Bayesian Analysis (2007) **2**, Number 3, pp. 479–482

Comment on Article by Jain and Neal

C.P. Robert*

From a stylistic point of view, I think this paper reads very much like a sequel to the important paper Jain and Neal (2004) and therefore it is not exactly self-contained since the main bulk of the paper is a commentary of the program provided in Section 4.2. Instead of the current version, I would thus have preferred a truly self-contained version with a more user-friendly introduction, for instance when reading and re-reading Sections 3 and 4.1...[1]

The central point of the paper is to extend Jain and Neal (2004) so that the lack of complete conjugacy of the prior does not prevent the algorithm from being run. Indeed, in Jain and Neal (2004), the model parameters are completely hidden in that the likelihood and the prior only depend on the cluster index vector $\mathbf{c}$, which means working in a finite set. The difficulty with priors G_0 that do not lead to closed form marginals is that the parameters must take part in the simulation process. The idea at the core of the current paper is to take advantage of the conditional conjugacy, i.e. the fact that the prior on a given parameter is still conjugate and thus manageable, conditional on all the other parameters, so that a Gibbs sampling version can be implemented.

At this stage, I understand the rationale of the partial conjugacy for the Metropolis-Hastings ratio to be computed (Section 4.1) but I wonder how difficult it would be to extend the idea to any type of prior distribution. I also note that at both split and merge stages the algorithm simulates new values of the parameter from the *prior* distribution, rather than from a more adapted distribution. This is as generic as it can be, but simulating from vague priors usually slows down algorithms and it is of course impossible for improper priors. It thus seems to me that the factor t directing the number of intermediate Gibbs (or Metropolis-Hastings) iterations in Step 3 must be influential in the overall behaviour of the algorithm and that large values of t may be necessary to overcome the dependence on the starting value.

More generally, I also wonder why a more global tempering strategy would not fare better than the local split-merge proposals used in the paper. For illustration purposes, I implemented below the regular Gibbs sampler in the [BetaBinomial] Example 1 of Jain and Neal (2004) and compared it with a naïve tempered version where the tempered likelihood L_τ is made of a product of $\tau \geq 1$ (sub)likelihoods based on a partition of the observations in τ random clusters, τ being itself uniform on $\{1, \ldots, n/2\}$. (The advantages of using this form of tempering are (a) that the same Gibbs sampler can be used for the sublikelihoods and (b) that the normalising constant of the tempered version is still available, as opposed to the choice of a power of the likelihood. The acceptance probability at the end of the tempered moves is then function of the likelihood ratio $L(\theta|x)/L_\tau(\theta|x)$ and can be directly computed.) As shown on Figure 1 *(bottom)*,

*CREST and CEREMADE, Uni. Paris Dauphine, France, mailto:xian@ceremade.dauphine.fr

[1]This may explain why the following reads more like an eloped referee's report than like a true discussion!

