Example: Large Scale Bayesian variable selection

data: $y \in \mathbb{R}^n \quad X \in \mathbb{R}^{n \times p}, \ p >> n$

model: $Y = X\beta + \varepsilon \ , \ \varepsilon \sim N(0, \sigma^2 I_n)$

Prior info: β is sparse

effective approach: spike + slab prior

Given $\delta = (\delta_1, \ldots, \delta_p) \in \mathbb{R}^p \quad \delta_j \in \{0, 1\}$

$$\beta_j | \delta_j \sim \begin{cases} N(0, \gamma_1) & \delta_i = 1 \\ N(0, \delta_0) & \delta_i = 0 \end{cases} \qquad \delta_j \sim \text{Bern}(\alpha)$$

Marginally $\beta_j \propto \alpha N(0, \delta_1) + (1-\alpha) N(0, \gamma_0)$

Typically we take $\delta_1 >> \delta_0$

Then we change the model $Y = X\beta_\delta + \varepsilon \ ; \ \beta_\delta = \beta \cdot \delta$

To summarize the hieraehial model

$$Y = X\beta_\delta + \varepsilon$$

$$\delta \overset{iid}{\sim} \text{Bern}(\alpha)$$

$$\beta_j | \delta_j \overset{ind}{\sim} \begin{cases} N(0, \gamma_1) & \delta_j = 1 \\ N(0, \gamma_0) & \delta_j = 0 \end{cases}$$

$$\pi(\sigma^2) = 1 \quad \text{(improper prior)}$$

Variation! Leave first level and include

$\beta_j \sim N(0, \gamma_0)$ to obtain an approx. sparse solution

$$\pi(\sigma^2, \beta, \delta | y) \propto \left(\frac{1}{\sigma^2}\right)^{n/2} e^{-\frac{1}{2\sigma^2} \|Y - X\beta_\delta\|_2^2} \prod_{j=1}^{p} \alpha^{\delta_i}(1-\alpha)^{1-\delta_i} f_1(\beta_j)^{\delta_i} f_0(\beta_j)^{1-\delta_i}$$

$$f_i(x) = \frac{1}{\sqrt{2\pi\gamma_i}} e^{-\frac{x_i^2}{2\gamma_i}}$$

Rmk: We could also use an IG prior on σ^2

MCMC: Given $(\beta, \delta) \Rightarrow \sigma^2 \sim IG\left(\frac{n}{2} - 1, \frac{\|Y - X\beta_\delta\|_2^2}{2}\right)$

Given (σ^2, δ) linear regression

Model $\begin{cases} Y \sim X D_\delta \beta + \varepsilon \\ \quad D_\delta = \text{diag}(\vec{\delta}) \\ \beta \sim N(0, m_\delta) \quad m_\delta = \begin{cases} \gamma_1 & \delta_1 = 1 \\ \gamma_0 & \delta_0 = 0 \end{cases} \end{cases}$

$\xrightarrow{\text{Posterior}}$ $\beta | \sigma^2, \delta, Y \sim N(m_\delta, \sigma^2 \Sigma_\delta)$

$$m_\delta = \Sigma_\delta D_\delta X^T Y$$

$$\Sigma'_\delta = (D_\delta X' X D_\delta + \sigma^{-} M_\delta^{-1})^{-1}$$

Given: (σ^2, β) update each δ_j

$$\|Y - X\beta_\delta\|_2^2 \Rightarrow \delta_j \beta_j^2 \|x_j\|_2^2 - 2\delta_j \beta_j \langle x_j, y - X(\beta\cdot\delta)_{-j} \rangle$$

$$\delta_j \mid \sigma^2, \beta, Y \sim \mathrm{Bern}(\alpha_j) \quad a_j = \frac{1}{1 + C_j}$$

$$C_j = \frac{1-\alpha}{\alpha} \sqrt{\frac{\gamma_1}{\gamma_0}} \exp\left\{ \frac{1}{2\sigma^2} \left(\beta_j^2 \|x_j\|_2^2 - 2\beta_j \langle x_j, y - (X\beta)_{-j} \rangle - \frac{\beta_j^2}{2} \left(\frac{1}{\gamma_0} - \frac{1}{\gamma_1} \right) \right) \right\}$$

Issues: {Expensive: need to invert $D_\delta X^T X D_\delta + \sigma^2 M_\delta^{-1}$

Variational Approximation:

$$q(\sigma^2, \delta, \beta) = \mathrm{IG}(\sigma^2; \gamma_1, \gamma_2) \prod_{j=1}^{p} \mathrm{Bern}(\delta_j, \alpha_j) N(\beta_j; \mu_j, \sigma_j^2)$$

Search $(\gamma, \alpha, \mu, \sigma^2)$ that minimize $KL(\pi, q_{(\gamma, \alpha, \mu, \sigma)})$

$$\mathrm{ELBO}(\gamma, \alpha, \mu, \sigma^2) = \int \Big[\frac{n}{2} \log \frac{1}{\sigma^2} - \frac{\|Y - X\beta_\delta\|_2^2}{2\sigma^2} + \sum_{j=1}^{p} \delta_j \log \alpha + (1-\delta_j) \log(1-\alpha_j)$$

$$+ \delta_j \log f_1(\beta_j) + (1-\delta_j) \log f_0(\beta_j)$$

$$- \log q_{(-\#)}(\sigma^2, \delta, \beta) \Big] q_{(-\#)}(\sigma^2, \delta, \beta)$$

<u>Idea:</u> Fix (α, μ, σ^2)

$$\mathrm{ELB}(\gamma) = \int \Big[\frac{n}{2} \log\left(\frac{1}{\sigma^2}\right) - \frac{\mathbb{E}[\|Y - X\beta\|_2^2]}{2\sigma^2} - \log q_\gamma(\sigma^2) \Big] q(\gamma)$$

$$\Rightarrow r_1 = \frac{n}{2} \qquad r_2 = \mathbb{E}_q \left[\frac{\|Y - X\beta\|_2^2}{2} \right]$$

because $IG(x; c, d) \propto \left(\frac{1}{x}\right)^{c+1} e^{-d/x}$

$$\log IG(x, c, d) = \text{const.} + (c+1) \log \frac{1}{x} - \frac{d}{x}$$

$$\mathbb{E}_q \left[\|Y - X\beta\|_2^2 \right] = \mathbb{E}_q \left[\|Y\|_2^2 + \beta_\delta^T X^T X \beta_\delta - 2 \langle Y, X\beta_\delta \rangle \right]$$

$$= \|y\|_2^2 + \mathbb{E}_q \left[\mathrm{Tr}(X^T X \beta_\delta \beta_\delta^T) - 2 \langle Y, X\beta_\delta \rangle \right]$$

$$= \|y\|_2^2 + \sum_{j=1}^{p} \alpha_j \|x_j\|_2^2 (\mu_j^2 + \sigma^2) + \sum_{j \neq i} \langle x_i, x_j \rangle \mu_i \mu_j \alpha_i \alpha_j$$

$$- 2 \langle y, X(\mu \cdot \alpha) \rangle$$

$$= \|y - X(\mu \cdot \alpha)\|_2^2 + \sum_{j=1}^{p} (*)$$