Discussion of the paper "Approximate Bayesian inference for latent Gaussian models by using integrated nested Laplace approximations" by H. Rue, S. Martino and N. Chopin, for the Journal of the Royal Statistical Society

Dr. Andrew Gelman, Department of Statistics and Department of Political Science, Columbia University, New York

Statisticians often discuss the virtues of simple models and procedures for extracting a simple signal from messy noise. But in my own applied research I constantly find myself in the opposite situation: fitting models that are simpler than I would like—models that clearly miss important features of the data and, more importantly, important features of the underlying system I am modeling—because of computational limitations.

In some sense, "computational limitations" correspond to limited CPU time and memory. But in this age of gigabytes and more, it's only fair to describe these as limitations on our computational procedures. I am routinely in the position of wanting to fit a model that can't be fit using existing software, even though I know—*know*—that a simple enough algorithm must be out there to fit it using much less than the capabilities of a modern desktop PC.

The sorts of models I'm talking about include hierarchical models for parallel time series (for example, trends in public opinion in each of 50 states, or models for stochastically aligning tree ring data) and varying-intercept, varying-slope logistic regressions (that is, models where several coefficients can vary by group, in which case a covariance matrix needs to be modeled for the group-level structure).

In practice when fitting such models I lurch between various approximate methods based on point estimates, and full Gibbs-Metropolis which can be slow if not guided well. These two approaches can meet in the middle: approximations can be iteratively adjusted, leading ultimately to a Gibbs-like stochastic procedure, and Markov chain simulation can be made more efficient and reliable when guided by approximations that have been tailored to the problem at hand.

I welcome the article by Rue, Martino, and Chopin because it provides a more general way to construct these approximations. I suspect that, in addition to being a competitor to Gibbs and Metropolis, this approach ultimately can be used to make these stochastic algorithms more efficient.

As noted in the article, a challenge remains with problems with many hyperparameters, which are often themselves modeled hierarchically. As with the EM algorithm, it appears to be tricky to apply this method to a hierarchy with more than three levels, and I look forward to these researchers' future efforts in this area. It might help to model the hyperparameters explicitly rather than to consider them as unconstrained in some potentially large space.

I conclude with a remark on the comment in Section 7 of the article, that MCMC is often perceived to be "exact" even though in practice it is not. Fifteen or twenty years ago, MCMC itself had to fight this misconception in another form. At the time, importance sampling was

viewed as an exact method with MCMC as a sometimes necessary but unfortunate approximation. There was much discussion of how MCMC and importance sampling could work together, and ideas about starting with MCMC and then finishing up with importance sampling to get an exact result. Fortunately these ideas have subsided, as computational statisticians realized that actually existing importance sampling is not exact but can instead be viewed as just another iterative simulation method, and one that has no particular advantages over the Metropolis algorithm or other more clearly iterative approaches (Gelman, 1991).

Additional reference

Gelman, A. (1991). Iterative and non-iterative simulation algorithms. *Computing Science and Statistics* **24**, 433-438.