A Short History of Markov Chain Monte Carlo

Christian Robert and George Casella

2010
Introduction

- Lack of computing machinery, or background on Markov chains, or hesitation to trust in the practicality of the method

- MCMC impact on Statistics was felt until the early 1990s with Gelfand and Smith (1990)

- The rapid emergence of the dedicated BUGS (Bayesian inference Using Gibbs Sampling) software as early as 1991 was another compelling argument for adopting MCMC algorithms.
Before the Revolution

- Monte Carlo methods were born in Los Alamos, New Mexico during World War II from mostly physicists working on mathematical physics and the atomic bomb.

- Result in the Metropolis algorithm published by Metropolis et al. (1953).

- While Monte Carlo methods were in use by that time, MCMC was brought closer to statistical practicality by the work of Hastings in the 1970s.
The Metropolis et al. (1953) paper

First MCMC algorithm: Associated with a second computer, called MANIAC, built in Los Alamos under the direction of Metropolis in early 1952 (first in 1947)

Published in June 1953 in the Journal of Chemical Physics, the primary focus of Metropolis et al. (1953) is the computation of integrals of the form

$$\mathcal{J} = \frac{\int F(\theta) \exp\{-E(\theta)/kT\} d\theta}{\int \exp\{-E(\theta)/kT\} d\theta}$$

on $\mathbb{R}^{2N}$, $\theta$ denoting a set of $N$ particles on $\mathbb{R}^2$ and $E$ denotes energy.
The Metropolis et al. (1953) paper

- Updates each particle individually $\Rightarrow$ Primitive Gibbs sampler

- In order to improve the efficiency of the Monte Carlo method, Metropolis et al. (1953) propose a random walk modification of the $N$ particles.

- The number of iterations of the Metropolis algorithm used in the paper is limited: 16 steps for burn-in and 48 to 64 subsequent iterations, which required four to five hours on the Los Alamos computer.
The Hastings (1970) paper

- Generalization of the Metropolis algorithm that overcomes the dimensionality issue of Monte Carlo methods.

- Hastings (1970) defines his methodology for finite and reversible Markov chains. He also warns against high rejection rates as indicative of a poor choice of transition matrix, but does not mention the opposite pitfall of low rejection rates, associated with a slow exploration of the target.

- Presents a completely overlooked Metropolis-within-Gibbs type algorithm.
Seeds of the Revolution

- Hammersley and Handscomb (1964) present their textbook on Monte Carlo methods: Crude Monte Carlo, importance sampling, control variates and Conditional Monte Carlo.

- Besag (1974) was credited to some extent of the (re-)discovery of the Gibbs sampler but he expressed doubt about the practicality of his method.

- Because of its use for missing data problems, the EM algorithm (Dempster et al. 1977) has early connections with Gibbs sampling (Broniatowski et al. 1984) and Celeux and Diebolt (1985).
Gibbs

Geman and Geman (1984) brought Gibbs sampling into the arena of statistical application. This paper is also responsible for the name Gibbs sampling, because it implemented this method for the Bayesian study of Gibbs random fields.

This original implementation of the Gibbs sampler was applied to a discrete image processing problem.

The extent to which Gibbs sampling and Metropolis algorithms were in use within the image analysis and point process communities is actually quite large (Ripley 1987).
The Revolution

- Gelfand and Smith (1990) wrote the paper that is the genuine starting point for an intensive use of MCMC methods by the mainstream statistical community.

- Tanner and Wong (1987) had essentially the same ingredients as Gelfand and Smith (1990) but they focused on data augmentation and approximating the posterior distribution.

- Two years later, at an MCMC conference at Ohio State University organized by Alan Gelfand, Prem Goel, and Adrian Smith, there were three full days of talks.

- Many of the talks were to become influential papers; including Albert and Chib (1993), Gelman and Rubin (1992), Geyer (1992), Gilks (1992), Liu et al. (1994, 1995), Tierney (1994).
Theory Advances

- Tierney (1994), relevant theory developments on assumptions and properties needed to analyze the Markov chains.

- Liu et al. (1994, 1995) analyzed the covariance structure of and formally establish the validity of Rao-Blackwellization in Gibbs sampling.

- Rosenthal (1995) obtained one of the earliest results on exact rates of convergence.

- Hobert and Casella (1996) were able to document the conditions needed for a convergent Gibbs chain.
After the Revolution

- Particle filters are simulation methods adapted to sequential settings where data are collected progressively in time (Doucet et al. (2001)).

- Perfect sampling, the ability to use MCMC methods to produce an exact simulation from the target (Propp and Wilson (1996)).

- The invention of the reversible jump algorithm in can be seen as the start of the second MCMC revolution (Green (1995)): formalization of a Markov chain that moves across models and parameter spaces.

- Hobert and Casella (1996) were able to document the conditions needed for a convergent Gibbs chain.