbf{Y} \mid \mathbf{X}$$

Impact of $\mathbf{X}$ on $\mathbf{Y}$ is concentrated in $\mathbf{P}_{\mathcal{E}}\mathbf{Y}$. Informally we refer

$\mathbf{Q}_{\mathcal{E}}\mathbf{Y}$: immaterial part of $\mathbf{Y}$ and $\mathbf{P}_{\mathcal{E}}\mathbf{Y}$: material part of $\mathbf{Y}$

The conditions $\mathbf{Q}_{\mathcal{E}}\mathbf{Y} \mid \mathbf{X} \sim \mathbf{Q}_{\mathcal{E}}\mathbf{Y}$ and $\mathbf{P}_{\mathcal{E}}\mathbf{Y} \perp \mathbf{Q}_{\mathcal{E}}\mathbf{Y} \mid \mathbf{X}$ hold if and only if (Cook 2010)

- $\mathcal{E}$ envelopes $\mathcal{B} = span\{\boldsymbol{\beta}\}$, i.e. $\mathcal{B} \subseteq \mathcal{E}$
- $\mathcal{E}$ is reducing subspace of $\boldsymbol{\Sigma}$, i.e. $\boldsymbol{\Sigma} = \mathbf{P}_{\mathcal{E}}\boldsymbol{\Sigma}\mathbf{P}_{\mathcal{E}} + \mathbf{Q}_{\mathcal{E}}\boldsymbol{\Sigma}\mathbf{Q}_{\mathcal{E}}$

The intersection of all subspace $\mathcal{E}$ with the above properties is called $\boldsymbol{\Sigma}$-Envelope of $\mathcal{B}$ and denoted by $\mathcal{E}_{\boldsymbol{\Sigma}}(\mathcal{B})$ with $u = dim(\mathcal{E}_{\boldsymbol{\Sigma}}(\mathcal{B}))$

Coordinate representation of Envelope model

$$\mathbf{Y} = \boldsymbol{\mu} + \boldsymbol{\Gamma}\boldsymbol{\eta}\mathbf{X} + \boldsymbol{\varepsilon}, \quad \text{where } \boldsymbol{\beta} = \boldsymbol{\Gamma}\boldsymbol{\eta}, \boldsymbol{\Sigma} = \boldsymbol{\Gamma}\boldsymbol{\Omega}\boldsymbol{\Gamma}^T + \boldsymbol{\Gamma}_0\boldsymbol{\Omega}_0\boldsymbol{\Gamma}_0^T\boldsymbol{\Omega}, \boldsymbol{\Omega}_0 > 0$$

$\boldsymbol{\Gamma}$ and $\boldsymbol{\Gamma}_0$ be a basis for the space $\mathcal{E}_{\boldsymbol{\Sigma}}(\mathcal{B})$ and $\mathcal{E}_{\boldsymbol{\Sigma}}^{\perp}(\mathcal{B})$. Note that the choice of $\boldsymbol{\Gamma}$ and $\boldsymbol{\Gamma}_0$ are not unique.

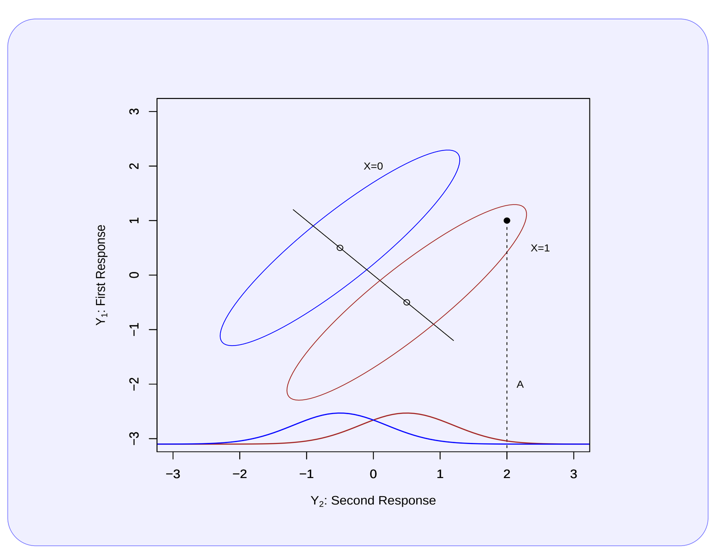
How Envelope Model Works? Toy Example

Consider the multivariate linear regression model with response Y_1, Y_2 and one predictor variable X with two label 0 and 1.

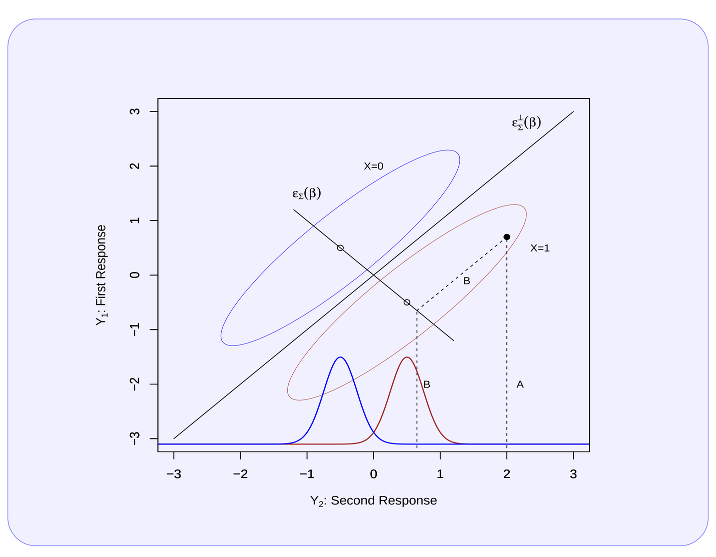
$$\mathbf{Y} = \begin{pmatrix} Y_1 \\ Y_2 \end{pmatrix} = \begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix} + \begin{pmatrix} \beta_1 \\ \beta_2 \end{pmatrix} X + \begin{pmatrix} \epsilon_1 \\ \epsilon_2 \end{pmatrix} \tag{1}$$

$\mu_1 = E(Y_1 \mid X = 0),\quad \beta_1 = E(Y_1 \mid X = 1) - E(Y_1 \mid X = 0)$
$\mu_2 = E(Y_2 \mid X = 0),\quad \beta_2 = E(Y_2 \mid X = 1) - E(Y_2 \mid X = 0)$

Schematic representation: standard model



Schematic representation: envelope model



Bayesian Envelope Model

Features that a Bayesian approach would offer are

- Comprehensive uncertainty characterization through the posterior distribution
- A framework to incorporate prior information
- Ability to deal with the case when $n < r$

A reparameterization of Envelope model

$$\mathbf{Y} = \boldsymbol{\mu} + \boldsymbol{\Gamma}\boldsymbol{\eta}\mathbf{X} + \boldsymbol{\varepsilon}, \qquad \boldsymbol{\Sigma} = \boldsymbol{\Gamma}\boldsymbol{\Omega}\boldsymbol{\Gamma}^T + \boldsymbol{\Gamma}_0\boldsymbol{\Omega}_0\boldsymbol{\Gamma}_0^T, \tag{2}$$

- $\boldsymbol{\beta} = \boldsymbol{\Gamma}\boldsymbol{\eta}$, $\boldsymbol{\Gamma} \in S_{r,u}^{+}$, $\boldsymbol{\Gamma}_0 \in S_{r,r-u}^{+}$ with $\boldsymbol{\Gamma}_0^T\boldsymbol{\Gamma} = 0$, $\boldsymbol{\eta} \in M_{u,p}$.
- $\boldsymbol{\Omega}$ and $\boldsymbol{\Omega}_0$ are diagonal matrices with diagonal entries arranged in decreasing order
- To estimate $(\boldsymbol{\mu}, \boldsymbol{\eta}, (\boldsymbol{\Gamma}, \boldsymbol{\Gamma}_0), \boldsymbol{\omega}, \boldsymbol{\omega}_0)$

If $u = r$, envelope model (2) is equivalent to the standard model by the one to one transformation $(\boldsymbol{\Gamma}, \boldsymbol{\eta}, \boldsymbol{\Omega}) \to (\boldsymbol{\beta} = \boldsymbol{\Gamma}\boldsymbol{\eta}, \boldsymbol{\Sigma} = \boldsymbol{\Gamma}\boldsymbol{\Omega}\boldsymbol{\Gamma}^T)$.

Prior Specification

- $\pi(\boldsymbol{\mu}) \propto 1$
- $\boldsymbol{\eta} \mid (\boldsymbol{\Gamma}, \boldsymbol{\Gamma}_0), \boldsymbol{\omega}, \boldsymbol{\omega}_0 \sim \mathcal{MN}_{u,p}(\boldsymbol{\Gamma}^T\mathbf{e}, \boldsymbol{\Omega}, \mathbf{C}^{-1}) \quad \boldsymbol{\Omega} := diag(\boldsymbol{\omega})$
- $\mathbf{O} \sim B_{r,r}(\mathbf{G}, D^{-1}) \quad \mathbf{O} := [\boldsymbol{\Gamma}, \boldsymbol{\Gamma}_0]$
- $(\boldsymbol{\omega}, \boldsymbol{\omega}_0)$ and $\mathbf{O}$ are apriori independent

The prior mode for $\mathbf{O}$ is an appropriately permuted version of the eigenvectors of $\mathbf{G}$

→ The entries of $\boldsymbol{\omega}$: order statistics of u i.i.d. observations from the Inverse-Gamma(α, λ) distribution

→ The entries of $\boldsymbol{\omega}_0$: order statistics of $r - u$ i.i.d observations from the Inverse-Gamma(α_0, λ_0) distribution

Uniform improper prior

The joint improper prior corresponding to uniform improper prior is given by $\pi(\boldsymbol{\mu}, \boldsymbol{\eta}, (\boldsymbol{\Gamma}, \boldsymbol{\Gamma}_0), \boldsymbol{\omega}, \boldsymbol{\omega}_0) \propto 1$

This corresponds to following hyperparameter choices

- $\mathbf{e} = \mathbf{0}, \mathbf{C} = 0, \quad \mathbf{G} = \mathbf{0}$
- $\alpha = -(1 + \frac{p}{2}), \lambda = 0, \quad \alpha_0 = -1, \lambda_0 = 0$

Emperical prior

$(\boldsymbol{\eta}^*, (\boldsymbol{\Gamma}^*, \boldsymbol{\Gamma}_0^*), \boldsymbol{\omega}^*, \boldsymbol{\omega}_0^*)$ obtained by a naive method

- Set $\mathbf{e} = \boldsymbol{\Gamma}^*\boldsymbol{\eta}^*$ and $\mathbf{C} = \mathbf{0}$
- α , λ and α_0 , λ_0: estimated by moment estimator using values $\boldsymbol{\omega}^*$ and $\boldsymbol{\omega}_0^*$
- $\mathbf{D}$ diagonal with diagonal elements $(\boldsymbol{\omega}^*, \boldsymbol{\omega}_0^*)$
- we employ a procedure to ensured that $\mathbf{O}^*$ is the prior mode of $\mathbf{O}$

Posterior Distribution

$$\begin{aligned}\pi((\boldsymbol{\mu}, \boldsymbol{\eta}, (\boldsymbol{\Gamma}, \boldsymbol{\Gamma}_0), \boldsymbol{\omega}, \boldsymbol{\omega}_0) \mid \mathbb{Y}) \propto\ & (2\pi)^{-(nr)/2}|\boldsymbol{\Omega}|^{-n/2}|\boldsymbol{\Omega}_0|^{-n/2} e^{-\frac{1}{2}\mathrm{tr}\left\{(\mathbb{Y} - \mathbf{1}_n\boldsymbol{\mu}^T - \mathbb{X}\boldsymbol{\eta}^T\boldsymbol{\Gamma}^T)(\boldsymbol{\Gamma}\boldsymbol{\Omega}\boldsymbol{\Gamma}^T + \boldsymbol{\Gamma}_0\boldsymbol{\Omega}_0\boldsymbol{\Gamma}_0^T)^{-1}(\mathbb{Y} - \mathbf{1}_n\boldsymbol{\mu}^T - \mathbb{X}\boldsymbol{\eta}^T\boldsymbol{\Gamma}^T)^T\right\}} \\ & \times |\boldsymbol{\Omega}|^{-p/2} e^{-\frac{1}{2}\mathrm{tr}\left(\boldsymbol{\Omega}^{-1}(\boldsymbol{\eta} - \boldsymbol{\Gamma}^T\mathbf{e})\mathbf{C}(\boldsymbol{\eta} - \boldsymbol{\Gamma}^T\mathbf{e})^T\right)} e^{-\frac{1}{2}\mathrm{tr}\left(\mathbf{D}^{-1}\mathbf{O}^T\mathbf{G}\mathbf{O}\right)} \prod_{i=1}^{u} \omega_i^{-\alpha-1} e^{-\frac{\lambda}{\omega_i}} \prod_{i=1}^{r-u} \omega_{0,i}^{-\alpha_0-1} e^{-\frac{\lambda_0}{\omega_{0,i}}}\end{aligned} \tag{3}$$

Theorem

The posterior density in (3) is proper under either of the following conditions.

- $n > \max(r, p + 3)$
- $n + 2\alpha > 1$, $\lambda, \lambda_0 > 0$ and $\mathbf{C}$ is positive definite

Sampling Scheme

Generalized Bingham distribution

A random matrix $\mathbf{Z} = [\mathbf{Z}_1 \ : \ \mathbf{Z}_2]$ is defined to have a generalized matrix Bingham distribution on $S_{2,2}$ with parameters $\mathbf{A}_1$ and $\mathbf{A}_2$ ($GB_{2,2}(\mathbf{A}_1, \mathbf{A}_2)$) if the probability density function of $\mathbf{Z}$ (w.r.t the Haar measure on $S_{2,2}$) is proportional to $e^{-\mathbf{Z}_1^T\mathbf{A}_1\mathbf{Z}_1 - \mathbf{Z}_2^T\mathbf{A}_2\mathbf{Z}_2}$.

An efficient algorithm has been developed to sample from $GB_{2,2}$.

- $\boldsymbol{\mu} \mid ((\boldsymbol{\Gamma}, \boldsymbol{\Gamma}_0), \boldsymbol{\eta}, \boldsymbol{\omega}, \boldsymbol{\omega}_0, \mathbb{Y})$ is Normal and $\boldsymbol{\eta} \mid ((\boldsymbol{\Gamma}, \boldsymbol{\Gamma}_0), \boldsymbol{\omega}, \boldsymbol{\omega}_0, \mathbb{Y})$ is Multivariate normal
- $\omega_i \mid (\boldsymbol{\Gamma}, \boldsymbol{\Gamma}_0), \boldsymbol{\omega}^{-i}, \boldsymbol{\omega}_0, \mathbb{Y}$ and $\omega_{0,i} \mid (\boldsymbol{\Gamma}, \boldsymbol{\Gamma}_0), \boldsymbol{\omega}, \boldsymbol{\omega}_0^{-i}, \mathbb{Y}$ are Truncated-Inverse-Gamma
- Sampling of O involves samling from $GB_{2,2}$ distribution

We start at a given initial value of the parameters, and repeat the following steps.

- Sampling from $((\boldsymbol{\Gamma}, \boldsymbol{\Gamma}_0), \boldsymbol{\omega}, \boldsymbol{\omega}_0) \mid \mathbb{Y}$

  → For $i = 1, 2, \cdots, u$, update ω_i

  → For $i = 1, 2, \cdots, r - u$, update $\omega_{0,i}$

  → For every pair or randomly chosen pair (i, j) such that $1 \le i < j \le r$, update $\mathbf{O}_{\cdot i}$ and $\mathbf{O}_{\cdot j}$

- Sample from full conditional distribution for $\boldsymbol{\eta} \mid ((\boldsymbol{\Gamma}, \boldsymbol{\Gamma}_0), \boldsymbol{\omega}, \boldsymbol{\omega}_0)$, $\mathbb{Y}$ and $\boldsymbol{\mu} \mid \mathbb{Y}$

The corresponding MC is Harris ergodic.

Model Selection: DIC Criteria

- Need to select $u \in 0, 1, \ldots, r$
- $\boldsymbol{\theta} := (\boldsymbol{\mu}, \boldsymbol{\eta}, (\boldsymbol{\Gamma}, \boldsymbol{\Gamma}_0), \boldsymbol{\omega}, \boldsymbol{\omega}_0)$ be the parameter vector
- For each u construct the appropriate Markov chain

→ $\{\theta^{(i)}\}_{i=1}^{M}$ the samples from the relevant posterior distribution (after an appropriate burn-in)

→ Calculate $DIC = \bar{D} + \frac{1}{2(M-1)}\sum_{i=1}^{M}\left(D\left(\theta^{(i)}\right) - \bar{D}\right)^2$ where $D(\boldsymbol{\theta}) := -2\log L(\boldsymbol{\theta})$ and $\bar{D} := \sum_{i=1}^{M} D\left(\theta^{(i)}\right)/M$.

- Select the value of u which corresponds to the minimum DIC.

Application: Analysis of Wheat Protein Data

A brief summary of the Wheat protein data, (Cook 2010) can be summarized as follows

- Consists of $r = 6$ responses, which measure the log infrared reflectance at six different wavelengths for 50 ground wheat samples
- The predictor is a binary indicator, taking 0 or 1 if a sample has high or low protein content
- There are 26 samples with high protein content, and 24 samples with low protein content

Model selection: DIC scores

u	Uniform	Emperical
0	1257.7	1254.5
1	1201.9	1197.5
2	1206.3	1374.3
3	1208	1245.8
4	1209.3	1266.2
5	1211.1	1333
6	1216	1668.1

Parameter estimation

Coefficient	Bayesian envelope model		Bayesian standard model	
	Post. mean	Post. SD	Post. mean	Post. SD
β_1	-1.039	0.378	2.934	10.479
β_2	4.406	0.498	7.745	8.630
β_3	3.630	0.417	7.219	9.273
β_4	-5.880	0.644	-2.395	10.157
β_5	0.594	0.224	2.799	14.601
β_6	-1.610	0.904	0.410	5.759

Extensive simulation has been conducted for: model selection accuracy, comparison with non-Bayesian envelope model and different versions of standard Bayesian models

Summary

- We developed a comprehensive Bayesian framework for estimation and model selection is in the context of envelope model
- A parameterization available for Bayesian analysis has been introduced
- Class of priors introduced has desirable proprieties:

  → flexible

  → sensible interpretation as well as specification of hyperparameters

  → easy to sample from the corresponding posterior

- conditions for posterior propriety have been investigated
- A new distribution $GB_{2,2}$ along with efficient sampling scheme is developed
- R package "BENV" is developed for data analysis
- The method is applied successfully on simulated and real datasets

References

- Khare K., Pal S. and Su Z. A Bayesian Approach for Envelope Models, Submitted to (and tentatively Accepted in) Annals of Statistics.
- Cook, R.D., Li, B. and Chiaromonte, F. (2010). Envelope Models for Parsimonious and Efficient Multivariate Linear Regression, Statistica Sinica, 20, 927-960.