1

An Introduction to MCMC

Petros Dellaportas
Gareth O. Roberts

1.1 MCMC and spatial statistics

Markov chain Monte Carlo (MCMC) algorithms are now widely used in virtually all areas of statistics. In particular, spatial applications featured very prominently in the early development of the methodology (Geman & Geman 1984), and they still provide some of the most challenging problems for MCMC techniques. It is not of a great surprise that most modern introductions to the subject emphasise the flexibility and generality of MCMC. We will do that here too, but we also have more targeted aims in supporting the other chapters in this book covering different areas of spatial statistics.

The problem which MCMC addresses is a familiar one throughout a broad range of diverse disciplines including physics, engineering, computer science and statistics. We are interested in simulating from a probability distribution, call it $\pi(x)$. Although the function $\pi(\cdot)$ can be written in closed form (at least up to a normalisation constant), suppose that we are interested in moments of π which are not computable analytically. Furthermore, direct simulation from π may be difficult or impossible, perhaps because of its high dimensionality, so that simple Monte Carlo evaluation of these moments of π is not possible. MCMC attacks this problem by instead simulating from a Markov chain whose invariant distribution is π.

As we shall see, there are only two basic techniques used in MCMC. One uses an accept/reject mechanism to 'correct' an arbitrary Markov chain in such a way as to make π its invariant distribution. This method is essentially the *Metropolis-Hastings algorithm*. The other method simplifies the high dimensional problem by successively simulating from different smaller-dimensional components of $\mathbf{X} = (X^{(1)}, \ldots X^{(d)})$. This idea is the essence of the *Gibbs sampler*. There is a wealth of different ways of applying these basic methods, and the two strategies can be combined in many different ways to suit different problems.

Within spatial statistics, π might define the distribution of a spatial process, or in other contexts, it might describe a Bayesian posterior distribution of parameters which govern an underlying spatial process. More complex distributions incorporating both these elements are also common. Since numbers of locations are typically very large, the dimensionality of

J. Møller (ed.), *Spatial Statistics and Computational Methods*

π is often extremely large, so that the need for MCMC methods in spatial statistics is particularly strong. Moreover, many spatial distributions exhibit strong and complex dependence structures, which can often severely inhibit MCMC performance.

In this chapter, we shall introduce MCMC methods in a general statistical context, although our illustrative example will have a simple spatial statistics structure that will be generalised considerably in later chapters. Our aim is not to give a comprehensive treatment of the subject, and interested readers wishing to learn more on this should read one of the currently available texts on the subject, see for example Gilks, Richardson & Spiegelhalter (1996) Tanner (1996), Gamerman (1997), Robert & Casella (1999), and Liu (2001). We will however provide sufficient details to give the reader an overview of the techniques, their immense generality and power, and some of the common pitfalls.

1.2 The Gibbs sampler

The most obvious way to break down a difficult high-dimensional simulation problem to a more tractable one, is to try to reduce it to a collection of more manageable smaller dimensional simulations. Many MCMC algorithms do this, but the simplest form of this strategy is an important special case called the *Gibbs sampler.*

Suppose that we wish to obtain a sample from the multivariate density $\pi(x^{(1)}, \ldots, x^{(d)})$. (We shall assume now and throughout this chapter, that distributions are continuous but everything extends immediately to the discrete distribution case.) We shall always use bracketed superscripts to index over components, so that a typical random variable distributed according to π might be written $\mathbf{X} = (X^{(1)}, \ldots X^{(d)})$, and the conditional distribution of the ith component given the rest according to π will simply be written as *the conditional* $X^{(i)}|(X^{(1)}, \ldots, X^{(i-1)}, X^{(i+1)}, \ldots, X^{(d)}$. The d conditional distributions of this type are termed the *full conditional distributions.*

The Gibbs sampler successively and repeatedly simulates from the conditional distributions of each component given the other components, see Algorithm 1 below. The realisations $\boldsymbol{X}_1, \boldsymbol{X}_2, \ldots$ obtained from iterating step 2 in Algorithm 1 are simulated from a Markov chain. Under mild regularity conditions (Roberts & Smith 1994), convergence of this chain to the stationary distribution $\pi(x^{(1)}, \ldots, x^{(d)})$ is guaranteed, so that for sufficiently large k, the values $(x_k^{(1)}, \ldots, x_k^{(d)})$, $\ldots, (x_n^{(1)}, \ldots, x_n^{(d)})$ can be regarded as realisations from this distribution.