Comment

G. O. Roberts, S. K. Sahu and W. R. Gilks

We congratulate the authors on a magnificent paper, providing a nicely paced introduction to Markov chain Monte Carlo and its applications, together with several new ideas. In particular the class of pairwise difference priors is bound to have a substantial impact on future applied work. Other ideas given less prominence in the paper are also valuable, for example, the construction of simultaneous credible regions based on MCMC output. There are several issues which we wish to comment on in detail.

MCMC ON IMPROPER POSTERIORS

We would like to consider the issues raised by possible impropriety of posterior distributions and the use of MCMC on such target posteriors. For instance, consider the logistic regression model in Section 4. The model specification in (4.1) together with the postulated priors make the model unidentifiable. So the resulting posterior distribution is improper. If the posterior is improper no notion of convergence in distribution is meaningful for the associated MCMC. However, we may ask if the associated sequence of draws of a lower-dimensional vector converges in distribution. When are we allowed to use samples from this nonconvergent MCMC to infer about our "identifiable" parameters of interest? To date there is no literature addressing all of these concerns in total generality, but in the context of generalized and normal linear models some of these issues have been addressed in Sahu and Gelfand (1994).

Improper Posteriors from Generalized Linear Models

Consider the usual linear model $\mathbf{Y} = X\boldsymbol{\beta} + \boldsymbol{\varepsilon}$, where $\mathbf{Y}$ is $n \times 1$, X is $n \times p$ $(n > p)$, $\boldsymbol{\beta}$ is $p \times 1$ and $\boldsymbol{\varepsilon} \sim N(\mathbf{0}, \sigma^2 I)$ with σ^2 known. Let X have column rank $r < p$. Assuming a flat prior for $\boldsymbol{\beta}$, the posterior for $\boldsymbol{\beta}$ is improper. However, the complete conditional distributions $\pi(\beta_l \mid \beta_j,\ j \neq l,\ \mathbf{Y})$ are all proper, so the Gibbs sampler can be implemented. Note also that $X\boldsymbol{\beta}$ has a singular normal posterior distribution given by

$$\pi(X\boldsymbol{\beta}|\mathbf{Y}) = N(X(X^TX)^{-}X^T\mathbf{Y}, \sigma^2X(X^TX)^{-}X^T). \tag{1}$$

Now we can choose a full-rank matrix R, $p - r \times p$, whose rows are linearly independent of the rows of X, that is, $R\boldsymbol{\beta}$ is a maximal set of nonestimables. Suppose we take as a prior $\pi(R\boldsymbol{\beta}) = N(\mathbf{0}, V)$, where V is a positive-definite matrix of appropriate order, and retain a flat prior for $X\boldsymbol{\beta}$. Then we can show that $\boldsymbol{\beta}$ has a proper posterior distribution given by

$$\pi(\boldsymbol{\beta}|\mathbf{y}) = N\Big((\sigma^{-2}X^TX + R^TV^{-1}R)^{-1}X^T\mathbf{y}/\sigma^2, (\sigma^{-2}X^TX + R^TV^{-1}R)^{-1}\Big).$$

It is easy to check that $\pi(X\boldsymbol{\beta}|\mathbf{Y})$ is exactly the same singular normal distribution as in (1). Further, the posterior of $R\boldsymbol{\beta}$ is the same as the prior, and $R\boldsymbol{\beta}$ is *a posteriori* independent of $X\boldsymbol{\beta}$. So any proper prior for $R\boldsymbol{\beta}$ does not alter the posterior for $X\boldsymbol{\beta}$ but makes the posterior distribution for $\boldsymbol{\beta}$ proper. If the rank of R is less than $p - r$, we do not have a proper posterior for $\boldsymbol{\beta}$. Thus the propriety of the posterior depends upon the propriety of the nonestimables $R\boldsymbol{\beta}$.

Much of the above can be extended to the case of structured generalized linear models (Sahu and Gelfand, 1994). With unknown scale parameters, checking propriety of posterior distributions is somewhat complex. See Hobert and Cassella (1993), Ibrahim and Laud (1991) for more in this regard.

Implications for MCMC

For the linear models discussed above, there are several possible choices for the prior specification of the nonestimables $R\boldsymbol{\beta}$. We consider three possibilities and examine the consequences for MCMC.

1. We could use a degenerate point prior, for example, $R\boldsymbol{\beta} \equiv \mathbf{0}$, which is equivalent to putting "usual constraints" in the classical analysis of linear models. Then we arrive at a lower-dimensional model with proper posterior, for which standard MCMC methods will work effectively.

G. O. Roberts and S. K. Sahu are Lecturer and Research Associate, respectively, at the Department of Pure Mathematics and Mathematical Statistics, University of Cambridge, United Kingdom. W. R. Gilks is Senior Scientist at the Medical Research Council Biostatistics Unit, Institute of Public Health, Cambridge, United Kingdom.