A Bayesian Approach for Envelope Models
Subhadip Pal
Department of Biostatistics and Bioinformatics, Emory University
Collaborators: Kshitij Khare and Zhihua Su

Envelope Approach for Multivariate Regression

Standard multivariate linear regression model: \( Y = \beta X + \varepsilon \), \( i = 1, 2, \ldots, n \)
\( Y \in \mathbb{R}^p, X \in \mathbb{R}^{n \times p}, \beta \in \mathbb{R}^p, \varepsilon \in \mathbb{R}^n \)

Envelopes arise by re-parameterization of the SLM in terms of the smallest subspace \( E \subseteq \mathbb{R}^p \) such that \( P_E \) is projection onto the space \( E \) and \( Q_\perp = I - P_E \)

\[ Q_E X = Q_\perp Y = Q_\perp Y | X \]

Impact of \( X \) on \( Y \) is concentrated in \( Q_E \). Formally, we refer to \( Q_E Y \) as the "envelope of \( Y \) after controlling for \( X \)."

Coordinate representation of Envelope model

\[ Y = \gamma_0 + \Gamma \beta + \varepsilon \]
\[ \xi = \Gamma \beta, \xi \in \Sigma \]

The conditions \( Q_E | X = Q_\perp = Q_\perp Y | X \) hold if and only if (Cook 2010)

- \( E \) envelopes \( B = \text{span}(\beta) \), i.e. \( B \subseteq E \)
- \( \varepsilon \) is reducing subspace of \( E \), i.e. \( \Sigma = P_E \Sigma P_E + Q_\perp \Sigma Q_\perp \)

The intersection of all subspaces \( E \) with the above properties is called \( \Sigma \)-Envelope of \( Y \) and denoted by \( Q_E \) with \( \alpha = \text{diag}(\alpha(\beta)) \)

Schematic representation: standard model

Schematic representation: envelope model

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

\[ Y = \begin{bmatrix} Y_1 \\ Y_2 \end{bmatrix} = \begin{bmatrix} \beta_0 + \beta_1 X + \varepsilon_1 \\ \beta_0 + \beta_2 X + \varepsilon_2 \end{bmatrix} \]

\[ \beta_0 = E | X = 0, \beta_1 = E | X = 1 - E | X = 0 \]

\[ \beta_2 = E | X = 0, \beta_0 = E | X = 1 - E | X = 0 \]

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 \( n < p \)

A reparameterization of Envelope model

\[ y = \mu + \Gamma \eta + \varepsilon, \quad \Sigma = \Gamma^T \Sigma \Gamma + \Gamma D \Gamma^T, \Gamma \Sigma \Gamma^T = \eta, \eta \in \Lambda_{p \times p} \]

Prior Specification

\[ \eta \sim \text{Normal}(0, \Sigma) \]

\[ 0 \leq \eta \leq \text{diag}(\eta(\beta)) \]

\[ O = \text{diag}(\eta(\beta)) \]

\( O \) and \( \text{diag}(\eta(\beta)) \) are asymptotically independent

- The entries of \( \eta \) order statistics of \( u \) i.e. observations from the Inverse-Gamma(0, \( \theta \)) distribution
- The entries of \( \eta \) order statistics of \( u \) i.e. observations from the Inverse-Gamma(0, \( \theta \)) distribution

Uniform improper prior

- The prior used to ensure \( O \) is as follows: \( \eta \sim \text{Uniform}(0, \text{diag}(\eta(\beta))) \)

Empirical prior

- \( \eta \sim \text{Normal}(0, \Sigma) \)

Construction of hyperparameters

- \( \sigma, \lambda, \alpha_1, \alpha_2 \)

Posterior Distribution

\[ p(\beta|y, \Sigma) \propto p(y|\beta, \Sigma) p(\beta|\Sigma) \]

Theorem

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

- \( n > \max(p, \beta) \)
- \( n > 2n + 1, \lambda_0 > 0 \) and \( C \) is positive definite

Sampling Scheme

Generalized Binomial distribution

A random matrix \( Z | \beta \) is defined to have a generalized binomial distribution on \( S_{p+1} \) with parameters \( A_i | \beta \) and \( A_0 | \beta \) such that the probability density function of \( Z \) is \( L \) the mean response on \( S_{p+1} \) is proportional to \( Z_\text{tr} D Z_\text{tr} \)

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

- Start from \( \beta = 0 \)
- \( \beta \sim \text{Normal}(0, \Sigma) \)

Model selection: DIC Criteria

- Need to select \( \omega \in \{ 1, \ldots, r \} \)
- \( \beta \sim \text{Normal}(m, \Sigma) \)
- For each such a Markov chain

Application: Analysis of Wheat Protein Data

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

- Consists of \( n = 6 \) responses, which measure the log-transformed levels at six different wavelengths for 50 ground wheat samples
- The predictor is a linear index, taking 0 or 1 if a sample has high or low protein content
- There are 20 samples with high protein content, and 24 samples with low protein content

Summary

- We developed a comprehensive Bayesian framework for estimation and model selection in the context of envelope model
- A parameterization available for Bayesian analysis has been introduced
- Class of priors introduced has desirable properties:
  - Flexible
  - Sensitive interpretation as well as specification of hyperparameters
  - Easy to sample from the corresponding posterior
- Conditions for posterior propriety have been investigated
- A new distribution \( \text{G} \) along with efficient sampling scheme is developed
- R package "BEUV" is developed for data analysis
- The method is applied successfully on simulated and real datasets

References