**General Posterior/MCMC Structure in Mixtures**

Observations $y_1, \ldots, y_n$ are i.i.d. from the model $\sum_{j=1}^{k} \pi_j f(\cdot|\theta_j)$. Here each $y_i$ is a $k-$vector. In the normal mixture model, for example, $f()$ is multivariate normal with mean and variance matrix parameters $\theta_j = \{\mu_j, \Sigma_j\}$ in component $j$. Classification indicators $z = \{z_1, \ldots, z_n\}$ where $z_i = j$ if observation $y_i$ comes from component $j$.

The structure of conditional posteriors for MCMC analysis is as follows. Choose values for classifiers $z$. Then run through the sequence of conditional posteriors to resample parameters $\theta_j$, probabilities $\pi$ and then new $z$ in sequence.

(A) Given $z$, data are classified as $k$ independent random samples, with $n_j$ in group $j$. Compute and simulate the posteriors for $\theta_j$ in each group. In the multivariate normal case with Jeffrey's prior on $\{\mu_j, \Sigma_j\}$ in each group, the calculations are standard. Write $\bar{y}_j$ for the mean of the data classified in component $j$, and $S_j$ for the "sum of squares matrix" in that component. Sample $\{\mu_j, \Sigma_j\}$ by composition (as in Assignment 4):
  - first $\Sigma_j$ from $IW_{n_j-1}(\Sigma_j|S_j^{-1})$ then
  - $\mu_j$ from $N(\mu_j|\bar{y}_j, \Sigma_j/n_j)$.

(B) Given $z$, data are classified as $k$ independent random samples, with $n_j$ in group $j$. Compute and simulate the posterior for $\pi$. With a Dirichlet prior the relevant posterior is Dirichlet too. Take the uniform prior so that the posterior given $z$ is $Dir(\pi|m)$ where $m = (1+n_1, \ldots, 1+n_k)$. Sample using the construction of the Dirichlet through independent gamma variates, a trivial computation.

(C) Now resample $z$ to reclassify the data, before going around again. Given data $y$ and *all* parameters $\phi = \{\pi, \theta_1, \ldots, \theta_k\}$, the $z_i$ are independent with conditional posteriors

$$\pi_{ij}^* \equiv Pr(z_i = j|y, \phi) \propto \pi_j \ f(y_i|\theta_j);$$

i.e., prior times likelihood. Compute and normalise to get the set of *posterior classification probabilities* $(\pi_{i1}^*, \ldots, \pi_{ik}^*)$ for $z_i$. Run through this for each observation $i = 1, 2, \ldots, n$, sampling each $z_i$ from $1, 2, \ldots, k$ using these probabilities (e.g., using the S-Plus function `sample()`). This reclassifies the data. In the normal mixture case,

$$\pi_{ij}^* = q_{ij} / \sum_{s=1}^{k} q_{is}$$

where

$$q_{ij} = \pi_j |\Sigma_j|^{-1/2} \exp(-(y_i - \mu_j)'\Sigma^{-1}(y_i - \mu_j)/2).$$

- Run through (A)-(C) iteratively, as usual in MCMC.

*Special Twist:* If some data are classified for sure (i.e., training data), then simply fix the corresponding $z_i$ values throughout – at part (C), just skip the resampling for these values.

**Classification**

To infer classification of the *observed data*, note that $\pi_{ij}^*$ is the *posterior probability that $y_i$ comes from component $j$*. These are uncertain quantities, but in the process of MCMC analysis above, we have generated a *sample from the posterior distribution for the* $\pi_{ij}^*$. The MC sample mean of $\pi_{ij}^*$ is, for example, an estimate of the true posterior mean, and might be used to classify the observed data.

**Predictive distributions**

Now predict a future observation $y_{n+1}$. To sample the predictive density $p(y_{n+1}|y)$ just once, take one of the simulations from the posterior – i.e., a sampled set $\{z, \pi, \theta_1, \ldots, \theta_k\}$. Given these values,

- sample a classifier $z_{n+1}$ from $1, 2, \ldots, k$ according to probabilities $\pi$, then
- sample $y_{n+1}$ from $f(\cdot|\theta_j)$.

Repeat this for each of the MCMC samples of $\{z, \pi, \theta_1, \ldots, \theta_k\}$ to build a sample from the predictive distribution.

**Predictive Classification**

Suppose a future value $y_{n+1}$ were observed. Which component did it come from? Answer this by computing the chances that $z_{n+1} = j$ given any specified value of $y_{n+1}$. Vary the value of $y_{n+1}$ here to see these *predictive classification probabilities* change according to where $y_{n+1}$ lies. Note that

$$Pr(z_{n+1} = j|y_{n+1}, \phi) = \pi_{n+1,j}^*$$

which are computed as above. So each MCMC sample of values of parameters $\phi$ generates a value of $\pi_{n+1,j}^*$, the set of which is a corresponding posterior sample. Again, taking the mean of this sample generates an approximate value of the posterior mean of $(\pi_{n+1,1}^*, \ldots, \pi_{n+1,k}^*)$. These can be used graphically in different ways. In the 2−dimensional normal mixture model, we can identify the regions of values of $y_{n+1}$ in which component 1 has highest classification probability, or component 2, or whatever; call this the *modal component plot*. It tells us which is the most likely component generating that observed value, and gives a nice classification summary.