1 Solving for the Posterior Distribution on $\boldsymbol{\beta}$: Motivating the problem

$$\mathsf{P}(\boldsymbol{\beta} \mid D) = \sum_{m=1}^{M} \mathsf{P}(\boldsymbol{\beta} \mid D, Z = m) Pr(Z = m|D) \tag{1}$$

If $\boldsymbol{\beta}$ results from a mixture of multivariate normals across many different prior covariance matrices $\mathbf{W}_m$ and clusters of 'types', we need to compute the posterior on $\boldsymbol{\beta}$. We might also assume that the observed variance in effect sizes, $\hat{\mathbf{V}}$, is a diagonal matrix.

1.1 Prior specification

Taking a Bayesian approach to inference, we place the following priors (a mixture of normal distributions) on $\boldsymbol{\beta}$

$$\boldsymbol{\beta} \mid \pi \sim \sum_{m=1}^{M} \pi_m^{\beta} \mathcal{N}(\mathbf{0}, \mathbf{W_m}) \tag{2}$$

where $\pi = (\pi_1, \ldots, \pi_M)$ are the mixture proportions which are constrained to be non-negative and sum to one, $\mathbf{W} = (\mathrm{W}_1, \ldots, W_M)$ are the S by S grid of effect-size prior variance matrices for each normal distribution. For now, we assume that the variance matrices $\mathbf{W}_m$ are known and estimates π by using an empirical Bayes procedure - that is we find the maximum-likelihood estimator.

1.2 BF and posterior prob approximation

MLE for π, the prior weight on $Pr(z_i = m)$ can be computed as described using the EM algorithm.

Let $= \mathbf{Var}(\hat{\boldsymbol{\beta}})$ and $\mathbf{W}_m = \mathbf{W_m}$, the group-specific variance. We exploit the fact that

$$\begin{aligned} E(\hat{\boldsymbol{\beta}}) &= E(E(\hat{\boldsymbol{\beta}}|\boldsymbol{\beta}) = E(\boldsymbol{\beta}) = 0 && (3)\\ Var(\hat{\boldsymbol{\beta}}) &= Var(E(\hat{\boldsymbol{\beta}}|\boldsymbol{\beta}) + E(Var(\hat{\boldsymbol{\beta}}|\boldsymbol{\beta})) && (4)\\ &= \mathbf{W}_m + \mathbf{V} && (5) \end{aligned}$$

Recognizing that we can use as a sufficient statistic to summarize our data:

$$
\begin{aligned}
\mathsf{P}(D \mid Z = m) &= \mathsf{P}(\hat{\beta} \mid Z = m) &(6)\\
\hat{\beta} \mid Z = m &\sim \mathcal{N}(0, \mathbf{W}_m + \hat{\mathbf{V}}) &(7)\\
\mathsf{P}(\hat{\beta} \mid Z = m) &= (2\pi)^{-k/2} |\hat{\mathbf{V}} + \mathbf{W}_m|^{\frac{-1}{2}} \exp{[-\frac{\hat{\boldsymbol{\beta}}'(\mathbf{W}_m + \mathbf{V})^{-1}\hat{\boldsymbol{\beta}}}{2}]}. &(8)
\end{aligned}
$$

And a posterior on $\boldsymbol{\beta}$ is

$$
\begin{aligned}
\mathsf{P}(\boldsymbol{\beta} \mid D, Z = m) &\propto \mathcal{L}(\boldsymbol{\beta})\mathsf{P}(\boldsymbol{\beta} \mid Z = m) &(9)\\
&\propto \exp{[-\frac{(\boldsymbol{\beta} - \hat{\boldsymbol{\beta}})'\hat{\mathbf{V}}^{-1}(\boldsymbol{\beta} - \hat{\boldsymbol{\beta}})}{2}]} \exp{[-\frac{\boldsymbol{\beta}'\mathbf{W}_m\boldsymbol{\beta}}{2}]}. &(10)
\end{aligned}
$$

leading to

$$
\boldsymbol{\beta} \mid D, Z = \;\sim\; \mathcal{N}((\hat{\mathbf{V}}^{-1} + \mathbf{W}_m^{-1})^{-1}\hat{\mathbf{V}}^{-1}\hat{\boldsymbol{\beta}}, (\hat{\mathbf{V}}^{-1} + \mathbf{W}_m^{-1})^{-1}).
$$

$$
(12)
$$

So, we will sum over all possible posteriors $\boldsymbol{\beta} \mid D, Z = m$ and weights :

$$
\begin{aligned}
Pr(\boldsymbol{\beta} \mid D) &= \sum_{m=1}^{M} Pr(Z = m|D) Pr(\boldsymbol{\beta} \mid D, Z = m) &(13)\\
&= \sum_{m=1}^{M} \frac{\pi_m Pr(D|Z = m)}{\sum_n \pi_n Pr(D|Z = n)} Pr(\boldsymbol{\beta} \mid D, Z = m). &(14)
\end{aligned}
$$

Where the posterior weights arise from (8) and the contributions of the posterior according to each combination of prior variances $\mathbf{W}_m$ arises from the individual multivariate normals specified in (12).