# Variational Bayesian Inference for a Mixture of Mixtures

## A Finite Mixture of Finite Gaussian Mixtures

Let $\mathbf{y} = (y_1, y_2, \ldots, y_N)$ be a sample from a finite mixture of finite Gaussian mixtures. We use shared kernels by letting $\mu_{bk} = \mu_k$ and $\sigma^2_{bk} = \sigma^2_k$ for all $k = 1, 2, \ldots, K$ and $b = 1, 2, \ldots, B$. Then their joint density is given by

$$\begin{aligned} p(\mathbf{y} \mid \boldsymbol{\mu}, \boldsymbol{\sigma^2}, \{\mathbf{p}\}_{b=1}^{B}, \boldsymbol{\phi}) &= \prod_{i=1}^{N} \sum_{b=1}^{B} \phi_b \sum_{k=1}^{K} p_{bk} N(y_i \mid \mu_{bk}, \sigma^2_{bk}) \\ &= \prod_{i=1}^{N} \sum_{b=1}^{B} \phi_b \sum_{k=1}^{K} p_{bk} N(y_i \mid \mu_k, \sigma^2_k). \end{aligned}$$

where $\boldsymbol{\phi} = (\phi_1, \phi_2, \ldots, \phi_B)'$ with $\sum_{b=1}^{B} \phi_b = 1$ and $\{\mathbf{p}\}_{b=1}^{B} = \{\mathbf{p}_1, \ldots, \mathbf{p}_B\}$ and $\mathbf{p}_b = (p_{b1}, p_{b2}, \ldots, p_{bK})'$ with $\sum_{k=1}^{K} p_{bk} = 1$.

Let the weights $\boldsymbol{\phi}$ and $\mathbf{p}_b$ have a Dirichlet$(\alpha_\phi \mathbf{1})$ and Dirichlet$(\alpha_p \mathbf{1})$ prior respectively. For each $y_i$ we can introduce auxiliary variable vectors $\mathbf{w} = (\mathbf{w}_1, \mathbf{w}_2, \ldots, \mathbf{w}_N)$ for the "parent" distribution where $\sum_{b=1}^{B} w_{ib} = 1$ and $p(w_{ib} = 1) = \phi_b$, Then given the "parent" distribution we can introduce another set of auxiliary variable vectors $\mathbf{z}_b = (\mathbf{z}_{b1}, \mathbf{z}_{b2}, \ldots, \mathbf{z}_{bN})$ for the "child" distribution where $\sum_{k=1}^{K} z_{ibk} = 1$ and $p(z_{ibk} = 1) = p_{bk}$. Then, $\mathbf{w} \mid \boldsymbol{\phi} \sim$ Multinomial$(N, \boldsymbol{\phi})$ and $\mathbf{z}_b \mid \mathbf{w}, \boldsymbol{p}_b \sim$ Multinomial$(N, \mathbf{p}_b)$.

Now the joint density can be written as

$$p(\mathbf{y} \mid \boldsymbol{\mu}, \boldsymbol{\sigma^2}, \mathbf{w}, \{\mathbf{z}_b\}_{b=1}^{B}) = \prod_{i=1}^{N} \prod_{b=1}^{B} \prod_{k=1}^{K} p(y_i \mid \mu_k, \sigma^2_k)^{w_{ib} z_{ibk}} p(\mathbf{z}_b \mid \mathbf{w}, \boldsymbol{p}_b) p(\mathbf{w} \mid \boldsymbol{\phi})$$

## A Finite Mixture of Infinite Gaussian Mixtures

Extending this to a finite mixture of infinite Gaussian mixtures model using the Stick-breaking representation of a Dirichlet Process.

$$\begin{aligned} p(\mathbf{y} \mid \boldsymbol{\mu}, \boldsymbol{\sigma^2}, \mathbf{w}, \{\mathbf{z}_b\}_{b=1}^{B}) &= \prod_{i=1}^{N} \prod_{b=1}^{B} \prod_{k=1}^{\infty} p(y_i \mid \mu_k, \sigma^2_k)^{w_{ib} z_{ibk}} p(\mathbf{z}_b \mid \mathbf{w}, \boldsymbol{p}_b) p(\mathbf{w} \mid \boldsymbol{\phi}) \\ p(\mathbf{z}_b \mid \mathbf{w}, \boldsymbol{p}_b) &= \text{Multinomial}(N, \mathbf{p}_b) \quad \text{where} \quad \mathbf{p}_b = (p_{b1}, p_{b2}, \ldots) \\ p_{bk} &= V_{bk} \prod_{k=1}^{k-1} (1 - V_{bk}) \quad \text{where} \quad V_{bk} \sim \text{Beta}(1, \beta_b) \quad b = 1, 2, \ldots, B \\ p(\mathbf{w} \mid \boldsymbol{\phi}) &= \text{Multinomial}(N, \boldsymbol{\phi}) \\ p(\boldsymbol{\phi}) &= \text{Dirichlet}(\alpha_\phi \mathbf{1}) \\ p(\mu_k) &= N(\mu_0, \sigma^2_0) \\ p(\sigma^2_k) &= \text{inverse-gamma}(A, B) \end{aligned}$$

**Posterior**

$$p(\boldsymbol{\mu}, \boldsymbol{\sigma^2}, \mathbf{w}, \{\mathbf{z}_b\}_{b=1}^{B} \mid \mathbf{y}) \propto \prod_{i=1}^{N}\prod_{b=1}^{B}\prod_{k=1}^{\infty} p(y_i \mid \mu_k, \sigma_k^2)^{w_{ib} z_{ibk}} p(z_{ibk} \mid \mathbf{w}, \boldsymbol{p}_b) p(w_{ib} \mid \boldsymbol{\phi}) p(p_{bk}) p(\phi_b) p(\mu_k) p(\sigma_k^2)$$

**Full Conditionals**

$$p(w_{ib} = 1 \mid \cdot) \propto \phi_b \prod_{k=1}^{\infty} p_{bk} N(y_i \mid \mu_k, \sigma_k^2) \quad \text{for} \quad b = 1, 2, \ldots, B, \quad i = 1, 2, \ldots, N$$

$$p(z_{ibk} = 1 \mid \cdot) \propto p_{bk} N(y_i \mid \mu_k, \sigma_k^2) \prod_{b=1}^{B} \phi_b \quad \text{for} \quad k = 1, 2, \ldots, \infty, \quad i = 1, 2, \ldots, N$$

$$p(V_{bk} \mid \cdot) = \text{Beta}\Big(1 + n_{bk}, \beta + \sum_{j=k+1}^{\infty} n_{bj}\Big) \quad \text{where} \quad n_{bk} = \sum_{i=1}^{N} z_{ibk} w_{ib} \quad k = 1, 2, \ldots$$

$$p(\boldsymbol{\phi} \mid \cdot) = \text{Dirichlet}(\boldsymbol{\alpha}) \quad \alpha_j = \alpha_{\phi_j} + \sum_{i=1}^{N} y_i w_{ib}$$

$$p(\mu_k \mid \cdot) = N(a^{-1}b, a^{-1}) \quad \text{where} \quad a = \frac{1}{\sigma_0^2} + \frac{\sum_{b=1}^{B} n_{bk}}{\sigma_k^2} \quad \text{and} \quad b = \frac{\mu_0}{\sigma_0^2} + \frac{\sum_{i=1}^{N}\sum_{b=1}^{B} y_i z_{ibk} w_{ib}}{\sigma_k^2} \quad k = 1, 2, \ldots$$

$$p(\sigma_k^2 \mid \cdot) = \text{inverse-gamma}\Big(A + \frac{\sum_{b=1}^{B} n_{bk}}{2}, B + \frac{\sum_{i=1}^{N}\sum_{b=1}^{B}(y_i - \mu_k)^2 z_{ibk} w_{ib}}{2}\Big) \quad k = 1, 2, \ldots$$

**Optimal Densities**

The optimal densities are given by

$$q_i^*(\theta_i) \propto \exp\{E_{-\theta_i} \log p(\theta_i \mid \cdot)\}.$$

**Optimal density for $w_{ib}$**

$$\begin{aligned}
\log q_w^*(w_{ib}) &\propto E_{\boldsymbol{\phi}, \mu_k, \sigma_k^2} \log p(w_{ik} = 1 \mid \cdot) \\
&= E_{\boldsymbol{\phi}}[\log \phi_b] + \sum_{k=1}^{\infty} \big[E_{\mathbf{p}}[\log p_{bk}] - \frac{1}{2}\log(2\pi) - \frac{1}{2} E_{\sigma_k^2}[\log \sigma_k^2] - \frac{1}{2E_{\sigma_k^2}[\sigma_k^2]} E_{\mu_k}[(y_i - \mu_k)^2]\big] \\
&= \log \tilde{\phi}_{ib_{q(\phi)}}
\end{aligned}$$

It follows that

$$q_w^*(w_i) = \text{categorical}(\boldsymbol{\phi}_{q(\phi)}) \quad \text{where} \quad \phi_{ib_{q(\phi)}} = \frac{\tilde{\phi}_{ib_{q(\phi)}}}{\sum_{b=1}^{B} \tilde{\phi}_{ib_{q(\phi)}}}$$

**Optimal density for $z_{ibk}$**

$$\begin{aligned}
\log q_z^*(z_{ibk}) &\propto E_{\boldsymbol{p}, \mu_k, \sigma_k^2} \log p(z_{ibk} = 1 \mid \cdot) \\
&= E_{\mathbf{p}}[\log p_{bk}] - \frac{1}{2}\log(2\pi) - \frac{1}{2} E_{\sigma_k^2}[\log \sigma_k^2] - \frac{1}{2E_{\sigma_k^2}[\sigma_k^2]} E_{\mu_k}[(y_i - \mu_k)^2] + \sum_{b=1}^{B} E_{\boldsymbol{\phi}} \log(\phi_b) \\
&= \log \tilde{p}_{ibk_{q(p)}}
\end{aligned}$$

It follows that

$$q_z^*(z_{ib}) = \text{categorical}(\boldsymbol{p}_{q(p)}) \quad \text{where} \quad p_{ibk_{q(p)}} = \frac{\tilde{p}_{ibk_{q(p)}}}{\sum_{k=1}^{\infty} \tilde{p}_{ibk_{q(p)}}}$$

**Optimal density for $V_{bk}$**

$$q_V^*(V_{bk}) = \text{Beta}\Big(\alpha_{q(V_{bk})}, \beta_{q(V_{bk})}\Big) \quad \text{where} \quad \alpha_{q(V_{bk})} = 1 + E_{w,z}[n_{bk}], \quad \beta_{q(V_{bk})} = \beta + \sum_{j=k+1}^{\infty} E_{w,z}[n_{bj}]$$

**Optimal density for $\phi$**

$$q_\phi^*(\phi) = \text{Dirichlet}(\boldsymbol{\alpha}_{q(\phi)}) \quad \text{where} \quad \alpha_{q(\phi_j)} = \alpha_{\phi_j} + \sum_{i=1}^{N} y_i E_w[w_{ib}]$$

**Optimal density for $\mu_k$**

$$\begin{aligned}
q_{\mu_k}^*(\mu_k) &= N(\mu_{q(\mu_k)}, \sigma^2_{q(\mu_k)}) \\
\sigma^2_{q(\mu_k)} &= \Big(\frac{1}{\sigma_0^2} + \frac{E_{w,z}[n_{bk}]}{E_{\sigma_k^2}[\sigma_k^2]}\Big)^{-1} \\
\mu_{q(\mu_k)} &= \sigma^2_{q(\mu_k)} * \Big(\frac{\mu_0}{\sigma_0^2} + \frac{\sum_{i=1}^{N}\sum_{b=1}^{B} y_i E_z[z_{ibk}] E_w[w_{ib}]}{E_{\sigma_k^2}[\sigma_k^2]}\Big)
\end{aligned}$$

**Optimal density for $\sigma_k^2$**

$$\begin{aligned}
q_{\sigma_k^2}^*(\sigma_k^2) &= \text{inverse-gamma}\Big(A_{q(\sigma_k^2)}, B_{q(\sigma_k^2)}\Big) \\
A_{q(\sigma_k^2)} &= A + \frac{E_{w,z}[n_{bk}]}{2}, \\
B_{q(\sigma_k^2)} &= B + \frac{\sum_{i=1}^{N}\sum_{b=1}^{B} E_{w,z,\mu_k}[(y_i - \mu_k)^2 z_{ibk} w_{ib}]}{2}
\end{aligned}$$

**Expectations**

$$\begin{aligned}
E_\phi[\log \phi_b] &= \psi(\alpha_{q(\phi_b)}) + \psi\Big(\sum_{b=1}^{B} \alpha_{q(\phi_b)}\Big) \\
E_p[\log p_{bk}] &= E_V[\log V_{bk}] + \sum_{j=1}^{k-1} E_V[\log(1 - V_{bj})] \\
&= \psi(\alpha_{q(V_{bk})}) - \psi(\alpha_{q(V_{bk})} + \beta_{q(V_{bk})}) + \sum_{j=1}^{k-1} [\psi(\beta_{q(V_{bj})}) - \psi(\alpha_{q(V_{bj})} + \beta_{q(V_{bj})})] \\
E_{\sigma_k^2}[\log \sigma_k^2] &= \log(B_{q(\sigma_k^2)}) - \psi(A_{q(\sigma_k^2)}) \\
E_{\sigma_k^2}[\sigma_k^2] &= \frac{B_{q(\sigma_k^2)}}{A_{q(\sigma_k^2)}} \\
E_{\mu_k}[\mu_k] &= \mu_{q(\mu_k)} \\
E_{\mu_k}[(y_i - \mu_k)^2] &= (y_i - \mu_{q(\mu_k)})^2 + \sigma^2_{q(\mu_k)}
\end{aligned}$$