19.2.5 *Barriers ad ergodicity breaking*

We conclude the study of the Markov chains by focusing once more on the problem of their convergence towards the equilibrium distribution. The problem of ergodicity breaking we discuss in this section, is often studied in connection with a specific model (especially those used in statistical mechanics). However, it is a much more general phenomenon, which we therefore study at the level of generic Markov chains.

Let us start from a concrete example. Suppose we have a Markov chain defined on integer numbers and we want the asymptotic probability distribution to be the sum of two Gaussians with unit variance, one centered in $-L$ and the other in L. For simplicity, we only allow for transitions between nearest neighbor states, i.e., between number differing by one. In the simulation, we obviously use the most efficient algorithm among those we know at present, namely the Metropolis algorithm.

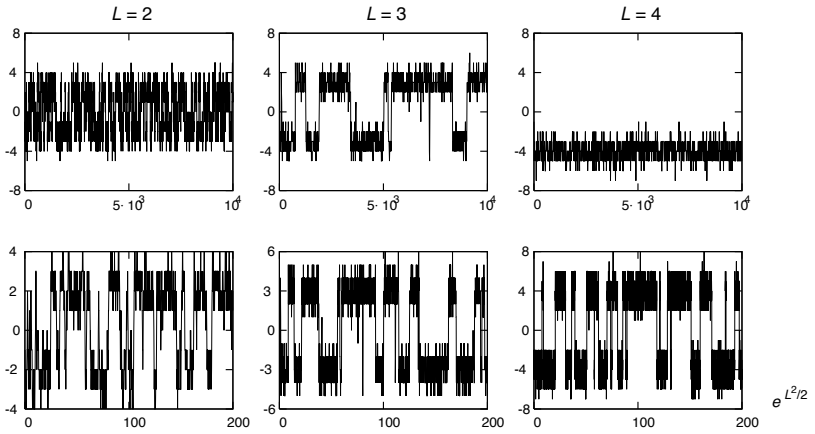

Fig. 19.5 A random walker's position in a Markov chain with an asymptotic probability distribution concentrated in L and $-L$, for three values of L. In the three plots on the bottom, the times are multiplied by a factor $e^{-L^2/2}$.

In the three plots on top of Figure 19.5 we show the random walker's position, initially starting from the origin, for the first 10^4 steps and for three different values of L. For $L = 2$ the walker passes relatively easily between the two maxima of the distribution. Instead, for $L = 3$ the time between jumps among maxima is already of the order of a thousand of steps (11 crossovers $L \leftrightarrow -L$ can clearly be identified). Finally, for $L = 4$, the walker remains during all the 10^4 steps around the Gaussian's maximum

with negative average. In the latter case, we need to extend the simulation and consider much longer times in order for the walker to visit both regions in which the probability $\vec{\pi}$ is sensibly different from zero.

The cause of this behavior is relatively simple. Any trajectory moving the walker from one maximum of $\vec{\pi}$ to the other one must pass through regions in which the asymptotic probability is extremely small. In this particular case, this statement might seem obvious and predictable (we are working in one dimension and to go from $-L$ to L or vice versa we need to pass through the origin). However, this is general valid also in more complicated cases, such as when dealing with multidimensional integration domains.

In the simple case of the two Gaussian distributions centered in L and $-L$ the asymptotic distribution is

$$\pi_j \propto e^{-(j+L)^2/2} + e^{-(j-L)^2/2} .$$

In this case we can estimate the time needed by the walker to pass from the negative to the positive states. Passing from negative to positive values (and the other way around) only occurs when the walker is at the origin and continues in the opposite direction with respect to where he came from. While the second of these events occurs with probability $1/2$, the first one has probability $\pi_0 = \mathcal{O}(e^{-L^2/2})$. So, this event only occurs every $\tau = 1/\pi_0$ steps and we expect the walker to move from one maximum to the other each $\mathcal{O}(e^{L^2/2})$ time steps.

This argument is formally correct, only in the limit $L \to \infty$ in which the probability π_0 tends to zero and the relaxation time diverges. Nevertheless, also for small values of L this leads to excellent results. Indeed, in the lower plots of Figure 19.5 we changed the time scales by a multiplicative factor $\exp(-L^2/2)$, as shown along the horizontal axes. The three resulting curves are very similar (apart from some larger fluctuations in the case $L = 2$, where the separation between the maxima is too small to clearly show the effect).

Often, the term *barrier* is used to indicate the region in which $\vec{\pi}$ is very small; the random walker needs many attempts before he actually manages to cross it.

Please note that in the upper right panel of Fig. 19.5 the number of steps in our simulation were not chosen accurately. Indeed, we cannot be sure to have overcome all barriers in the Markov chain under study. By looking at that plot, we might believe to have reached the asymptotic state, though the walker is only visiting half of the true asymptotic distribution's

relevant states. Sometimes, the only way to make sure the entire Markov chain has been visited is to increase the length of the numerical simulation. In Section 20.2.1 we explain in more detail how to study the convergence towards equilibrium of a Markov chain Monte Carlo method.

Hands on 4 - Barrier crossing times

Consider a Markov chain with an asymptotic distribution consisting of two Gaussian distributions centered in L and $-L$ and a barrier around the origin. We just saw that, for the Metropolis algorithm, the transition from one side of the barrier to the other occurs each $\tau = \mathcal{O}(e^{L^2/2})$ steps, i.e., it is a rare dynamical process for large L. Moreover, when this event takes place, it is very quick, as is clear from the rightmost plot on the bottom of Figure 19.5.

Study numerically the time it takes to cross a barrier. To this purpose, compute the length of the random walker's trajectories starting from the position $-L$ and reaching the position L without ever passing by the position $-L$ again, and the length of those following the opposite path. Considering that these trajectories last at least $2L$ steps, you should be able to conclude from this analysis, that when the walker decides to cross the barrier, he does so without any delays, following an almost ballistic trajectory!

For any value of L the Markov chain we studied in this section is ergodic, i.e., it is possible to move from any state to any other one (even though the time needed might be very long in some cases). On the contrary, in the limit $L \to \infty$, the relaxation time τ really diverges and the chain decomposes into two subchains, one consisting of the positive states and the other of the negative ones. Indeed, in the limit $L \to \infty$ an *ergodicity breaking* occurs.

The system showing ergodicity breaking we have discussed here is very simple. Indeed, it is not difficult to figure out that a different choice of the probabilities u_{ij}, i.e., a different connectivity between the states, allows to avoid the ergodicity breaking phenomenon. For example, by including the transitions which are symmetric with respect to the origin, i.e., $j \leftrightarrow -j$, among the allowed one, all dynamical phenomena we described in this section (from the slowing down due to the barrier to the ergodicity breaking) vanish. The transition $j \leftrightarrow -j$ represents a kind of shortcut around the barrier eliminating all these effects. Note how the transition between two states which are symmetric with respect to the origin is always accepted in

the Metropolis algorithm, because these states have the same asymptotic probability, $\pi_j = \pi_{-j}$.

The search for these "shortcuts" is one of the most interesting aspects in the development of Markov chain Monte Carlo algorithms. Unfortunately, for more complicated systems than the one discussed in this section, the shortcuts are not at all obvious. For example, the so-called "cluster" algorithms [Newman and Barkema (1999)] do apply some kind of shortcut, but only because of a deep knowledge of the physical-mathematical properties of the Ising model, described in Section 19.3.

19.3 The Ising model

The Ising model is characterized by a set of binary variables placed at the vertices of a regular lattice. These variables are called *spins* because in a typical application of the model they represent elementary magnetic moments. In this section, we study the model defined on a two-dimensional square lattice, such as the one in Figure 19.6. The spins of nearest neighboring vertices, i.e., those connected by an edge, interact with each other, by means of a ferromagnetic interaction attempting to align thse two spins in the same direction. The energy of a configuration $\boldsymbol{s} \equiv (s_1, \ldots, s_N)$ of the system's N spins is given by

$$E(\boldsymbol{s}) = -\sum_{(i,j)} J\, s_i\, s_j \ . \tag{19.20}$$

In the expression (19.20) the spin variables can only take on two values $s_i = \pm 1$, the sum is over the pairs of indices $i, j \in \{1, \ldots, N\}$ corresponding to adjacent vertices and the coupling constant[5] J is positive, because the ferromagnetic interaction fosters the alignment of the spins. The minus sign preceding the sum is needed because the preferred configurations in Nature are those minimizing the energy. Indeed, two aligned spins have energy $-J$, while two spins in the opposite direction have energy $+J$. It is easy to generalize this model to other interaction topologies. We can take any graph, place spins on its vertices and have each couple of spins connected by an edge to interact with a coupling term $-Js_is_j$.

We follow the advice of Section 10.4 to use in a simulation the simplest measurement units. As the energy (19.20) is an integer multiple of J (which

[5]The symbol J we use for the coupling constant should not be confused with the one of the Joule, the measurement unit of energy (which might might actually be the measurement unit of the coupling constant J, which is an energy).