

Monte Carlo Methods - Lecture 2

Why is metropolis right? Is it the only update that works?

Consider again the Boltzmann probability of finding a state (configuration), c , in a classical system,

$$P_c^{eq} = e^{-\beta E_c} / Z \quad (1)$$

where Z is the partition function $Z = \sum_c e^{-\beta E_c}$. The key thing to note is that it is easy to calculate $e^{-\beta E_c}$ for one configuration, but finding Z is hard as we have to sum over all configurations. We thus look for update rules that do not use Z . Moreover we want to make sure that the probability (1) is recovered by our update procedure.

In general we can write the time evolution of the probability of finding a state using the “master” equation,

$$P_k(t+1) = P_k(t) + \sum_{l \neq k} (w_{l \rightarrow k} P_l(t) - w_{k \rightarrow l} P_k(t)). \quad (2)$$

This is a restricted class of master equation as the configuration or probability at time $t+1$ only depends on the probability at time t . Processes of this type are called Markov systems. In modern terminology they are also sometimes called cellular automata and have a very rich behavior - see e.g. Wolfram.

In an equilibrium MC procedure we want to choose the transition probabilities w so that the state probability $P_k(t)$ approaches the equilibrium probability P_k^{eq} of Eq. (1). Lets assume that our dynamics reaches equilibrium, then as $t \rightarrow \infty$, $P_k(t+1) = P_k(t) = P_k^{eq}$, so that,

$$\sum_l (w_{l \rightarrow k} P_l^{eq} - w_{k \rightarrow l} P_k^{eq}) = 0. \quad (3)$$

This important balance equation defines the transition rates w that are consistent with equilibrium. A simple way to ensure that this equation holds is to demand that every term in the sum is zero, which gives the detailed balance condition,

$$\frac{w_{l \rightarrow k}}{w_{k \rightarrow l}} = \frac{P_k^{eq}}{P_l^{eq}} = e^{-(E_k - E_l)/k_B T}. \quad (4)$$

A very important feature of this equation is that Z cancels on the RHS, so the ratio of transition rates depends only on the energy difference between two states. It also implies that if we develop an algorithm with transition rates which obey this equation, then according to the master equation (2), we will generate states with a probability characteristic of equilibrium, at long times. Note that although only the balance condition (3), not detailed balance (4), is required, almost all algorithms use a detailed balance to ensure compatibility with equilibrium. Even with this restriction there are a large number of possible update rules. One simple choice of update rules, where $a_{i \rightarrow l}$ is the probability that we change state i to l , is given by $a_{i \rightarrow l} \propto w_{i \rightarrow l}$. However this is not useful as we can't calculate Z .

Now we define $x = e^{-(E_k - E_l)/k_B T}$ then Eq. (4) is satisfied for any $0 < F(x) = w_{k \rightarrow l} < 1$ which satisfies,

$$\frac{F(x)}{F(1/x)} = x \quad (5)$$

By direct substitution it is easy to show that two popular algorithms which satisfy this equation are the Metropolis method,

$$F(x) = \min(x, 1) \quad (6)$$

and the heat bath method,

$$F(x) = \frac{x}{1+x} \quad (7)$$

Note that this analysis shows that if we use these update rules in the master equation, then we will reach thermal equilibrium. But we don't want to solve the whole master equation as it operates in the full $2^N \times 2^N$ configuration space. Instead we randomly sample to find the typical behavior it predicts.

Convergence issues

The discussion above demonstrated that the Metropolis procedure satisfies detailed balance. Now we need to understand how many time steps we need to take in order to find good averages for the quantities that we are interested in. What limits the convergence of this type of process??

In the ideal case, a system is ergodic, which means that it is able to explore all of the important phase space provided it is allowed sufficient computational time. In these cases, the time average leads to results characteristic of the thermodynamic average, provided we exclude early time transients and provided we use a sampling method that obeys detailed balance. However many (most) problems of physical and engineering interest are non-ergodic over the accessible time scales, leading to memory, history, symmetry breaking, or hysteretic effects. The Ising system that we are studying is ergodic for $T > T_c$ and non-ergodic for $T < T_c$, for large enough samples. You can observe non-ergodicity for low temperatures by starting with a random spin state.

Mathematically a system is only ergodic at very low temperatures if its free energy landscape is purely convex. In our Ising system the energy landscape is convex at high temperature, while at low temperature there is symmetry breaking and the free energy landscape has two degenerate minima, at a finite positive magnetization and a finite negative magnetization. Many (most) interesting computational problems are non-ergodic for most initial conditions due to the fact that they have a more complex "energy (or free energy) landscape". In our example we can easily find the initial condition that leads to behavior representative of experiment, however in most problems this is much more difficult, unless we have the correct prior information. Examples include structural optimization, ab-initio quantum problems, np-complete problems, spin glasses,

The discussion above makes it clear that we need to first decide what we want to extract from the simulations - in our case we focus on the magnetization $M = \langle \sum_i S_i \rangle$, where the brackets indicate an average - averaging is discussed further below. From the configurations that we generate, we can calculate pretty much anything we want. However the hardest thing to find is the Entropy, which of course means that the Free Energy is also hard. The internal energy is relatively easy as are correlation functions.

As noted in the first lecture, as the temperature increases we expect to go from a magnetized state at low temperature to a non-magnetic state at high temperature. To simulate this experimental process, we start with a magnetized state and allow it to relax to a state characteristic of the value of $J/k_B T$ which we set. Note that if we run the simulation for long enough at $T < T_c$ we can still get $m = 0$ as the system may have time to flip into the other orientation of magnetization ie. it becomes ergodic. The non-ergodic behavior is actually the interesting aspect of the problem and represents the true experimental behavior. Nevertheless, if the system flips to the opposite orientation, we can simply take the average of the modulus of the magnetization to find behavior that is similar to the non-ergodic or symmetry breaking case.

Theoretical predictions $N \rightarrow \infty, t \rightarrow \infty$.

At long enough times, the system equilibrates so the the magnetization reaches a steady state value. No plot the steady state value as a function of temperature to observe the transition from the magnetized to non-magnetized state. The behavior near the Curie point, T_c , is of the form,

$$m \approx |T_c - T|^\beta \tag{8}$$

and is mathematically singular, at least in an infinite system and at infinite time. At the curie point the domain structure is very interesting, exhibiting a power law distribution of domain sizes. If we cool the system from a high temperature domain state, the MC procedure finds it very difficult to restore the magnetized state. This is also true in real magnets where a demagnetized magnet needs to be cooled in a field to restore its magnetic alignment.

Rounding due to finite lattice size

The mathematical singularity in the magnetization behavior only occurs for large samples at equilibrium, so that smaller samples have a rounded behavior near T_c . This rounding behaves as,

$$m(T_c, L) \approx L^{-\beta/\nu} \quad (9)$$

Rounding due to slow relaxation

The larger a sample is, the longer it takes to relax to its equilibrium value, in general we write $m = m_0 + \exp^{-t/\tau}$, where τ is the relaxation time. Away from critical points τ depends on the correlation length $\tau \propto \xi^z$, where $z \approx 2$. The situation is special near the critical point where large clusters have to be overturned - in that case, the relaxation is algebraic instead of exponential - The relaxation time thus diverges near the critical point, a phenomenon called critical slowing down. If the MC simulations runs (or experimental measurements) are not over sufficiently long times, the critical behavior is rounded.

How to average the data - Law of large numbers

If we make a measurement of a quantity, y , a number of times, N , so that we have a set of measurements $y_1 \dots y_N$, then if we take the average of the measurements we get,

$$\langle y \rangle_{N=} = \frac{1}{N} \sum_{i=1}^N y_i, \quad (10)$$

then the value of the average obeys $\langle y \rangle_{N=} = y_\infty + O(\sigma/N^{1/2})$, where y_∞ is the true value of the quantity and σ is the true standard deviation of the quantity. You can see that the error in the average quantity decreases as the square root of the number of measurements. We need to keep this in mind in MC work as it shows that to obtain good average values, we need to make many measurements.

Application to drunk student: Consider that the student takes steps drawn from a random distribution $P(y)$. After N steps, the student's average position is given by $N \langle y \rangle_{N=}$. The typical distance a student reaches is then of order $N^{1/2}$. In the case of the ink drop the width of the ink drop grows as $\langle |x| \rangle \approx Dt$. Other applications include the conformations of random coil states of polymers and proteins where the size of the coil is $R_g \approx N^{1/2}$. Actually due to self repulsion the coil is often further extended to $R_g \approx N^{3/5}$.

Central limit theorem

In statistical physics, thermodynamic quantities obey the central limit theorem, for example the internal energy per particle obeys the law of large numbers $M/N \epsilon + O(1/N^{1/2})$, in the thermodynamic limit $N \rightarrow \infty$. The central limit theory is a more complete statement of the statistics in the large N limit,

$$P(M_N) = A e^{-b(M_N - M_\infty)^2/N} \quad (11)$$

where b does not depend on N , to leading order.

Some empirical guidance

For a system of size $L \times L$, allow a warm up time of L^2 sweeps and then calculate the magnetization every L sweeps of the lattice. Run the simulation for greater than L^3 sweeps. This means the computational time is greater than L^5 . For a sample of size $L = 20$ this means quite long computational times. You also need to run for a range of temperatures.