1. ~~$g(y|\theta) = C^n(\theta) e^{\sum_{j=1}}$~~

1. ~~$g(y|\theta) = \frac{1}{a(\theta)} e^{\theta t^T(y)} b(y)$~~ $\longrightarrow$ form of exponential family

~~where $a(\theta) = \int_y e^{\theta t^T(y)} b(y)\, dy.$~~

$$Z_{i,k} = \begin{cases} 1 & \text{if } y_i \text{ is from distribution } k. \\ 0 & \text{o.w.} \end{cases}$$

$$f_i(y_i | Z_i, \theta) = \prod_{k=1}^{K} g_k(y_i | \theta_K)^{Z_{i,k}}.$$

$$p_i(y_i, Z_i | \theta, \pi) = f_i(y_i | Z_i, \theta) \cdot q(Z_i | \pi) = \prod_{k=1}^{K} \{\pi_k g_k(y_i | \theta_k)\}^{Z_{i,k}}$$

E-step:

$$Q(\theta, \pi | \theta_p, \pi_p) = \sum_{i=1}^{n} E_{Z|y}(\log p_i(y_i, Z_i | \theta, \pi) \mid \theta_p, \pi_p)$$

$$= \sum_{i=1}^{n} E_{Z|y}[\log f_i(y_i | Z_i, \theta) | \theta_p, \pi_p] + \sum_{i=1}^{n} E_{Z|y}[\log q(Z_i | \pi) \mid \theta_p, \pi_p]$$

$$E_{Z|y}[\log f_i(y_i | Z_i, \theta) | \theta_p, \pi_p] = E_{Z|y}\left[\sum_{k=1}^{K} \log g_k(y_i | \theta_K) \cdot Z_{ik} \mid \theta_p, \pi_p\right]$$

$$= \sum_{k=1}^{K} \log g_k(y_i | \theta_K) E_{Z|y}(Z_{ik} \mid \theta_p, \pi_p) \quad ①$$

$$E_{Z|y}[\log q(Z_i | \pi) | \theta_p, \pi_p] = E_{Z|y}\left[\sum_{k=1}^{K} Z_{ik} \log(\pi_k) \mid \theta_p, \pi_p\right]$$

$$= \sum_{k=1}^{K} \log \pi_k E_{Z|y}(Z_{ik} | \theta_p, \pi_p) \quad ②.$$

~~$E_{Z|y}(Z_{ik} | \theta_p, \theta,)$~~

$$E_{Z|y}(Z_i | \theta_p, \pi_p) = \sum_{Z_i \in \Omega_Z} \frac{\prod_{k=1}^{K} \{\pi_k g_k(y_i | \theta_k)\}^{Z_{i,k}}}{\sum_{k=1}^{K} \pi_k g_k(y_i | \theta_K)} Z_i$$

$$\Rightarrow E_{Z|y}(Z_{ik} | \theta_p, \pi_p) = \frac{\pi_{kp} g_k(y_i | \theta_{kp})}{\sum_{j=1}^{K} \pi_{jp} g_j(y_i | \theta_{jp})}$$

① + ②

$$\Rightarrow Q(\theta, \pi | \theta_p, \pi_p) = \sum_{i=1}^{n} \sum_{k=1}^{K} [\log g_k(y_i | \theta_K) + \log(\pi_K)] \cdot \frac{\pi_{kp} g_k(y_i | \theta_{kp})}{\sum_{j=1}^{K} \pi_{jp} g_j(y_i | \theta_{jp})}.$$

M-step: Since $\sum_{k=1}^{K} \pi_k = 1$, use a Lagrange multiplier.

$$\frac{\partial}{\partial \pi_k} \left(Q(\theta, \pi | \theta_p, \pi_p) - \lambda \left(\sum_{k=1}^{K} \pi_k - 1\right)\right) = \sum_{i=1}^{n} \frac{1}{\pi_k} \frac{\pi_{kp} g_k(y_i | \theta_{kp})}{\sum_{j=1}^{K} \pi_{jp} g_j(y_i | \theta_{jp})} - \lambda$$

$$\Rightarrow \quad \hat{\pi}_k = \frac{1}{n}\sum_{i=1}^{n} \frac{\pi_{kp}\, g_k(y_i \mid \theta_{kp})}{\sum_{j=1}^{k} \pi_{jp}\, g_j(y_i \mid \theta_{jp})}$$

Use

$\frac{\partial}{\partial \theta} Q(\theta, \pi \mid \theta_p, \pi_p)$ to get θ_{p+1}

$$\frac{\partial}{\partial \theta_k} Q(\theta, \pi \mid \theta_p, \pi_p) = \sum_{i=1}^{n} \left(\frac{\pi_{kp}\, g_k(y_i \mid \theta_{kp})}{\sum_{j=1}^{k} \pi_{jp}\, g_j(y_i \mid \theta_{jp})} \cdot \frac{\frac{\partial}{\partial \theta_k} g_k(y_i \mid \theta_k)}{g_k(y_i \mid \theta_k)} \right)$$

Now suppose we have a Beta distribution.

$$g(y; \theta) = \frac{y^{\alpha-1}(1-y)^{\beta-1}}{B(\alpha, \beta)}$$

$$\hat{\pi}_k = \frac{1}{n}\sum_{i=1}^{n} \frac{\pi_{k0}\, g_k(y_i \mid \theta_{k0})}{\sum_{j=1}^{k} \pi_{k0}\, g_j(y_j \mid \theta_{j0})}$$

For α and β, there is no closed form since there is a gamma function in the likelihood, we can use gradient descent to solve.

2. $y_j \sim N(X_j\beta, \underbrace{\tau^2 Z_j Z_j^T + \sigma_j^2 I}_{\Sigma})$ $\quad b_j | y_j \sim N(\tau^2 Z_j^T \Sigma^{-1}(y_j - X_j\beta),\ \tau^2 I - \tau^4 Z_j \Sigma_j^{-1} Z_j)$

$$\begin{pmatrix} b_j \\ y_j \end{pmatrix} \sim N\left(\begin{bmatrix} 0 \\ X_j\beta \end{bmatrix}, \begin{bmatrix} \tau^2 I & \tau^2 Z_j \\ \tau^2 Z_j & \tau^2 Z_j Z_j^T + \sigma_j^2 I \end{bmatrix} \right)$$

E-step:

$$Q = \sum_{j=1}^{m} E[\log(y_j, b_j) | y_j, \theta^p]$$

$$= \sum_{j=1}^{m} E\Big(-\frac{n_j}{2}\log(2\pi\sigma_j^2) - \frac{1}{2}\log\tau^2 - \frac{1}{2\sigma_j^2}\big[(y_j - X_j\beta)^T(y_j - X_j\beta) + (Z_j b_j)^T(Z_j b_j) + (Z_j b_j)^T(y_j - X_j\beta)\big]$$

$$- \frac{1}{2\tau^2} b_j^T b_j \ \Big|\ y_j, \theta^p \Big),$$

where $\theta = \{\beta, \tau, \sigma_j\}, j \in \{1 \dots m\}$

M-step: $\frac{\partial Q}{\partial \beta}, \frac{\partial Q}{\partial \tau}, \frac{\partial Q}{\partial \sigma_j}$

$$Q = \sum_{j=1}^{m}\Big[-\frac{n_j}{2}\log(2\pi) - \log\tau - n_j\log\sigma_j - \frac{(y_j - X_j\beta)^T(y_j - X_j\beta)}{2\sigma_j^2} - \frac{(y_j - X_j\beta)^T}{\sigma_j^2}(\tau^p)^2 Z_j Z_j^T (\Sigma^p)^{-1}(y_j - X_j\beta^p)$$

$$- \frac{1}{2\sigma_j^2}\, \text{trace}\big(Z_j^T Z_j (\tau^p)^2 - (\tau^p)^4 Z_j^T (\Sigma)^{-1} Z_j \big) + (\tau^p)^4 (y_j - X_j\beta^p)^T \Sigma^{p-1} Z_j Z_j^T Z_j Z_j^T (\Sigma^p)^{-1}(y_j - X_j\beta^p)$$

$$- \frac{1}{2\tau^2}\, \text{trace}\big((\tau^p)^2 I - (\tau^p)^4 Z_j^T \Sigma^{-1} Z_j \big) + (\tau^p)^4 (y_j - X_j\beta^p)^T \Sigma^{-1} Z_j Z_j^T \Sigma^{p-1}(y_j - X_j\beta^p) \Big]$$

$$\Rightarrow \frac{\partial Q}{\partial \beta} = -\sum_{j=1}^{m} \frac{(\tau^p)^2 Z_j Z_j^T (\Sigma^p)^{-1}(y_j - X_j\beta^p)}{(\sigma_j^p)^2} X_j - \frac{1}{2}\sum_{j=1}^{m}\frac{1}{\sigma_j^p}\big(2\beta(X_j^T X_j) - 2y_j^T X_j\big)$$

$$\hat{\beta}^{p+1} = \sum_{j=1}^{m} \frac{X_j^T X_j}{(\sigma_j^p)^2} \cdot \sum_{j=1}^{m} \frac{X_j^T\big(y_j - (\tau^p)^2 Z_j Z_j^T (\Sigma^p)^{-1}(y_j - X_j\beta^p)\big)}{(\sigma_j^p)^2}$$

$$\Rightarrow \frac{\partial Q}{\partial \tau} = \sum_{j=1}^{m} \frac{A_j^p}{\tau^3} - \frac{m}{\tau}$$

$$\hat{\tau}^{p+1} = \sqrt{\frac{\sum_{j=1}^{m} A_j^p}{m}}$$

$$\Rightarrow \frac{\partial Q}{\partial \sigma_j} = \frac{1}{(\sigma_j)^3}\big[G_j^p - (y_j - X_j\beta^p)^T(y_j - X_j\beta^p) - 2(y_j - X_j\beta^p)^T(\tau^p)^2 Z_j Z_j^T (\Sigma^p)^{-1}(y_j - X_j\beta^p)\big]$$

$$- \frac{n_j}{\sigma_j}$$

$$\hat{\sigma}_j^{p+1} = \sqrt{\frac{G_j^p - (y_j - X_j\beta^p)^T(y_j - X_j\beta^p) - 2(y_j - X_j\beta^p)^T(\tau^p)^2 Z_j Z_j^T (\Sigma^p)^{-1}(y_j - X_j\beta^p)}{n_j}}$$