Gibbs Sampling algorithm for GMM with unknown Rotation

Michael Habeck [1,2], **Nima Vakili**[2]

1- Statistical Inverse Problems in Biophysics, Max Planck Institute for Biophysical Chemistry, Am Fassberg 11, 37077 Göttingen, Germany
2- Felix Bernstein Institute for Mathematical Statistics in the Biosciences, Georg August University Göttingen, Germany

Computing posterior by using conditional and prior states

The full posterior for Gibbs sampler is:

$$p(\theta, Z|D) \propto p(D, Z|\theta)p(\theta)$$

where the full likelihood would be:

$$p(D, Z|\theta) = p(x, z|\mu, \tau, \omega, R)$$

and

$$p(\theta) \propto p(\mu|\tau)p(\tau)p(\omega)p(R)$$

First we have to compute the whole priors.
Prior for the means is:

$$\begin{aligned} p(\mu|\tau) \propto \prod_{k=1}^{K} p(\mu_k|\tau) &= \prod_{k=1}^{K} N(\mu_k|\mu_0, \tau\tau_0 I) \\ &= \prod_{k=1}^{K} \frac{\tau\tau_0}{2\pi}^{\frac{D+1}{2}} \exp\left(\frac{-\tau\tau_0(\mu_k - \mu_0)^2}{2}\right) \end{aligned}$$

and prior for precision is:

$$p(\tau|\alpha_0, \beta_0) = \frac{\beta_0^{\alpha_0}}{\Gamma(\alpha_0)} \alpha^{\alpha_0 - 1} \exp(-\beta_0\tau)$$

prior for latent variables is:

$$p(z|\omega) = \prod_{k}^{K} \omega_k^{-z}$$

prior for weights is:

$$p(\omega|\omega_0) = \prod_{k=1}^{K} \omega_k^{\omega_0 - 1}$$

and finally we set the prior for rotations to constant parameter $p(R) = 1$.

Now we have to compute the conditional posteriors for each of the components. For the latent variable we have:

$$\begin{aligned}
p(z|x,\mu,\tau,\omega,R) &\propto p(x|z,\mu,\tau,R)p(z|\omega) \\
&\propto \prod_{ink} N(x_{ik}|P(R_i\mu_k),\tau^{-1}I)^{\beta z_{ink}} \prod_{ink} \omega^{z_{ink}} \\
&= \prod_{ink} [\omega_k N(x_{ik}|P(R_i\mu),\tau^{-1}I]^{\beta z_{ink}} = \prod_{in} p(z_{in}|x_{ik},\mu,\tau,\omega,R)
\end{aligned}$$

where

$$\begin{aligned}
p(z_{in}|x_{in},\mu,\tau,\omega,R) &\propto \prod_k [\omega_k N(x|P(R\mu_k),\tau^{-1}I)]^{\beta} \\
&= \frac{\prod_k [\omega_k N(x|P(R\mu_k),\tau^{-1}I)]^{\beta}}{\sum_k [\omega_k N(x_i|P(R\mu_k),\tau^{-1}I)]^{\beta}}
\end{aligned}$$

The next parameter is weights:

$$\begin{aligned}
p(\omega|x,z,\mu,\tau,R) &\propto p(x,z|\omega)p(\omega) \\
&\propto \prod_{ik} \omega_k^{z_{ik}} \cdot \prod_k w_k^{\alpha_0 - 1}
\end{aligned}$$

and for the means we have:

$$\begin{aligned}
p(\mu|\omega,z,\tau,R) &\propto p(x,z|\mu,\tau,R)p(\mu|\tau) \\
&\propto \prod_{ink} N(x_{in}|R_i\mu_k,\tau^{-1}I)^{\beta z_{ink}} \cdot \prod_k N(\mu_k|\mu_0,\tau_0^{-1}\tau^{-1}I) \\
&= \prod_k p(\mu_k|x,z,\tau,R)
\end{aligned}$$

where

$$p(\mu_k|x,z,\tau,R) \propto \prod_{ink} N(x_{in}|R_i\mu_k,\tau^{-1}I)^{\beta} \cdot N(\mu_k|\mu_0,\tau_0^{-1}\tau^{-1}I)$$

likelihood for μ :

$$\begin{aligned}
\mu_k &\propto \exp(-\frac{\beta\tau}{2}\sum_{in} z_{ink}\|X_{in} - PR_i\mu_k\|^2) \\
&= \exp(const + \sum_{in} -\frac{\beta\tau}{2} z_{ink}\|PR_i\mu_k\|^2 + \sum_{in} \beta\tau z_{ink} X_{in}^T PR_i\mu_k) \\
&= \exp(\mu_k^T[\sum_{in} -\frac{\beta\tau}{2} z_{ink} R_i^T P^T PR_i]\mu_k + \mu_k^T \sum_{in} \beta\tau z_{ink} R_i^T P^T X_{in})
\end{aligned}$$

we set:

$$A_k = \beta \sum_i N_{ik}(PR_i)^T(PR_i), \qquad N_{ik} = \sum_n z_{ink}, \qquad b_k = \beta \sum_{in} z_{ink} R_i^T P^T X_{in}$$

and the posterior is:

$$\begin{aligned}
\mu_k &\propto \exp(-\frac{\tau}{2}[\mu_k^T A_k \mu_k - 2\mu_k^T b_k] - \frac{\tau\tau_0}{2}\|\mu_k - \mu_0\|^2) \\
\tilde{A_k} &= A_k + \tau_0 I \\
\tilde{b_k} &= b_k + \tau_0 \mu_0 \\
\mu_k &\propto \exp(-\frac{\tau}{2}[\mu_k^T \tilde{A}_k \mu_k - 2\mu_k^T \tilde{b_k}]) \\
\Sigma &= \tilde{A}_k^{-1}, \qquad \mu = \Sigma \tilde{b_k}
\end{aligned}$$

and for computing precision we have:

$$\begin{aligned}
p(\tau|\mu, \omega, Z, R, D) &\propto p(D, Z|\mu, \omega, R)p(\mu|\tau)p(\tau) \\
&\propto \tau^{\frac{Nd}{2}} \exp(\frac{-\beta\tau}{2} \sum_{ink} z_{ink}\|x_{in} - PR_i\mu_k\|^2) \\
&\cdot \quad \tau^{\frac{k(d+1)}{2}} \exp(-\frac{\tau\tau_0}{2} \sum_k \|\mu_k - \mu_0\|^2) \\
&\cdot \quad \tau^{\alpha_0 - 1} \exp(-\beta_0 \tau)
\end{aligned}$$

by taking the logarithm on both sides:

$$\begin{aligned}
\log p(\tau|\mu, \omega, z, R, D) &= \sum_{i=1}^N \sum_{k=1}^K \beta z_{ik} \log\left[\frac{\tau}{2\pi}^{\frac{d}{2}} \|x_i - PR\mu_k\|^2\right] + (\alpha_0 + 1)\log\tau - \beta_0\tau \\
&+ \sum_{k=1}^K \log \tau^{\frac{d+1}{2}} - \sum_{k=1}^K \frac{\tau\tau_0}{2}\|\mu_k\|^2 \\
&= \frac{\beta dN}{2}\log\tau - \frac{\tau}{2}\sum_{n=1}^N \sum_{k=1}^K \beta z_{ik}\|x_i - PR_i\mu_k\|^2 + (\alpha_0 - 1)\log\tau - \beta_0\tau \\
&+ \frac{dK}{2}\log\tau - \tau\tau_0 \sum_{k=1}^K \|\mu_k\|^2 \\
&= \left[\frac{\beta dN + (d+1)K}{2} + \alpha_0 - 1\right]\log\tau - \left[\frac{1}{2}\sum_{n=1}^N \sum_{k=1}^K \beta z_{ik}\|x_i - PR\mu_k\|^2 + \beta_0 - \tau_0 \sum_{k=1}^K \|\right. \\
&= (\tilde{a} - 1)\log\tau - \tilde{b}\tau
\end{aligned}$$

which implies that:

$$p(\tau|\mu, \omega, z, R, D) = Gamma(\tau|\tilde{a}, \tilde{b})$$

where

$$2\tilde{a} = 2\alpha_0 + \beta dN + (d+1)K$$

$$2\tilde{b} = 2\beta_0 + \sum_{i=1}^{N}\sum_{k=1}^{K} \beta z_{ik}\|x_i - PR\mu_k\|^2 + \tau\sum_{k=1}^{K}|\mu_k\|^2$$

and finally the rotations:

$$\begin{aligned} p(R|x, z, \mu, \tau, \omega) &\propto p(x, z|\mu, \tau, \omega, R)p(R) \\ &\propto \prod_{ink}[\omega_k N(x_{in}|R_i, \mu_k, \tau^{-1}I)]^{\beta z_{ink}} \\ &= \prod_i p(R_i|x, z, \mu, \tau, \omega) \end{aligned}$$

where

$$\begin{aligned} p(R_i|x, z, \mu, \tau, \omega) &\propto \exp(-\frac{\tau}{2}\sum_{nk}\beta z_{ink}\|x_{in} - PR_i\mu_k\|^2) \\ &\propto \exp(\tau tr([\sum_{nk}\beta z_{ink}\mu_k x_{in}^T P]R_i)) - \frac{\tau}{2}tr(\sum_{n,k}\beta z_{ink}\mu_k\mu_k^T R_i^T P^T PR_i) \\ A_i &= \tau\sum_{nk}\beta z_{ink}\mu_k x_{in}^T P \\ C &= P^T P \\ B_i &= -\frac{\tau}{2}\sum_{nk}\beta z_{ink}\mu_k\mu_k^T = -\frac{\tau}{2}\sum_k \beta N_{ik}\mu_k\mu_k^T \end{aligned}$$

where

$$N_{ik} = \sum_n z_{ink}$$

you can find the numerical computational process for the rotations in appendix.
General form of the likelihood for GMM:

$$p(D|\mu, \tau, \omega, R) = \prod_{n=1}^{N}\sum_{k=1}^{k}\omega_k N(x_n|PR\mu_k, \tau)$$

The Augmented likelihood is:

$$\begin{aligned} p(D|\mu, \tau, z, R) &= \prod_n\prod_k[N(X_n|\mu_k, \tau)]^{\beta z_{nk}} \\ p(z|\omega) &= \prod_n M(z_n|1, \omega) = \prod_n\prod_k \omega^{z_{nk}}\delta(1 - \sum_k z_{nk}) \end{aligned}$$

[if we set, $\omega_k = \frac{1}{K}$, we will have:

$$\begin{aligned}
\log p(z|\omega) &= \sum_n \sum_k z_{nk} \log w_k \\
&= \log \omega \sum_n \sum_k z_{nk} \\
&= N \log \omega = -N \log K]
\end{aligned}$$

$$\begin{aligned}
M(n|N,P) &= \left(\frac{N!}{n_{1!},\dots,n_{k!}} \prod_k P_k^{n_k} \delta(N - \sum_k n_k)\right) \\
p(D|\mu,\tau,\omega,R) &= \sum_z p(D|\mu,\tau,z,R)p(z|w) \\
p(D,z|\mu,\tau,\omega,R) &= p(D|\mu,\tau,z,R)p(z|\omega)
\end{aligned}$$

The likelihood and the priors in total:

$$[p(D|\mu,\tau,z,R)p(z|\omega)] \left[p(\omega|\omega_0) \prod_k p(\mu_k|\mu_0,\tau_0,\tau) p(\tau|\alpha,\beta) p(R) \right]$$

1 Appendix

Computation of the Rotations :

$$tr(BR^TCR) + tr(AR)$$

where

$$\begin{aligned}
B^T &= B, \qquad -B > 0, \qquad X^T(-B)X \geq 0 \\
-B &= U\Lambda U^T \\
\Lambda &= \begin{bmatrix} \lambda_1 & 0 & 0 \\ 0 & \lambda_2 & 0 \\ 0 & 0 & \lambda_3 \end{bmatrix}
\end{aligned}$$

and

$$\lambda_i \geq 0, \qquad U^TU = UU^T = I_{3\times 3}$$

so we have:

$$\begin{aligned}
-tr(BR^TCR) + tr(AR) &= tr(U\Lambda U^T R^T CR) + tr(ARUU^T) \\
&= tr(\Lambda (RU)^T C(RU)) + tr((U^TA)(RU)) \\
&= tr(\Lambda \tilde{R}^T C\tilde{R}) + tr(\tilde{A}\tilde{R})
\end{aligned}$$

where

$$\tilde{R} = RU, \qquad \tilde{A} = U^T A$$

and

$$\tilde{R}^T \tilde{R} = I$$

and

$$C = P^T P = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix} = I - e_3 e_3^T$$

$$\begin{aligned} tr(\Lambda \tilde{R}^T (I - e_3 e_3^T) \tilde{R}) &= tr(\Lambda) - tr(\Lambda \tilde{R}^T e_3 e_3^T \tilde{R}) \\ &= tr(\Lambda) - e_3^T \tilde{R} \lambda \tilde{R}^T e_3 \\ &= tr(\Lambda) - \tilde{r}_3^T \Lambda \tilde{r}_3 \end{aligned}$$

where

$$\begin{aligned} \tilde{r}_i &= \tilde{R}^T e_i, \qquad \tilde{r}_i \in R^3 \\ tr(AR) &= tr(\tilde{A}\tilde{R}) \end{aligned}$$

where

$$\begin{aligned} \tilde{A} &= U^T A, \qquad A = D_{3\times 2} P_{2\times 3} \\ D &= \tau \sum_{nk} z_{ink} \mu_k X_{in}^T \\ \tilde{A} &= U^T D P, \qquad \tilde{D} = U^T D \\ \tilde{A} &= \tilde{D} P \end{aligned}$$

where

$$\begin{aligned} P &\in R^{2\times 3}, \qquad D \in R^{3\times 2} \\ tr(AR) &= tr(\tilde{D} P \tilde{R}) \\ P\tilde{R} &= \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix} \begin{pmatrix} \tilde{r_1}^T \\ \tilde{r_2}^T \\ \tilde{r_3}^T \end{pmatrix} = \begin{pmatrix} \tilde{r_1}^T \\ \tilde{r_2}^T \end{pmatrix} \in R^{2\times 3} \\ \tilde{D}^T &= \begin{pmatrix} \tilde{d_1}^T \\ \tilde{d_2}^T \end{pmatrix} \\ tr(AR) &= tr(\tilde{D} \begin{pmatrix} \tilde{r_1}^T \\ \tilde{r_2}^T \end{pmatrix}) = \tilde{d_1}^T \tilde{r_1}^T + \tilde{d_r}^T \tilde{r_2}^T \end{aligned}$$

so finally we have:

$$
\begin{aligned}
-tr(BR^TCR) \quad &+ \quad tr(AR) \\
&= \quad tr\Lambda - \tilde{r_3}^T \Lambda \tilde{r_3} + \tilde{d_1}^T \tilde{r_1}^T + \tilde{d_2}^T \tilde{r_2}^T
\end{aligned}
$$

subject to,

$$\tilde{r_i} \perp \tilde{r_j}$$