“How absurdly simple!”, I cried.
“How absurdly simple!”, said he, a little nettled. “Every problem becomes very childish when once it is explained to you.”

Arthur Conan Doyle
The Adventure of the Dancing Men

Reader’s guide

This chapter is the first of a series of two on simulation methods based on Markov chains. Although the Metropolis–Hastings algorithm can be seen as one of the most general Markov chain Monte Carlo (MCMC) algorithms, it is also one of the simplest both to understand and explain, making it an ideal algorithm to start with.

This chapter begins with a quick refresher on Markov chains, just the basics needed to understand the algorithms. Then we define the Metropolis–Hastings algorithm, focusing on the most common versions of the algorithm. We end up discussing the calibration of the algorithm via its acceptance rate in Section 6.5.
6.1 Introduction

For reasons that will become clearer as we proceed, we now make a fundamental shift in the choice of our simulation strategy. Up to now we have typically generated \textit{iid} variables directly from the density of interest $f$ or indirectly in the case of importance sampling. The Metropolis–Hastings algorithm introduced below instead generates \textit{correlated} variables from a Markov chain. The reason why we opt for such a radical change is that Markov chains carry different convergence properties that can be exploited to provide easier proposals in cases where generic importance sampling does not readily apply. For one thing, the requirements on the target $f$ are quite minimal, which allows for settings where very little is known about $f$. Another reason, as illustrated in the next chapter, is that this Markov perspective leads to efficient decompositions of high-dimensional problems in a sequence of smaller problems that are much easier to solve.

Thus, be warned that this is a pivotal chapter in that we now introduce a totally new perspective on the generation of random variables, one that has had a profound effect on research and has expanded the application of statistical methods to solve more difficult and more relevant problems in the last twenty years, even though the origins of those techniques are tied with those of the Monte Carlo method in the remote research center of Los Alamos during the Second World War. Nonetheless, despite the recourse to Markov chain principles that are briefly detailed in the next section, the implementation of these new methods is not harder than those of earlier chapters, and there is no need to delve any further into Markov chain theory, as you will soon discover. (Most of your time and energy will be spent in designing and assessing your MCMC algorithms, just as for the earlier chapters, not in establishing convergence theorems, so take it easy!)

6.2 A peek at Markov chain theory

This section is intended as a minimalist refresher on Markov chains in order to define the vocabulary of Markov chains, nothing more. In case you have doubts or want more details about these notions, you are strongly advised to check a more thorough treatment such as Robert and Casella (2004, Chapter 6) or Meyn and Tweedie (1993) since no theory of convergence is provided in the present book.

A Markov chain $\{X^{(t)}\}$ is a sequence of dependent random variables $X^{(0)}, X^{(1)}, X^{(2)}, \ldots, X^{(t)}, \ldots$ such that the probability distribution of $X^{(t)}$ given the past variables depends only on $X^{(t-1)}$. This conditional probability distribution is called a \textit{transition kernel} or a \textit{Markov kernel} $K$; that is,