Spatially dependent mixture models with a random number of components

Modelli mistura spazio-dipendenti con un numero aleatorio di componenti

Matteo Gianella, Mario Beraha and Alessandra Guglielmi

Abstract In finite mixture models, the choice of the number of components is crucial. From the Bayesian perspective, the correct approach is assuming such number unknown and random. In this work, we set such a prior on a finite mixture model for areal data, assuming that, within each area, data are iid from area-specific densities and we introduce spatial dependence in their joint distribution. We propose a transdimensional sampler via reversible jump which exploits optimal proposals that improve chain mixing and sampler efficiency. The approach is validated on a simulated scenario.

Abstract *Nei modelli mistura finiti, la scelta del numero di componenti è fondamentale. Da un punto di vista bayesiano, l'approccio corretto è assumere tale numero incognito e aleatorio. In questo lavoro, definiamo tale prior in un modello mistura finito per dati areali, assumendo che, in ogni area, i dati siano iid da una densità ad essa specifica ed introduciamo una dipendenza spaziale nella loro distribuzione congiunta. Proponiamo un sampler transdimensionale per mezzo di reversible jump che sfrutta l'introduzione di proposte ottimali che migliorano il mixing della catena e l'efficienza del sampler. L'approccio è validato su dati simulati.*

Key words: Bayesian model selection, reversible jump MCMC, spatial mixtures

1 Introduction

Mixture models provide a natural framework for model-based clustering as well as for approximating densities that are not suitably modeled by standard parametric families. For a recent review, see [2]. Even though mixtures are often used under

Matteo Gianella[1], Mario Beraha[1, 2] and Alessandra Guglielmi[1]
[1]Department of Mathematics, Politecnico di Milano, Milano, Italy
[2]Department of Computer Science, Università degli Studi di Bologna, Bologna, Italy
e-mail: matteo1.gianella@mail.polimi.it, {mario.beraha, alessandra.guglielmi}@polimi.it

the assumption of exchangeable samples from a unique unknown distribution, there are cases in which such models have been adopted to model data that show spatial dependence. A novel work in this field is [1], where the problem of modeling areal data is considered. In particular, [1] assumes a finite mixture with a fixed number of components H for each area and introduce spatial dependence via a suitable prior on the weights of the mixtures, i.e. the *logistic multivariate CAR prior*.

A common issue with finite mixtures is the choice of an appropriate number of mixture components, particularly important when the analysis requires an interpretation of the clusters induces by the mixture. Two strategies are commonly adopted to deal with such problem. The first one consists in fixing H to a reasonably large upper bound and assume a *sparse* prior on the weights, so that, asymptotically, only $k < H$ components result allocated (see [8]). However, such *sparse* priors have been studied only for classical mixture models, i.e., when data are exchangeable from a single unknown distribution.

The second strategy is straightforward under the Bayesian approach and is the one we assume here: the number of components H is unknown and considered random. Despite its conceptual simplicity, this latter approach is characterized by computational difficulties since a *transdimensional* Markov Chain Monte Carlo (MCMC) algorithm should be designed for posterior inference. Examples of such transdimensionals MCMC algorithms include the reversible jump MCMC in [7] and the MCMC based on birth-and-death processes in [9]. More recently, by exploiting the notion of exchangeable partition probability function (EPPF) [5] has proposed a "marginal" MCMC sampling scheme.

In this work, we extend the the spatial mixture model defined in [1] by assuming a prior on the number H of components and we propose a transdimensional sampler via a reversible jump MCMC algorithm. This sampling strategy is forced by the model itself, since theoretical results about *sparse* priors in [8] are not available in this more complex setting, and EPPF of our model is not known in analytical form so far.

The rest of this article is organized as follows. In Section 2 we introduce the model. Then, Section 3 describes and motivates the reversible jump move step. Finally, Section 4 presents a simulation study to check the correctness and efficiency of our algorithm.

2 The Bayesian model

In this section we introduce a model that extends the Bayesian mixtures in [1] by assuming a prior on the number of components.

Likelihood for areal data. Consider data $\mathbf{y}_1, \dots, \mathbf{y}_I$, where $\mathbf{y}_i = (y_{i1}, \dots, y_{iN_i})^T$ are exchangeable observations from the areal unit i, for $i = 1, \dots, I$. Assume that a neighbouring structure G between the I different areal units is known. We assume G as a $I \times I$ matrix, where its entries G_{ij} indicates whether i and j are neighbours ($G_{ij} = 1$) or not ($G_{ij} = 0$). For each i in $1, \dots, I$, the conditional distribution of our data is

specified as follows:

$$y_{ij} \mid \mathbf{w}_i,\ \boldsymbol{\tau},\ H \overset{\text{iid}}{\sim} \sum_{h=1}^{H} w_{ih}\mathcal{N}(\cdot \mid \tau_h) \quad j = 1,\ldots,N_i, \tag{1}$$

where $\mathbf{w}_i = (w_{i1},\ldots,w_{iH})^T$ is a H-dimensional vector in the simplex S^H, i.e., $w_{ih} \geq 0$ and $\sum_h w_{ih} = 1$ and $\mathcal{N}(\cdot \mid \tau_h)$ denotes the the Gaussian density with parameters $\tau_h = (\mu_h, \sigma_h^2)$. Observe how in (1) the (μ_h, σ_h^2)'s are shared across all the spatial locations. Thus, this model allows to introduce dependency between mixtures associated to different areas only through the prior for the weights $(\mathbf{w}_1,\ldots,\mathbf{w}_I)$.

Logistic MCAR prior. The prior introduced to induce spatial dependence among mixtures from different areas is defined through a multivariate CAR distribution on a transformation of the weights. In particular, the weights are transformed via the additive log ratio map, defined from S^H to $\mathbb{R}^{H-1}$ and such that:

$$\widetilde{w}_h = \log(w_h/w_H) \quad h = 1,\ldots,H-1.$$

Once the transformed weights have been defined, then a multivariate CAR prior is imposed on $\widetilde{\mathbf{w}} = (\widetilde{\mathbf{w}}_1,\ldots,\widetilde{\mathbf{w}}_I)$ as:

$$\widetilde{\mathbf{w}}_i \mid \widetilde{\mathbf{w}}_{-i},\ \boldsymbol{\Sigma},\ \rho,\ H \sim \mathcal{N}_{H-1}\left(\rho \textstyle\sum_j G_{ij}\widetilde{\mathbf{w}}_j, \boldsymbol{\Sigma}\right) \quad i = 1,\ldots,I. \tag{2}$$

This model defines a unique joint distribution for $\widetilde{\mathbf{w}}$ when $\rho \in (-1,1)$; see [3].

Prior on H. Given the likelihood for the data and the prior for the weights, the Bayesian model is then extended adding priors on the hyperparameters. In our context, we model the atoms τ_h conditioning to H, independently from a $Normal - InvGamma(\mu_0, a, b, \lambda)$, the matrix $\boldsymbol{\Sigma}$ is assumed to be diagonal, i.e. $\boldsymbol{\Sigma} = \sigma^2\boldsymbol{I}$ with $\sigma^2 \sim InvGamma(\alpha,\beta)$, ρ has a uniform prior in $(0,1)$. Finally, we assume a shifted Poisson distribution on the number of components H, i.e., $H - 1 \sim Poi(\lambda)$.

3 Reversible Jump computation via recursive auxiliary priors

The reversible jump MCMC sampler [4] provides a general framework for transdimensional simulation schemes. It can be viewed as an extension of the Metropolis-Hastings algorithm. As it happens in standard Metropolis-Hastings, given the current state of the chain $\boldsymbol{\theta} = (H, \boldsymbol{\theta}_H)$ (where we make explicit the "dimension" H), the next state $\boldsymbol{\theta}' = (H', \boldsymbol{\theta}_{H'})$ is (i) sampled from a proposal distribution $q(\boldsymbol{\theta}, \boldsymbol{\theta}')$, and (ii) accepted with probability $\alpha(\boldsymbol{\theta}, \boldsymbol{\theta}')$ equal to

$$\alpha(\boldsymbol{\theta}, \boldsymbol{\theta}') = \min\left\{1, \frac{\pi(\boldsymbol{\theta}' \mid \mathbf{y})q(\boldsymbol{\theta}', \boldsymbol{\theta})}{\pi(\boldsymbol{\theta} \mid \mathbf{y})q(\boldsymbol{\theta}, \boldsymbol{\theta}')}\right\}.$$

Usually, the proposal distribution is defined in two steps. If $\boldsymbol{\theta}_H \in \mathbb{R}^{n_H}$ and $\boldsymbol{\theta}' \in \mathbb{R}^{n_{H'}}$, with $n_{H'} > n_H$ and $d = n_{H'} - n_H$, first a random vector $\mathbf{u} \in \mathbb{R}^d$ is sampled from a distribution $q_d(\mathbf{u})$ and then $\boldsymbol{\theta}_{H'}$ is defined as $g_{H\to H'}(\boldsymbol{\theta}_H, \mathbf{u})$ for a suitable mapping function $g_{H\to H'}$. Since both the proposal distribution $q_d(\mathbf{u})$ and the mapping function $g_{h\to h'}$ are arbitrary, the definition of a suitable reversible jump move is usually a difficult task. The approach we follow to design a reversible jump move for the model described in Section 2 is based on [6], where the author defines optimal auxiliary priors and proposals for generic nested models indexed by an integer $H \in \mathbb{N}^+$ with unknown parameter $\boldsymbol{\theta}_H$ and prior of the form $\pi(\boldsymbol{\theta}_H \mid H)\pi(H)$. Let us introduce key ideas in [6], that we will adapt to our context.

Since the models are nested, the unknown parameters are nested as well, i.e., if $H' > H$, the first H elements of $\boldsymbol{\theta}_{H'}$ correspond to vector $\boldsymbol{\theta}_H$, the one indexed by H. Given the current state $(\boldsymbol{\theta}_H, H)$, consider moving to $(\boldsymbol{\theta}_{H'}, H')$ with $H' = H+1$ (the case $H' = H-1$ is identical but with H and H' swapped). The joint distribution for $(\mathbf{y}, \boldsymbol{\theta}_{H'}, H)$ is defined as:

$$f(\mathbf{y}, \boldsymbol{\theta}_{H'}, H) = \tilde{\pi}_H([\boldsymbol{\theta}_\infty]_{H'} \mid \boldsymbol{\theta}_H, \mathbf{y}) f(\mathbf{y} \mid H, \boldsymbol{\theta}_H)\pi(\boldsymbol{\theta}_H)\pi(H),$$

where $[\boldsymbol{\theta}_\infty]_{H'}$ represents the H'-th element of $\boldsymbol{\theta}_\infty$ and $\tilde{\pi}_H([\boldsymbol{\theta}_\infty]_{H+1} \mid \boldsymbol{\theta}_H, \mathbf{y})$ needs to be defined. Choosing such quantity as the conditional posterior

$$\pi([\boldsymbol{\theta}_\infty]_{H+1} \mid \mathbf{y}, H+1, \boldsymbol{\theta}_H) \propto f(\mathbf{y} \mid H+1, \boldsymbol{\theta}_{H+1})\pi(\boldsymbol{\theta}_{H+1} \mid H+1) \tag{3}$$

guarantees optimal conditions in terms of overall chain mixing and minimization of the estimated variance. Nonetheless, this optimal posterior is not known a priori, so we need to estimate it. We see how this is possible in our specific case, i.e. for the spatial mixture model.

First of all, at a fixed dimension H, the unknown parameter vector $\boldsymbol{\theta}_H$ is $vec(\widetilde{\mathbf{w}}, \boldsymbol{\tau})$, where $\widetilde{\mathbf{w}} = vec(\{\widetilde{\mathbf{w}}_i\}_{i=1:I})$ and vec indicates the vectorization of the given quantity. Thus, in case the algorithm propose to add a new component, it is required to sample $[\boldsymbol{\theta}_\infty]_{H+1} = (w_{1H+1}, \ldots, w_{IH+1}, \mu_{H+1}, \sigma^2_{H+1})$. The great novelty of this approach is the fact that the posterior distribution $\pi([\boldsymbol{\theta}_\infty]_{H+1} \mid \mathbf{y}, H+1, \boldsymbol{\theta}_H)$ is a direct proposal distribution for the new added component. In this way, we side step the artificial construction of proposal distributions and mapping functions, whose definition is totally arbitrary and does not ensure any particular property in terms of sampling performance. The acceptance rate $\alpha[(H, \boldsymbol{\theta}_H), (H', \boldsymbol{\theta}_{H'})]$ is given by $\min\{1, A\}$, with A equal to

$$\begin{aligned} A \quad &= \frac{f(\mathbf{y} \mid alr^{-1}(\tilde{\boldsymbol{w}})_{i=1:I}, \boldsymbol{\tau}, H')\pi(\boldsymbol{\tau} \mid H')\pi(\tilde{\boldsymbol{w}}_{i=1:I} \mid \rho, \boldsymbol{\Sigma}, H')\pi(H')}{f(\mathbf{y} \mid alr^{-1}(\tilde{\boldsymbol{w}})_{i=1:I}, \boldsymbol{\tau}, H)\pi(\boldsymbol{\tau} \mid H)\pi(\tilde{\boldsymbol{w}}_{i=1:I} \mid \rho, \boldsymbol{\Sigma}, H)\pi(H)} \\ &\times \left(\frac{\chi_{\{H'=H+1\}}}{\tilde{\pi}([\boldsymbol{\theta}_\infty]_{H'} \mid \boldsymbol{\theta}_H, \mathbf{y})} + \chi_{\{H'=H-1\}} \tilde{\pi}([\boldsymbol{\theta}_\infty]_H \mid \boldsymbol{\theta}_{H'}, \mathbf{y}) \right). \end{aligned} \tag{4}$$

Note that the marginal priors for $\boldsymbol{\Sigma}$ and for ρ do not appear in (4) since they does not depend on the number of components H. Finally, since the proposal posterior distribution is not known, in case of addition of a new component we approximate it with a multivariate Gaussian centered in $\boldsymbol{\theta}^* = \arg\max \pi([\boldsymbol{\theta}_\infty]_{H+1} \mid \mathbf{y}, H+1, \boldsymbol{\theta}_H)$

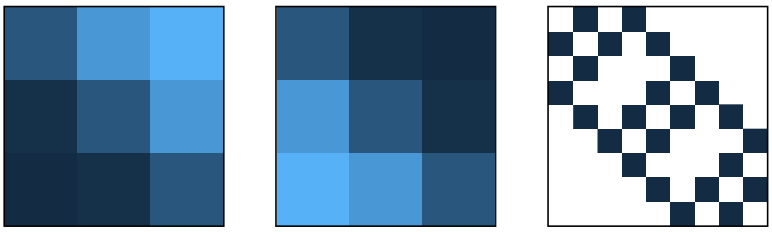

Fig. 1 Simulation from spatially dependent weights, from left to right: w_{i1}, w_{i2} and G.

(see (3)) and variance equal to the Hessian of $\pi([\boldsymbol{\theta}_\infty]_{H+1} \mid \mathbf{y}, H+1, \boldsymbol{\theta}_H)$ evaluated in $\boldsymbol{\theta}^*$. On the other hand, when we reduce the dimension of the state, the parameters of the approximated optimal posterior are computed using as maximizer the selected component to drop.

4 Simulation study

We present a simple simulation study to assess the performance of our reversible jump MCMC algoritm. Code for posterior simulation has been implemented in `C++` and is available as an `R` package at https://github.com/TeoGiane/SPMIX . We run the MCMC chain for a total of $10,000$ iterations, discarding the first $5,000$ as burn-in and thinning the chain every two iterations, so that the final sample size is $2,500$.

We consider 9 areas numbered in lexicographical order as in Figure 1 (left). In the i-th area, we draw a sample from

$$y_{ij} \overset{\text{iid}}{\sim} w_{i1}\mathcal{N}(-5,1) + w_{i2}\mathcal{N}(0,1) + w_{i3}\mathcal{N}(5,1) \quad j = 1,\dots,25.$$

The weights are computed as $alr^{-1}(\widetilde{\mathbf{w}})$ and $\widetilde{\mathbf{w}}$ is defined as

$$\tilde{w}_{i1} = 3(x_i - \bar{x}) + 3(y_i - \bar{y}) \qquad \tilde{w}_{i2} = -3(x_i - \bar{x}) - 3(y_i - \bar{y}), \tag{5}$$

being (x_i, y_i) and $(\bar{x}, \bar{y})$ the coordinates of the center of area i and of the grid center. In this context, G is defined so that two areas are close if they share a common edge, as in Figure 1 (right).

Note that the number of samples in each location is extremely small, so that the sharing of information between neighboring mixtures is essential. Moreover, observe that (5) induces a different kind of spatial dependency from (2).

Figure 2 (left) shows the posterior distribution of the number of components. The remaining panels in Figure 2 shows the estimated and true densities in two areas. The sampler is extremely effective in retrieving the correct number of components and in estimating densities. It also provides remarkable results even with few observations in each area.

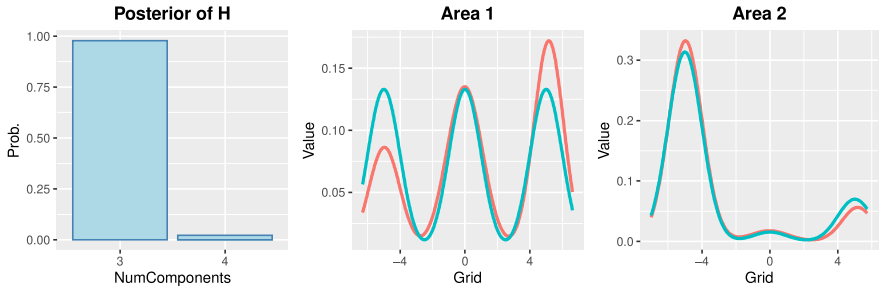

Fig. 2 Some results obtained for the considered scenario. To the left, the posterior traceplot of H and the comparison between theoretical (light blue) and estimated (pink) density for two areas.

5 Conclusions

In this paper, we have extended the model introduced in [1] by adding a prior on the number of components. From the computational point of view, the model becomes very challenging and we have proposed a reversible jump algorithm. We have selected a suitable reversible jump move to directly build the optimal proposal and avoid the artificial construction of mapping functions and proposal distributions. We have set up an efficient sampling scheme able of sampling from such model to show the goodness of fit of our model in retrieving the right number of components and capturing the spatial dependency among different areas.

References

1. Beraha, M., Pegoraro, M., Peli, R., Guglielmi, A.: Spatially dependent mixture models via the Logistic Multivariate CAR prior. arXiv preprint arXiv:2007.14961 (2020)
2. Fruhwirth-Schnatter, S., Celeux, G., Robert, C.P.: Handbook of mixture analysis. CRC press (2019)
3. Gelfand, A.E., Vounatsou, P.: Proper multivariate conditional autoregressive models for spatial data analysis. Biostatistics **4**(1), 11–15 (2003)
4. Green, P.J.: Reversible jump Markov chain Monte Carlo computation and Bayesian model determination. Biometrika **82**(4), 711–732 (1995)
5. Miller, J.W., Harrison, M.T.: Mixture models with a prior on the number of components. Journal of the American Statistical Association **113**(521), 340–356 (2018)
6. Norets, A.: Optimal auxiliary priors and reversible jump proposals for a class of variable dimension models. Econometric Theory p. 1–33 (2020)
7. Richardson, S., Green, P.J.: On Bayesian analysis of mixtures with an unknown number of components. Journal of the Royal Statistical Society: Series B (Statistical Methodology) **59**(4), 731–792 (1997)
8. Rousseau, J., Mengersen, K.: Asymptotic behaviour of the posterior distribution in overfitted mixture models. Journal of the Royal Statistical Society: Series B (Statistical Methodology) **73**(5), 689–710 (2011)
9. Stephens, M.: Bayesian analysis of mixture models with an unknown number of components - an alternative to reversible jump methods. Annals of statistics pp. 40–74 (2000)