# Investigating the Influence of Different Prior Choices for Mixtures of Gaussian Mixtures

Gertraud Malsiner-Walli[1], Sylvia Frühwirth-Schnatter[2], Bettina Grün[1]

*Second Bayesian Young Statisticians Meeting (BAYSM 2014)*
*Vienna, September 18–19, 2014*

[1] Johannes Kepler Universität, Department of Applied Statistics, Linz, Austria
<Gertraud.Malsiner_Walli,Bettina.Gruen>@jku.at

[2] Wirtschaftsuniversität WU, Institute for Statistics and Mathematics, Wien, Austria
Sylvia.Fruehwirth-Schnatter@wu.ac.at

**Abstract**

In the framework of model-based clustering, using a standard finite normal mixture model can lead to overestimating the number of data clusters if the shape of a data cluster deviates from that generated by a normal distribution, and more than one normal component is needed to fit the data cluster. Our approach to capture non-Gaussian data clusters consists in specifying a mixture model of cluster distributions where each cluster distribution is itself a mixture of normal subcomponents.

In the Bayesian framework, we investigate how priors on the mixture parameters can be specified in order to obtain both a sparse solution regarding the estimated number of data clusters and a good semiparametric fit of the data cluster densities using mixtures of normals. Our approach is investigated in two simulation studies and on various benchmark data sets.

**Keywords**: Cluster analysis; Finite mixture models; Dirichlet prior; Normal gamma prior; Number of clusters.

We employ a finite mixture model to cluster observations into homogeneous groups and to estimate the number of these groups. However, specifying the standard finite normal mixture model can lead to overestimation of the number of data clusters if the shape of a data cluster deviates from that generated by a normal distribution, and more than one normal component is needed to fit the data cluster.

Our approach consists in specifying a mixture model of cluster distributions where each cluster distribution is itself a mixture of normal subcomponents. In this way, we aim at capturing non-Gaussian data clusters by fitting a normal mixture to each cluster density in a semiparametric way. The purpose of our modeling is to identify the number of clusters and the cluster distributions in the case of possibly well-separated and dense data clusters, however, without making further assumptions on the cluster shapes.

In order to guide the estimation of this highly over-parameterized mixture of mixtures model, in the Bayesian framework we specify strongly informative priors for the mixture parameters to induce certain characteristics in our model we are interested in.

Through suitable prior specifications we want to induce sparsity in regard to the number of mixture clusters, abundance in regard to the number of subcomponents forming a cluster, shrinkage toward the cluster mean in regard to the subcomponent means and flexibility in regard to the subcomponent shapes.

For estimating the number of data clusters, we adapt the concept of the "sparse finite mixture model", see [2]. We specify a mixture of mixtures model where the number of specified clusters clearly overfits the number of true clusters. Then, following [3] we define a sparse weight prior for the clusters in order that during MCMC sampling superfluous clusters are emptied and use the most frequent number of non-empty clusters occurring during MCMC sampling as an estimator for the number of true clusters. Additionally, the result of [3] is used to ensure that within a cluster all subcomponents are filled with observations during MCMC sampling. Since our goal is to approximate an arbitrary cluster distribution through a mixture of normals, we want to avoid empty subcomponents. Therefore, we specify a redundant prior on the subcomponent weights to ensure that to all subcomponents observations are assigned during MCMC sampling and a good fit of the cluster distribution is achieved.

Regarding the modeling of the subcomponents forming the cluster distributions, we are facing a non-identifiability problem as it cannot be decided by the likelihood which subcomponents are responsible for which cluster. In the Bayesian framework, our strategy to decide which subcomponents should be regarded as belonging to the same cluster consists in specifying highly informative priors for the subcomponent parameters in such a way that subcomponents within a cluster are forced to have flat, strongly overlapping densities whereas different clusters are encouraged to be well-separated and wildly spread apart. We use the variance-covariance decomposition of a mixture of mixtures model to determine the prior parameters of the subcomponent means and subcomponent variance-covariance matrices.

Identification of the cluster distributions is obtained by clustering a functional of the subcomponent means in the point process representation using $k$-means clustering as suggested and explained in [1] and [2], respectively.

The performance of our approach is investigated in two simulation studies. Additionally, our approach is illustrated on benchmark data sets.

## References


[1] Frühwirth-Schnatter, S. (2011). "Label Switching Under Model Uncertainty." In K. Mengerson and C. Robert and D.M. Titterington (Eds.), *Mixtures: Estimation and Application* , pp. 213–239. Wiley.

[2] Malsiner-Walli, G., Frühwirth-Schnatter, S., Grün, B. (2014). "Model-based clustering based on sparse finite Gaussian mixtures". *Unpublished manuscript.*

[3] Rousseau, J., Mengersen, K. (2011). "Asymptotic Behaviour of the Posterior Distribution in Overfitted Mixture Models." *Journal of the Royal Statistical Society B*, **73**(5), 689–710.