

ISI feature on highly-cited paper

Reversible jump Markov chain Monte Carlo computation and Bayesian model determination, by Peter J Green, *Biometrika* **82** (4): 711-732 Dec 1995

(Interview conducted in about 2005)

Why, in your view, is your paper highly cited?

The paper introduces Reversible Jump Markov chain Monte Carlo, a new simulation-based methodology for fitting statistical models that have variable-dimension parameters. It caught on simply because people found this useful in their statistical modelling work, and indeed even now it must be the most widely used method for treating these problems. The paper appeared in 1995, just the right time to make an impact, since the extraordinary rise of interest in computational Bayesian methods that started in the late 80's had by that time reached just about every area of application of statistical analysis, and researchers were starting to tackle more challenging problems. Curiously, the problems that Reversible Jump solves were probably not very widely regarded as interesting before the paper appeared, but the example applications in the paper, and early applications of the ideas by other authors, seem to have generated broader interest in these problems and showed people that solutions were possible, so the topic became self-reinforcing. Not very much of this was foreseen when I wrote the paper!

What are the circumstances which led you to your work?

In October 1993, Ulf Grenander and Michael Miller presented a paper on a computer vision problem, at a Royal Statistical Society discussion meeting in London. Papers for such meetings are available a week or two in advance so that others can prepare contributions for the discussion, which is eventually published along with the authors' rejoinder, and the paper itself. I was studying the paper the night before the meeting, and was particularly intrigued by the computational algorithm they developed, involving stochastic simulation, for inference about objects in microscopy images. It occurred to me both (1) that this kind of algorithm could equally well address a huge class of inference problems of which theirs was just one special example, and (2) that the algorithm itself could be made both more general, and easier to use and understand. I presented these points in a five-minute slot in the discussion (well, I probably ran over time by a minute or two).

It didn't make a very big impact, but I worked on the idea over the next month or two, verified the mathematical details, attempted to codify the method into a recipe that was easier to use, and implemented three fairly substantial examples – to multiple change-point analysis for point processes, to image segmentation, and to a partition problem for binomial data. I thought these would help convince a broad range of readers that the

ideas would be relevant to them. This package formed the paper that I submitted to *Biometrika*, and which appeared in 1995. By the time the paper had appeared, however, the work had already being quite widely taken up. The idea was propagated both through preprint servers, and through conference and workshop talks; there was a lot of workshop activity in the general area of stochastic systems in Europe at the time, thanks to the European Science Foundation funded network on Highly Structured Stochastic Systems. It was a very exciting time for research in this area, and I was lucky that my paper was ready at exactly the right time to get noticed.

Would you describe the significance of this work for your field?

Many problems in statistical inference, in all sorts of field of application, have the character that “the number of things you don’t know is one of the things you don’t know”. In microscopy on cell tissue, you want to count the number of cells, and measure the properties of each. In the analysis of genetic data on quantitative trait loci you want to estimate the number of loci and the quantitative effect at each. In multiple regression you want to select relevant variables and estimate effects for the selected variables. Problems like this are ubiquitous. I was interested in how you could make simultaneous inference about all unknowns in such problems, given noisy data. This is a bit of a challenge technically if you want a statistically rigorous solution, since most statistical methods deal with vectors of unknowns, parameters, of fixed dimension. The philosophically appealing way to do this inference is though fully probabilistic, or Bayesian, analysis. Bayes theorem tells you the theoretical solution, but you need to compute it in practice.

Reversible jump Markov chain Monte Carlo is a class of methods for doing that. which is feasible even in very complex models, and which is not too cumbersome to set up and use. So the immediate significance of the work was that people could now solve these problems. The interesting thing is that secondarily but ultimately more significantly, people started noticing that these ‘variable dimension’ problems were all around them, and started thinking that estimating all the unknowns simultaneously, and doing other tricks like Bayesian model averaging, might be good things to do. The methodology embraces both model-choice situations, where several competing explanations for the data are being entertained, and the analysis needs to simultaneously choose and fit a model, and settings where there is really a single model which has a variable dimension parameter, for example a series representation of a function where the number of terms is not fixed.

Where has this research gone since the publication of your paper?

Most of the citations of the paper are by researchers who are implementing the method to solve their own inference problems – the citations appear in over 160 different journals, in many different fields. They will often have tailored details of the algorithm to

their specific situation, and so particular idioms and rules of thumb have emerged over time. Almost all of the papers citing mine that do not deal with specific applications introduce novel generic statistical methodology, for example in model selection or nonparametric regression, building on Reversible Jump. Only a very few authors have really developed the idea itself and made significant additions to the simulation methodology. The most significant papers to mention here are by Tierney (Ann. Appl. Probab., 1998), Godsill (J. Comput. Graph. Stat., 2001), by Mira and myself (Biometrika, 2001), by Brooks, Giudici and Roberts (J. Roy. Statist. Soc. B, 2003), and by Cappé, Robert and Rydén (J. Roy. Statist. Soc. B, 2003).

Where do you see it going 10 years from now?

On the practical side, the methods are not yet easy enough to use. You can make mistakes setting them up, and even when you do get it right, the methods usually need to be tuned, and that can be laborious. So there would be tremendous benefit if the process of implementation could be automated, and I hope we see work in this direction. It has started; for example my PhD student David Hastie has devised a very nice adaptive/automatic approach based on normal mixtures.

More fundamentally, there are a host of challenging theoretical questions about how to summarise inference in complex variable-dimension models: Reversible Jump helps you compute what you want, but can't tell you what to compute. I hope my work will assist researchers in testing out in practical contexts their developing ideas about these important issues.