ECM algorithm for two clusters, with constraints on the cluster means and variances, and known data variances.

Roger Day

May 26, 2013

We assume that $Y_k \sim N(\psi_{G(k)}, \tau_{G(k)k})$ where $\tau_{G(k)k} = {\sigma^2}_k + V_{G(k)}$, and $\Pr[G(k) = g] = \pi_g$ for $g = 0, 1$.

The group membership vector G is regarded as missing data for purposes of the "expectation/conditional maximization" algorithm (ECM).

With a normal approximation, the complete data likelihood per observation k is:

$$\Pr[Y_k, G(k)|\psi, V, \pi] = \Pr[Y_k|G(k)] \cdot \Pr[G(k)]$$
$$\propto \exp(-(Y_k - \psi_{G(k)})^2/2\tau_{G(k)k}) \cdot (\tau_{G(k)k})^{-1/2} \cdot \pi_{G(k)}$$

Now suppose that the variances σ_k^2 are known (approximately). Define the free variable $\phi = (\psi_0, \psi_1, \pi_1, V_0, V_1)$, and $\pi_0 = 1 - \pi_1$, and similarly for the current fixed estimate ϕ^*. Following standard calculations, the expectation of the complete-data log likelihood is

$$Q(\phi, \phi*) = E^* \left(- \sum\nolimits_k \left((Y_k - \psi_{G(k)})^2/2\tau_{G(k)k}(V) - 1/2 \log \tau_{G(k)k}(V)\right) + N_0 \log \pi_0 + N_1 \log \pi_1\right)$$
$$= {}^1\!/\!_2 \sum\nolimits_k \sum\nolimits_g \pi^*_{gk} \left(-(Y_k - \psi_g)^2/(\sigma_k^2 + V_g) - \log(\sigma_k^2 + V_g)\right) + {N^*}_0 \log \pi_0 + {N^*}_1 \log \pi_1$$

with

$$\frac{{\pi^*}_{0k}}{{\pi^*}_{1k}} = \frac{\Pr(Y_k, G(k) = 0|\psi_0^*, V_0^*, \pi_0^*)}{\Pr(Y_k, G(k) = 1|\psi_1^*, V_1^*, \pi_1^*)} = \frac{{\pi^*}_0}{{\pi^*}_1} \frac{\exp\left(-(Y_k - \psi_0^*)^2/2(V_0^* + \sigma_k^2)\right)/\sqrt{V_0^* + \sigma_k^2}}{\exp\left(-(Y_k - \psi_1^*)^2/2(V_1^* + \sigma_k^2)\right)/\sqrt{V_1^* + \sigma_k^2}}$$

so that $E^* N_g = \sum\limits_k \pi^*_{gk}$ for g=0,1.

To maximize Q, we set its partial derivatives to zero.

$$\partial Q/\partial \psi_g = -\sum\nolimits_k {\pi^*}_{gk} \left((Y_k - \psi_g)\left({\sigma^2}_k + \hat{V}_g\right)^{-1}\right) = 0$$

and

$$\partial Q/\partial V_g = \frac{1}{2} \sum\nolimits_k {\pi^*}_{gk} \left(+(Y_k - \psi_g)^2 \left({\sigma^2}_k + V_g\right)^{-2} - \left({\sigma^2}_k + V_g\right)^{-1}\right) = 0$$

We cannot solve these two equations simultaneously. However, fixing $V_g = \hat{V}_g$ we can solve the first:

$$\hat{\psi}_g = \frac{\sum_k \pi^*{}_{gk} Y_k \left(\sigma^2{}_k + \hat{V}_g\right)^{-1}}{\sum_k \pi^*{}_{gk} \left(\sigma^2{}_k + \hat{V}_g\right)^{-1}}$$

and fixing $\psi_g = \hat{\psi}_g$ we can solve the second:

$$\begin{aligned}
&\sum\nolimits_k \pi^*{}_{gk} \left((Y_k - \psi_g)^2 - \left(\sigma^2{}_k + V_g\right)\right) = 0 \\
&\sum\nolimits_k \pi^*{}_{gk} (Y_k - \psi_g)^2 = \sum\nolimits_k \pi^*{}_{gk} \left(\sigma^2{}_k + V_g\right) = \sum\nolimits_k \pi^*{}_{gk} \sigma^2{}_k + N_g^* V_g \\
&\hat{V}_g = \max\left(0, \sum\nolimits_k \pi^*{}_{gk} (Y_k - \hat{\psi}_g)^2 - \sum\nolimits_k \pi^*{}_{gk} \sigma^2{}_k\right) / N_g^*
\end{aligned}$$

So in the M step we iterate between these formulas for $\hat{\psi}_g$ and for $\hat{V}_g$. This constitutes an ECM algorithm.

The ECM context makes it easy to account for constraints. In a mixture of correlation distributions, it may be assumed that $\psi_0 = 0$. We might also want to assume $V_0 = 0$, or at least to test it to see if the variation in the g=0 component is due only to noise. These constraints are inserted into the respective formulas.