VARIABLE SELECTION FOR MULTICOLINEARITY

1. Model and Posterior Distribution

Let us assume the linear model is

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

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 Curtis and Ghosh (2011), using the finite approximation of the $\mathcal{P}^*$:

$$\beta_j \overset{i.i.d.}{\sim} \pi_0\delta_0 + (1-\pi_0)\mathcal{P}^* = \pi_0\delta_0 + (1-\pi_0)\sum_{h=1}^{L} p_h\delta_{\theta_h},$$

where $(p_1, \cdots, p_L) \sim Dir(\alpha/L, \cdots, \alpha/L)$ and $\theta_h \overset{i.i.d}{\sim} \mathcal{P}_0 = \mathcal{N}(0, \eta^{-1})$. Moreover, if we assume $\theta_0 = 0$ and define $\tilde{p}_0 = \pi_0$ and $\tilde{\pi}_h = (1-\pi_0)p_h$, then

$$\beta_j \overset{i.i.d.}{\sim} \sum_{h=0}^{L} \tilde{p}_h\delta_{\theta_h}.$$

The other way I am thinking is to do the modeling like this way:

$$\begin{aligned}
\beta_j &\overset{i.i.d.}{\sim} \mathcal{P},\ \mathcal{P} \sim DP(\alpha_0, \mathcal{P}_0) \\
\theta_h &\overset{i.i.d.}{\sim} \pi\delta_0 + (1-\pi)\mathcal{N}(0, \eta^{-1}) := \mathcal{P}_0,\ \forall h, \\
\pi &\sim Beta(a_\pi, b_\pi).
\end{aligned}$$

1.2. Likelihood for the Linear Model.

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

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

where $y_i \sim \mathcal{N}(\mathbf{x}_i'\boldsymbol{\beta}, \sigma^{-1})$, $\mathbf{x}_i = (x_{i1}, \cdots, x_{ip})'$ and $\mathbf{X} = (\mathbf{x}_1, \cdots, \mathbf{x}_n)'$.

1.3. **Posterior Distribution.** Then the model becomes,

$$\begin{aligned}
y_i &\sim \mathcal{N}(\mathbf{x}_i'\boldsymbol{\beta}, \sigma^{-1}),\ i=1,\cdots,n \\
\beta_j \mid \tilde{\boldsymbol{\pi}} = (\tilde{p}_0,\cdots,\tilde{p}_L)' &\overset{i.i.d.}{\sim} \sum_{h=0}^{L} \tilde{p}_h \delta_{\theta_h},\ \forall j=1,\cdots,p \\
\pi_0 &\sim \mathcal{U}(0,1) \\
\mathbf{p} = (p_1,\cdots,p_h)' &\sim Dir(\alpha/L,\cdots,\alpha/L) \\
\theta_h &\overset{i.i.d.}{\sim} \mathcal{N}(0,\eta^{-1}), \forall h=1,\cdots,L \\
\eta &\sim \mathcal{G}(a_\eta, b_\eta) \\
\sigma &\sim \mathcal{G}(a_\sigma, b_\sigma).
\end{aligned}$$

Let us introduce latent indexes variables $\mathbf{S}$, then the full hierarchical Bayesian model is like

$$\begin{aligned}
y_i &\sim \mathcal{N}(\mathbf{x}_i'\boldsymbol{\beta}, \sigma^{-1}),\ i=1,\cdots,n \\
(\beta_j \mid \theta_h, S_j = h) &= \theta_h,\ \forall j=1,\cdots,p,\ \forall h=0,\cdots,L \\
S_j &\sim Multinom(\tilde{p}_0,\cdots,\tilde{p}_L),\ \forall j=1,\cdots,p \\
\pi_0 &\sim \mathcal{U}(0,1) \\
\mathbf{p} = (p_1,\cdots,p_h)' &\sim Dir(\alpha/L,\cdots,\alpha/L) \\
\theta_h &\overset{i.i.d.}{\sim} \mathcal{N}(0,\eta^{-1}),\ \forall h=1,\cdots,L \\
\eta &\sim \mathcal{G}(a_\eta, b_\eta) \\
\sigma &\sim \mathcal{G}(a_\sigma, b_\sigma).
\end{aligned}$$

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

$$\begin{aligned}
& \pi(\boldsymbol{\beta}, \boldsymbol{\theta}, \mathbf{S}, \pi_0, \mathbf{p}, \eta, \sigma \mid \mathbf{X}, \mathbf{y}, \alpha_0, a_\eta, b_\eta, a_\sigma, b_\sigma) \\
\propto\ & \mathcal{L}(\mathbf{y}; \mathbf{X}, \boldsymbol{\beta}, \sigma^{-1}) \prod_{i=1}^{n} \pi(\beta_j \mid \pi_0, \mathbf{p}, \mathbf{S}, \boldsymbol{\theta}) \prod_{h=1}^{L} \mathcal{N}(\theta_h; 0, \eta^{-1}) \prod_{j=1}^{p} \pi(S_j \mid \pi_0, \mathbf{p}) \\
\times\ & Unif(\pi_0; 0,1) Dir(\mathbf{p}; \alpha/L, \cdots, \alpha/L) \pi(\eta^{-1}) \pi(\sigma^{-1}) \\
\propto\ & \sigma^{n/2} \exp\left\{ -\frac{\sigma \sum_{i=1}^{n} (y_i - x_{i1}\theta_{S_1} - \cdots - x_{ip}\theta_{S_p})^2}{2} \right\} \prod_{j=1}^{p} \prod_{h=0}^{L} \tilde{\pi}_h^{\mathbf{1}(S_j=h)} \\
\times\ & \left(\prod_{h=1}^{L} \eta^{1/2} \exp\left\{ -\frac{\eta \theta_h^2}{2} \right\} \right) \prod_{h=1}^{L} p_h^{(\alpha/L-1)} \eta^{a_\eta - 1} \exp(-b_\eta \eta)\, \sigma^{a_\sigma - 1} \exp(-b_\sigma \sigma)
\end{aligned}$$

1.4. **Gibbs Sampling Procedure.**

1.4.1. *Full conditional distribution of* $\boldsymbol{\theta}$. Let us define $m_h = \{j : S_j = h\}$ and $|m_h|$ is the total number of elements in the set m_h. define

$$
\begin{aligned}
z_i &= y_i - \sum_{j \notin m_h} x_{ij}\theta_{S_j}, \\
x_{im_h} &= \sum_{j \in m_h} x_{ij}.
\end{aligned}
$$

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

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

If $|m_h| = 0$, then

$$[\theta_h \mid \eta] \sim \mathcal{N}(0, \eta^{-1}).$$

1.4.2. *Full conditional distribution of* $\mathbf{S}$. For $j = 1, \cdots, p$,

$$[S_j = h \mid S_{[-j]}, \boldsymbol{\theta}, \sigma, \pi_0, \mathbf{p}] \propto \tilde{p}_h \exp\left\{-\frac{\sigma \sum_{i=1}^n (y_i - \sum_{k \neq j} x_{ik}\theta_{S_k} - x_{ij}\theta_h)^2}{2}\right\}$$

1.4.3. *Full conditional distribution of* π_0. Assume $m_0 = \{j : S_h = 0\}$ and $|m_0|$ be the total number of elments in the set m_0. Thus,

$$[\pi_0 \mid \mathbf{p}, \mathbf{S}] \sim Beta(|m_0| + 1, p - |m_0| + 1)$$

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

$$[\mathbf{p} \mid \mathbf{S}, \pi_0, \alpha] \sim Dir(\alpha/L + |m_1|, \cdots, \alpha/L + |m_L|).$$

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

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

1.4.6. *Full conditional distribution of* σ.

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

where $\boldsymbol{\beta} = (\theta_{S_1}, \cdots, \theta_{S_p})'$.

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.**