Jeffreys Priors for Mixture Models

Distribuzioni a priori di Jeffreys per i modelli mistura

Clara Grazian and Christian P. Robert

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.
Abstract *I modelli mistura sono uno strumento utile e flessibile per descrivere dati dalla struttura complicata, ad esempio multimodale o asimmetrica. In ambito Bayesiano, è un fatto noto in letteratura che sia necessario essere attenti con l'utilizzo di distribuzioni a priori improprie, dal momento che la distribuzione a posteriori potrebbe non essere propria. Purtroppo, questa caratteristica rende difficile un approccio Bayesiano oggettivo. In questo lavoro, verrà presentata un'analisi dei risultati ottenuti utilizzando distribuzioni a priori (non informative) di Jeffreys.*

Key words: Jeffreys prior, Mixture models, Objective Bayes

Clara Grazian
CEREMADE, Université Paris Dauphine, Place du Marchal de Lattre de Tassigny, 75016 Paris, France
Dipartimento di Scienze Statistiche, Sapienza Università di Roma, Rome, Italy
CREST, Paris, France
email:clara.grazian@ceremade.dauphine.fr

Christian P. Robert
CEREMADE, Université Paris Dauphine, Place du Marchal de Lattre de Tassigny, 75016 Paris, France
CREST, Paris, France
University of Warwick, Coventry, UK
email:xian@ceremade.dauphine.fr

1 Introduction

Using a mixture of distributions to model a random variable provides great flexibility. Mixture models have an obvious application when the sampling population consists of several sub-populations, nevertheless a mixture model provides also a mean to model different zones of support of the true distribution.

Mixture models are defined as follows:

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

where $w_i \in (0,1)$ and $\sum_{i=1}^{K} w_i = 1$, K is the number of components and θ_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 mixture of normal distributions, as shown in [1]. In general, maximum likelihood estimation in this setting is obtained through the EM algorithm. For a comprehensive review see [4].

In a Bayesian setting, [3] shows that it is not possible to have noninformative priors and obtain proper posterior distributions because it's 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 [5] 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 a Jeffreys prior is applied.

We remind that the Jeffreys prior (see [2]) is defined as:

$$\pi^J(\theta) \propto \sqrt{\det(\mathfrak{I}(\theta))} \tag{2}$$

where $\mathfrak{I}(\theta)$ is the expected Fisher information matrix. This prior distribution as some appealing properties, in particular it is invariant by reparameterization, nevertheless it may be proper or improper depending on the model.

This prior has been applied to different parameters of finite mixture models. We have always considered different scenarios, with normal and not normal components, and the results obtained are always comparable.

We will analyze the results when only the weights are unknown in Section 2 and the results for the case where only the means of a mixture of normal distributions are unknown in Section 3. In Section 4 there will be a discussion, with some suggestions for future work.

2 Jeffreys prior for the weights of a 3-component mixture model

In the general case of a 3-component mixture model, the Jeffreys prior for the weights of the components is a function of two parameters, since there is a constraint on the sum of the weights. The expected Fisher information matrix is then

$$\mathfrak{I}(w_1,w_2)=\begin{bmatrix} \int_{\mathfrak{X}} \frac{[f_1(x;\theta)-f_3(x;\theta)]^2}{\sum_{i=1}^3 w_i f_i(\mathbf{x}|\theta_i)}\,dx & \int_{\mathfrak{X}} \frac{[f_1(x;\theta)-f_3(x;\theta)][f_2(x;\theta)-f_3(x;\theta)]}{\sum_{i=1}^3 w_i f_i(\mathbf{x}|\theta_i)}\,dx \\ \int_{\mathfrak{X}} \frac{[f_1(x;\theta)-f_3(x;\theta)][f_2(x;\theta)-f_3(x;\theta)]}{\sum_{i=1}^3 w_i f_i(\mathbf{x}|\theta_i)}\,dx & \int_{\mathfrak{X}} \frac{[f_2(x;\theta)-f_3(x;\theta)]^2}{\sum_{i=1}^3 w_i f_i(\mathbf{x}|\theta_i)}\,dx \end{bmatrix} \tag{3}$$

and the Jeffreys prior is

$$\pi(w_1,w_2) \propto \sqrt{\det(\mathfrak{I}(w_1,w_2))} \tag{4}$$

and $w_3 = 1 - w_1 - w_2$.

This prior distribution may be only numerically managed.

2.1 Prior distribution

The marginal Jeffreys prior approximations for w_1, w_2 and w_3 are described in Figure 1 which shows that this prior gives a higher weight to values close to 0. The simulations seem to suggest that this prior distribution is proper, since the chains have neither stopped because of problems of numerical integration nor got stuck at values at the limit of the parameter space.

2.2 The posterior distribution

Figure 2 shows the histograms approximating the marginal posterior distributions of the weight w_1 of a mixture of normal distributions with well-separated components (the histograms for w_2 are completely analogous). The chains always seem to reach convergence and the histograms show that the approximated posterior distribution is concentrated around the true values and closely follow the numerical (not probabilistic) approximation.

3 Jeffreys prior for the means of a 3-component mixture model

Consider a mixture of 3 normal distributions, with known weights and variances.

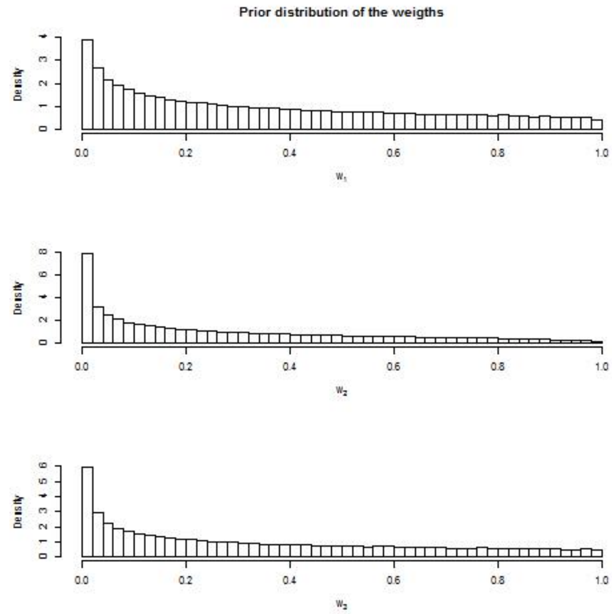

Fig. 1 Histograms of the approximations of the Jeffreys prior distribution for the weigths of a normal mixture model $f(\mathbf{y} \mid w_1, w_2, w_3) = w_1 \mathcal{N}(-10,1) + w_2 \mathcal{N}(0,5) + w_3 \mathcal{N}(15,0.5)$ obtained with a Metropolis-Hastings random walk algorithm with 10^6 simulations.

The Jeffreys prior for the means depends on the expected Fisher information matrix, whose elements are

$$\frac{\partial^2 \log f}{\partial \mu_i^2} = \frac{w_i}{\sqrt{2\pi}\sigma_i^2} \left\{ \frac{\left[\exp\left(-\frac{1}{2}\left(\frac{x-\mu_i}{\sigma_i}\right)^2\right)\left(\frac{x-\mu_i}{\sigma_i}\right)^2 - \exp\left(-\frac{1}{2}\left(\frac{x-\mu_i}{\sigma_i}\right)^2\right)\right] \sum_{l=1}^3 w_l \mathcal{N}\left(\mu_l, \sigma_l^2\right)}{\left(\sum_{l=1}^3 w_l \mathcal{N}\left(\mu_l, \sigma_l^2\right)\right)^2} \right\}$$
$$- \left(\frac{w_i}{\sqrt{2\pi}\sigma_i^2}\right)^2 \left\{ \frac{\exp\left(-\left(\frac{x-\mu_i}{\sigma_i}\right)^2\right)(x-\mu_i)^2}{\left(\sum_{l=1}^3 w_l \mathcal{N}\left(\mu_l, \sigma_l^2\right)\right)^2} \right\} \quad (5)$$

and

$$\frac{\partial^2 \log f}{\partial \mu_i \partial \mu_j} = -\frac{w_i}{\sqrt{2\pi}\sigma_i^2} \frac{w_j}{\sqrt{2\pi}\sigma_j^2} \left\{ \frac{\exp\left(-\frac{1}{2}\left(\frac{x-\mu_i}{\sigma_i}\right)^2\right)(x-\mu_i)\exp\left(-\frac{1}{2}\left(\frac{x-\mu_j}{\sigma_j}\right)^2\right)(x-\mu_j)}{\left(\sum_{l=1}^3 w_l \mathcal{N}\left(\mu_l, \sigma_i^2\right)\right)^2} \right\} \quad (6)$$

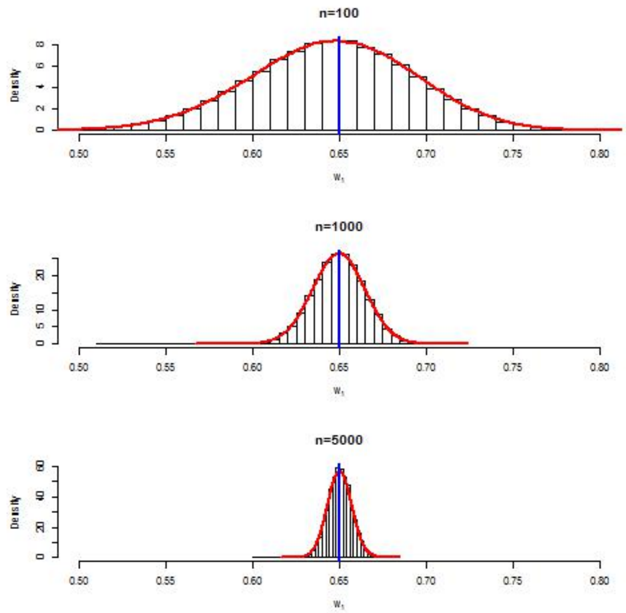

Fig. 2 Histograms of the approximations of the marginal posterior distribution of w_1 in a normal mixture model $f(\mathbf{y} \mid w_1, w_2, w_3) = w_1 \mathcal{N}(-10, 1) + w_2 \mathcal{N}(0, 5) + w_3 \mathcal{N}(15, 0.5)$ with a Jeffreys prior for the weigths, the numerical approximation of the posterior distribution obtained with the integrate function in R (red line) and the true value (blue line) for different sample sizes.

3.1 The posterior distribution

In this case the prior distribution seems to be improper. See [5] for a proof that in the case of 2-component normal mixture models with only one unknown mean the Jeffreys prior on the unknown mean leads to an improper posterior distribution.

Neverthless, the simulation study in Figure 3 shows that the chains approximating the marginal posterior distributions of the means reach convergence around the true values, even if problems of label swtiching appear in this context. Figure 3 shows the boxplots of the approximated posterior means for 10 different simulation studies.

4 Discussion

An objective Bayesian analysis is advisible in some contexts, in particular when the experimenter's influence has to be reduced. Unfortunately non-informative prior distributions are often improper and this could lead to improper posterior distribu-

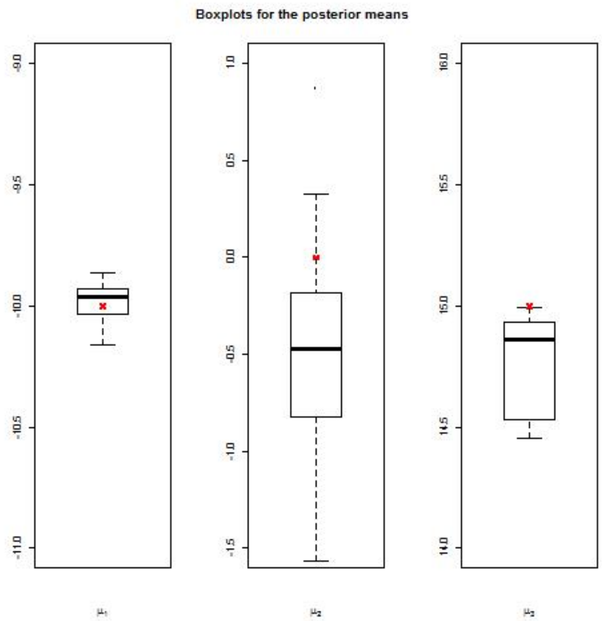

Fig. 3 Boxplots for the approximated posterior means for the mean paramters of a normal mixture model $f(\mathbf{y} \mid w_1, w_2, w_3) = w_1 \mathscr{N}(-10,1) + w_2 \mathscr{N}(0,5) + w_3 \mathscr{N}(15,0.5)$
with a Jeffreys prior. The boxplots represents 10 simulation study with a sample size equal to 100 and 10^6 simulations. The true values are in red.

tion in some settings, for instance this often happens in the case of mixture models. We have presented some encouraging examples of the use of the Jeffreys prior with these models. The situation where only the standard deviations are unknown or the one where both the means and the weights are unknown have been already tested, with good results in terms of properness of the posterior distribution. These first results may suggest some applications of the Jeffreys prior in the case of model selection when comparing mixture models with different types or different numbers of components.

References

1. Basford, K.E., McLachlan, G.J.: Likelihood estimation with normal mixture models. Appl. Statist. **34**(3), 282–289 (1985)
2. Jeffreys, H.: *The theory of probability*. Oxford University Press, 1961.
3. McLachlan, G.J., Peel, D.: *Finite mixture models*. Wiley, 2004.
4. Titterington, D.M., Smith, A.F.M., Makov, U.E.: *Statistical analysis of finite mixture distributions*. Vol. 7. Wiley, 1985.
5. Wasserman, L.: Asymptotic inference for mixture models using data-dependent priors. J. R. Statist. Soc. B **62**(1), 159–180 (2000)