

**Review of:**  
**“An Efficient Interpolation Technique for Jump Proposals in Reversible-Jump Markov Chain Monte Carlo Calculations” by Farr, Mandel and Stevens**

This paper addresses an interesting and important issue in Bayesian model selection: Developing efficient jump proposals in reversible-jump Markov-chain Monte Carlo (RJMCMC) algorithms to transition between the models under consideration (high rejection rates for inter-model proposals represent the limiting factor in many applications). The approach proposed here—to exploit sampling of the posterior probability density (PPD) in the individual model spaces, characterized as a  $k$ D tree structure—seems like a very promising idea which could justify publication. However, there are a number of significant issues that must be addressed before publication is recommended.

1. The history of this work is unclear and concerning. The first paragraph of page 4 says “We have successfully applied this RJMCMC technique to a 10-way model selection among alternative mass distribution models for black-hole X-ray binaries (Farr et al. 2010)”. This reference is again cited on page 13 as a successful application of the method developed in the present manuscript. But this raises an obvious question: Has the method described in the present manuscript already been published by the authors? If so, publication of the present manuscript does not seem justified since the it considers only a trivial toy problem which adds nothing of significance. Further, the reference section indicates that Farr et al. 2010 is in the submission stage to the *Astrophysical Journal*. What’s the story here—is this a typo, or a badly out-of-date reference, or has the paper actually been in-review for five years? Is the method described in the present manuscript the same as in the 2010 paper (sounds like it is)? This must be sorted out, and unless the present manuscript contains new concepts compared to earlier work, publication is not justified.
2. The explanation of the inter-model proposal and acceptance criterion in Section 3c is inadequate. While well-known material on Bayesian analysis (Section 3a) and MCMC (Section 3b) is presented at length with equations, the new RJMCMC approach is skimmed over in a single sentence (no equations) as “We perform MCMC in the super-model parameter space just like a regular MCMC; we propose jumps to different parameters within a model (intra-model jumps) and jumps to a different model with different parameters (inter-model jumps)” (page 6). In fact, using the manuscript’s notation, the Metropolis-Hastings-Green acceptance probability for a jump to a proposed model  $M_i$  with parameters  $\vec{\theta}_i^p$  from a current model  $M_j$  with parameters  $\vec{\theta}_j$  is given by

$$p_{\text{accept}} = \min \left( 1, \frac{p(\vec{\theta}_i^p | d, M_i)}{p(\vec{\theta}_j | d, M_j)} \frac{Q(\vec{\theta}_j \rightarrow \vec{\theta}_i^p)}{Q(\vec{\theta}_i^p \rightarrow \vec{\theta}_j)} |\mathbf{J}| \right),$$

where  $|\mathbf{J}|$  is the determinant of the Jacobian matrix for the diffeomorphism from  $(j, M_j)$  to  $(i, M_i)$ . This equation should be given and explained in the paper. RJMCMC algorithms are often formulated such that  $|\mathbf{J}| = 1$ ; if that is the case here it must be explained (it is not generally true of all proposals, and neglecting to mention this could give wrong ideas about RJMCMC). Further, the inter-model proposal distribution should be given explicitly (an equation), either in Section 3c or Section 4 where  $k$ D trees are described. This may be relatively simple: if the proposal probability density is just the  $k$ D-tree PPD approximation for the proposed model (independent of the current model), then I’m guessing that  $Q(\vec{\theta}_j \rightarrow \vec{\theta}_i^p) = 1/(N_i V_i)$  where  $N_i$  is the number of samples for the  $i$ th model space and  $V_i$  is the volume of the  $k$ D-tree box containing  $\vec{\theta}_i^p$ . But either way, the inter-model proposal must be clearly defined in the manuscript.

3. The examples given in the manuscript seem simplified to the point of triviality and/or are poorly explained. The only example given of model selection (the potentially important contribution of this paper) only involves sampling between 1D Gaussian and Cauchy distributions (two unknown parameters each). While this does illustrate the method (barely), it is of no practical relevance and is so much simpler than most realistic model-selection problems that it is not clear how efficiency and performance would scale, or how useful the method could be in practice. I assume the authors are researching real-world model-selection problems of practical interest (like alternative mass distribution models for black-hole X-ray binaries, mentioned on page 4). Including a meaningful example such as this would greatly improve the paper and increase its significance. The other two examples represent only single-model sampling and are of much less interest, but even so they should be properly explained. In the examples in Figures 3 and 4, what are the problems, what are the parameters, and what do the results mean? This current presentation is not satisfying to the reader.
4. Claims are made of “dramatically improved convergence” (page 3), that “convergence is generally rapid” (page 4), and “the interpolation method described above can dramatically improve the runtime of an RJMCMC” (page 10). While I believe the method developed here could substantially improve efficiency in some cases, this is not actually demonstrated in the paper. The trivial example in Figure 2 appears to increase inter-model acceptance by a factor of  $\sim 8$  or so, but no results on RJMCMC sampling convergence are given. The example considered in Figure 4 claims to have improved convergence by a factor of two (not dramatic), but there is no discussion of what definition is applied for convergence—and as noted on page 6 (third paragraph), defining convergence for MCMC sampling is a tricky business. Even applying a reasonable working definition of convergence, given the random nature of MCMC sampling, averages over several MCMC runs are often required to meaningfully compare convergence rates—was that done here? While the claims of “dramatic” improvements in conver-

gence may be true in some cases, they don't appear to be properly established in the paper.

5. Minor points:

- (a) Notation: While notation differs between disciplines, I am used to boldface (rather than over-arrows) being used to represent vectors in published work—the authors could consider this. Further, why is the set of parameters written as a vector ( $\vec{\theta}$ ), but the set of data not written as a vector ( $\vec{d}$ )?
- (b) Page 5, first line after Section 2b heading: Should be  $j = 1, \dots$  not  $i = 1, \dots$ .
- (c) Page 5, third line after Section 2b heading: Should be “set of parameters” not “set of parameter”.
- (d) Figure 1 would be clearer if parameter axes were added indicating the origin, with axes scales in terms of standard deviations in  $x$  and  $y$ .
- (e) The  $x$  axis of Figure 2 is labeled  $N_{\text{boxing}}$  while it should be  $N_{\text{box}}$ .
- (f) Figure 3 caption:  $kd$  should be  $kD$ .
- (g) Equation (5.1):  $P$  should be  $Q$  (this is the proposal distribution). Also,  $V$  might be better written as  $V_{\text{box}}$  to indicate that it is the volume of the current box, like  $N_{\text{box}}$  indicates the number of samples in the current box ( $V$  is not a constant but varies with boxes).
- (h) Page 11: Inconsistent use of math or Roman font for some subscripts; e.g.,  $N_{\text{box}}$  and  $N_{\text{box}}$ , and  $N_{\text{crit}}$ . When the subscript is a label rather than an index, Roman font is preferred.
- (i) Page 11, after Eq. (5.1): Please give an equation for  $V$ , the volume of the intersection of the two (potentially high-dimensional) boxes.
- (j) Page 12, first paragraph discusses violations of detailed balance arising from using the sampling history to define proposals. Such methods are well known, including the requirements which must be met to avoid biased sampling (so-called diminishing adaptations). References (such as Brooks et al., 2011 *Handbook of Markov chain Monte Carlo*, New York: Springer) should be included here to be clear this is not a new idea.
- (k) Figure 4's caption states “The acceptance probability rate asymptotes to the steady-state solution once sufficient samples are accumulated for the  $kD$  tree to allow the sample density or be accurately interpolated.” This asymptotic convergence appears to occur by about step  $2.0 \times 10^6$ ; hence, all samples collected prior to this point should be discarded as burn-in or the sampling is potentially biased. This should be mentioned explicitly in the paper.

- (1) Figure 4: Why denote the  $x$  axis divisions from 0 to 25 000 and then include “( $\times 100$ )” in the axis label? Why not denote the divisions from 0 to 2.5 and include “( $\times 10^6$ )” in the label—this is the common use of exponential notation.

## List of Figures