Description of Basic Cluster Model

Probability model: Let $X_{ij} \in \{0, \dots, k_{max}\}$. Then for $0 \leq k \leq k_{max}$,

$$\text{logit}(P(X_{ij} = k)) = \theta_j + \beta_i + \tau_{i,k} - ||z_j - w_{i,k}||_2^2$$

where we enforce the constraint $\tau_{i,0} = 0$ to avoid confounding with baseline question difficulty β_i. Note that this differs slightly from the standard ordinal model in that we are modeling logits of $P(X_{ij} = k)$ instead of $P(X_{ij} < k)$. This is necessary when the ordinal responses are given different latent positions, as otherwise we would need the constraint $||z_k - w_{i,k}||_2^2 \geq ||z_j - w_{i,k+1}||_2^2$ for all i, which would not be meaningful to enforce. For calculating likelihoods, we have

$$P(X_{ij} = k) \propto \text{sigmoid}(\theta_j + \beta_i + \tau_{i,k} - ||w_{i,k} - z_j||_2^2)$$

This is not equality, as we need to normalize over k to ensure that the probabilities sum to one.

Priors:

$$z_j, w_{i,k} \sim \sum_{g=1}^{G} \lambda_g MVN_d(\mu_g, \sigma_g I_d)$$

$$\lambda \sim \text{Dirichlet}(\nu)$$

$$\forall g: \ \ \sigma_g \sim \sigma_0^2 \text{Inv}\chi_\alpha^2$$

$$\forall g: \ \ \mu_g \sim MVN_d(0, \omega^2 I_d)$$

$$\theta \sim N(0, \sigma_\theta^2 I_{n_z}), \beta \sim N(0, \sigma_\beta^2 I_{n_w}), \tau_{i,k} \sim N(0, \sigma_\tau^2)$$

$$\sigma_\theta^2 \sim \text{Inv-Gamma}(a, b); \ \ \sigma_\beta \text{ kept fixed}$$

For hyper-parameter values we let $a = 1000$, $b = 100$, $\sigma_\beta^2 = 10$, ν_g (for all g_z, g_w). We also let $G = 4$, although these can be tuned through model selection procedures.

Proposal distributions. For z, w, β, θ, and τ, proposed values are drawn from normal distributions centered around the previous value with tunable proposal standard deviations, which has the advantage of being symmetric so terms such as $p(z^{(1)} \to z')/p(z' \to z^{(1)}) = 1$. For example,

$$p(z^{(1)} \to z') \sim N(z^{(1)}, \sigma_{z_{prop}}^2)$$

For testing the algorithm we set proposal standard deviations to 5.

For cluster parameters K, λ, σ, and μ we do not use MCMC as the posteriors can be sampled directly, as can be seen below.

Posteriors: For z, w, β, and θ, the posterior distributions cannot be computed analytically so we use the following relationships for MCMC sampling:

$$\pi(\beta_i|\mathbf{X}, \mathbf{Z}, \mathbf{W}, \theta) \propto \pi(\beta_i) \prod_j P(X_{ik}|\beta_i, \theta_k, z_j, w_{i,k}, \tau_{i,k})$$

$$\pi(\theta_k|\mathbf{X}, \mathbf{Z}, \mathbf{W}, \theta, \sigma_\theta^2, \tau_{i,k}) \propto \pi(\theta_k|\sigma_\theta^2) \prod_i P(X_{ik}|\beta_i, \theta_k, z_j, w_{i,k}, \tau_{i,k})$$

$$\pi(w_{i,k}|\mathbf{X}, \mathbf{Z}, \mathbf{W}, \theta, \sigma_w^2) \propto \pi(w_{i,k}|\sigma_w^2) \prod_j P(X_{ik}|\beta_i, \theta_k, z_j, w_{i,k}), \tau_{i,k}$$

$$\pi(z_j|\mathbf{X}, \mathbf{W}, \beta, \theta, \sigma_z^2) \propto \pi(z_j|\sigma_z^2) \prod_i P(X_{ik}|\beta_i, \theta_k, z_j, w_{i,k}, \tau_{i,k})$$

$$\pi(\tau_{i,k}|\mathbf{X}, \mathbf{W}, \beta, \theta, \sigma_z^2) \propto \pi(\tau_{i,k}|\sigma_z^2) \prod_i P(X_{ik}|\beta_i, \theta_k, z_j, w_{i,k}, \tau_{i,k})$$

For the cluster parameters, the posteriors have closed forms so they can be sampled directly. This method is analogous to that used in Handcock, Raftery, and Tantrum (2007).

$$\mu_g|\text{others} \sim MVN\left(\frac{m_g \overline{z}_g}{m_g + \sigma_g^2/\omega^2}, \frac{\sigma_g^2}{m_g + \sigma_g^2/\omega^2}\right)$$

$$\sigma_g^2|\text{others} \sim (\sigma_0^2 + ds_g)$$

$$P(K_{z_j} = g|\text{others}) = \frac{\lambda_g \phi_d(z_j; \mu_g, \sigma_g^2 I_d)}{\sum_{g=1}^G \lambda_g \phi_d(z_j; \mu_g, \sigma_g^2 I_d)}$$

$$P(K_{w_{i,k}} = g|\text{others}) = \frac{\lambda_g \phi_d(w_{i,k}; \mu_g, \sigma_g^2 I_d)}{\sum_{g=1}^G \lambda_g \phi_d(w_{i,k}; \mu_g, \sigma_g^2 I_d)}$$

where

$$m_g = \sum_{k:K_{z_j}=g} 1 + \sum_{i:K_{w_{i,k}}=g} 1$$

$$s_g^2 = \frac{1}{d}\left(\sum_{k:K_{z_j}=g} (z_j - \mu_g)^T (z_j - \mu_g) + \sum_{i:K_{w_{i,k}}=g} (w_{i,k} - \mu_g)^T (w_{i,k} - \mu_g)\right)$$

$$\overline{z}_g = \frac{1}{m_g}\left(\sum_{k:K_{z_j}=g} z_j + \sum_{k:K_{z_j}=g} z_j\right)$$

and ϕ_d is the d-dimensional multivariate normal density.

Rejection probabilities are

$$r(z'_k, z_j^{(t)}) = \frac{\pi(z'_j|z_{-j}, \beta, \theta, \mathbf{W})}{\pi(z_j^{(t)}|z_{-j}, \beta, \theta, \mathbf{W})}$$

$$r(w'_i, w_{i,k}^{(t)}) = \frac{\pi(w'_{i,k}|w_{-i}, \beta, \theta, \mathbf{Z})}{\pi(w_{i,k}^{(t)}|w_{-i}, \beta, \theta, \mathbf{Z})}$$

Rejection probabilities are analogous for θ and β.

Overall procedure. We preform the steps in the following order:

1. Propose and update $\mathbf{z}$.
2. Propose and update $\mathbf{w}$.
3. Sample K, μ_g, σ_g^2, and λ_g from their posterior distributions.
4. Propose and update θ.
5. Update σ_θ^2 through MCMC.
6. Propose and update β.