Discussion of 'The art of data augmentation'

by David A. van Dyk and Xiao-Li Meng

Antonietta Mira*
Università degli Studi dell'Insubria, Italy.

Peter J. Green†
University of Bristol, UK.

November 2000

Introduction

As long-time enthusiasts for auxiliary variables methods in Bayesian MCMC, we are glad to have the opportunity to discuss this interesting paper.

The authors have a reputation for picturesque and metaphorical titles, and this paper is no exception. We were intrigued by the 'artistic' aspirations here, but remain in some doubt as to whether this should be read in the sense of 'non-scientific' or 'aesthetically pleasing'! One of our objectives in preparing our discussion was to see whether more closely relating the ideas here to existing auxiliary variables methods might both diminish the former characteristic and enhance the latter. This was only partially successful, and we will conclude with some questions for the authors, which we hope they will address in their rejoinder: is our intention doomed, and can they do better?

Graphical models

Existing MCMC methods are set in the context of a full probability model for a (usually structured) collection of variables; in Bayesian MCMC we are interested in the distribution of some of these (the unknowns) given the remainder (the data). Auxiliary variables methods augment this collection of variables and elaborate the corresponding probability model, so that the original model is obtained either as the *marginal* for the original variables, or as their *conditional*, given the values of the auxiliaries. In either case, a simulation of the augmented model can be used in obvious ways to extract information about the original one.

In fact, the term 'auxiliary variables' is commonly reserved for the marginal case, while the conditional one is usually called 'simulated tempering'. Both originate in statistical physics: see Edwards and Sokal (1988) and Marinari and Parisi (1992) respectively. Note that these are not treatments of special problems but quite general formulations.

* *Department*, Università degli Studi dell'Insubria, *Street*, Varese, Italy.
Email: `anto@aim.unipv.it`.

† Department of Mathematics, University of Bristol, Bristol BS8 1TW, UK.
Email: `P.J.Green@bristol.ac.uk`.

In our perspective, the 'data' are simply the (original) variables that are to be conditioned upon, and are not otherwise logically distinguished from the rest of the variables. Thus the term 'data augmentation' seems something of an historical accident: the additional variables need have no direct interpretation as any kind of 'data'.

With this in mind, we examined the conditional and marginal forms of data augmentation discussed in this paper, especially in the light of the examples, hoping that the conditional independence structure of the augmented models would prove revealing about the intrinsic ideas here.

We find it helpful to regard the observed data Y as a sub-vector of what the authors call augmented data, rather than as a deterministic function $\mathcal{M}$ of it. The authors themselves do this in their Section 3. Thus we write Y instead of Y_{obs} and (Y, Z) instead of Y_{aug}; they would call Z 'missing data'. Our target is the posterior $p(\theta|Y)$ derived from a given joint model $p(\theta, Y)$: what is the structure of the augmented models $p^\star(\theta, \alpha, Z, Y)$ that the authors are studying?

This seems elusive. The basis for conditional augmentation, equation (2.1), says simply that $p^\star(Y|\theta, \alpha) = p(Y|\theta)$. Superficially, this looks like the definition of simulated tempering, but note that this must hold for *all* α. It in fact asserts that Y and α are conditionally independent given θ, as well as requiring that the augmented model agrees with the original for the distribution of Y given θ. In marginal augmentation, (3.1) tells us only that the augmented model should again satisfy $p^\star(Y|\theta) = p(Y|\theta)$.

Thus both forms of data augmentation discussed here are examples of auxiliary variables, and no particular structure relating (θ, α, Z, Y) is assumed other than that just mentioned.

This lack of structure is reinforced by examination of the examples, where the joint distribution $p^\star(\theta, \alpha, Z, Y)$ variously factors as $p^\star(\alpha)p(\theta)p^\star(Z|\alpha)p^\star(Y|f(Z, \alpha), \theta)$ (section 6), $p^\star(\alpha)p(\theta)p^\star(Z|\alpha, \theta)p(Y|\theta)$ (section 7) and $p^\star(\alpha)p(\theta)p^\star(Z|\alpha, \theta)p^\star(Y|Z, g(\alpha, \theta))$ (section 8), where f and g are particular functions.

Our concern here about lack of structure is not because cherished prejudices about the universal utility of conditional independence ideas are being threatened, but because it does seem to highlight what we see as a major problem with the ideas advocated in the paper. While the authors' results do assist us in *tuning* a given augmentation scheme to give optimal performance, they do not seem to say much about how to devise the scheme in the first place. The lack of commonality among the examples does seem to leave us with little on which to build a scheme for an entirely new model. It is an extreme contrast with the general ability of MCMC to provide usually workable algorithms on the basis of standard recipes, not requiring subtle analysis of the model.

Eigenvalues

[Here could go your observations about interleaving and positive eigenvalues.]

Criteria for conditional DA

[Here could go your observations about relating criteria 2 and 3.]

References

Edwards, R. G. and Sokal, A. D. (1988). Generalization of the Fortuin-Kastelyn-Swendsen-Wang representation and Monte Carlo algorithm. *Physical Review*, D, **38**, 2009–2012.

Marinari, E. and Parisi, G. (1992). Simulated tempering: a new Monte Carlo scheme. *Europhysics Letters*, **19**, 451–458.