

Supplemental Material for: Bayesian Clustered Coefficients Regression with Auxiliary Covariates Assistant Random Effects

Guanyu Hu, Yishu Xue, and Zhihua Ma

S1 Proof for Proposition 1

Proof. By definition, the matrix Σ is positive definite if and only if for any vector $\mathbf{x} \neq \mathbf{0}$, we have

$$\mathbf{x}^\top \Sigma \mathbf{x} > 0.$$

Expanding Σ as the linear combination of \mathbf{I}_n and $\Sigma_1, \dots, \Sigma_J$, we have

$$\begin{aligned} \mathbf{x}^\top \Sigma \mathbf{x} &= \mathbf{x}^\top \left(\alpha_0 \mathbf{I}_n + \sum_{i=1}^J \alpha_i \Sigma_i \right) \mathbf{x} \\ &= \alpha_0 \mathbf{x}^\top \mathbf{x} + \sum_{i=1}^J \alpha_i \mathbf{x}^\top \Sigma_i \mathbf{x}. \end{aligned}$$

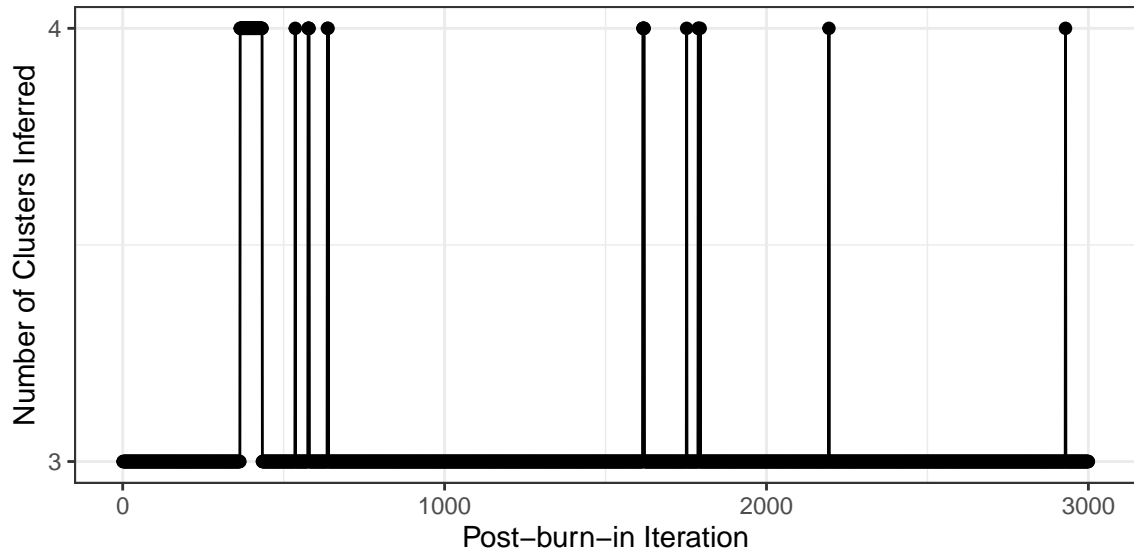
We know $\mathbf{x}^\top \mathbf{x} > 0$ as $\mathbf{x} \neq \mathbf{0}$, and since Σ_i is positive definite,

$$\mathbf{x}^\top \Sigma_i \mathbf{x} > 0.$$

Since $\sum_{j=0}^J \alpha_j = 1$ and $0 \leq \alpha_j \leq 1$ for $j = 1, \dots, J$, the α_j 's cannot be all 0's, and therefore

$$\alpha_0 \mathbf{x}^\top \mathbf{x} + \sum_{i=1}^J \alpha_i \mathbf{x}^\top \Sigma_i \mathbf{x} > 0,$$

which concludes that Σ is positive definite. □



Supp Figure 1: Trace plot for the number of inferred clusters in each post-burn-in iteration.

S2 Convergence Check for Real Data Application

We examine the 3000×159 matrix of posterior chains for the latent variables z_1, \dots, z_{159} to check if the posterior number of clusters, 3, is a stable estimate. The traceplot in Supp Figure 1, verifying that the uncertainty in the final reported value is minimal. Out of the 3,000 post-burn-in iterations, only 87 take the value 4, and the rest 2,913 iterations yield 3 clusters.

As an additional verification, we use the Dahl’s method (Dahl, 2006), which is an alternative approach to infer the final clustering configuration, and check how many clusters it inferred. The function `salso()` in R package `salso` (Dahl et al., 2020) provides a greedy search method that returns a partition that summarizes a partition distribution. A loss function is minimized during the greedy search. With the default loss function, variation of information (see, e.g., Wade and Ghahramani, 2018), the

minimizer partition is composed of three clusters of size 151, 5, and 3. Despite the minor differences in cluster sizes, the number of clusters agrees with our conclusion obtained using the mode, presented in Section 5 of the manuscript.

References

- Dahl, D. B. (2006). Model-based clustering for expression data via a Dirichlet process mixture model. *Bayesian Inference for Gene Expression and Proteomics*, **4**, 201–218.
- Dahl, D. B., Johnson, D. J., and Müller, P. (2020). *salso: Search Algorithms and Loss Functions for Bayesian Clustering*. URL <https://CRAN.R-project.org/package=salso>. R package version 0.2.5.
- Wade, S. and Ghahramani, Z. (2018). Bayesian cluster analysis: Point estimation and credible balls (with discussion). *Bayesian Analysis*, **13**(2), 559–626.