EM algorithm

1 Normal-normal model, with group-specific variances as fixed-effects

1.1 The model

The model is:

1. $Z_g \sim \pi$.
2. $M_{g,j,\mathscr{T}} \sim \text{Normal}(\mu_0, \tau_0^2)$.
3. $X_{i,g,t}|\{Z_g = j, M_{g,j,\mathscr{T}} = \mu_{g,j,\mathscr{T}}\} \sim \text{Normal}(\mu_{g,j,\mathscr{T}}, \tau^2_{g,j,\mathscr{T}})$.
4. $S_{i,g} = \sum_{t=1}^{T} p_{\sigma_i,t} X_{i,g,t}$.

$S_{i,g}$ is observed. The parameters $\theta = (p_{\sigma,t}, \pi_j, \tau_{i,g,\mathscr{T}}, \tau_0, \mu_0)$.

Let

1. $S_{i,g,\mathscr{T}} = \sum_{t\in\mathscr{T}} p_{\sigma_i,t} X_{i,g,t}$.
2. $p_{\sigma,\mathscr{T}} = \sum_{t\in\mathscr{T}} p_{\sigma,t}$.
3. $\text{var}_{\sigma,g,j} = (\sum_{\mathscr{T}\in\mathbb{T}_j} p^2_{\sigma,\mathscr{T}})\tau_0^2 + \sum_{\mathscr{T}\in\mathbb{T}_j}(\sum_{t\in\mathscr{T}} p^2_{\sigma,t})\tau^2_{g,j,\mathscr{T}}$.

Note that:

1. $S_{i,g,\mathscr{T}}|\{Z_g = j, M_{g,j,\mathscr{T}} = \mu_{g,j,\mathscr{T}}\} \sim \text{Normal}(p_{\sigma_i,\mathscr{T}}\mu_{g,j,\mathscr{T}}, (\sum_{t\in\mathscr{T}} p^2_{\sigma_i,t})\tau^2_{g,j,\mathscr{T}})$ due to Fact NLC.
2. $p_{\sigma_i,\mathscr{T}} M_{g,j,\mathscr{T}}|\{Z_g = j\} \sim \text{Normal}(p_{\sigma_i,\mathscr{T}}\mu_0, p^2_{\sigma_i,\mathscr{T}}\tau_0^2)$, due to Fact NSC.
3. $S_{i,g,\mathscr{T}}|\{Z_g = j\} \sim \text{Normal}(p_{\sigma_i,\mathscr{T}}\mu_0, p^2_{\sigma_i,\mathscr{T}}\tau_0^2 + (\sum_{t\in\mathscr{T}} p^2_{\sigma_i,t})\tau^2_{g,j,\mathscr{T}})$, due to fact NCP and the above.
4. $S_{i,g}|\{Z_g = j\} \sim \text{Normal}(\mu_0, \text{var}_{\sigma_i,g,j})$, due to Fact NLC and the above.

Thus we can rewrite the model as:

1. $Z_g \sim \pi$.
2. $S_{i,g}|\{Z_g = j\} \sim \text{Normal}(\mu_0, \text{var}_{\sigma_i,g,j})$.

1.2 Likelihood

The likelihood is

$$\begin{aligned} L(\theta) = f_\theta(s) &= E[f_\theta(s|Z)] \\ &= \prod_{g=1}^{G}\sum_{j=1}^{J} P(Z_g = j) f(s_g|Z_g = j) \\ &= \prod_{g=1}^{G}\sum_{j=1}^{J} \pi_j \prod_{i=1}^{n} f_{\text{Normal}(\mu_0,\text{var}_{\sigma_i,g,j})}(s_{i,g}) \end{aligned}$$

The loglikelihood is

$$\ell(\theta) = \log f_\theta(s) = \sum_{g=1}^{G} \text{logsumexp}_{j=1}^{J}\Big(\log(\pi_j) + \sum_{i=1}^{n} \log f_{\text{Normal}(\mu_0,\text{var}_{\sigma_i,g,j})}(s_{i,g})\Big)$$

insert TMP1

1.3 EM algorithm

Note

$$\begin{aligned} J(\theta|\tilde{\theta}) &:= \sum_z f_{\tilde{\theta}}(z|s) \\ &= \sum_{g=1}^{G}\sum_{j=1}^{J} f_{\tilde{\theta}}(j|s_g)\log f_\theta(j,s_g) \quad \text{(EMN 40)} \end{aligned}$$

where

1. $f(j|s_g) = \frac{f(j)f(s_g|j)}{\sum_{j'=1}^{J} f(j')f(s_g|j')}$ (EMN 39, 40).
2. $f(s_g|j) = \prod_{i=1}^{n} f(s_{i,g}|j) = \prod_{i=1}^{n} f_{\text{Normal}(\mu_0,\text{var}_{\sigma_i,g,j})}(s_{i,g})$.
3. $f(j) = \pi_j$.
4. $\log f(j,s_g) = \log f(j) + \log f(s_g|j) = \log \pi_j + \sum_{i=1}^{n} \log f_{\text{Normal}(\mu_0,\text{var}_{\sigma_i,g,j})}(s_{i,g})$.

So

$$\begin{aligned} J(\theta|\tilde{\theta}) &= \sum_{g=1}^{G}\sum_{j=1}^{J} f_{\tilde{\theta}}(j|s_g)\Big[\log\pi_j + \sum_{i=1}^{n}\log f_{\text{Normal}(\mu_0,\text{var}_{\sigma_i,g,j})}(s_{i,g})\Big] \\ &= \sum_{g=1}^{G}\sum_{j=1}^{J} f_{\tilde{\theta}}(j|s_g)\Big[\log\pi_j + \sum_{i=1}^{n}\Big(-\log\sqrt{2\pi} - \log\sqrt{\text{var}_{\sigma_i,g,j}} - \frac{(s_{i,g}-\mu_0)^2}{2\text{var}_{\sigma_i,g,j}}\Big)\Big] \end{aligned}$$

where

$$f_{\tilde{\theta}}(j|s_g) = \Big[\frac{\tilde{\pi}_j \prod_{i=1}^{n} f_{\text{Normal}(\tilde{\mu}_0,\tilde{\text{var}}_{\sigma_i,g,j})}(s_{i,g})}{\sum_{j'=1}^{J}\tilde{\pi}_{j'}\prod_{i=1}^{n} f_{\text{Normal}(\tilde{\mu}_0,\tilde{\text{var}}_{\sigma_i,g,j'})}(s_{i,g})}\Big]$$

The Lagrangian

$$\Lambda(\theta) = J(\theta|\tilde{\theta}) + \lambda(1-\sum_{j=1}^{J}\pi_j) - \sum_{\sigma=1}^{\Sigma}\lambda_\sigma(1-\sum_{t=1}^{T}p_{\sigma,t})$$

The EM algorithm is to repeat:

1. Compute $f_{\tilde{\theta}}(j|s_g)$.
2. Maximize $\Lambda(\theta)$.

1.4 Maximization of $\Lambda(\theta)$

First, note that

$$\begin{aligned} \frac{\partial\Lambda}{\partial\pi_j} &= \sum_g \frac{f_{\tilde{\theta}}(j|s_g)}{\pi_j} - \lambda \\ \frac{\partial\Lambda}{\partial\lambda} &= 1 - \sum_j \pi_j \end{aligned}$$

so the maximizer is π_j such that

$$\begin{cases}\sum_g \frac{f_{\tilde{\theta}}(j|s_g)}{\pi_j} = \lambda \\ \sum_j \pi_j - 1 = 0\end{cases} \qquad \text{i.e.,} \quad \begin{cases}\sum_g \frac{f_{\tilde{\theta}}(j|s_g)}{\lambda} = \pi_j \\ \sum_j \pi_j - 1 = 0\end{cases} \qquad \text{i.e.,} \quad \pi_j = \frac{\sum_g f_{\tilde{\theta}}(j|s_g)}{\sum_{j',g} f_{\tilde{\theta}}(j'|s_g)}$$

Next, note that

$$\frac{\partial \text{var}_{\sigma,g,j}}{\partial p_{\sigma,t}} = 2p_{\sigma,\mathscr{T}(j,t)}\tau_0^2 + 2p_{\sigma,t}\tau_{g,j,\mathscr{T}(j,t)}^2$$
$$\frac{\partial \text{var}_{\sigma,g,j}}{\partial \tau_0} = 2\tau_0 \sum_{\mathscr{T}\in\mathbb{T}_j} p_{\sigma,\mathscr{T}}^2$$
$$\frac{\partial \text{var}_{\sigma,g,j}}{\partial \tau_{g,j,\mathscr{T}}} = 2\tau_{g,j,\mathscr{T}} \sum_{t\in\mathscr{T}} p_{\sigma,t}^2$$

so

$$\begin{aligned}
\frac{\partial \Lambda}{\partial p_{\sigma,t}} &= \sum_{i,g,j} f_{\tilde{\theta}}(j|s_g)\frac{(s_{i,g}-\mu_0)^2}{\text{var}_{\sigma_i,g,j}^2}\frac{\partial \text{var}_{\sigma_i,g,j}}{\partial p_{\sigma,t}} - \lambda_\sigma \\
&= \sum_{i:\sigma_i=\sigma}\sum_{g,j} f_{\tilde{\theta}}(j|s_g)\frac{(s_{i,g}-\mu_0)^2}{\text{var}_{\sigma,g,j}^2}\left[2p_{\sigma,\mathscr{T}(j,t)}\tau_0^2 + 2p_{\sigma,t}\tau_{g,j,\mathscr{T}(j,t)}^2\right] - \lambda_\sigma \\
\frac{\partial \Lambda}{\partial \tau_0} &= \sum_{i,g,j} f_{\tilde{\theta}}(j|s_g)\frac{(s_{i,g}-\mu_0)^2}{\text{var}_{\sigma_i,g,j}^2}\left[2\tau_0 \sum_{\mathscr{T}\in\mathbb{T}_j} p_{\sigma_i,\mathscr{T}}^2\right] \\
\frac{\partial \Lambda}{\partial \tau_{g,j,\mathscr{T}}} &= \sum_{i} f_{\tilde{\theta}}(j|s_g)\frac{(s_{i,g}-\mu_0)^2}{\text{var}_{\sigma_i,g,j}^2}\left[2\tau_{g,j,\mathscr{T}} \sum_{t\in\mathscr{T}} p_{\sigma_i,t}^2\right]
\end{aligned}$$

Also

$$\frac{\partial \Lambda}{\partial \mu_0} = \sum_{i,g,j} f_{\tilde{\theta}}(j|s_g)\frac{s_{i,g}-\mu_0}{\text{var}_{\sigma_i,g,j}}$$
$$\frac{\partial \Lambda}{\partial \lambda_\sigma} = 1 - \sum_{t=1}^{T} p_{\sigma,t}$$

It is not obvious how to solve the above analytically, so instead we use a numerical method.

2 Normal-normal model, with shared variance as fixed-effect

2.1 The model

The model is:

1. $Z_g \sim \pi$.
2. $M_{g,j,\mathscr{T}} \sim \text{Normal}(\mu_0, \tau_0^2)$.
3. $X_{i,g,t}|\{Z_g = j, M_{g,j,\mathscr{T}} = \mu_{g,j,\mathscr{T}}\} \sim \text{Normal}(\mu_{g,j,\mathscr{T}}, \tau^2)$.
4. $S_{i,g} = \sum_{t=1}^{T} p_{\sigma_i,t}X_{i,g,t}$.

$S_{i,g}$ is observed. The parameters $\theta = (p_{\sigma,t}, \pi_j, \tau, \tau_0, \mu_0)$.

Let

1. $S_{i,g,\mathscr{T}} = \sum_{t\in\mathscr{T}} p_{\sigma_i,t}X_{i,g,t}$.
2. $p_{\sigma,\mathscr{T}} = \sum_{t\in\mathscr{T}} p_{\sigma,t}$.
3. $\text{var}_{\sigma,j} = (\sum_{\mathscr{T}\in\mathbb{T}_j} p_{\sigma,\mathscr{T}}^2)\tau_0^2 + (\sum_{t=1}^{T} p_{\sigma,t}^2)\tau^2$.

Note that:

1. $S_{i,g,\mathscr{T}}|\{Z_g = j, M_{g,j,\mathscr{T}} = \mu_{g,j,\mathscr{T}}\} \sim \text{Normal}(p_{\sigma_i,\mathscr{T}}\mu_{g,j,\mathscr{T}}, (\sum_{t\in\mathscr{T}} p_{\sigma_i,t}^2)\tau^2)$ due to Fact NLC.
2. $p_{\sigma_i,\mathscr{T}} M_{g,j,\mathscr{T}}|\{Z_g = j\} \sim \text{Normal}(p_{\sigma_i,\mathscr{T}}\mu_0, p_{\sigma_i,\mathscr{T}}^2\tau_0^2)$, due to Fact NSC.
3. $S_{i,g,\mathscr{T}}|\{Z_g = j\} \sim \text{Normal}(p_{\sigma_i,\mathscr{T}}\mu_0, p_{\sigma_i,\mathscr{T}}^2\tau_0^2 + (\sum_{t\in\mathscr{T}} p_{\sigma_i,t}^2)\tau^2)$, due to fact NCP and the above.
4. $S_{i,g}|\{Z_g = j\} \sim \text{Normal}(\mu_0, \text{var}_{\sigma_i,j})$, due to Fact NLC and the above.

Thus we can rewrite the model as:

1. $Z_g \sim \pi$.
2. $S_{i,g}|\{Z_g = j\} \sim \text{Normal}(\mu_0, \text{var}_{\sigma_i,j})$.

2.2 Likelihood

The likelihood is

$$L(\theta) = f_\theta(s) = \prod_{g=1}^{G}\sum_{j=1}^{J}\pi_j\prod_{i=1}^{n} f_{\text{Normal}(\mu_0,\text{var}_{\sigma_i,j})}(s_{i,g})$$

and the loglikelihood is

$$\ell(\theta) = \log f_\theta(s) = \sum_{g=1}^{G}\text{logsumexp}_{j=1}^{J}\Big(\log(\pi_j) + \sum_{i=1}^{n}\log f_{\text{Normal}(\mu_0,\text{var}_{\sigma_i,j})}(s_{i,g})\Big)$$

2.3 EM algorithm

Note

$$\begin{aligned} J(\theta|\tilde{\theta}) &= \sum_{g=1}^{G}\sum_{j=1}^{J} f_{\tilde{\theta}}(j|s_g)\log f_\theta(j,s_g) \\ &= \sum_{g=1}^{G}\sum_{j=1}^{J} f_{\tilde{\theta}}(j|s_g)\Big[\log\pi_j + \sum_{i=1}^{n}\log f_{\text{Normal}(\mu_0,\text{var}_{\sigma_i,j})}(s_{i,g})\Big] \\ &= \sum_{g=1}^{G}\sum_{j=1}^{J} f_{\tilde{\theta}}(j|s_g)\Big[\log\pi_j + \sum_{i=1}^{n}\Big(-\log\sqrt{2\pi} - \log\sqrt{\text{var}_{\sigma_i,j}} - \frac{(s_{i,g}-\mu_0)^2}{2\text{var}_{\sigma_i,j}}\Big)\Big] \end{aligned}$$

where

$$f_{\tilde{\theta}}(j|s_g) = \Big[\frac{\tilde{\pi}_j\prod_{i=1}^{n} f_{\text{Normal}(\tilde{\mu}_0,\tilde{\text{var}}_{\sigma_i,j})}(s_{i,g})}{\sum_{j'=1}^{J}\tilde{\pi}_{j'}\prod_{i=1}^{n} f_{\text{Normal}(\tilde{\mu}_0,\tilde{\text{var}}_{\sigma_i,j'})}(s_{i,g})}\Big]$$

The Lagrangian

$$\Lambda(\theta) = J(\theta|\tilde{\theta}) + \lambda(1 - \sum_{j=1}^{J}\pi_j) - \sum_{\sigma=1}^{\Sigma}\lambda_\sigma(1 - \sum_{t=1}^{T} p_{\sigma,t})$$

The EM algorithm is to repeat:

1. Compute $f_{\tilde{\theta}}(j|s_g)$.
2. Maximize $\Lambda(\theta)$.

2.4 Maximization of $\Lambda(\theta)$

First, note that

$$\frac{\partial \Lambda}{\partial \pi_j} = \sum_g \frac{f_{\tilde{\theta}}(j|s_g)}{\pi_j} - \lambda$$
$$\frac{\partial \Lambda}{\partial \lambda} = 1 - \sum_j \pi_j$$

so the maximizer is π_j such that

$$\begin{cases} \sum_g \frac{f_{\tilde{\theta}}(j|s_g)}{\pi_j} = \lambda \\ \sum_j \pi_j - 1 = 0 \end{cases} \qquad \text{i.e.,} \quad \begin{cases} \sum_g \frac{f_{\tilde{\theta}}(j|s_g)}{\lambda} = \pi_j \\ \sum_j \pi_j - 1 = 0 \end{cases} \qquad \text{i.e.,} \quad \pi_j = \frac{\sum_g f_{\tilde{\theta}}(j|s_g)}{\sum_{j',g} f_{\tilde{\theta}}(j'|s_g)}$$

Next, note that

$$\frac{\partial \mathrm{var}_{\sigma,j}}{\partial p_{\sigma,t}} = 2p_{\sigma,\mathscr{T}(j,t)}\tau_0^2 + 2p_{\sigma,t}\tau^2$$
$$\frac{\partial \mathrm{var}_{\sigma,j}}{\partial \tau_0} = 2\tau_0 \sum_{\mathscr{T}\in\mathbb{T}_j} p_{\sigma,\mathscr{T}}^2$$
$$\frac{\partial \mathrm{var}_{\sigma,j}}{\partial \tau} = 2\tau \sum_{t=1}^{T} p_{\sigma,t}^2$$

so

$$\begin{aligned}
\frac{\partial \Lambda}{\partial p_{\sigma,t}} &= \sum_{i,g,j} f_{\tilde{\theta}}(j|s_g) \frac{(s_{i,g}-\mu_0)^2}{\mathrm{var}_{\sigma_i,j}^2} \frac{\partial \mathrm{var}_{\sigma_i,j}}{\partial p_{\sigma,t}} - \lambda_\sigma \\
&= \sum_{i:\sigma_i=\sigma} \sum_{g,j} f_{\tilde{\theta}}(j|s_g) \frac{(s_{i,g}-\mu_0)^2}{\mathrm{var}_{\sigma,j}^2} \left[2p_{\sigma,\mathscr{T}(j,t)}\tau_0^2 + 2p_{\sigma,t}\tau^2\right] - \lambda_\sigma \\
\frac{\partial \Lambda}{\partial \tau_0} &= \sum_{i,g,j} f_{\tilde{\theta}}(j|s_g) \frac{(s_{i,g}-\mu_0)^2}{\mathrm{var}_{\sigma_i,j}^2} \left[2\tau_0 \sum_{\mathscr{T}\in\mathbb{T}_j} p_{\sigma_i,\mathscr{T}}^2\right] \\
\frac{\partial \Lambda}{\partial \tau} &= \sum_{i,g,j} f_{\tilde{\theta}}(j|s_g) \frac{(s_{i,g}-\mu_0)^2}{\mathrm{var}_{\sigma_i,j}^2} \left[2\tau \sum_{t=1}^{T} p_{\sigma_i,t}^2\right]
\end{aligned}$$

Also

$$\frac{\partial \Lambda}{\partial \mu_0} = \sum_{i,g,j} f_{\tilde{\theta}}(j|s_g) \frac{s_{i,g}-\mu_0}{\mathrm{var}_{\sigma_i,j}}$$
$$\frac{\partial \Lambda}{\partial \lambda_\sigma} = 1 - \sum_{t=1}^{T} p_{\sigma,t}$$

It is not obvious how to solve the above analytically, so instead we use a numerical method.

3 Appendix: Facts

Fact NSC: If $X \sim \mathrm{Normal}(\mu, \tau^2)$, $\alpha > 0$, then $Y := \alpha X \sim \mathrm{Normal}(\alpha\mu, \alpha^2\tau^2)$.

Proof: If $\phi(y) = y/\alpha$ then $f(y) = f(\phi(y))\phi'(y) \propto \exp(-\frac{(y/\alpha-\mu)^2}{2\tau^2}) = \exp(-\frac{(y-\alpha\mu)^2}{2\alpha^2\tau^2})$. ■

Fact NSUM: If $X_i \sim \text{Normal}(\mu_i, \tau_i^2)$, the X_i are independent, and $Y := \sum_{i=1}^n X_i$, then $Y \sim \text{Normal}(\sum_{i=1}^n \mu_i, \sum_{i=1}^n \tau_i^2)$.

Fact NLC: If $X_i \sim \text{Normal}(\mu_i, \tau_i^2)$, the X_i are independent, and $Y := \sum_{i=1}^n \alpha_i X_i$, then $Y \sim \text{Normal}(\sum_{i=1}^n \alpha_i\mu_i, \sum_{i=1}^n \alpha_i^2\tau_i^2)$.

Fact NCP: If $M \sim \text{Normal}(\mu_0, \tau_0^2)$ and $X|\{M = \mu\} \sim \text{Normal}(\mu, \tau^2)$, then $X \sim \text{Normal}(\mu_0, \tau_0^2 + \tau^2)$.

Proof: $f(x) = \int f(x|\mu)f(\mu)\, d\mu \propto \int \exp(-\frac{(x-\mu)^2}{2\tau^2} - \frac{(\mu-\mu_0)^2}{2\tau_0^2})\, d\mu \propto \exp(-\frac{(x-\mu_0)^2}{2(\tau_0^2+\tau^2)})$. ■

TMP1:

$\langle$*compute loglikelihood*$\rangle\equiv$

```
{
  int i, sig, g, t, j;
  list_t *calT;

  // compute sd
  mat_t *sd = gsl_matrix_calloc(Sigma, J);
  \for (sig \in [Sigma]) {
    double sos = \sum_{t \in [T]} pow2(theta->p[sig,t]);
    \for (j \in [J]) {
      double sos_0 =
        \sum_{calT \in bbT[j]}
          pow2( \sum_{t \in calT} theta->p[sig,t] );
      sd[sig,j] = sqrt(pow2(theta->tau_0) * sos_0
                       + pow2(theta->tau) * sos);
    }
  }

  // compute log likelihood
  double loglik =
    \sum_{g \in [G]}
    \logsumexp_{j \in [J]} (
      theta->pi[j] + \sum_{i \in [n]}
        gaussian_log_pdf(s[i,g], theta->mu_0, sd[sigma[i],j]));

  gsl_matrix_free(sd);
}
```