# Supporting Information for "Integrating Sample Similarities into Latent Class Analysis: A Tree-Structured Shrinkage Approach" by Li, Park, Aziz, Liu, Price and Wu

Mengbing Li[1], Daniel E. Park[3], Maliha Aziz[3], Cindy M. Liu[3], Lance B. Price[3], and Zhenke Wu[1,2*]

[1]Department of Biostatistics, University of Michigan, Ann Arbor, MI 48109, USA; E-mail: *zhenkewu@umich.edu

[2]Michigan Institute for Data Science (MIDAS), University of Michigan, Ann Arbor, MI 48019, USA;

[3]Environmental and Occupational Health, Milken Institute School of Public Health, The George Washington University, Washington, DC 20052, USA

## A1 Details of the Variational Inference Algorithm

In the following, let $q_t(A)$ represent a generic variational distribution for unknown quantities in $A$ at iteration $t$; Let $q_t(-A)$ represent the variational distribution for all but the random quantities in $A$. Let $\text{pr}(A)$ represent a generic true joint distribution of the quantities in $A$. $[Q] := \{1, \ldots, Q\}$ represents the set of positive integers smaller than or equal to an integer $Q$.

*Step 0.* Initialize the variational distribution $q_t(\cdot)$ at $t = 0$. Because the updates for each component of the variational distribution in Equation (14) of the Main Paper has a closed form that is fully determined by the first and second moments, it is sufficient to initialize these moments. In addition, because the sigmoid functions are bounded by Gaussian kernels that depend on additional tuning parameters $(\boldsymbol{\psi}, \boldsymbol{\phi})$, we need to initialize them too. Finally, we initialize hyperparameters $(\boldsymbol{\tau}_1, \boldsymbol{\tau}_2)$. In particular,

- Additive components of the logistic stick-breaking parameters $\alpha_{uk}$ given $s_u = 1$: $\{(\mu^{(t)}_{\alpha_{uk},1}, \sigma^{2,(t)}_{\alpha_{uk},1}) = (E_{q_t}[\alpha_{uk} \mid s_u = 1], V_{q_t}[\alpha_{uk} \mid s_u = 1]) : k \in [K-1], u \in \mathcal{V}\}$. The mean and variance determine the optimal variational distribution for $\alpha_{uk}$ given $s_u = 1$, which can be shown to be a Gaussian distribution;

- Logit-transformed response probabilities: $\{(E_{q_t}[\gamma_{jk}], V_{q_t}[\gamma_{jk}]) : j \in [J], k \in [K]\}$;
- Probability of spike-and-slab indicators: $\{p_u^{(t)} = E_{q_t}[s_u] : u \in \mathcal{V}\}$;
- Tuning parameters in the Jaakkola-Jordan lower bounding technique: $\{\psi_{jk}^{(t)},\ j \in [J], k \in [K]\}$, $\{\phi_k^{(v),(t)},\ v \in \mathcal{V}_L,\ k \in [K-1]\}$, and
- The hyperparameters $\{\tau_{1k\ell}^{(t)}, k \in [K-1], \ell \in [L]\}$, $\{\tau_{2jk}^{(t)}, j \in [J], k \in [K]\}$.

Compute additional first and second moments as follows:

$$E_{q_t}[\eta_{vk}^2] = \sum_{u \in a(v)} \left\{ p_u^{(t)} (\sigma^{2,(t)}_{\alpha_{uk,1}} + (1 - p_u^{(t)})[\mu^{(t)}_{\alpha_{uk,1}}]^2) \right\} + E^2_{q_t}[\eta_{vk}], v \in \mathcal{V}_L$$
$$E_{q_t}[\alpha_{uk}^2] = p_u^{(t)} (\sigma^{2,(t)}_{\alpha_{uk,1}} + [\mu^{(t)}_{\alpha_{uk,1}}]^2) + (1 - p_u^{(t)}) \sigma^{2,(t)}_{\alpha_{uk,0}},$$

where $\sigma^{2,(t)}_{\alpha_{uk,0}} = \tau_{1k\ell_u} w_u$ is the variance of $\alpha_{uk}$ in its variational distribution given $s_u = 0$ (this will be derived in Step 1b below according to the VI update for $\alpha_{uk}$). Finally, compute $E_{q_t}[\gamma_{jk}^2] = \sigma^{2,(t)}_{\gamma_{jk}} + [\mu^{(t)}_{\gamma_{jk}}]^2$, $E_{q_t}[\eta_{vk}] = \sum_{u \in a(v)} E_{q_t}[\xi_{uk}]$, $E_{q_t}[\xi_{uk}] = E_{q_t}[s_u \alpha_{uk}] = p_u^{(t)} \mu^{(t)}_{\alpha_{uk,1}}$. Calculate the initial $\mathcal{E}^*(q_t) = 0$

At Step $t+1$, iterate between Step 1 to 4 until convergence:

*Step 1a.* Update $q_{t+1}(\boldsymbol{Z}_i^{(v)})$, $i \in [n_v]$, $v \in \mathcal{V}_L$ by a multinomial distribution with probabilities $\boldsymbol{r}_i^{(v),(t+1)} = (r_{i1}^{(v),(t+1)}, \ldots, r_{iK}^{(v),(t+1)})^\mathsf{T}$:

$$r_{ik}^{(v),(t+1)} \propto$$
$$\exp\left[ \sum_{j=1}^{J} \log \sigma(\psi_{jk}^{(t)}) + \left[ X_{ij}^{(v)} E_{q_t}(\gamma_{jk}) - \psi_{jk}^{(t)} \right]/2 - g(\psi_{jk}) \left\{ E_{q_t}(\gamma_{jk}^2) - [\psi_{jk}^{(t)}]^2 \right\} \right.$$
$$+ \sum_{m<k} \left( \log \sigma(\phi_m^{(v),(t)}) + \left[ -E_{q_t}(\eta_{vm}) - \phi_m^{(v),(t)} \right]/2 - g(\phi_m^{(v),(t)}) \left\{ E_{q_t}(\eta_{vm}^2) - [\phi_m^{(v),(t)}]^2 \right\} \right)$$
$$\left. + \mathbf{1}\{k < K\} \left( \log \sigma(\phi_k^{(v),(t)}) + \left[ E_{q_t}(\eta_{vk}) - \phi_k^{(v),(t)} \right]/2 - g(\phi_k^{(v),(t)}) \left\{ E_{q_t}(\eta_{vk}^2) - [\phi_k^{(v),(t)}]^2 \right\} \right) \right].$$

*Step 1b.* Update $q_{t+1}(\boldsymbol{\gamma})$ and $q_{t+1}(s_u, \boldsymbol{\alpha}_u)$, $u \in \mathcal{V}$. We first do the update for the root node $u = u_0$ when $\boldsymbol{\gamma}$ gets updated; for non-root nodes, the $\boldsymbol{\gamma}$ is not updated. We follow a topological ordering in $\mathcal{T}_w$ when updating $(s_u, \boldsymbol{\alpha}_u)$, $u \in \mathcal{V}$; Random ordering also

works. In particular, the update is

$$\begin{aligned}
&\log q_{t+1}(\boldsymbol{\gamma}, s_u, \boldsymbol{\alpha}_u) = E_{q_t(-(\boldsymbol{\gamma},s_u,\boldsymbol{\alpha}_u))}\left[\log\{h^*(\mathbf{X},\boldsymbol{\psi},\boldsymbol{\gamma},\mathbf{Z})h^{**}(\boldsymbol{\phi},\boldsymbol{s},\boldsymbol{\alpha},\mathbf{Z})\mathrm{pr}(\boldsymbol{s},\boldsymbol{\gamma},\boldsymbol{\alpha},\boldsymbol{\varrho})\}\right] + const \\
&= -\sum_{j=1}^{J}\sum_{k=1}^{K}\left\{\frac{1}{2\sigma^{2,(t+1)}_{\gamma_{jk}}}\left(\gamma_{jk} - \mu^{(t+1)}_{\gamma_{jk}}\right)^2 + \frac{1}{2}\log\left[2\pi\sigma^{2,(t+1)}_{\gamma_{jk}}\right]\right\} \\
&\quad - s_u\sum_{k=1}^{K-1}\left\{\frac{1}{2\sigma^{2,(t+1)}_{\alpha_{uk},1}}(\alpha_{uk} - \mu^{(t+1)}_{\alpha_{uk},1})^2 + \frac{1}{2}\log[2\pi\sigma^{2,(t+1)}_{\alpha_{uk},1}]\right\} \\
&\quad - (1-s_u)\sum_{k=1}^{K-1}\left\{\frac{1}{2\tau^{(t)}_{1k\ell_u}w_u}\alpha^2_{uk} + \frac{1}{2}\log\{2\pi\tau^{(t)}_{1k\ell_u}w_u\}\right\} + s_u\zeta^{(t+1)}_u + const
\end{aligned}$$

where const does not depend on $s_u$, $\boldsymbol{\gamma}$ and $\boldsymbol{\alpha}_u$; $\mu^{(t+1)}_{\gamma_{jk}} = B^{(t)}_{jk}/A^{(t)}_{jk}$, $\sigma^{2,(t+1)}_{\gamma_{jk}} = 1/A^{(t)}_{jk}$, $\mu^{(t+1)}_{\alpha_{uk},1} = D^{(t)}_{uk}/C^{(t)}_{uk}$, $\sigma^{2,(t+1)}_{\alpha_{uk},1} = 1/C^{(t)}_{uk}$ where

$$A^{(t)}_{jk} = \frac{1}{\tau^{(t)}_{2jk}} + 2\sum_{v\in\mathcal{V}_L}\sum_{i=1}^{n_v} r^{(v),(t+1)}_{ik} g(\psi^{(t)}_{jk}), \tag{S1}$$

$$B^{(t)}_{jk} = \sum_{v\in\mathcal{V}_L}\sum_{i=1}^{n_v}\left\{r^{(v),(t+1)}_{ik}X^{(v)}_{ij}/2\right\}, \tag{S2}$$

$$C^{(t)}_{uk} = \frac{1}{\tau^{(t)}_{1k\ell_u}w_u} + 2\sum_{v\in\mathcal{V}_L\cap d(u)}\sum_{i=1}^{n_v}\sum_{m=k}^{K} r^{(v),(t+1)}_{im} g(\phi^{(v),(t)}_k), \text{ for } k\in[K-1], \tag{S3}$$

$$D^{(t)}_{uk} = \sum_{v\in\mathcal{V}_L\cap d(u)}\sum_{i=1}^{n_v}\frac{1}{2}r^{(t+1)}_{ik} - \frac{1}{2}\sum_{m=k+1}^{K} r^{(t+1)}_{im} - 2\left(\sum_{m=k}^{K} r^{(t+1)}_{im} g(\phi^{(v),(t)}_k)\sum_{w\in a(v)\setminus u} E_{q^*_t}[s_w\alpha_{wk}]\right), \tag{S4}$$

$$\zeta^{(t+1)}_u = E_{q_t}\left[\log\frac{\rho_{\ell_u}}{1-\rho_{\ell_u}}\right] - \frac{1}{2}\sum_{k=1}^{K-1}[\log(\tau^{(t)}_{1k\ell_u}w_u) + \log(C^{(t)}_{uk})] + \sum_{k=1}^{K-1}\frac{D^{(t),2}_{uk}}{2C^{(t)}_{uk}}, \tag{S5}$$

where $E_{q^*_t} = E_{q_{t+1}}$ if the node $w \in a(v)\setminus u$ has already been updated; $E_{q^*_t} = E_{q_t}$ otherwise.

It is readily recognized that the update for $\boldsymbol{\gamma}$ is Gaussian and the update for $(s_u, \boldsymbol{\alpha}_u)$

is a two-component Gaussian mixture:

$$\underbrace{\prod_{j=1}^{J}\prod_{k=1}^{K}\mathcal{N}\left(\gamma_{jk} \mid \mu_{\gamma_{jk}}^{(t+1)}, \sigma_{\gamma_{jk}}^{2,(t+1)}\right)}_{(I)}$$
$$\cdot \underbrace{\prod_{k=1}^{K-1}\mathcal{N}\left(\alpha_{uk} \mid s_u\mu_{\alpha_{uk}}^{(t+1)}, s_u\sigma_{\alpha_{uk},1}^{2,(t+1)} + (1-s_u)\sigma_{\alpha_{uk},0}^{2,(t+1)}\right) \cdot \mathsf{Bernoulli}\left(s_u; p_u^{(t+1)}\right)}_{(II)}, \tag{S6}$$

where $\sigma_{\gamma_{jk}}^{2,(t+1)} = \tau_{2jk}^{(t)}$, $\sigma_{\alpha_{uk},0}^{2,(t+1)} = \tau_{1k\ell_u}^{(t)} w_u$, and $p_u^{(t+1)}$ satisfies $\log\left\{\frac{p_u^{(t+1)}}{1-p_u^{(t+1)}}\right\} = \zeta_u^{(t+1)}$.

Of note, we have induced factorization $\prod_{j=1}^{J}\prod_{k=1}^{K} q_{t+1}(\gamma_{jk})$ and $\prod_{k=1}^{K-1} q_{t+1}(\alpha_{uk} \mid s_u)$. In the variational family, we did not assume this factorization. However, we do obtain updates that factorize. This phenomenon is determined by the underlying generative model (the graph structure that determines the joint distribution) and the variational assumption (which determines blocks of parameters to iteratively update) (see, e.g., Bishop, 2006, Section 10.2.5). We update $q_{t+1}(\boldsymbol{\gamma})$ according to component $(I)$ in (S6) and $q_{t+1}(s_u, \boldsymbol{\alpha}_u)$ according to $(II)$ in (S6) when $u$ is the root node; for a non-root node, we only update $q_{t+1}(s_u, \boldsymbol{\alpha}_u)$ according to component $(II)$ in (S6).

*Step 1c.* Update $q_{t+1}(\rho_\ell)$, $\ell = 1, \ldots, L$ by $\mathsf{Beta}(e_\ell^{(t+1)}, f_\ell^{(t+1)})$, where $e_\ell^{(t+1)} = \sum_{u\in\mathcal{V}:\ell_u=\ell} p_u^{(t+1)} + a_\ell$ and $f_\ell^{(t+1)} = \sum_{u\in\mathcal{V}:\ell_u=\ell}(1-p_u^{(t+1)}) + b_\ell$.

For every $d$ steps above, do Step 2-4:

*Step 2.* Update variational parameters $\{\psi_{jk}, j \in [J], k \in [K]\}$ and $\{\phi_k^{(v)}, v \in \mathcal{V}_L, k \in [K-1]\}$ by optimizing the lower bound $\mathcal{E}^*(q_{t+1})$ which leads to the updates:

$$\psi_{jk}^{(t+1)} = \sqrt{E_{q_{t+1}}[\gamma_{jk}^2]}, \qquad \phi_k^{(v),(t+1)} = \sqrt{E_{q_{t+1}}[\eta_{vk}^2]}. \tag{S7}$$

*Step 3.* Update the hyperparameters $\tau_{1k\ell}$ by

$$\tau_{1k\ell}^{(t+1)} = \frac{1}{\sum_{u:\ell_u=\ell} 1} \sum_{u\in\mathcal{V}:\ell_u=\ell} E_{q_{t+1}}[\alpha_{uk}^2/w_u], \tag{S8}$$

and update $\tau_{2jk}$ by $\tau_{2jk}^{(t+1)} = E_{q_{t+1}}[\gamma_{jk}^2]$.

*Step 4.* Compute $\mathcal{E}^*(q_{t+1})$ according to Appendix A1.1. Stop the iteration once the absolute change in $\mathcal{E}^*(q_{t+1})$ is less than a tolerance `tol=1e-8`. The hyperparameter updates are often slower than the variational parameter updates to converge in terms of the $\mathcal{E}^*(q_{t+1})$. In practice, we can separate the tolerance levels for the hyperparameter updates (`hyper_tol=1e-4`) and VI parameter updates (e.g., `tol=1e-8`). One may update the hyperparameters every $d$ steps of the updates of the variational parameters.

In practice, we can adjust $d$ to speed up the convergence. In this paper, we use $d = 10$ which works well in simulations and data analysis.

We access the approximate posterior densities when needed by plugging in the relevant moments at convergence. Finally, different initializations may lead to distinct converged values of the parameters, some of which are local optima. In practice, we initialize $M$ times and select the converged set of parameters (among $M$ sets) that produces the highest $\mathcal{E}^*(q_{t+1})$; In this paper, we used $M = 5$.

## A1.1 Computing $\mathcal{E}^*(q)$

For ease of presentation, we omit the iterator $t$ during the VI updates.

$$\mathcal{E}^*(q) = \mathcal{E}^*(q; \boldsymbol{\phi}, \boldsymbol{\psi}, \boldsymbol{\tau}_1, \boldsymbol{\tau}_2) =$$

$$= \sum_{v \in \mathcal{V}_L} \sum_{i=1}^{n_v} \sum_{k=1}^{K} r_{ik}^{(v)} \Bigg\{ \sum_{m<k} \Bigg( \log \sigma(\phi_m^{(v)}) + \left[(-1) E_{q_t}(\eta_{vm}) - \phi_m^{(v)}\right]/2 - g(\phi_m^{(v)}) \left\{ E_{q_t}(\eta_{vm}^2) - [\phi_m^{(v)}]^2 \right\} \Bigg)$$

$$+ \mathbf{1}\{k < K\} \Bigg( \log \sigma(\phi_k^{(v)}) + \left[ E_{q_t}(\eta_{vk}) - \phi_k^{(v)} \right]/2 - g(\phi_k^{(v)}) \left\{ E_{q_t}(\eta_{vk}^2) - [\phi_k^{(v)}]^2 \right\} \Bigg)$$

$$+ \sum_{j=1}^{J} \log(\sigma(\psi_{jk})) + E_q \left[ \{X_{ij}^{(v)} \gamma_{jk} - \psi_{jk}\}/2 \right] - g(\psi_{jk}) E_q \left( \{X_{ij}^{(v)} \gamma_{jk}\}^2 - \psi_{jk}^2 \right) \Bigg\} \tag{S9}$$

$$+ \sum_{u \in \mathcal{V}} E_q[\log \rho_{\ell_u}] E_q[s_u] + E_q[\log(1 - \rho_{\ell_u})] E_q[1 - s_u] \tag{S10}$$

$$+ \sum_{\ell=1}^{L} (a_\ell - 1) E_q[\log \rho_\ell] + (b_\ell - 1) E_q[\log(1 - \rho_\ell)] - \log Beta(a_\ell, b_\ell) \tag{S11}$$

$$- \sum_{u \in \mathcal{V}} \sum_{k=1}^{K-1} \left( \frac{E_q[\alpha_{uk}^2]}{2\tau_{1k\ell_u} w_u} + \frac{1}{2} \log(2\pi \tau_{1k\ell_u} w_u) \right) \tag{S12}$$

$$- \sum_{j=1}^{J} \sum_{k=1}^{K} \left( \frac{E_q[\gamma_{jk}^2]}{2\tau_{2jk}} + \frac{1}{2} \log(2\pi \tau_{2jk}) \right) \tag{S13}$$

$$+ \frac{1}{2} \sum_{j=1}^{J} \sum_{k=1}^{K} E_q[(\gamma_{jk} - \mu_{jk})^2 / \sigma_{\gamma_{jk},1}^2] + \log(2\pi \sigma_{\gamma_{jk},1}^2) \tag{S14}$$

$$- \sum_{u \in \mathcal{V}} E_q[s_u] \log(p_u) + E_q[1 - s_u] \log(1 - p_u) \tag{S15}$$

$$+ \frac{1}{2} \sum_{u \in \mathcal{V}} \sum_{k=1}^{K-1} E_q[s_u (\alpha_{uk} - \mu_{\alpha_{uk}})^2 / \sigma_{\alpha_{uk},1}^2] + E_q[s_u] \log(2\pi \sigma_{\alpha_{uk},1}^2) \tag{S16}$$

$$+ \frac{1}{2} \sum_{u \in \mathcal{V}} \sum_{k=1}^{K-1} [E_q[1 - s_u] + E_q[1 - s_u] \log(2\pi \tau_{1k\ell_u} w_u)] \tag{S17}$$

$$- \sum_{l=1}^{L} \{(e_\ell - 1) E_q[\log(\rho_\ell)] + (f_\ell - 1) E_q[\log(1 - \rho_\ell)] - \log \mathsf{Beta}(e_\ell, f_\ell)\} \tag{S18}$$

$$- \sum_{v \in \mathcal{V}_L} \sum_{i=1}^{n_v} \sum_{k=1}^{K} r_{ik}^{(v)} \log(r_{ik}^{(v)}) \tag{S19}$$

# A2 Additional Simulation Details and Results

*Multiple true leaf groups.* In the Main Paper, we set $G = 3$ leaf groups in the tree shown in Figure 3a of the Main Paper (with indexed internal nodes and leaves): Group 1 ($\{6, 7, 8\}$), Group 2 ($\{9, 10, 11\}$) and Group 3 ($\{12, 13, 14, 15, 16\}$). The true grouping of leaves is obtained by setting $s_1 = s_2 = s_3 = 1$ in Equation (5) of the Main Paper and $s_u = 0, u \neq 1, 2, 3$. We set $\boldsymbol{\pi}_v = (0.197, 0.303, 0.500)^\mathsf{T}$, $(0.644, 0.221, 0.134)^\mathsf{T}$ and $(0.4, 0.3, 0.3)^\mathsf{T}$ for leaf $v \in G_1, G_2, G_3$, respectively. For $K = 3$ classes, we set the response probability profiles $\boldsymbol{\Theta} = [\boldsymbol{\theta}_{\cdot 1}, \ldots, \boldsymbol{\theta}_{\cdot K}]^\mathsf{T}$ to be

$$\begin{aligned}
\boldsymbol{\theta}_{\cdot,k=1} &= \{(\theta_0, 1-\theta_0, 1-\theta_0), \ldots, (\theta_0, 1-\theta_0, 1-\theta_0)\}, \\
\boldsymbol{\theta}_{\cdot,k=2} &= \{(1-\theta_0, \theta_0, 1-\theta_0), \ldots, (1-\theta_0, \theta_0, 1-\theta_0)\}, \\
\boldsymbol{\theta}_{\cdot,k=3} &= \{\underbrace{(1-\theta_0, 1-\theta_0, \theta_0), \ldots, (1-\theta_0, 1-\theta_0, \theta_0)}_{J}\},
\end{aligned}$$

for $J = 21, 84$ binary measurements per subject, and $\theta_0 = 0.95, 0.8$ to represent stronger and weaker between-class signal strengths. We simulated for $N = 1000, 4000$ observations, under balanced or unbalanced leaf-specific sample sizes (see Section 5.1 in the Main Paper).

*A single true leaf group.* We also simulated $R = 200$ independent replicate data sets corresponding to the truth with a single leaf group (equivalent to a single vector of latent class probabilities $\boldsymbol{\pi}_v = \boldsymbol{\pi}_0$), under which a classical latent class model would be perfectly appropriate. We fitted the proposed model and compared the RMSE against a few alternative models as in the Main Paper. Figure S3 shows, by learning the posterior node-specific slab-versus-spike selection probabilities, the proposed model produced similar or smaller RMSEs for estimating the population latent class probabilities. Here we have set $\boldsymbol{\pi}_0 = (0.4, 0.3, 0.3)^\mathsf{T}$ and otherwise identical scenario setup as in the above setting of multiple true leaf groups.

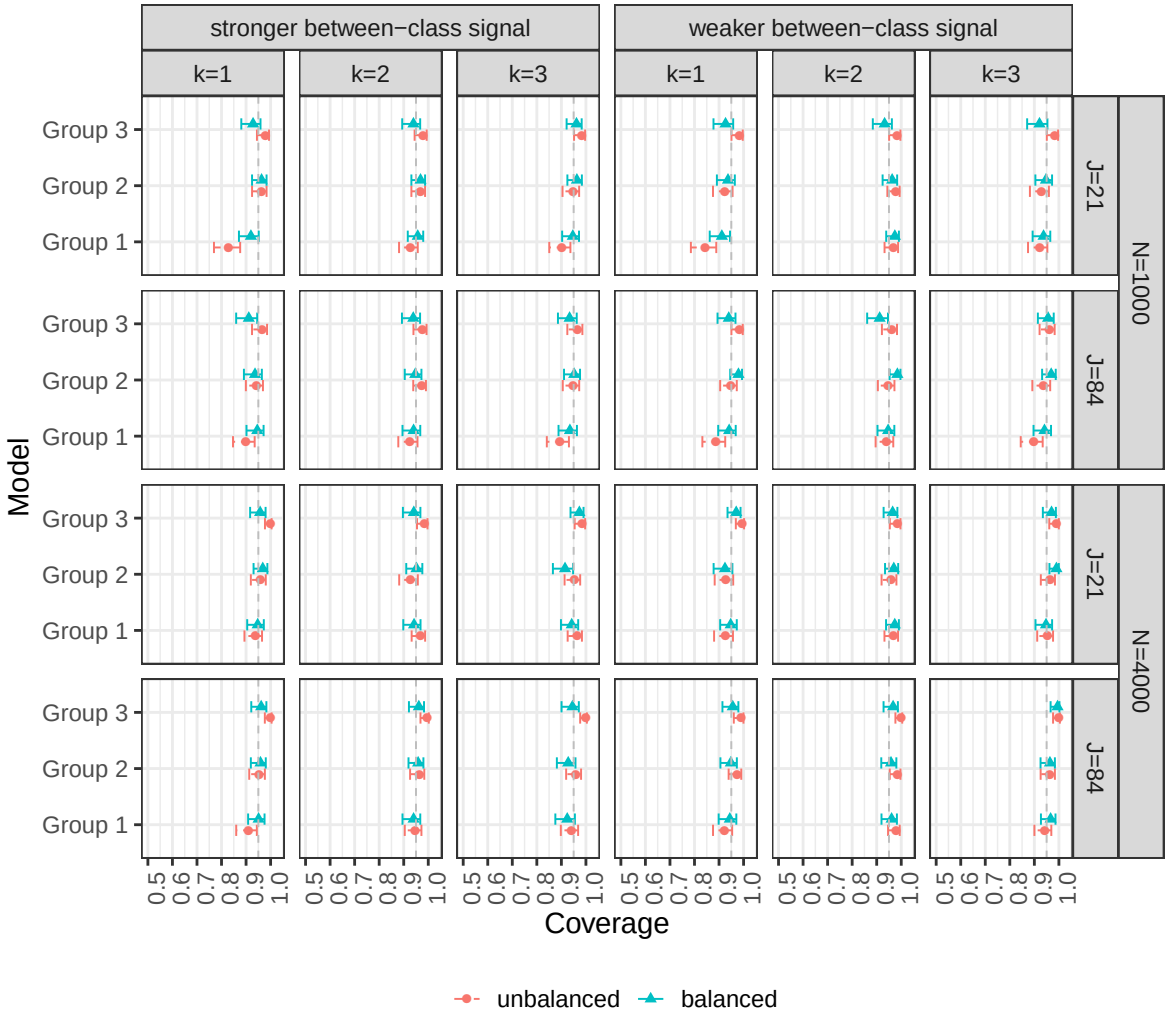


Figure S1: Empirical coverage rates ("•") of the approximate 95% credible intervals (CrIs) based on the proposed grouped estimates from 200 replications (the intervals reflect Monte Carlo uncertainty). The vertical dashed lines indicate 95%.

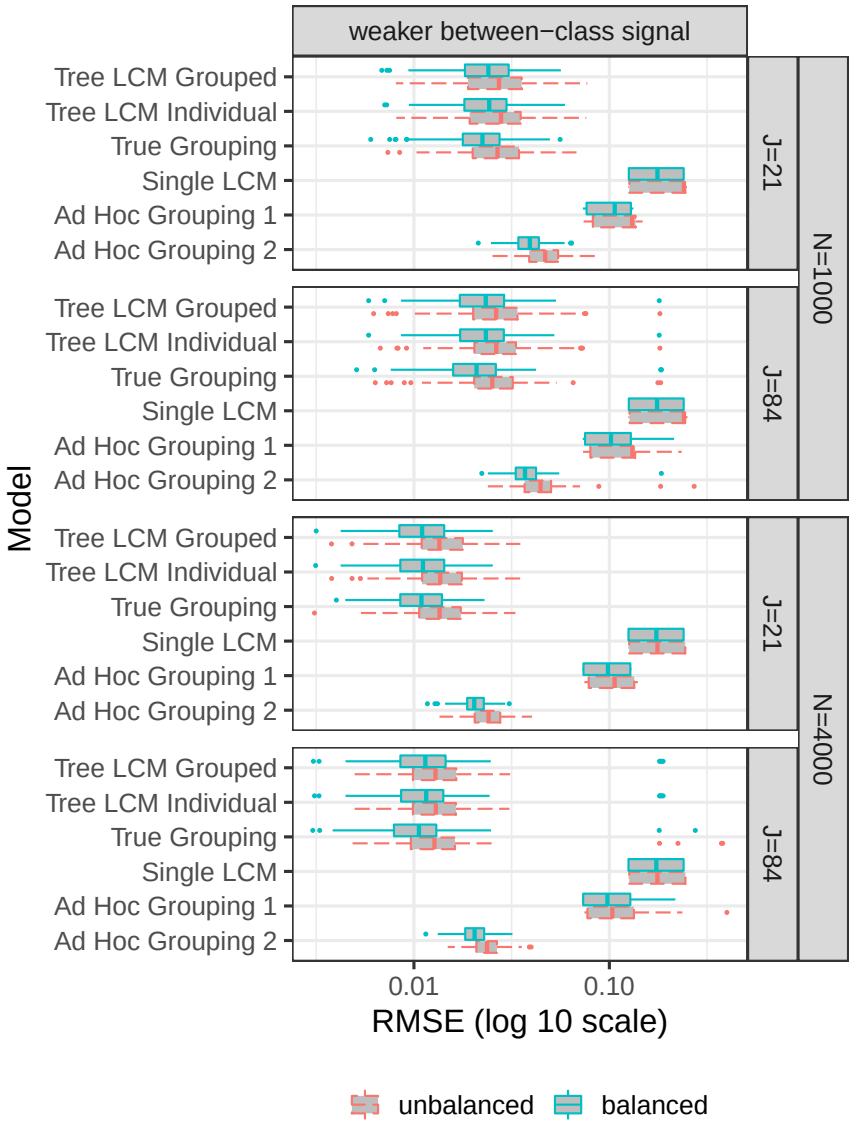


Figure S2: Under less discrepant class-specific response probabilities, simulation studies show the proposed model produces grouped estimates $\widehat{\boldsymbol{\pi}}_v^{\text{dgrp}}$ with similar or smaller RMSEs compared to alternatives (see Section 5 in the Main Paper).

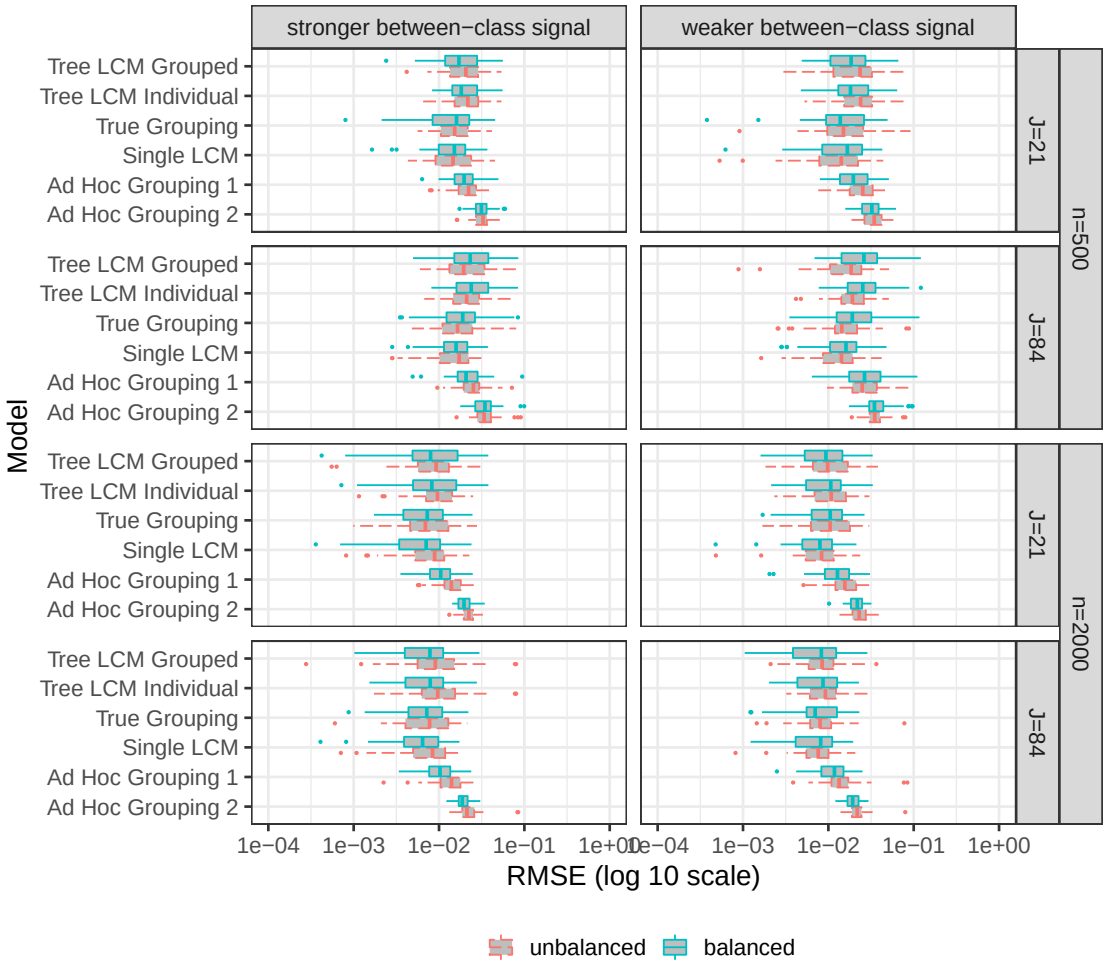


Figure S3: Under the simulation truth of a single vector of latent class probabilities $\boldsymbol{\pi}_v = \boldsymbol{\pi}_0$ (equivalent to a single leaf group), simulation results show that the proposed model produces grouped estimates $\widehat{\boldsymbol{\pi}}_v^{\text{dgrp}}$ with similar or smaller RMSEs compared to alternatives considered in Section 5.1 of the Main Paper.

# A3 Appendix Figures

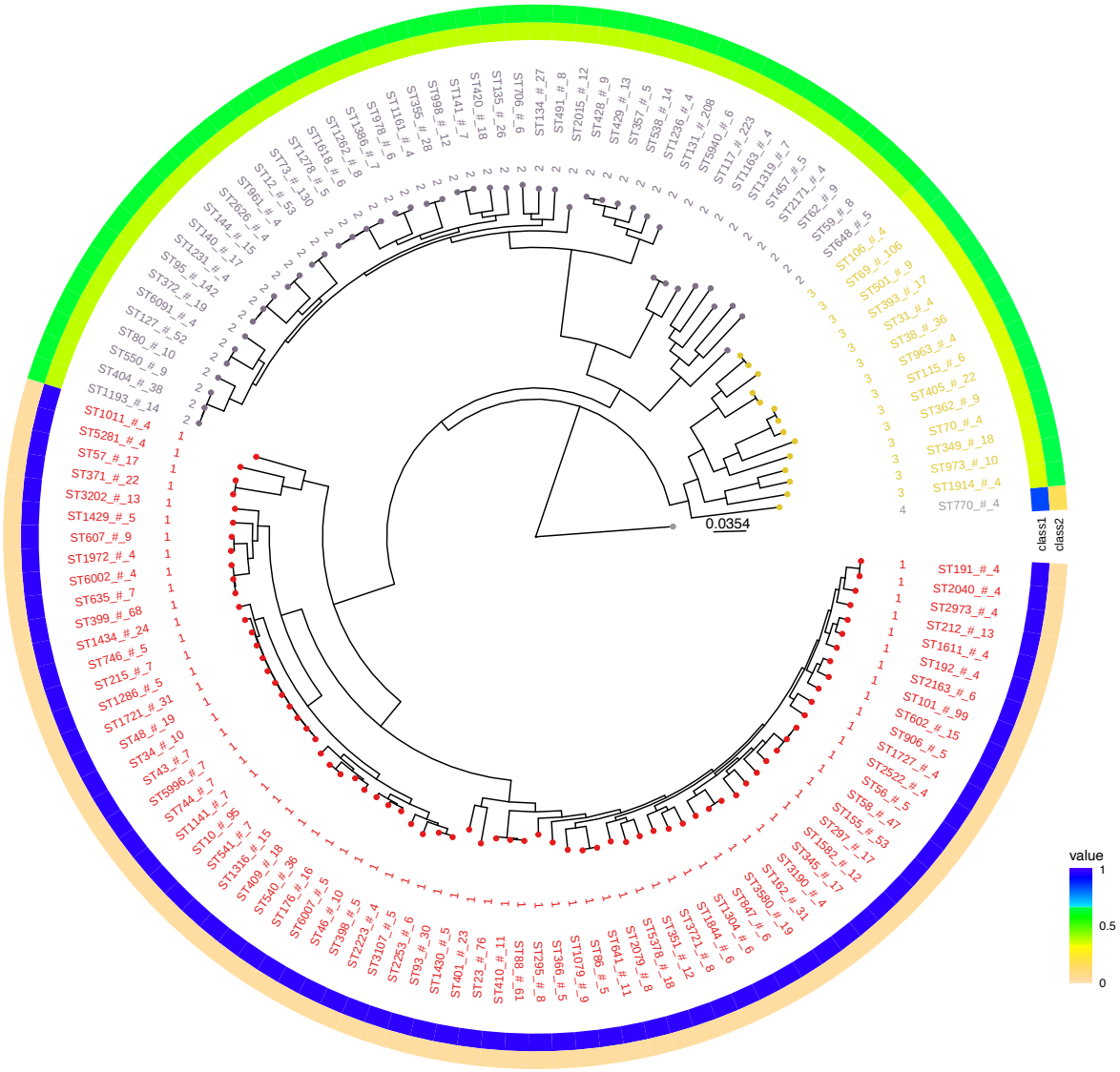


Figure S4: Data results: estimated groups and class probabilities based on fixed ad hoc leaf groupings as indicated by the group id's near the tips of the circular tree.

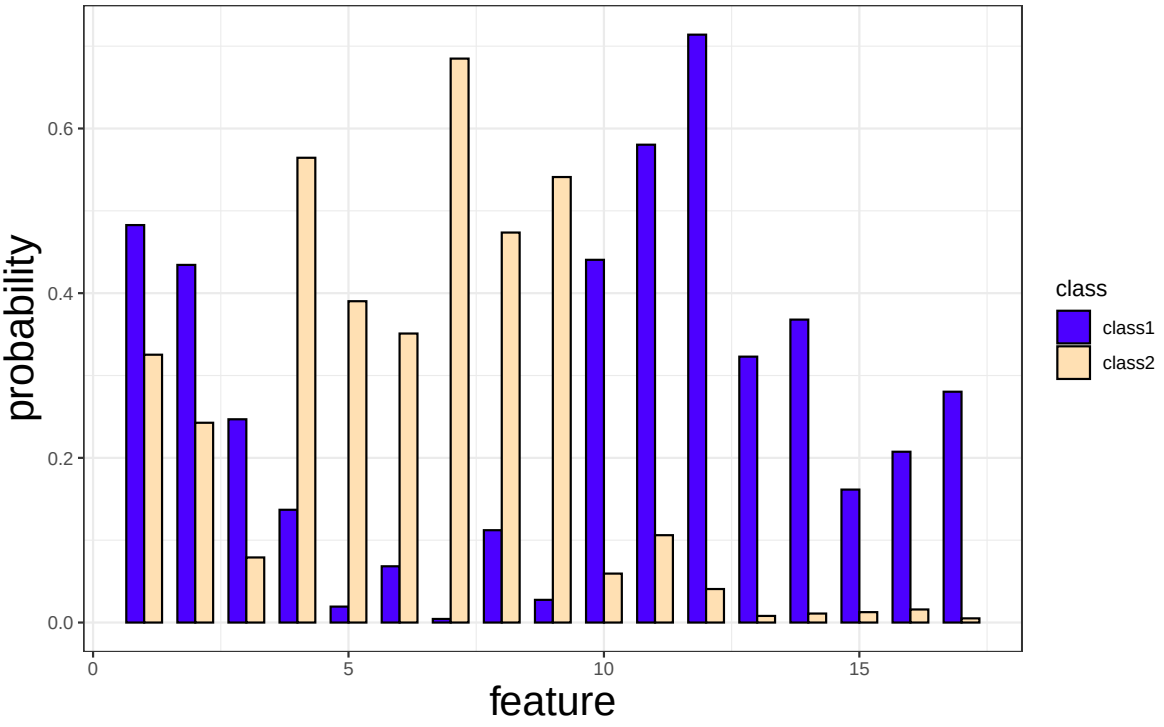


Figure S5: Data results: estimated class-specific response profiles based on fixed ad hoc leaf groupings as in Figure S4.

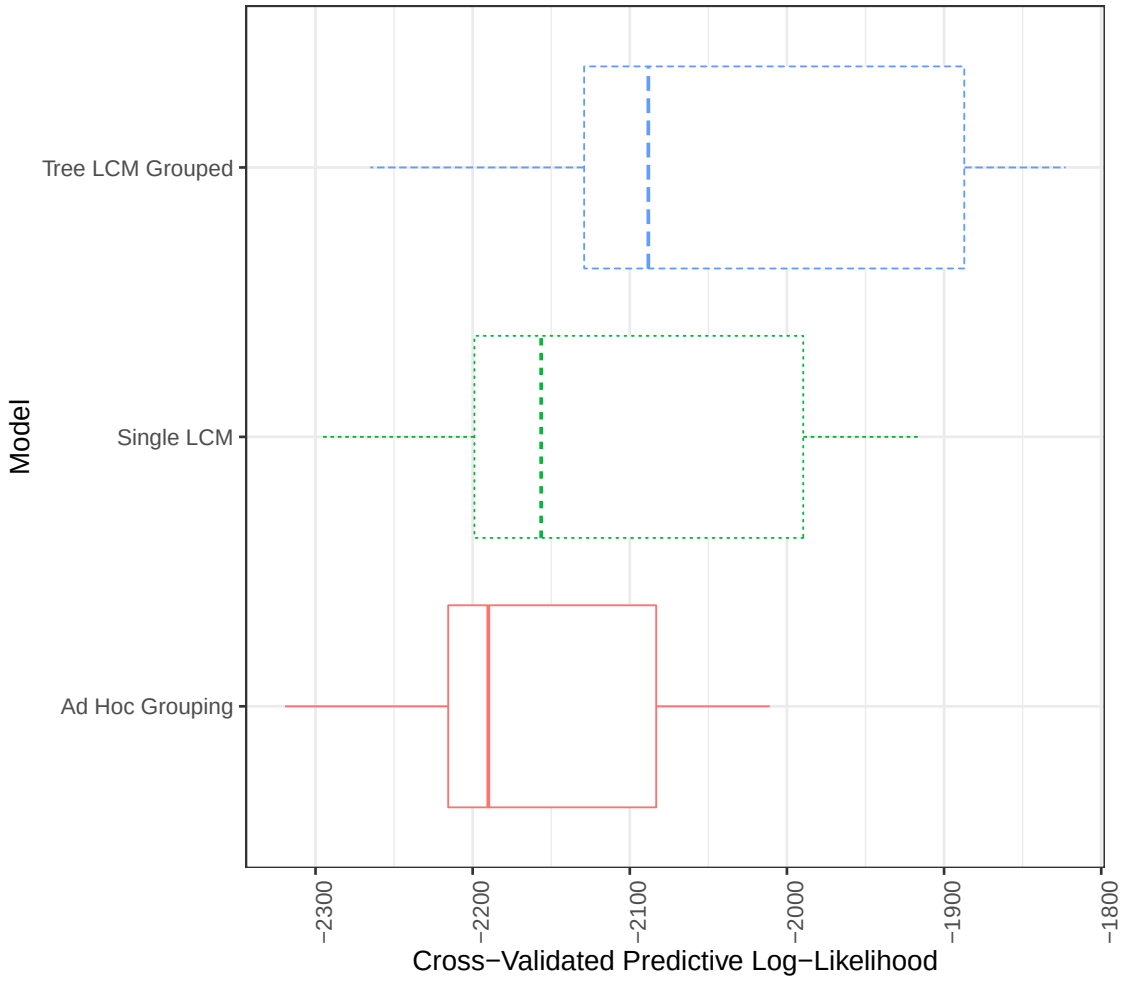


Figure S6: Data results: the proposed model produces the highest 10-fold cross-validated log-likelihood than the single LCM and the model based on an ad hoc grouping (as in Figure S4).

# References

Bishop, C. M. (2006). *Pattern Recognition and Machine Learning (Information Science and Statistics)*. Springer-Verlag, Berlin, Heidelberg.