Jeffreys Priors for Mixture Models

Clara Grazian[1], Christian P. Robert[2]

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

[1] CEREMADE Université Paris-Dauphine, Paris, France
Dipartimento di Scienze Statistiche, Sapienza Università di Roma, Rome, Italy
CREST, Paris, France
clara.grazian@ceremade.dauphine.fr

[2] CEREMADE Université Paris-Dauphine, Paris, France
University of Warwick, Coventry, UK
CREST, Paris, France
xian@ceremade.dauphine.fr

Abstract

Mixture models may be a useful and flexible tool to describe data with a complicated structure, for instance characterized by multimodality or asymmetry. In a Bayesian setting, it is a well established fact that one need to be careful in using improper prior distributions, since the posterior distribution may not be proper. This feature leads to problems in carry out an objective Bayesian approach. In this work an analysis of Jeffreys priors in the setting of finite mixture models will be presented.

Keywords: non-informative priors; Fisher information; Metropolis-Hastings.

1 Introduction

Mixture models are defined as follows:

$$f\left(\mathbf{x} \mid \boldsymbol{\theta}, \mathbf{w}\right) = \sum_{i=1}^{K} w_i f_i\left(\mathbf{x} \mid \boldsymbol{\theta_i}\right) \tag{1}$$

where $w_i \in (0,1)$ and $\sum_{i=1}^{K} w_i = 1$, K is the number of components and $\boldsymbol{\theta}_i$ is the vector of parameters of the ith-component.

In this setting, the maximum likelihood estimation may be problematic, even in the simple case of Gaussian mixture models, as shown in [1]. For a comprehensive review see [6].

In a Bayesian setting, [2] and [4] suggest to be careful when using improper priors because it is always possible that the sample does not include observations for one or more components, so the data are not informative about the model. One can refers to [7] for a proposal prior distribution in the setting of general mixture models.

In this work we want to analyze the posterior distribution for the parameters of a mixture model with a finite number of components when the Jeffreys prior (see [3]) is applied.

In particular, we want to assess the convergence of the posterior distribution when using the Jeffreys prior for the parameters of a Gaussian mixture model, even when the prior for some parameters is improper conditionally on the others.

2 Conditional Jeffreys priors

Consider a 2-component mixture model. The Jeffreys prior for the distribution's weigths, by considering the components known, is a function of just one parameter because of the constraint on the sum of the weigths:

$$\pi^J(w_1) \propto \sqrt{\int_{\mathcal{X}} \frac{\left(f(\mathbf{x};\boldsymbol{\theta_1}) - f(\mathbf{x};\boldsymbol{\theta_2})\right)^2}{w_1 f(\mathbf{x};\boldsymbol{\theta_1}) + (1-w_1) f(\mathbf{x};\boldsymbol{\theta_1})} d\mathbf{x}} \tag{2}$$

$$\leq \sqrt{\int_{\mathcal{X}} \frac{\left(f(\mathbf{x};\boldsymbol{\theta_1}) - f(\mathbf{x};\boldsymbol{\theta_2})\right)^2}{w_1 f(\mathbf{x};\boldsymbol{\theta_1})} d\mathbf{x}} \tag{3}$$

$$= \sqrt{\frac{1}{w_1} c_1} \tag{4}$$

where c_1 is a positive constant. The resulting prior may be easily generalized to the case of K components. It can be easily shown that this prior is proper and convex (in the general case of K components, it can be shown the prior is still proper and the marginals are still convex). The form of the prior depends on the type of components.

Now consider a 2-component Gaussian mixture model. The conditional Jeffreys prior for the mean parameters depends on the following derivatives:

$$\frac{\partial^2 \log f}{\partial \mu_i^2} = \frac{w_i}{\sqrt{2\pi}\sigma_i} \left\{ \frac{\left[\exp\left(-\frac{1}{2}\left(\frac{x-\mu_i}{\sigma_i}\right)^2\right)\left(\frac{x-\mu_i}{\sigma_i^2}\right)^2 - \exp\left(-\frac{1}{2}\left(\frac{x-\mu_i}{\sigma_i}\right)^2\right)\frac{1}{\sigma_i^2}\right]}{w_1\mathcal{N}(\mu_1;\sigma_1) + (1-w_1)\mathcal{N}(\mu_1,\sigma_1)} \right\} - \left\{ \frac{w_i}{\sqrt{2\pi}\sigma_i} \frac{\exp\left(-\frac{1}{2}\left(\frac{x-\mu_i}{\sigma_i}\right)\right)^2 \frac{x-\mu_i}{\sigma_i^2}}{w_1\mathcal{N}(\mu_1;\sigma_1) + (1-w_1)\mathcal{N}(\mu_1,\sigma_1)} \right\}^2 \tag{5}$$

and

$$\frac{\partial^2 \log f}{\partial \mu_1 \partial \mu_2} = -\frac{w_1}{\sqrt{2\pi}\sigma_1}\frac{w_2}{\sqrt{2\pi}\sigma_2} \left\{ \frac{\exp\left(-\frac{1}{2}\left(\frac{x-\mu_1}{\sigma_1}\right)^2\right)\frac{x-\mu_1}{\sigma_1^2}\exp\left(-\frac{1}{2}\left(\frac{x-\mu_2}{\sigma_2}\right)^2\right)\frac{x-\mu_2}{\sigma_2^2}}{w_1\mathcal{N}(\mu_1;\sigma_1) + (1-w_1)\mathcal{N}(\mu_1,\sigma_1)} \right\} \tag{6}$$

With a simple change of variable, it is easy to see that each element depends only on the difference between the means, therefore it is flat with respect to each μ_i. The generalization to K components comes directly.

A very similar argument applies for the conditional prior distribution of the standard deviations, which is again improper.

The results for a 2-component Gaussian mixture model are shown in Figure 1.

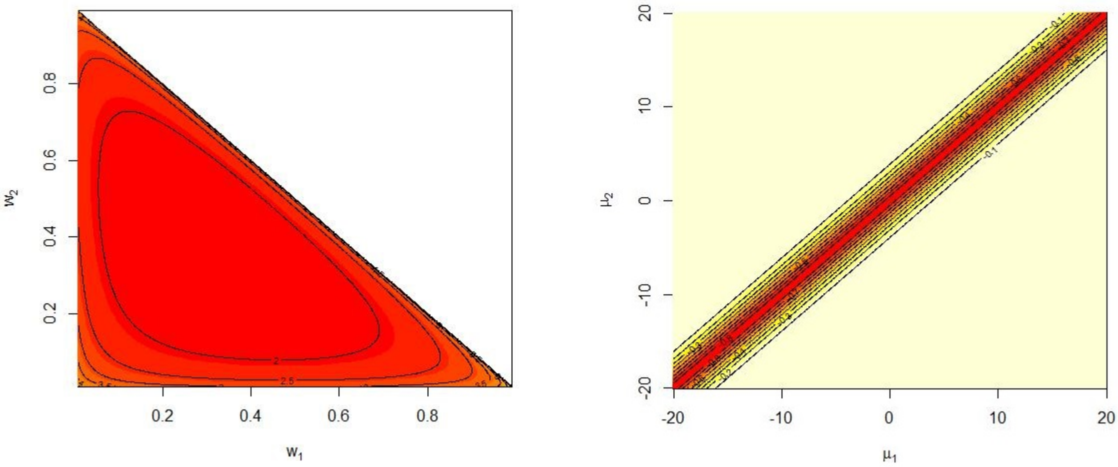

Figure 1: Conditional Jeffreys prior for the weigths of a 2-component Gaussian mixture model

3 The posterior distribution for a mixture model when Jeffreys prior is used

It is a well-established fact in the literature that using independent improper priors lead to improper posterior distributions. One may use improper priors in mixture models by introducing some form of dependence between the components, as shown in [5]. Actually the Jeffreys prior does that by considering the Fisher information matrix.

The results obtained via simulations for the posterior distribution of the parameters of a mixture model, shown in Figure 2, are encouraging: all the chains always converge. Figure 2 shows the results for a 3-component Gaussian mixture model, neverthless other scenarios have been considered, with different types of components and the results are always comparable in terms of convergence.

4 Discussion

There are two important drawbacks when using the Jeffreys prior for mixture model.

First, the prior depends on integrals which have to be approximated.

Second, the approximation of the determinant of the Fisher information matrix could be hard to compute in terms of computational time. A way to use parallel computation to reduce the running time has been considered, even if it is beyond the scope of this paper.

The possibility of defining a noninformative prior in the case of mixture models could be useful. In particular, further research may be focused on the case of a non-fixed number of components and to understand the impact of such a prior in model choice.

References

[1] Basford, K.E., McLachlan, G.J. (1985). "Likelihood estimation with normal mixture models." *Applied Statistics*, **34**(3), 282–289.

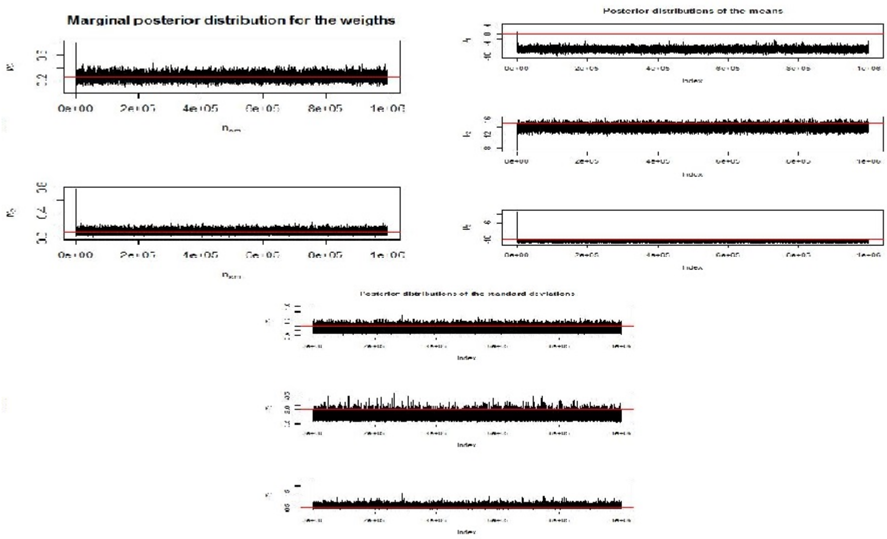

Figure 2: Marginal posterior distributions (chains obtained via Metropolis-Hastings) for the parameters of a 3-components Gaussian mixture model (weigths on the left top, means on the rigth top and standard deviations on the bottom.

[2] Frhwirth-Schnatter, S. (2006). *Finite Mixture and Markov Switching Models: Modeling and Applications to Random Processes.* 1st edition. Springer, New York.

[3] Jeffreys, H. (1961). *The theory of probability.* 3rd edition. Oxford University Press, Oxford.

[4] McLachlan, G.J., Peel, D. (2000). *Finite mixture models.* 1st edition. Wiley, Newark.

[5] Mengersen, K. and Robert, C. (1996). "Testing for mixtures: A Bayesian entropic approach (with discussion)." In Berger, J., Bernardo, J., Dawid, A., Lindley, D., and Smith, A., editors, *Bayesian Statistics 5*, pages 255276, Oxford University Press, Oxford.

[6] Titterington, D.M., Smith, A.F.M., Makov, U.E. (1985). *Statistical analysis of finite mixture distributions.* Vol. 7. Wiley, Newark.

[7] Wasserman, L. (2000). "Asymptotic inference for mixture models using data-dependent priors." *Journal of the Royal Statistical Society, Series B*, **62**(1), 159–180.