VARIABLE SELECTION FOR MULTICOLINEARITY

1. Model and Posterior Distribution

Let us assume the linear model is

$$y_i = x_{1i}\beta_1 + \cdots + x_{pi}\beta_p + \epsilon_i, \;\; \epsilon_i \sim \mathcal{N}(0, \sigma^2). \tag{1}$$

1.1. Priors on $\boldsymbol{\beta} = (\beta_1, \cdots, \beta_p)'$.

Suppose

$$\beta_j \overset{i.i.d}{\sim} \pi_0\delta_0 + (1-\pi_0)\mathcal{P}, \;\; \mathcal{P} \sim DP(\alpha_0, \mathcal{P}_0),$$

and $\mathcal{P}_0 = \mathcal{N}(0, \eta^2)$. To obtain more insight into the variable selection and clustering process, we consider some theoretical properties implied by the DP prior on $\mathcal{P}$. First, following Sethuraman's (1994) stick-breaking representation:

$$\mathcal{P} = \pi_0\delta_0 + (1-\pi_0)\mathcal{P}^* = \pi_0\delta_0 + (1-\pi_0)\sum_{h=1}^{\infty}\pi_h\delta_{\theta_h},$$

where $\{\pi_h, h = 1, \cdots, \infty\}$ is an infinite sequence of random weights, with

$$\pi_h = V_h\prod_{\ell=1}^{h-1}(1 - V_\ell) \text{ and } V_h \overset{i.i.d.}{\sim} Beta(1, \alpha_0), \text{ for } h = 1, \cdots, \infty.$$

and $\Theta = \{\theta_h, h = 1, \cdots, \infty\}$ is an infinite sequence of random atoms, with $\theta_h \overset{i.i.d}{\sim} \mathcal{P}_0 = \mathcal{N}(0, \eta^2)$. Moreover, if we assume $\theta_0 = 0$ and define $\tilde{\pi}_0 = \pi_0$ and $\tilde{\pi}_h = (1-\pi_0)\pi_h$, then $\mathcal{P}$ can be rewritten as

$$\mathcal{P} = \sum_{h=0}^{\infty}\tilde{\pi}_h\delta_{\theta_h}.$$

1.2. Likelihood for the Linear Model.

From the assumption of (1), we get the likelihood function of the linear model as

$$\mathcal{L}(\mathbf{y}; \mathbf{X}, \boldsymbol{\beta}, \sigma^2) = \prod_{i=1}^{n} f(y_i) = \frac{1}{(2\pi\sigma)^n}\exp\left\{-\frac{\sum_{i=1}^{n}(y_i - \mathbf{x}_i'\boldsymbol{\beta})^2}{2\sigma^2}\right\},$$

where $y_i \sim \mathcal{N}(\mathbf{x}_i'\boldsymbol{\beta}, \sigma^2)$, $\mathbf{x}_i = (x_{i1}, \cdots, x_{ip})'$ and $\mathbf{X} = (\mathbf{x}_1, \cdots, \mathbf{x}_n)'$.

1.3. **Posterior Distribution.** In order to simplify the computation issue, we truncated the stick-breaking representation of $\mathcal{P}^*$ to be finite with large value L and set $V_L = 1$. Then the model becomes,

$$\begin{aligned}
y_i &\sim \mathcal{N}(\mathbf{x}_i'\boldsymbol{\beta}, \sigma^2),\ i = 1, \cdots, n \\
\beta_j \mid \tilde{\boldsymbol{\pi}} = (\tilde{\pi}_0, \cdots, \tilde{\pi}_L)' &\overset{i.i.d.}{\sim} \sum_{h=0}^{L} \tilde{\pi}_h \delta_{\theta_h} \\
\pi_0 &\sim \mathcal{U}(0,1) \\
\pi_h &= V_h \prod_{\ell=1}^{h-1}(1 - V_\ell) \text{ with } V_h \overset{i.i.d.}{\sim} Beta(1, \alpha_0) \\
\theta_h &\overset{i.i.d.}{\sim} \mathcal{N}(0, \eta^2) \\
\eta^2 &\sim Inv\text{-}\mathcal{G}(a_\eta, b_\eta) \\
\sigma^2 &\sim Inv\text{-}\mathcal{G}(a_\sigma, b_\sigma).
\end{aligned}$$

Based on this, it is easy to write out the posterior distribution of this linear model

$$\begin{aligned}
& \pi(\boldsymbol{\beta}, \boldsymbol{\theta}, \pi_0, \mathbf{V}, \eta^2, \sigma^2 \mid \mathbf{X}, \mathbf{y}, \alpha_0, a_\eta, b_\eta, a_\sigma, b_\sigma) \\
\propto\ & \mathcal{L}(\mathbf{y}; \mathbf{X}, \boldsymbol{\beta}, \sigma^2) \prod_{i=1}^{n} \pi(\beta_j \mid \pi_0, \mathbf{V}, \boldsymbol{\theta}) \pi(\pi_0) \prod_{h=1}^{L} \pi(V_h \mid \alpha_0) \prod_{h=1}^{L} \pi(\theta_h \mid \eta^2) \pi(\eta^2) \pi(\sigma^2) \\
\propto\ & \sigma^{-n} \exp\left\{ -\frac{\sum_{i=1}^{n}(y_i - \mathbf{x}_i'\boldsymbol{\beta})^2}{2\sigma^2} \right\} \prod_{j=1}^{p} \Bigg\{ \pi_0 \mathbf{1}(\beta_j = 0) + \sum_{h=1}^{L} (1 - \pi_0) \\
\times\ & \left[V_h \prod_{\ell=1}^{h-1}(1 - V_\ell) \right] \mathbf{1}(\beta_j = \theta_h) \Bigg\} \prod_{h=1}^{L} (1 - V_h)^{\alpha_0 - 1} \eta^{-L} \left(\prod_{h=1}^{L} \exp\left\{ -\frac{\theta_h^2}{2\eta^2} \right\} \right) \\
\times\ & \eta^{-2a_\eta - 2} \exp\left(-\frac{b_\eta}{\eta^2} \right) \sigma^{-2a_\sigma - 2} \exp\left(-\frac{b_\sigma}{\sigma^2} \right)
\end{aligned}$$

1.4. **Gibbs Sampling Procedure.**

1.4.1. *Full conditional distribution of $\boldsymbol{\beta}$.*

$$[\beta_j \mid \boldsymbol{\beta}_{[-j]}, \boldsymbol{\theta}, \mathbf{V}, \pi_0] \sim \sum_{h=0}^{L} \tilde{\pi}_h^* \delta_{\theta_h},\ \forall j = 1, \cdots, p$$

where

$$\tilde{\pi}_h^* = \frac{\tilde{\pi}_h \mathcal{N}\left(\theta_h; \frac{\sum_{i=1}^{n}(y_i - \mathbf{x}_{[-i]}'\boldsymbol{\beta}_{[-j]}) x_{ij}}{\sigma^{-2}\sum_{i=1}^{n} x_{ij}^2}, \frac{\sigma^2}{\sum_{i=1}^{n} x_{ij}^2} \right)}{\sum_{k=1}^{L} \tilde{\pi}_k \mathcal{N}\left(\theta_k; \frac{\sum_{i=1}^{n}(y_i - \mathbf{x}_{[-i]}'\boldsymbol{\beta}_{[-j]}) x_{ij}}{\sigma^{-2}\sum_{i=1}^{n} x_{ij}^2}, \frac{\sigma^2}{\sum_{i=1}^{n} x_{ij}^2} \right)}.$$

1.4.2. *Full conditional distribution of $\boldsymbol{\theta}$.* Let us define $m_h = \#\{j : \beta_j = \theta_h\}$ and define

$$\begin{aligned} z_i &= y_i - \sum_{\{j:\beta_j \neq \theta_h\}} x_{ij}\beta_j, \\ x_{im_h} &= \sum_{\{j:\beta_j = \theta_h\}} x_{ij}. \end{aligned}$$

Then for $h = 1, \cdots, L$,

$$[\theta_h \mid \boldsymbol{\beta}, \boldsymbol{\theta}_{[-h]}, \sigma^2, \eta^2] \sim \mathcal{N}\left(\frac{\sigma^{-2}\sum_{i=1}^n z_i x_{im_h}}{\sigma^{-2}\sum_{i=1}^n x_{im_h}^2 + \eta^{-2}}, \frac{1}{\sigma^{-2}\sum_{i=1}^n x_{im_h}^2 + \eta^{-2}}\right).$$

Here $\boldsymbol{\theta}_{[-h]}$ is useful to group and calculate the number of coefficients $\beta_j = \theta_h$ for $j = 1, \cdots, p$.

1.4.3. *Full conditional distribution of π_0.* Assume $m_0 = \#\{j : \beta_h = 0\}$, then $(p - m_0) = \#\{j : \beta_h \neq 0\}$. Thus,

$$[\pi_0 \mid \boldsymbol{\beta}, \boldsymbol{\theta}, \mathbf{V}] \sim Beta(m_0 + 1, p - m_0 + 1)$$

1.4.4. *Full conditional distribution of $\mathbf{V}$.*

$$[V_h \mid \boldsymbol{\beta}, \boldsymbol{\theta}, \pi_0, \alpha_0] \sim Beta(m_h + 1, \sum_{\ell=h+1}^{L} m_\ell + \alpha_0),\ \forall h = 1, \cdots, L-1,$$

and $V_L = 1$.

1.4.5. *Full conditional distribution of η^2.*

$$[\eta^2 \mid \boldsymbol{\theta}, a_\eta, b_\eta] \sim Inv\text{-}\mathcal{G}\left(\frac{L}{2} + a_\eta, \frac{\sum_{h=1}^L \theta_h^2}{2} + b_\eta\right)$$

1.4.6. *Full conditional distribution of σ^2.*

$$[\sigma^2 \mid \boldsymbol{\beta}, a_\sigma, b_\sigma] \sim Inv\text{-}\mathcal{G}\left(\frac{n}{2} + a_\sigma, \frac{\sum_{i=1}^n (y_i - \mathbf{x}_i'\boldsymbol{\beta})^2}{2} + b_\sigma\right)$$

At the first attempt, let us use $a_\sigma = b_\sigma = a_\eta = b_\eta = 0.1$ and $\alpha = 1$.

1.5. **Varying the choice of hyperparamters.**