Bayesian Analysis (2010) **5**, Number 1, pp. 41–44

Rejoinder

Charles R. Hogg*, Joseph B. Kadane†, Jong Soo Lee‡ and Sara A. Majetich§

We thank the discussants for their valuable commentary on our paper. A common theme was the value in making modern statistical methods more widely known in the physical sciences. Below, we comment on specific issues raised by the discussants.

1 Algorithm Efficiency

Skilling and Sivia rightly point out the poor efficiency of our MCMC algorithm. We've continued to attack the speed problem, and are pleased to report significant progress.

The main culprit was the proposal strategy for bounded parameters. We wanted distributions which vanish outside the allowed region, since this greatly simplifies correcting for unequal proposal probabilities. The scaled Beta fits nicely; moreover, there are simple analytic expressions for the parameter values which give a target mean and variance,

$$\begin{aligned}\alpha &= \left[\frac{\mu(1-\mu)}{\sigma^2} - 1\right]\mu \\ \beta &= \left[\frac{\mu(1-\mu)}{\sigma^2} - 1\right](1-\mu),\end{aligned} \tag{1}$$

as long as $\alpha > 0$ and $\beta > 0$.[1] However, these formulae are non-unimodal when $\alpha < 1$ or $\beta < 1$, exhibiting proposal probabilities which diverge near the boundaries. This caused a computational instability, since the most extreme values were the most often proposed. We circumvented this by requiring both α and β to be greater than one, yielding a unimodal distribution.

More seriously, our original approach is suboptimal even without this numerical instability. The support of the distribution covered the entire domain, leading to huge proposed jumps when near the boundaries. We eliminated this problem with a new paradigm for bounded proposals. Pick a "jump" distance J: the limit is either J, or the boundary, whichever is closer. Explicitly, we generate candidates X' from current value X as follows:

$$\frac{X' - (X - JA)}{J(A+B)} \sim \text{Beta}(\alpha = A+1, \beta = B+1) \tag{2}$$

*Department of Physics, Carnegie Mellon University, Pittsburgh, PA, mailto:chogg@andrew.cmu.edu

†Department of Statistics, Carnegie Mellon University, Pittsburgh, PA, mailto:kadane@stat.cmu.edu

‡Department of Statistics, Carnegie Mellon University, Pittsburgh, PA and Department of FREC, University of Delaware, Newark, DE, mailto:jslee@udel.edu

§Department of Physics, Carnegie Mellon University, Pittsburgh, PA,, mailto:sara@cmu.edu

[1] These conditions can always be met by requesting a smaller target variance.

 DOI:10.1214/10-BA501REJ

with A the maximum allowed negative jump, and B the maximum allowed positive jump, both as fractions of J. This new approach eliminates extreme proposals which waste computation, while still respecting the boundaries of parameters.

We found further room for improvement with strongly correlated parameters. Originally, we proposed changes to parameters one at a time, and made the accept/reject decision before generating the next parameter's proposals. We now generate simultaneous proposals for correlated parameters, empirically reducing the number of steps to equilibrium by a factor of 2. Moreover, the steps themselves are considerably quicker, since multiple accept/reject decisions are condensed into one. We expect our algorithms could converge still more quickly, either with proposals skewed towards the direction of correlation, or by reparameterizing the correlated parameters.

In addition to algorithmic improvements, we reduced the computation time by recoding our MCMC loop in `C++` instead of `R`. We output the chain to a textfile, then use `boa` to analyze the results just as in the previous case.

2 Bayes factors

It is worth noting that unlike posterior distributions, likelihood functions are not probability distributions over the parameter space. In order to interpret them, some prior distribution must be assumed.

We think, however, that Bayes factors are overemphasized. In the very special case in which there are only two possible "states of the world", Bayes factors are sufficient. However in the typical case in which there are many possible states of the world, Bayes factors are sufficient only when the decision-maker's loss has only two values: 0 if the decision is correct and 1 otherwise. Thus the use of Bayes factors involves the rather unsatisfactory idea that if I decide that a parameter $\theta = 0$ and I am wrong, it doesn't matter how wrong I am: $\theta = e^{17}$ and $\theta = e^{-17}$ have the same loss for me. Few situations in any science satisfy this criterion (see Kadane and Dickey (1980)).

However, we take the issue of robustness seriously. Prior robustness is rather simple in the MCMC context. An MCMC sample can be reweighted by the ratio of the prior of interest to the prior used in the MCMC, whether the prior incorporates independence or dependence among parameters.

3 Modeling, Calibration, Design

We also thank Nick Hengartner for his constructive thoughts about how to use hierarchical Bayesian models flexibly. We agree that further collaboration between physicists and statisticians could push the field forward. Unlike in Rutherford's day, machine time is precious, so efficient use of the data produced seems well worthwhile.

References

Kadane, J. B. and Dickey, J. M. (1980). "Bayesian Decision Theory and the Simplification of Models." In Kmenta, J. and Ramsey, J. (eds.), *Evaluation of Econometric Models*, NBER Chapters, 245–268. National Bureau of Economic Research, Inc. URL http://ideas.repec.org/h/nbr/nberch/11704.html 42