1 Mixture Model with Mixture Normal Error (Univariate)

First we consider the univariate case. Suppose we observe responses $Y_i, i = 1, ..., n$ and missingness indicator of Y_{i2}: $R_i = 1, 0$ for $i = 1, ..., n$ with covariates: $\boldsymbol{x}_i$. What we are interested is the marginal quantile regression of Y_i. Suppose the model is

$$\begin{aligned}
Y_i|R_i = 1 &\sim \Delta_i + \boldsymbol{x}_i\boldsymbol{\beta}^{(1)} + \epsilon_i^{(1)} \\
Y_i|R_i = 0 &\sim \Delta_i + \boldsymbol{x}_i\boldsymbol{\beta}^{(0)} + \epsilon_i^{(0)} \\
R_i &\sim \text{Bernoulli}(\pi) \\
\epsilon_i^{(R)}|G_i = j, R_i &\sim N(\mu_j, \sigma_j^2), j = 1, ..., K \\
G_i|R_i &\sim \text{Multinomial}(\boldsymbol{\omega}_{(R)}) \\
p(Y_i \leq \boldsymbol{x}_i\gamma) &= \tau_2,
\end{aligned}$$

where Δ_i is a function of other parameters which will be explained later.

Here we suppose $\epsilon_i^{(1)}$ and $\epsilon_i^{(0)}$ share the same mixture distributions (μ, σ), but with different probabilities ($\omega_{(1)}$ and $\omega_{(0)}$). We can weaken the assumption with different mixture distributions ($\mu^{(1)}, \mu^{(0)}, \sigma^{(1)}, \sigma^{(0)}$).

Priors 0

$$\begin{aligned}
\boldsymbol{\gamma} &\sim N(\boldsymbol{\gamma}^p, \boldsymbol{V}_\gamma^p) \\
\boldsymbol{\beta} &\sim N(\boldsymbol{\beta}^p, \boldsymbol{V}_\beta^p) \\
\pi &\sim \text{Beta}(\alpha_{\pi 1}, \alpha_{\pi 2}) \\
&\propto \pi^{\alpha_{\pi 1}-1}(1-\pi)^{\alpha_{\pi 2}-1} \\
\boldsymbol{\omega}_{(R)} &\sim \text{Dirichlet}(\boldsymbol{\alpha})
\end{aligned}$$

Priors 1a

$$\begin{aligned}
\mu_j &\sim N(m, r\sigma_j^2) \\
1/\sigma_j^2 &\sim \text{Gamma}(a/2, b/2) \\
\text{Hypepriors:} \quad 1/r &\sim \text{Gamma}(c/2, d/2) \\
m &\sim N(m_0, \tau_m)
\end{aligned}$$

where $a, b, c, d, m_0, \tau_m, \boldsymbol{\alpha}$ are fixed and known. Fix mean of μ_j at m_0 because of identifiability problem.

Prior 1b To simplify the problem, another priors settings can be specified as

$$\begin{aligned}
\mu_j &\sim N(0, 1) \\
\sigma_j &= 1
\end{aligned}$$

To summarize responses and paramters:

Observed: $Y_i, \boldsymbol{x}_i, R_i$ for $i = 1, \ldots, n$,

Updatable: **parameters:** denote $\boldsymbol{\theta} = (\gamma, \boldsymbol{\beta}, \pi, \boldsymbol{\omega}, \boldsymbol{\mu}, \boldsymbol{\sigma})$;

latent variables: G_i

1.1 Full Conditional

The full conditional distribution is

$$p_i(Y_i, G_i = j, R_i|\boldsymbol{\theta}, \boldsymbol{x}_i) = p(Y_i|R_i, G_i = j, \boldsymbol{x}_i, \boldsymbol{\theta})p(G_i = j|\boldsymbol{\theta}, R_i)p(R_i|\boldsymbol{\theta}). \tag{1}$$

Later we will see Δ_i is a function of $\boldsymbol{\theta} = (\gamma, \boldsymbol{\beta}, \pi, \boldsymbol{\omega}, \boldsymbol{\mu}, \boldsymbol{\sigma}), \boldsymbol{x}_i$, but not related to G_i and R_i.

Expand RHS in (1),

$$\begin{aligned} p(Y_i|R_i, G_i = j, \boldsymbol{x}_i, \boldsymbol{\theta}) &= \phi_N(Y_i; \Delta_i + \mu_j + \boldsymbol{x}_i\boldsymbol{\beta}^{(R_i)}, \sigma_j^2) \\ p(G_i = j|\boldsymbol{\theta}, R_i) &= \omega_{1(R_i)}^{I(G_i=1)} \cdots \omega_{K(R_i)}^{I(G_i=K)} \\ p(R_i|\boldsymbol{\theta}) &= \pi^{R_i}(1-\pi)^{1-R_i}. \end{aligned}$$

1.2 Calculation of Δ_i

Δ_i is determined by $\boldsymbol{\theta} = (\boldsymbol{\gamma}, \boldsymbol{\beta}, \boldsymbol{\pi}, \boldsymbol{\omega}, \boldsymbol{\mu}, \boldsymbol{\sigma})$ and $\boldsymbol{x}_i$.

$$\begin{aligned} \tau &= P(Y_i \leq \boldsymbol{x}_i\boldsymbol{\gamma}) \\ &= \pi P(Y_i \leq \boldsymbol{x}_i\boldsymbol{\gamma}|R = 1) + (1-\pi)P(Y_i \leq \boldsymbol{x}_i\boldsymbol{\gamma}|R = 0) \\ &= \pi \left[\sum_{j=1}^{K} \omega_{j(1)} P(Y_i \leq \boldsymbol{x}_i\boldsymbol{\gamma}|R = 1, G_i = j)\right] \\ &\quad + (1-\pi)\left[\sum_{j=1}^{K} \omega_{j(0)} P(Y_i \leq \boldsymbol{x}_i\boldsymbol{\gamma}|R = 0, G_i = j)\right] \\ &= \pi \left(\sum_{j=1}^{K} \omega_{j(1)} \Phi\left(\frac{\boldsymbol{x}_i\boldsymbol{\gamma} - (\Delta_i + \boldsymbol{x}_i\boldsymbol{\beta}^{(1)} + \mu_j)}{\sigma_j}\right)\right) \\ &\quad + (1-\pi)\left(\sum_{j=1}^{K} \omega_{j(0)} \Phi\left(\frac{\boldsymbol{x}_i\boldsymbol{\gamma} - (\Delta_i + \boldsymbol{x}_i\boldsymbol{\beta}^{(0)} + \mu_j)}{\sigma_j}\right)\right) \end{aligned}$$

Thus $\Delta_i = h(\boldsymbol{\gamma}, \boldsymbol{\beta}, \pi, \boldsymbol{\omega}, \boldsymbol{\mu}, \boldsymbol{\sigma}, \boldsymbol{x}_i)$.

1.3 MCMC

All the unknown parameters are (for prior 1a)

$$\boldsymbol{\theta} = (\gamma, \boldsymbol{\beta}, \pi, \boldsymbol{\omega}, \mu, \sigma^2, m, r), G_i$$

and for prior 1b,

$$\boldsymbol{\theta} = (\gamma, \boldsymbol{\beta}, \pi, \boldsymbol{\omega}, \mu), G_i.$$

We use block Gibbs sampling method to sample posterior distribution of $\boldsymbol{\theta}$ and $\boldsymbol{G}$.

$\boldsymbol{\theta}|\boldsymbol{Y}, \boldsymbol{X}, \boldsymbol{R}, \boldsymbol{G}$:

$$\begin{aligned}
p(\boldsymbol{\theta}|\boldsymbol{Y}, \boldsymbol{X}, \boldsymbol{R}, \boldsymbol{G}) \propto\ & p(\boldsymbol{Y}, \boldsymbol{G}, \boldsymbol{R}|\boldsymbol{\theta}, \boldsymbol{X})\pi(\boldsymbol{\theta}) \\
=\ & \prod_{i=1}^{n} \left[p_i(Y_i, G_i = j, R_i|\boldsymbol{\theta}, \boldsymbol{x}_i)\right] \pi(\boldsymbol{\theta}) \\
=\ & \prod_{i=1}^{n} \left[p(Y_i|R_i, G_i = j, \boldsymbol{x}_i, \boldsymbol{\theta}) p(G_i = j|\boldsymbol{\omega}, R_i) p(R_i|\pi)\right] \pi(\boldsymbol{\theta}) \\
=\ & \prod_{i=1}^{n} \left[\phi_N(Y_i; \Delta_i + \boldsymbol{x}_i \boldsymbol{\beta}^{(R_i)} + \mu_{G_i}, \sigma_{G_i})\right] \left[\omega_{1(R_i)}^{\sum I(G_i=1)} \cdots \omega_{K(R_i)}^{\sum I(G_i=K)}\right] \\
& \left[\pi^{\sum R_i}(1-\pi)^{n-\sum R_i}\right]
\end{aligned}$$

Metropolis-Hasting sampling algorithm is needed. We may apply the following candidate distribution:

γ: normal random walk

$\boldsymbol{\beta}$: normal random walk

π: normal random walk ?

$\boldsymbol{\omega}$: normal random walk ?

$G|\boldsymbol{Y}, \boldsymbol{R}, \boldsymbol{X}, \boldsymbol{\theta}$:

$$\begin{aligned}
p(G_i|\boldsymbol{Y}_i, \boldsymbol{x}_i, R_i, \boldsymbol{\theta}) \propto\ & p(\boldsymbol{Y}_i, \boldsymbol{G}_i, R_i|\boldsymbol{\theta}, \boldsymbol{x}_i) \\
=\ & p(Y_i|R_i, G_i, \boldsymbol{x}_i, \boldsymbol{\theta}) p(G_i = j|\boldsymbol{\omega}, R_i) \\
=\ & \phi_N(Y_i; \Delta_i + \boldsymbol{x}_i \boldsymbol{\beta}^{(R_i)} + \mu_{G_i}, \sigma_{G_i}) \omega_{G_i(R_i)}
\end{aligned}$$

Thus the posterior of G_i is still multinomial, but with different parameters $\boldsymbol{\omega}^*$:

$$\begin{aligned}
p(G_i = j|\boldsymbol{Y}_i, \boldsymbol{x}_i, R_i, \boldsymbol{\theta}) =\ & \frac{\omega^*_{j(R_i)}}{\sum_{k=1}^{K} \omega^*_{k(R_i)}}, \\
\omega^*_{j(R_i)} =\ & \omega_{j(R_i)} \phi_N(Y_i; \Delta_i + \boldsymbol{x}_i \boldsymbol{\beta}^{(R_i)} + \mu_j, \sigma_j^2).
\end{aligned}$$

2 Mixture model with mixture normal error (Bivariate)

Consider the bivariate case. Suppose we observe responses response $Y_i = (Y_{i1}, Y_{i2}), i = 1, ..., n$, missingness indicator of Y_{i2}: $R_i = 1, 0; i = 1, ..., n$ and covariates $\boldsymbol{x}_i$.

Using pattern mixture model settings,

Y_{i1}:

$$
\begin{aligned}
Y_{i1}|R_i = 1 &\sim \Delta_{i1} + \boldsymbol{x}_i\boldsymbol{\beta}_1^{(1)} + \epsilon_{i1}^{(1)} \\
Y_{i1}|R_i = 0 &\sim \Delta_{i1} + \boldsymbol{x}_i\boldsymbol{\beta}_1^{(0)} + \epsilon_{i1}^{(0)} \\
R_i &\sim \text{Bernoulli}(\pi) \\
\epsilon_{i1}^{(R_i)}|G_{i1} = j, R_i &\sim N(\mu_{1j}, \sigma_{1j}^2), j = 1, ..., K \\
G_{i1}|R_i &\sim \text{Multinomial}(\boldsymbol{\omega}_{1(R_i)}) \\
\boldsymbol{\omega}_{1(R_i)} &= (\omega_{11(R_i)}, ..., \omega_{1K(R_i)})
\end{aligned}
$$

where Δ_{i1} is a function of other parameters which will be explained later.

Y_{i2}:

$$
\begin{aligned}
Y_{i2}|Y_{i1}, R_i = 1 &\sim \Delta_{i2} + \beta_y Y_{i1} + \epsilon_{i2}^{(1)} \\
Y_{i2}|Y_{i1}, R_i = 0 &\sim \Delta_{i2} + \boldsymbol{x}_i\boldsymbol{\beta}_2^{(0)} + \beta_{ySP} Y_{i1} + \epsilon_{i2}^{(0)} \\
\epsilon_{i2}^{(R_i)}|G_{i2} = j, R_i &\sim N(\mu_{2j}, \sigma_{2j}^2), j = 1, ..., K \\
G_{i2} &\sim \text{Multinomial}(\boldsymbol{\omega}_{2(R_i)}) \\
\boldsymbol{\omega}_{2(R_i)} &= (\omega_{21(R_i)}, ..., \omega_{2K(R_i)})
\end{aligned}
$$

where Δ_{i2} is a function of other parameters which will be explained later.

Here we suppose $\epsilon_{i2}^{(1)}$ and $\epsilon_{i2}^{(0)}$ share the same mixture distributions (μ_2, σ_2), but with different probabilities ($\omega_{2(1)}$ and $\omega_{2(0)}$).

Thus $\boldsymbol{\beta}_2^{(0)}$, β_{ySP} and $\boldsymbol{\omega}_{2(0)}$ are sensitivity parameters. When $\beta_2^{(0)} = \mathbf{0}$, $\beta_{ySP} = \beta_y$ and $\boldsymbol{\omega}_{2(0)} = \boldsymbol{\omega}_{2(1)}$, MAR condition satisfies.

2.1 Calculation of Δ_{i1} and Δ_{i2}

Δ_{i1}**:**

$$\begin{aligned}
\tau &= P(Y_{i1} \leq \boldsymbol{x}_i\gamma_1) \\
&= \pi P(Y_{i1} \leq \boldsymbol{x}_i\gamma_1|R_i = 1) + (1-\pi)P(Y_{i1} \leq \boldsymbol{x}_i\gamma_1|R_i = 0) \\
&= \pi\left[\sum_{j=1}^{K}\omega_{1j(1)}P(Y_{i1} \leq \boldsymbol{x}_i\gamma_1|R_i = 1, \mu_{1j}, \sigma_{1j})\right] \\
&\quad + (1-\pi)\left[\sum_{j=1}^{K}\omega_{1j(0)}P(Y_{i1} \leq \boldsymbol{x}_i\gamma|R_i = 0, \mu_{1j}, \sigma_{1j})\right] \\
&= \pi\left(\sum_{j=1}^{K}\omega_{1j(1)}\Phi\left(\frac{\boldsymbol{x}_i\gamma_1 - (\Delta_{i1} + \boldsymbol{x}_i\boldsymbol{\beta}_1^{(1)} + \mu_{1j})}{\sigma_{1j}}\right)\right) \\
&\quad + (1-\pi)\left(\sum_{j=1}^{K}\omega_{1j(0)}\Phi\left(\frac{x_i\gamma_1 - (\Delta_{i1} + x_i\beta_1^{(0)} + \mu_{1j})}{\sigma_{1j}}\right)\right)
\end{aligned}$$

Thus $\Delta_{i1} = h_1(\boldsymbol{\gamma_1}, \boldsymbol{\beta_1^{(R_i)}}, \pi, \boldsymbol{\omega}_{1(R_i)}, \boldsymbol{\mu}_1, \boldsymbol{\sigma}_1, \boldsymbol{x}_i)$.

Every term in the equation is closed form in terms of standard normal CDF. And it is monotone for Δ_{i1}.

Δ_{i2}**:**

$$\begin{aligned}
\tau &= P(Y_{i2} \leq \boldsymbol{x}_i\gamma_2) \\
&= \pi P(Y_{i2} \leq \boldsymbol{x}_i\gamma_2|R_i = 1) + (1-\pi)P(Y_{i2} \leq \boldsymbol{x}_i\gamma_2|R_i = 0) \\
&= \pi\left(\int P(Y_{i2} \leq \boldsymbol{x}_i\gamma_2|R_i = 1, Y_1)dF(Y_1|R_i = 1)\right) + \\
&\quad (1-\pi)\left(\int P(Y_{i2} \leq \boldsymbol{x}_i\gamma_2|R_i = 0, Y_1)dF(Y_1|R_i = 0)\right) \\
&= \pi\left[\int\sum_{j=1}^{K}\omega_{2j(1)}P(Y_{i2} \leq \boldsymbol{x}_i\gamma_2|R_i = 1, Y_1, G_2 = j)d(\sum_{j'=1}^{K}\omega_{1j'(1)}F(Y_1|R_i = 1, G_1 = j'))\right] \\
&\quad + (1-\pi)\left[\int\sum_{j=1}^{K}\omega_{2j(0)}P(Y_{i2} \leq \boldsymbol{x}_i\gamma_2|R_i = 0, Y_{i1}, G_2 = j)d(\sum_{j'=1}^{K}\omega_{1j'(0)}F(Y_1|R_i = 0, G_1 = j'))\right] \\
&= \pi\left[\sum_{j,j'}^{K}\omega_{2j(1)}\omega_{1j'(1)}\int\Phi\left(\frac{\boldsymbol{x}_i\gamma_2 - (\Delta_{i2} + \beta_y Y_1 + \mu_{2j})}{\sigma_{2j}}\right)dF(Y_1; \Delta_{i1} + x_i\beta_1^{(1)} + \mu_{1j'}, \sigma_{1j'}^2)\right] \\
&\quad + (1-\pi)\left[\sum_{j,j'}^{K}\omega_{2j(0)}\omega_{1j'(0)}\int\Phi\left(\frac{\boldsymbol{x}_i\gamma_2 - (\Delta_{i2} + \boldsymbol{x}_i\boldsymbol{\beta}_2^{(0)} + \beta_{ySP}Y_1 + \mu_{2j})}{\sigma_{2j}}\right)\right. \\
&\quad\quad \left. dF(Y_1; \Delta_{i1} + x_i\beta_1^{(0)} + \mu_{1j'}, \sigma_{1j'}^2)\right]
\end{aligned}$$

As noted before, each term in above equation is a closed form in terms of standard normal CDF. Given all the identifiable parameters and sensitivity parameters, Δ_{i2} can be determined and a function of

$$\Delta_{i2} = h_2(\gamma_1, \gamma_2, \boldsymbol{\beta}_1^{(R_i)}, \boldsymbol{\beta}_2^{(0)}, \pi, \boldsymbol{\omega}_1, \boldsymbol{\omega}_2, \tau, \boldsymbol{\mu}_1, \boldsymbol{\mu}_2, \boldsymbol{\sigma}_1, \boldsymbol{\sigma}_2, \beta_y, \beta_{ySP}, \boldsymbol{x}_i).$$