Published in final edited form as:

Multivariate Behav Res. 2015 ; 50(1): 127. doi:10.1080/00273171.2014.988987.

Abstract: Advantages of a Bayesian Approach for Examining Class Structure in Finite Mixture Models

Sierra A. Bainter

Department of Psychology, University of North Carolina

In the social sciences, finite mixture models (FMMs) are being used increasingly to describe unobserved population heterogeneity in psychological constructs and processes. In most applications, maximum likelihood (ML) estimation is used, and the number of latent classes is determined by estimating solutions with different numbers of classes and selecting the optimal number based on various statistical criteria. However, no statistic works well in all cases for all models, and different criteria often give contradictory results. Further difficulties arise in estimation and identification of mixture models. The likelihood surface is highly multimodal as well as infinite at the boundary of the parameter space—problematic for ML estimation because the EM algorithm only finds local modes. When a FMM is estimated with too many classes or insufficient restrictions, identification is violated, and the asymptotic behavior of the ML estimator is unstable. Unidentified models can appear to be estimated normally using popular software packages, as parameters are fixed to boundary values. All of these challenges can lead to convergence problems and spurious solutions and threaten the reliability and validity of FMM applications using ML.

In this study I investigate how a Bayesian approach can help overcome current challenges with ML-based FMM applications in social science. I analytically compare Bayesian and ML approaches to determining the number of classes and present potential advantages of the fully Bayesian approach as well as challenges that arise in practice.

Whereas ML estimation determines a point estimate for each parameter in the model, a Bayesian approach estimates the distribution for each parameter. In a Bayesian framework, more information is available for determining the number of classes as well as key advantages for estimating FMMs. The number of latent classes can be directly estimated, allowing for examination of the posterior distribution of the number of classes. This fully Bayesian approach to mixture modeling uses a reversible jump MCMC algorithm developed by Richardson and Green (1997) to allow jumps between the parameter subspaces associated with different numbers of classes. Information about the distribution of the number of classes allows researchers to determine the weight of evidence for solutions with different numbers of latent classes and demonstrate the degree to which alternative interpretations of the data are tenable. Further, whereas the behavior of the MLE is sensitive to singularities, over-fitting, and violated identification, the posterior distribution using Bayesian estimation is stable and asymptotically tends to empty unneeded components.

Along with advantages to Bayesian estimation come a number of challenges. A high level of statistical expertise is required to ensure that appropriate priors are used and the MCMC algorithm is exploring the entire posterior correctly. Posterior inference of mixtures is also complicated by label switching.

Besides analytical comparison of Bayesian and ML approaches to examining class structure in FMMS, I empirically demonstrate the fully Bayesian approach to determine the number of classes using classic data and simulated examples. I conclude with recommendations for further quantitative developments and empirical applications of this promising framework for applications of FMMS in psychology.

Acknowledgments

I gratefully acknowledge support from NIDA grant F31DA035523. I would like to thank my SMEP sponsor, Patrick Curran, and David Dunson for their help.

REFERENCE

Richardson S, & Green P (1997). On Bayesian analysis of mixtures with an unknown number of components. *Journal of the Royal Statistical Society: Series B (Methodological)*, 59(4), 731–792.