Discussion of the John Geweke and Sylvia Frühwirth-Schnatter talks about Label Switching in mixtures

Gilles Celeux

Inria Saclay-Île-de-France

March 3 2010

John :"The multimodal posterior introduces no complications for MCMC posterior simulators."

- I agree that virtual permutation-augmented simulators solve the problem for well-separated component mixtures since the MCMC algorithms stay in a single modal region.
- And multimodality does not provoke MCMC slow convergence (but, no multimodal MCMC sequences indicate slow convergence).

- Otherwise reordering the parameters is an important and not easy task directly related to the labeling problem.
- For instance, labeling is an important issue for RJMCMC samplers.

Relabeling MCMC draws

Various methods have been proposed to reorder MCMC draws.

- A reference paper is Stephens (2000) which proposes a method among the most efficient one in practice. It is relying upon specifying a loss function and is minimising its posterior expected loss with a k-means like algorithm on the conditional probabilities than a point arises from one of the mixture components.
- An interesting extension is the probabilistic approach of Sperrin *et al.* (2009) which allows for incorporation of the uncertainty in the relabeling process.
- But all these methods require to identify $K!$ clusters.

The Point process representation of the MCMC draws

- It is a great interest of the Sylvia k-means clustering method in the point process representation of the MCMC draws to identify K clusters instead of $K!$ clusters.
- Since the vector parameters to be clustered could be of high dimension, Sylvia suggests to consider a subset of the component-specific parameters to obtain the clustering sequences.

I agree to think that this reduction of dimension is highly desirable.

- But, choosing such a subset could be rather influential and delicate...
- Some "data analyst" talents coud be required to make a sensible choice, while the Stephens labeling method avoids to work in the mixture parameter space...

The Role of the Prior on the Weight Distribution

In the Part II of the Sylvia's talk, I have been impressed by her demonstration of the great influence of the Prior on the mixing weights to select the number of mixture components.

- The example that she considered $(K^{\text{true}} = 2, \mu_1 = -4, \mu_2 = 3, \sigma_1^2 = \sigma_2^2 = .5, \eta_1 = .2)$ with $n = 1000$ is a mixture with quite well separated components.
- Any penalized information criteria (BIC, ICL and even AIC) will sharply choose $K = 2$ for this example.
- My feeling, after her bright demonstration, is that choosing "non informative" Priors for mixture models remains quite challenging (see also Aitkin 2002).

Is Bayesian model-based clustering possible ?

Model-based cluster analysis is one of the most important domain of application of mixture models. In this setting, n, d are large and K could be large as well.
To day, I see no reason to favor the Bayesian view point for model-based cluster analysis...

- It is clear that the Bayesian point of view can be fruitful for estimating mixtures for ill-posed problems (same sample sizes, overlapping components) (see Dias and Wedel 2004) but only for low dimensions and small number of mixture components.
- But, in most situations, the advantage of Bayesian inference over maximum likelihood inference is unclear for estimating mixture models.

References

Aitken, M. (2001), Likelihood and Bayesian analysis of mixtures, *Statistical Modelling*, 1, 287-304.

Dias, J. G. and Wedel, M. (2004), An empirical comparison of EM, SEM and MCMC performance for problematic Gaussian mixture, *Statistics and Computing*, 14, 323, 332.

Sperrin, M., Jaki, T. and Wit, E. (2010) Probabilistic relabeling strategies for the label switching problem in Bayesian mixture models, *Statistics and Computing* (to appear) available on line.

Stephens, M. (2000). Dealing with label switching in mixture models, *Journal of the Royal Statistical Society, B*, 62, 795–809.