Monte Carlo Simulations in Statistical Physics

In these notes I discuss Monte Carlo simulations for the study of *classical* models in statistical mechanics. I include a *simple and direct* proof that the method *converges* to the Boltzmann distribution. Usually, physics articles discuss this important point by just giving a reference to the mathematical literature on "Markov chains", where the proof is rather abstract. In these notes I give a proof of convergence which is self contained and uses only elementary algebra.

In statistical mechanics one computes averages of a quantity A from the Boltzmann distribution, *i.e.*

$$\langle A \rangle = \sum_l P_l^{\rm eq} A_l, \tag{1}$$

where l denotes a state, A_l is the value of A in that state, and $P_l^{\rm eq}$ is the equilibrium (Boltzmann) probability distribution for the system to be in state l, *i.e.*

$$P_l^{\rm eq} = \frac{e^{-\beta E_l}}{\sum_m e^{-\beta E_m}}, \tag{2}$$

where E_l is the energy of state l and

$$\beta = \frac{1}{k_B T}, \tag{3}$$

with T the temperature and k_B Boltzmann's constant.

The number of states is exponentially large in the number of degrees of freedom, N, and so it is quite impractical to perform the sum in Eq. (1) except if N is really tiny. However, we are generally interested in large N, especially the thermodynamic limit, $N \to \infty$. Monte Carlo simulations provide a means of studying large (though still not infinite) systems numerically.

In Monte Carlo methods, rather than summing over *all* the states in Eq. (1), one samples a small fraction of these states. This leads to an *estimate* of the average, which will not be exact but will have statistical errors. We generate states using an iterative procedure, discussed below, which (after some initial transient) generates states with the *correct Boltzmann probability distribution*. Hence the estimate of the average is

$$\langle A \rangle_{\rm est} = \frac{1}{t_0} \sum_{t=1}^{t_0} A(t), \tag{4}$$

where t, which we shall think of as being like time, denotes a configuration generated by the algorithm, $A(t)$ is the value of A in the configuration at "time" t, and t_0 is the number of measurements. The exponential factors in Eq. (2) do not appear in Eq. (4) because they are incorporated in the probabilities that the states are generated.

From standard statistical arguments, the difference between the estimate, $\langle A \rangle_{\rm est}$ in Eq. (4), and the exact value $\langle A \rangle$ in Eq. (1) is a random variable whose size is proportional to $n^{-1/2}$ where n is the number of *statistically independent* measurements. The configurations generated by the algorithm will be *correlated* in general up to a certain *relaxation time*, τ, and so n will be less than t_0, generally $n \simeq t_0/\tau$.

An introduction to the theory of Monte Carlo simulations is in the books by Newman and Berkema[1] and Landau and Binder[2] A more advanced and thorough discussion is given in article by Sokal[3].

The simulation will begin with the system in some state, l_0 say. We then generate stochastically (*i.e.* in a random manner) a subsequent set of states, to which we will give a "time label" $t, = 0, 1, 2, \cdots$. As a simple example, consider the Ising model where we have a set of N interacting "spins", S_i, which take values ± 1. A typical way to generate a subsequent state would be to pick a spin at random and then either flip it or leave it alone with a certain probability (discussed below).

At $t = 0$ the system is definitely in state l_0 but at later times it can be in different states with non-zero probability $P_l(t)$. We desire that at long times $P_l(t)$ approaches the equilibrium distribution $P_l^{\rm eq}$, *i.e.*

$$\lim_{t\to\infty} P_l(t) = P_l^{\rm eq}. \tag{5}$$

Clearly the initial distribution is very different from this,

$$P_l(0) = \delta_{l,l_0}. \tag{6}$$

The initial distribution is made to converge to the equilibrium distribution after a certain "time" by a judicious choice of the (non-negative) "transition rates", $w_{l\to m}$, where $w_{l\to m}$ is the probability that the system it will be in state m ($\neq l$) at time $t+1$, given that it is in state l at time t.

The evolution of the probabilities $P_l(t)$ follows the "master equation",

$$P_l(t+1) - P_l(t) = \sum_{m\neq l} \left[P_m(t)\, w_{m\to l} \, - P_l(t)\, w_{l\to m} \right]. \tag{7}$$

The first term on the right hand side describes transitions into state l from m (which therefore increases P_l and so has a plus sign) while the second term describes transitions out of state l, which decreases $P(l)$. Note that only terms with $m \neq l$ contribute. We can also define $w_{l\to l}$ to be the probability that the system stays in state l, i.e. $w_{l\to l} = 1 - \sum_{m\neq l} w_{l\to m}$, which means that

$$\sum_m w_{l\to m} = 1, \tag{8}$$

if the term $m = l$ is included. Eq. (8) implies that the master equation, Eq. (7), can be written

$$P_l(t+1) = \sum_m P_m(t)\, w_{m\to l}, \tag{9}$$

where the term $m = l$ is included.

Clearly Eq. (7), preserves the normalization of probabilities

$$\sum_l P_l(t+1) = \sum_l P_l(t) = 1, \tag{10}$$

since $\sum_{l,m} P_m(t)\, w_{m\to l} = \sum_{l,m} P_l(t)\, w_{l\to m}$. To see this, just interchange the dummy labels, l and m, in one expression and you get the other.

A necessary condition for the method to work is that the Boltzmann distribution, $P^{\rm eq}$, is a *stationary* distribution, *i.e.* if, somehow, we have got the system to equilibirum at time t, $P_l(t) = P_l^{\rm eq}$, then it is still in equilibriumat time $t+1$, $P_l(t+1) = P_l^{\rm eq}$. From Eq. (7) this requires that

$$\sum_m \left(P_l^{\rm eq}\, w_{l\to m} \, - P_m^{\rm eq}\, w_{m\to l} \right) = 0\,, \tag{11}$$

or equivalently, from Eq. (8), that

$$P_l^{\rm eq} = \sum_m P_m^{\rm eq}\, w_{m\to l}\,. \tag{12}$$

In practice, stationarity is usually accomplished by making *each* term in Eq. (11) vanish, *i.e.*

$$P_l^{\rm eq}\, w_{l\to m} = P_m^{\rm eq}\, w_{m\to l}\,, \tag{13}$$

which is known as the *detailed balance* condition. Because the equilibrium distribution is given by Eq. (2), the detailed balance condition can be written

$$\frac{w_{l\to m}}{w_{m\to l}} = e^{-\beta(E_m - E_l)}. \tag{14}$$

Notice that the detailed balance condition only determines a ratio of transition rates; there are many possible choices for the $w_{l\to m}$ which satisfy both this condition and the requirement that that $w_{l\to m} \geq 0$.

A common way of implementing an elementary Monte Carlo move (also known as an "update") is to first choose a "trial" state m as the possible state for the system at time $t+1$. The probability that the trial state is m if the state at time t was l, is given by a "proposal matrix" U_{lm}. This satisfies the condition $\sum_m U_{lm} = 1$, and is usually chosen to be symmetric. State m is then accepted as the state at $t+1$ with some probability $a_{l\to m}$ (explained below), *i.e.*

$$w_{l\to m} = U_{lm} a_{l\to m}. \tag{15}$$

Otherwise the state at $t+1$ is the old state, l.

For example, in an Ising problem, state m is frequently chosen to be a state in which one of the spins (chosen at random) in state l has been reversed. In this case $U_{lm} = 1/N$, where N is the number of sites, if l and m differ by a single spin flip, and 0 otherwise. One generates a random number, r, with a uniform distribution between 0 and 1, and if $r < a_{l\to m}$ the move is accepted, *i.e.* the state at time $t+1$ is m, and otherwise the move is rejected, *i.e.* the state at $t+1$ is l, the *same* as at time t. Testing a single spin is called an "*update*". We also use the term "*sweep*" to denote the updating of N spins, where N is the total number. The sweep is the natural unit in which to describe the length of a simulation.

It is also possible to pass sequentially through the lattice (*i.e.* to test spins $1, 2, 3, \cdots, N$ in that order) rather than to go through the lattice in a random sequence. This saves the generation of one random number per site. In sequential updating, one sweep corresponds to the updating of each spin once. (In random updating, each spin is updated once on average, but some may be updated more than once and some not at all.)

Going back to the general discussion, the energy difference is $\Delta E = E_m - E_l$ and the detailed balance condition for a is clearly

$$\frac{a_{l\to m}}{a_{m\to l}} = \frac{w_{l\to m}}{w_{m\to l}} = e^{-\beta\Delta E}, \tag{16}$$

(for a symmetric proposal matrix). This is satisfied by

$$a_{l\to m} = F\left(e^{-\beta\Delta E}\right), \tag{17}$$

where F is *any* function which satisfies $0 \leq F(x) \leq 1$ (since probabilities cannot be negative or greater than unity) and

$$\frac{F(x)}{F(1/x)} = x \qquad \text{for all } x. \tag{18}$$

There are an infinite number of possible choices for $F(x)$, and two convenient ones are:

1. The Metropolis algorithm,

$$F(x) = \min(x, 1), \tag{19}$$

In this approach one always accepts the move if it gains energy, but only accepts it with probability $\exp(-\beta\Delta E)$ if it costs energy, *i.e.* if $\Delta E > 0$.

2.

$$F(x) = \frac{x}{1+x}, \tag{20}$$

which corresponds to an acceptance probability of

$$\frac{1}{e^{\beta\Delta E}+1}, \tag{21}$$

irrespective of the sign of ΔE. For an Ising model, where each spin can only be in one of two states, this is an example of the "heat-bath" method, where, after the update, the probability of the variable being altered is independent of its value before the update and just corresponds to a local thermal equilibrium for that variable in its instantaneous environment.

It is not enough to show that $P^{\rm eq}$ is a stationary distribution. We also need to show that the probabilities *converge* to $P^{\rm eq}$ starting from an arbitrary initial distribution. I give here a direct, elementary proof, which presumably exists elsewhere in the literature though I have been unable to find it (or a similar derivation). We assume the detailed balance condition, though the standard derivations make clear that this is not essential, but only that $P^{\rm eq}$ be a stationary distribution[3,4]. For convergence, the algorithm must also be "ergodic", *i.e.* starting from a given state at time $t = 0$ then, for any sufficiently large time t, the system can be in *any* state. This prevents the system being trapped in a subset of states.

We start with the following quantity, which is a measure of the deviation from equilibrium,

$$G = \sum_l \frac{1}{P_l^{\rm eq}} (P_l - P_l^{\rm eq})^2 = \sum_l \left(\frac{P_l^2}{P_l^{\rm eq}}\right) - 1 \tag{22}$$

evaluated at time t, where the last expression follows because P and $P^{\rm eq}$ are normalized.

At time $t+1$ we indicate (for compactness of notation) the probabilities by P_l' and the corresponding value of G by G'. We will show that G monotonically decreases, *i.e.*

$$\Delta G \equiv G' - G \leq 0, \tag{23}$$

where the equality only holds if the system is in equilibrium. Eq. (23) shows that the system will eventually approach arbitrarily close to the equilibrium distribution.

To prove Eq. (23) we start by using Eqs. (9) and (22) to write ΔG as

$$\Delta G = \sum_{l,m,n} \left[w_{m\to l} w_{n\to l} \frac{P_m P_n}{P_l^{\rm eq}} \right] - \sum_l \frac{P_l^2}{P_l^{\rm eq}} \quad (= \Delta G_1 + \Delta G_2). \tag{24}$$

For ΔG_1, the first term on the right hand side of Eq. (24), we use the detailed balance condition, Eq. (13), to replace $w_{m\to l}$ by $w_{l\to m}P_l^{\rm eq}/P_m^{\rm eq}$, and similarly for $w_{n\to l}$, which gives

$$\Delta G_1 = \sum_{l,m,n}\left[w_{l\to m}w_{l\to n}P_l^{\rm eq}\frac{P_mP_n}{P_m^{\rm eq}P_n^{\rm eq}}\right]. \tag{25}$$

In the second term in Eq. (24), ΔG_2, we use Eq. (8) to insert a factor of $\sum_m w_{l\to m}$ (and interchange the indices l and m), which gives

$$\Delta G_2 = -\sum_{l,m} w_{m\to l}\frac{P_m^2}{P_m^{\rm eq}} = -\sum_{l,m,n} w_{l\to m}w_{l\to n}P_l^{\rm eq}\left(\frac{P_m}{P_m^{\rm eq}}\right)^2. \tag{26}$$

where the second equality is obtained by applying the detailed balance relation again and incorporating a factor of $\sum_n w_{l\to n}$ (which is equal to unity). Taking the half the sum of the last expression for ΔG_2 and the same expression with m replaced by n, and including ΔG_1, we finally get

$$\Delta G = -\frac{1}{2}\sum_{l,m,n} w_{l\to m}w_{l\to n}\,P_l^{\rm eq}\left(\frac{P_m}{P_m^{\rm eq}}-\frac{P_n}{P_n^{\rm eq}}\right)^2, \tag{27}$$

where terms with $m=l$ and $n=l$ are included.

Eq. (27) is our main result. It shows that ΔG is definitely negative unless, for every state l, all states which can be reached from l in a single update (which will generically include l itself) have probabilities proportional to the equilibrium probabilities. The most natural scenario is that *all* states satisfy this with the same proportionality constant (which must be unity) i.e. the system is in equilbrium. However, ΔG also vanishes if $P_l = 0$ for some states which are "inaccessible" at time t and $P_m \propto P_m^{\rm eq}$ for the remaining states which are "accessible" at time t. Hence, to achieve full equilibrium, *i.e.* $\lim_{t\to\infty} P_l(t) = P_l^{\rm eq}$, the algorithm must also be *ergodic*, *i.e.* starting from a given state, after a sufficiently long time there is non-zero probability, $P_l(t)$, for the system to be in *any* state.

I conclude that if the algorithm is ergodic and satisfies detailed balance, then $\Delta G < 0$ unless the system is in equilibrium. This shows that the system will eventually converge to equilibrium, although no estimate is given for the time to converge.

I should mention that the condition of detailed balance, though generally used, is strictly not necessary, and the more relaxed condition of "balance", Eq. (11) is all that is needed. However, I haven't been able to find a *simple* derivation, using just elementary algebra, of convergence if the transition probabilities satisfy balance but not detailed balance.

Finally I give a brief discussion of sequential, rather than random, updating. For random updating the probability of making a transition is the same for every update, *i.e.* writing

$$P_l(t+1) = \sum_m \Gamma_{lm}P_m(t), \tag{28}$$

where Γ, the transition matrix, (related to w by $\Gamma_{lm} = w_{m\to l}$), is the same for each "time" t. However, for sequential updating, the transition matrix depends on which site is being updated, so, for a complete sweep, we have

$$\Gamma = \Gamma^{(1)}\Gamma^{(2)}\cdots\Gamma^{(N)}, \tag{29}$$

where $\Gamma^{(i)}$ is the transition matrix for updating spin i. Although the $\Gamma^{(i)}$ individually satisfy the detailed balance condition, the transition matrix for the whole sweep, Γ, does *not*,[3]

though it does preserve $P^{\rm eq}$ as a stationary distribution. This is because the probability of the reverse transition, $m \to l$ say, for a whole sweep, is related to the probability of transition $l \to m$ in the desired way only if the spins are updated in the *reverse* order.

Despite lack of detailed balance, convergence to the equilibrium distribution is still obtained. To see this note that it is trivial to generalize our proof of convergence to the case where the transition probabilities depend on "time". As long as each set of transition probabilities, $\Gamma^{(i)}$ here, satisfies the detailed balance condition, convergence to equilibrium must be obtained since G decreases at *each* step.

Probably the simplest "traditional" derivation of convergence based on the theory of Markov chains is in Ref. 4; I thank Onuttom Narayan for bringing this reference to my attention. This derivation does *not* require detailed balance, only that the algorithm has $P^{\rm eq}$ as a stationary distribution and is ergodic. It would be interesting to try to generalize the proof of convergence given above so as to avoid making the assumption of detailed balance.

This discussion of Monte Carlo simulations in statistical mechanics has been very brief. To apply the Monte Carlo method usefully it is necessary to (i) understand the statistical errors, see Ref. 3, and (ii) to ensure that enough Monte Carlo steps are done that the system has come to thermal equilibrium. This is also discussed by Sokal[3].

---

[1] M.E.J. Newman and G.T. Barkema, "*Monte Carlo Methods in Statistical Physics*", Oxford University Press, (1999).

[2] D.P. Landau and K. Binder, *A Guide to Monte Carlo Simulations in Statistical Physics* Cambridge University Press, 2nd. Ed. (2005).

[3] A.D. Sokal *Bosonic Algorithms* in "Quantum Fields on the Computer", ed. M. Creutz, 1992; *Monte Carlo Methods in Statistical Mechanics: Foundations and New Applications*, Cours de Troisième Cycle de la Physique en Suisse Romande (Lausanne, June 1989).

[4] J. G. Kemeny and J. L. Snell, "Finite Markov Chains", Van Nostrand (Princeton) (1960).