



Comments on: CliPS: How to identify cluster distributions in Bayesian mixture models

Nadja Klein¹ 

Received: 20 September 2026 / Accepted: 5 October 2026

© The Author(s) 2026

I congratulate the authors on this timely contribution. While my own research has not primarily focused on model-based clustering, it is a pleasure to discuss this paper by Gertraud Malsiner-Walli, Sylvia Frühwirth-Schnatter, and Bettina Grün, whose research I have admired and whom I greatly value as colleagues and members of our community. In Malsiner-Walli et al. (2026), Malsiner-Walli, Frühwirth-Schnatter, and Grün introduce the CliPS (Clustering in the Parameter Space) procedure for Bayesian mixture models. To my understanding, the central idea is to identify cluster distributions by clustering posterior draws of component-specific parameters, or suitable low-dimensional functionals thereof, in the point process representation (PPR) of a mixture model.

The motivation arises from two longstanding challenges in Bayesian mixture modeling:

1. Determining an appropriate number of clusters, and
2. addressing label switching in posterior simulation.

Rather than treating these as separate problems, CliPS addresses them simultaneously. The method proceeds by fitting a Bayesian mixture model, collecting posterior MCMC draws, projecting component-specific parameters into a suitable parameter space, and clustering these draws. A unique labeling is obtained whenever the resulting clustering yields a permutation of component labels within an iteration. Iterations for which this is not possible are discarded, and the frequency of such failures defines the non-permutation rate, which serves as a diagnostic measure of cluster separability and identifiability.

What I found appealing is that the procedure seems largely independent of the chosen clustering kernel. The authors demonstrate this through applications to mixtures of multivariate Gaussian distributions, latent class models for categorical data, and mixtures of Markov chains. They further show how the framework can be combined with finite mixtures with known numbers of components, mixture-of-finite-mixtures models, sparse finite mixtures, and Dirichlet process mixtures.

The paper ultimately advocates a shift from focusing solely on posterior partitions toward statistical inference on the underlying cluster-generating distributions. In this

✉ Nadja Klein
nadja.klein@kit.edu

¹ Scientific Computing Center, Karlsruhe Institute of Technology, Karlsruhe, Germany

sense, CliPS provides both a relabeling mechanism and a principled way to assess whether a fitted mixture model genuinely supports a cluster interpretation, thereby allowing one to characterize the population rather than only the data at hand.

I particularly appreciate the authors' reasoning that not every fitted mixture model should automatically be interpreted as evidence for meaningful clusters. The proposed non-permutation rate operationalizes this principle in a simple and transparent manner. As clustering methods become increasingly flexible and are applied in settings with limited domain knowledge, tools against overinterpreting spurious structure are needed.

From cluster distributions to conditional distributions. The CliPS framework shifts attention from clustering observations toward identifying and validating the distributions that generate them. In the models considered in the paper, each cluster corresponds to a fixed probability distribution indexed by parameters θ_k , and cluster identification is achieved through posterior separation of suitably chosen functionals $\phi(\theta_k)$. This naturally raises the question of how far this idea can be extended. In many modern statistical learning problems, the primary object of interest is no longer a fixed distribution but an entire family of distributions indexed by covariates. Distributional regression models (Klein 2024), conditional transformation models (Hothorn et al. 2014), and, more recently, generative approaches to distributional learning are designed precisely to characterize how the full conditional distribution

$$\mathbb{P}(Y \mid X = x)$$

changes with predictors. Similarly, in mixture-of-experts models (Jordan and Jacobs 1994), both cluster membership probabilities and cluster-specific models may depend on covariates. In such settings, latent groups are no longer characterized by static parameter vectors alone but rather by regression functions, conditional densities, or more generally distributional properties that vary over the predictor space. From this perspective, I wonder whether the central idea underlying CliPS may be generalized from identifying clusters of distributions to identifying clusters of conditional distributions. In a mixture-of-experts model (Jordan and Jacobs 1994),

$$p(y \mid x) = \sum_{k=1}^K \eta_k(x) f(y \mid x, \theta_k),$$

both cluster membership probabilities and cluster-specific models depend on covariates. Consequently, it is no longer immediately clear to me what the analogue of the PPR should be, nor what constitutes an appropriate notion of separation between latent components. Should one seek posterior separation of regression coefficients, function evaluations on a suitable reference grid, fitted conditional densities, predictive distributions, or learned low-dimensional representations thereof? Each choice induces a different interpretation of similarity between latent mechanisms and therefore a different notion of “conditional” clustering. More broadly, one may view CliPS as a framework for identifying low-dimensional representations that permit posterior separation of latent structures. Rather than asking whether two observations arise from

the same cluster distribution, one would ask whether they share the same conditional distributional properties. In this context, hierarchical Bayesian models may offer a promising route toward jointly learning cluster assignments and cluster-specific distributions while borrowing strength across related components. Exploring how such hierarchical structures interact with the posterior separation ideas underlying CliPS could provide an interesting direction for future research.

Concluding remark A final thought I would like to share concerns the choice of the functional $\varphi(\theta_k)$. One of the strengths of the paper is its emphasis on selecting scientifically meaningful summaries that distinguish clusters. At the same time, contemporary statistical learning increasingly relies on highly flexible models for which the defining parameters are no longer directly interpretable. In such settings, instead of clustering low-dimensional parameter summaries, one could imagine clustering latent representations, predictive distributions, regression functions, or other derived objects. Then, rather than hand-crafting informative summaries, one might seek latent representations that maximize posterior separation between cluster-specific distributions while preserving relevant predictive information. Such ideas could connect CliPS to the rapidly growing literature on representation learning and probabilistic embeddings (Bengio et al. 2013).

Funding Open Access funding enabled and organized by Projekt DEAL. The author acknowledges funding from the German Research Foundation (DFG) through the Emmy Noether Grant KL3037/1-1.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Bengio Y, Courville A, Vincent P (2013) Representation learning: A review and new perspectives. *IEEE Trans Pattern Anal Mach Intell* 35(8):1798–1828
- Hothorn T, Kneib T, Bühlmann P (2014) Conditional transformation models. *J Roy Stat Soc B* 76(1):3–27
- Jordan MI, Jacobs RA (1994) Hierarchical mixtures of experts and the EM algorithm. *Neural Comput* 6(2):181–214
- Klein N (2024) Distributional regression for data analysis. *Annual Review of Statistics and Its Application* 11:321–346
- Malsiner-Walli G, Frühwirth-Schnatter S, Grün B (2026) CliPS: How to identify cluster distributions in Bayesian mixture models. *TEST*, Invited paper

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.