Relabelling MCMC Algorithms in Bayesian Mixture Learning

R. Bardenet, Balázs Kégl, Gersende Fort

Relabelling MCMC Algorithms in Bayesian Mixture Learning

Rémi Bardenet
LAL, LRI (Computer Science Lab)
Paris-Sud XI University
91405 Orsay (France)
bardenet@lri.fr

Balázs Kégl
LAL, LRI (Computer Science Lab)
Paris-Sud XI University & CNRS
91405 Orsay (France)
balazs.kegl@gmail.com

Gersende Fort
LTCI
Telecom ParisTech & CNRS
46, rue Barrault, 75013 Paris (France)
gfort@telecom-paristech.fr

Markov Chain Monte Carlo (MCMC) sampling has demonstrated to be a powerful and versatile method for Bayesian inference in a variety of models, e.g. large treewidth graphical models in Machine Learning or intricate models in experimental Physics. MCMC methods simulate a Markov chain $(\theta^{(t)})$ that approximates independent draws from a previously defined target distribution π. A major problem arises however when the target π presents symmetries, for example in the case of a mixture model, where the likelihood of the observations x given the parameters $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$

$$p(x|\boldsymbol{\alpha},\boldsymbol{\beta}) = \sum_{m=1}^{M} \alpha_m f(x;\beta_m)$$

is invariant to relabellings of the variables, i.e. $p(x|\nu(\boldsymbol{\alpha}),\nu(\boldsymbol{\beta})) = p(x|\boldsymbol{\alpha},\boldsymbol{\beta})$ for any permutation ν of $\{1,...,M\}$ acting on the components of $\boldsymbol{\alpha}$ and the columns $\beta_m, m = 1..M$ of $\boldsymbol{\beta}$. If the prior $p(\boldsymbol{\alpha},\boldsymbol{\beta})$ also shows this permutation invariance, then the posterior $\pi(\boldsymbol{\alpha},\boldsymbol{\beta}) \propto p(\boldsymbol{\alpha},\boldsymbol{\beta}|x)p(\boldsymbol{\alpha},\boldsymbol{\beta})$ inherits it. As a consequence, the same permutation invariance is observed on the chain $(\theta^{(t)})$, in which components can switch from a time step to the following, resulting in poor marginal estimates; this phenomenon has been called *label switching* [4]. A related motivating problem exhibiting this pathologic behavior is counting events occurring independently at random times; a commonly used technique to learn the distribution of these events is Reversible Jump MCMC for mixture modelling [2], for which label switching is known to be an important issue [6].

The first solution consists in modifying the prior in order to make it select a single permutation of the variables, introducing an *identifiability constraint* [6]. This solution is however known to sometimes cause artificial biases by not respecting the topology of the posterior [5]. An effort has then been made to adapt to the estimated posterior surface with the design of relabelling algorithms [8][5] that process the MCMC sample *after the completion of the simulation run* by finding for each individual realization $\theta^{(t)}, t = 1..T$ a permutation $\nu_t(\theta^{(t)})$ of its components that minimizes

a posterior-based criterion depending on the whole chain history. [1] proposed an online version of the relabelling procedure, in which the simulation of each $\theta^{(t)}$ is followed by a permutation $\nu_t(\theta^{(t)})$ of its components, the permutation being chosen to minimize a user-defined criterion that depends only on the *past history* of the chain up to time t. We think the major advantage of this online approach is to be compatible with more efficient sampling schemes, like adaptive MCMC methods [3]. In both batch and online relabelling algorithms, inference is then carried out by means of ergodic averages using the *relabelled sample*. However, the introduction, be it online or batch, of permutation steps in the MCMC procedure modifies the distribution of the chain $(\theta^{(t)})$. Consequently, the target distribution is modified. Though empirical intuition that relabelling induces the learning of an appropriate identifiability constraint has been exposed [4], the existence of a target distribution and its relation with the original π is still an open problem, which has confined yet relabelling algorithms only to a successful heuristics.

We propose a new online relabelling procedure based on an adaptive MCMC algorithm [3][7] that tunes its design parameters on the fly to improve its efficiency. We prove the convergence of our algorithm and identify the link between the new target measure and the original distribution of interest π. We also study different mechanisms for the selection of the relabelling at each time step that are inspired by usual clustering techniques, and their influence on the convergence of the global MCMC algorithm. Finally, we demonstrate our algorithm on a problem inspired by a real counting issue encountered in experimental particle physics.

References

[1] G. Celeux. Bayesian inference for mixtures: The label-switching problem. In R. Payne and P. Green, editors, *COMPSTAT 98*. Physica-Verlag, 1998.

[2] P. J. Green. Reversible jump Markov chain Monte Carlo computation and Bayesian model determination. *Biometrika*, 82(4):711–732, 1995.

[3] H. Haario, E. Saksman, and J. Tamminen. An adaptive Metropolis algorithm. *Bernoulli*, 7:223–242, 2001.

[4] A. Jasra, C. C. Holmes, and D. A. Stephens. Markov Chain Monte Carlo methods and the label switching problem in Bayesian mixture modelling. *Statistical Science*, 20(1):50–67, 2005.

[5] J.M. Marin, K. Mengersen, and C.P. Robert. Bayesian modelling and inference on mixtures of distributions. *Handbook of Statisics*, 25, 2004.

[6] S. Richardson and P. J. Green. On Bayesian analysis of mixtures with an unknown number of components. *Journal of the Royal Statistical Society*, 59(4):731–792, 1997.

[7] E. Saksman and M. Vihola. On the ergodicity of the adaptive Metropolis algorithm on unbounded domains. *Annals of Applied Probability*, 20:2178–2203, 2010.

[8] M. Stephens. Dealing with label switching in mixture models. *Journal of the Royal Statistical Society, Series B*, 62:795–809, 2000.

Topic: Learning Algorithms
Preference: Oral/Poster