Gibbs sampler for progression model - Using Normal

1 The hierarchical model

Random variables:

For $g = 1, \ldots, G$:
  $Z_g \sim \text{Categorical}(\boldsymbol{\pi})$.
  For $j = 1, \ldots, J$:
    For $\mathscr{T} \in \mathbb{T}_j$:
      $M_{g,j,\mathscr{T}} \sim \text{Normal}(\mu_0, \sigma_0)$.
      For $i = 1, \ldots, n$:
        For $t \in \mathscr{T}$:
          $X_{i,g,t} | \{Z_g = j, M_{g,j,\mathscr{T}} = \mu_{g,j,\mathscr{T}}\} \sim \text{Normal}(\mu_{g,j,\mathscr{T}}, \sigma)$.
  For $i = 1, \ldots, n$:
    $S_{i,g} = \sum_{t=1}^{T} p_{\sigma_i,t} X_{i,g,t}$.

Parameters are:

1. $\mu_0 \in \mathbb{R}, \sigma_0 \in \mathbb{R}_+, \sigma \in \mathbb{R}_+$.
2. $\boldsymbol{\pi} \in$ standard J-simplex in $\mathbb{R}^J$.
3. $\mathbf{p}_\sigma \in$ standard T-simplex in $\mathbb{R}^T$ for $\sigma = 1, \ldots, \Sigma$.

Call $\boldsymbol{\theta} = (\mu_0, \sigma_0, \sigma, \boldsymbol{\pi}, \mathbf{p})$.

2 Change of variables

It will be useful to avoid having $S_{i,g} = \sum_{t=1}^{T} p_{\sigma_i,t} X_{i,g,t}$ as part of the model, so we change variables as follows. Let

$$\phi_{i,g,t} : y_{i,g,t} \mapsto y_{i,g,t} s_{i,g} / p_{\sigma_i,t}$$

Note that $\frac{\partial \phi_{i,g,t}}{\partial y_{i',g',t'}} = s_{i,g}/p_{\sigma_i,t}$ if $i = i', g = g', t = t'$ and 0 otherwise, so the Jacobian determinant

$$|J_\phi| = \prod_{i=1}^{n} \prod_{g=1}^{G} \prod_{t=1}^{T} s_{i,g}/p_{\sigma_i,t}$$

Thus the conditional density

$$\begin{aligned}
f(y_{i,g,t} | j, \mu_{g,j,\mathscr{T}}) &= (s_{i,g}/p_{\sigma_i,t}) f_{\text{Normal}(\mu_{g,j,\mathscr{T}}, \sigma)}(y_{i,g,t} s_{i,g}/p_{\sigma_i,t}) \\
&= \frac{s_{i,g}/p_{\sigma_i,t}}{\sqrt{2\pi\sigma^2}} \exp\Big(- \frac{(y_{i,g,t} s_{i,g}/p_{\sigma_i,t} - \mu_{g,j,\mathscr{T}})^2}{2\sigma^2} \Big) \\
&= \frac{s_{i,g}/p_{\sigma_i,t}}{\sqrt{2\pi\sigma^2}} \exp\Big(- \frac{(y_{i,g,t} - \mu_{g,j,\mathscr{T}} p_{\sigma_i,t}/s_{i,g})^2}{2\sigma^2 (p_{\sigma_i,t}/s_{i,g})^2} \Big) \\
&= f_{\text{Normal}(\mu_{g,j,\mathscr{T}} p_{\sigma_i,t}/s_{i,g}, \sigma p_{\sigma_i,t}/s_{i,g})}(y_{i,g,t})
\end{aligned}$$

3 Likelihood

The full likelihood is

$$
\begin{aligned}
f_{\boldsymbol{\theta}}(\mathbf{s},\mathbf{z},\mathbf{x},\boldsymbol{\mu}) &= \prod_{g=1}^{G} f(z_g) \prod_{i=1}^{n} f(s_{i,g}|x_{i,g,t}) \prod_{\mathscr{T}\in\mathbb{T}_{z_g}} f(\mu_{g,z_g,\mathscr{T}}) \prod_{t\in T} f(x_{i,g,t}|z_g,\mu_{g,z_g,\mathscr{T}}) \\
&= \prod_{g=1}^{G} \pi_{z_g} \prod_{i=1}^{n} 1_{\{\sum_{t=1}^{T} p_{\sigma_i,t} x_{i,g,t} = s_{i,g}\}} \prod_{\mathscr{T}\in\mathbb{T}_{z_g}} f_{\mathrm{Normal}(\mu_0,\sigma_0)}(\mu_{g,z_g,\mathscr{T}}) \prod_{t\in T} f_{\mathrm{Normal}(\mu_{g,z_g,\mathscr{T}},\sigma)}(x_{i,g,t})
\end{aligned}
$$

The marginal likelihood is

$$
f_{\boldsymbol{\theta}}(\mathbf{s}) = \mathbb{E}(f_{\boldsymbol{\theta}}(\mathbf{s},\mathbf{Z},\mathbf{X},\mathbf{M})) \quad (\mathbf{M} \text{ is upper case } \boldsymbol{\mu})
$$

However, we will change variables

We will use Monte Carlo EM to estimate $\boldsymbol{\theta}$. Define

$$
J(\boldsymbol{\theta}|\tilde{\boldsymbol{\theta}}) = \mathbb{E}_{\tilde{\boldsymbol{\theta}}} \log f_{\boldsymbol{\theta}}(\mathbf{s},\mathbf{Z},\mathbf{X},\mathbf{M})
$$

Sample $(\mathbf{z}^{(k)},\mathbf{x}^{(k)},\boldsymbol{\mu}^{(k)})$, $k=1,\ldots,K$, according to $\mathbb{P}_{\tilde{\theta}}$. Then

$$
J(\boldsymbol{\theta}|\tilde{\boldsymbol{\theta}}) \approx J_K(\boldsymbol{\theta}|\tilde{\boldsymbol{\theta}}) := \sum_{k=1}^{K} \log f_{\boldsymbol{\theta}}(\mathbf{s},\mathbf{z}^{(k)},\mathbf{x}^{(k)},\boldsymbol{\mu}^{(k)})
$$

The MCEM algorithm is to repeat $\tilde{\boldsymbol{\theta}} \leftarrow \arg\max_{\boldsymbol{\theta}} J_K(\boldsymbol{\theta}|\tilde{\boldsymbol{\theta}})$, taking new samples at each iteration, until convergence.