Physics 115/242
Monte Carlo simulations in statistical mechanics

Peter Young

(Dated: May 8, 2003)

I.   INTRODUCTION TO STATISTICAL MECHANICS

For additional information on the non-Monte Carlo part of this handout see *Thermal Physics* by Kittel and Kroemer.

In equilibrium statistical mechanics we consider the behavior of a system in equilibrium with its surroundings (a "heatbath"). Its energy is not exactly fixed because it can exchange energy with the bath, and the state of the system is not fixed. Statistical mechanics tells us that the system has a certain probability that the system can be in any of the states. The probability of being in state $|n\rangle$ with energy E_n is given by the "Boltzmann distribution"

$$P(n) = \frac{1}{Z} e^{-E_n/k_B T}, \tag{1}$$

where T is the absolute temperature and k_B is called Boltzmann's constant. The normalizing factor Z, given by

$$Z = \sum_\ell e^{-E_\ell/k_B T}, \tag{2}$$

is called the "partition function". The partition function provides a convenient link between the microscopic picture involving the states and energy levels and the macroscopic (or thermodynamic) picture since the free energy of the system, F, is related to it by

$$F = -k_B T \ln Z. \tag{3}$$

One calculates observable quantities as averages over states with the Boltzmann weight. The average of A say, (which could be the energy or the magnetization) is given by

$$\langle A \rangle = \sum_n P(n) A_n = \frac{\sum_n A_n e^{-E_n/k_B T}}{\sum_\ell e^{-E_\ell/k_B T}}, \tag{4}$$

where A_n is value of A in state-n (to be precise it is the quantum mechanical expectation value $\langle n|A|n\rangle$).

II.   A TOY MODEL

In order to illustrate the Monte Carlo method it is useful to have a simple example where things can be worked out explicitly. A good model to take is the *Ising* model of magnetism. The magnetic moment (spin) of an individual atom can be one of two possibilities: $\uparrow$ ("up") and $\downarrow$ ("down"). It is convenient to assign a numerical value to these two states by a variable S_i (for site i) which takes value $+1$ in the up state and -1 in the down state, i.e.

$$S_i = \begin{cases} +1 & (\uparrow) \\ -1 & (\downarrow) \end{cases} \tag{5}$$

There are N spins, S_i, $i = 1, 2, \cdots, N$ and we are interested in the situation where N is large. The total number of states is clearly 2^N, which is *enormous* if N is large. In general, and not just for this model, it is true that the number of states grows exponentially with the number of degrees of freedom in the problem. This will turn out to be important.

For example, in a solid N might be of order Avogadro's number, 6×10^{26}. Two to this number is so large that it is essentially impossible to visualize it. Even if we are interested in smaller sizes, then to enumerate all the states on a PC would take more than the age of the universe (about 4×10^{17} seconds) for N greater than about 80, not a very large number. Clearly, then, enumerating all the states is not feasible for large N. If we are going to determine averages in Eq. (4) numerically we need to do a sampling of some (small) fraction of the states. How to do this sampling is the topic of this handout.

The total magnetization of the system M is given by

$$M = \sum_{i=1}^{N} S_i. \tag{6}$$

For now we will forget about the energy of the states and the Boltzmann distribution, and just average giving *equal weight* to all states. Later on we will put in the Boltzmann factor.

We would like to know how many states have a particular value of the magnetization M. First of all we note that the *largest* value of M is N and, and the least is $-N$. We can get an idea of the *typical* value of M by calculating averages of M and M^2 (giving equal weight to all states as mentioned above). Now

$$\langle M \rangle = \sum_i \langle S_i \rangle = 0 \tag{7}$$

since S_i takes values ± 1 and so, giving equal weight to these two possibilities, we have

$$\langle S_i \rangle = 0 \tag{8}$$
$$\langle S_i^2 \rangle = 1. \tag{9}$$

The average of M^2 is given by

$$\langle M^2 \rangle = \sum_{i,j} \langle S_i S_j \rangle. \tag{10}$$

Now if $i \neq j$ the average is zero since the four possibilities for $S_i S_j$

$$S_i = +1,\ S_j = +1, \qquad S_i S_j = +1 \qquad \text{probability } 1/4 \tag{11}$$
$$S_i = -1,\ S_j = +1, \qquad S_i S_j = -1 \qquad \text{probability } 1/4 \tag{12}$$
$$S_i = +1,\ S_j = -1, \qquad S_i S_j = -1 \qquad \text{probability } 1/4 \tag{13}$$
$$S_i = -1,\ S_j = -1, \qquad S_i S_j = +1 \qquad \text{probability } 1/4 \tag{14}$$

average to zero. However, if $i = j$ then $\langle S_i^2 \rangle = 1$ (see Eq. (9)) and so Eq. (10) gives

$$\langle M^2 \rangle = N, \tag{15}$$

which means that $\langle M^2 \rangle^{1/2} = N^{1/2}$. Hence a *typical* value of M is of order $N^{1/2}$, which is very much less than the maximum value, of N.

One can easily write down the total number of states with a given M. If there are $N/2 + k$ states with spin up and $N/2 - k$ states with spin down, then
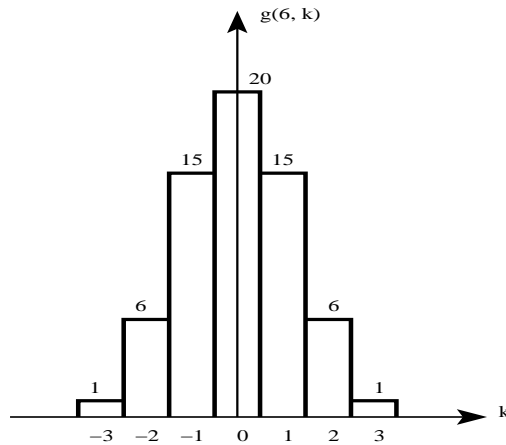
$$M = 2k. \tag{16}$$

Furthermore, the number of states with a given value of the spin inbalance factor k is equal to the number of ways arranging the up spins (say) among the N sites. This is given by the binomial coefficient

$$g(N, k) = \frac{N!}{\left(\frac{N}{2} + k\right)! \left(\frac{N}{2} - k\right)!}. \tag{17}$$

For example, for $N = 6$ we have

k	M	$g(n, k)$
-3	-6	1
-2	-4	6
-1	-2	15
0	0	20
1	2	15
2	4	6
3	6	1

$$\tag{18}$$

The distribution is clearly symmetric, and peaked about zero. From the central limit theorem we expect it to become Gaussian for large N and, from Eq. (15), have a width of order $\sqrt{N}$. This is indeed the case. Taking the log of $g(N, k)$, i.e.

$$\ln g(N, k) = \ln(N!) - \ln\left[\left(\frac{N}{2} + k\right)!\right] - \ln\left[\left(\frac{N}{2} - k\right)!\right] , \tag{19}$$

and using Stirling's approximation

$$\ln N! \simeq N \ln N - N + \tfrac{1}{2}\ln(2\pi N), \tag{20}$$

which is valid for large N, one finds, after a bit of algebra, that

$$g(N, k) = 2^N \sqrt{\frac{2}{\pi N}} \ \exp\left(\frac{-2k^2}{N}\right), \tag{21}$$

for large N. As expected, this is a Gaussian with zero mean and a width of order $\sqrt{N}$. Summing over all k,

$$\int_{-\infty}^{\infty} g(N, k)\, dk = 2^N, \tag{22}$$

correctly gives the total number of states as 2^N. Note that we have replaced the discrete sum by an integral, which is valid for large N since the integrand only changes slowly when k is increased by unity, and extended the range of integration from $\pm N/2$ to $\pm\infty$, which produces negligible error for large N because $g(N, k)$ is tiny for $|k| > N/2$. (In fact it is tiny for $k \gg \sqrt{N}/2$, which is much less than $N/2$.)

Eq. (21) clearly gives $\langle k \rangle = 0$ (since $g(N, k)$ is an even function of k) and we also find from standard properties of Gaussian integrals that

$$\langle k^2 \rangle = \frac{\int_{-\infty}^{\infty} g(N, k)\, k^2\, dk}{\int_{-\infty}^{\infty} g(N, k)\, dk} = \frac{N}{4}, \tag{23}$$

which implies $\langle M^2 \rangle = N$, in agreement with Eq. (15).

It is now useful to consider the magnetization per spin

$$m = \frac{M}{N} = \frac{2k}{N} = \frac{1}{N}\sum_{i=1}^{N} S_i \ . \tag{24}$$

The largest possible value of m is 1, and yet we see from Eqs. (21) and (24) that for large N the overwhelming majority of states have m very close to zero, more precisely within of order $1/\sqrt{N}$ of zero. This result will be important in developing a Monte Carlo method, since it shows that states generated randomly must have m very close to zero. For large N you would have to generate a huge number of configurations at random to find even one which had m significantly different from zero.

Now we need to make the problem a bit less trivial by putting in an expression for the energy. The simplest case, for which the model is still easily exactly solvable, is to include a magnetic field, h. The Hamiltonian (energy) is then given by

$$\mathcal{H} = -h \sum_i S_i = -hM. \tag{25}$$

Hence the Boltzmann weight of a state of magnetization M is proportional to $\exp(\beta h M) = \exp(N \beta h m)$, where we use the conventional notation

$$\beta = \frac{1}{k_B T}. \tag{26}$$

Hence *individual* states with a larger m have a larger Boltzmann weight than those with smaller m. Because the energy is proportional to N, the Boltzmann factor varies hugely depending on the value of m. However, the *number* of states is sharply peaked near $m = 0$ and so decreases with increasing m (for $m > 0$).

Hence we need to consider the *total* Boltzmann weight of all states with a given magnetization per spin m which is a product of two factors: (a) the number of states with that value of m, Eq.(21) with $k = Nm/2$, and (b) and the Boltzmann factor $\exp(N \beta h m)$. It follows that the probability $P(m)$ that the system has magnetization per spin m is given by

$$P(m) \propto \exp\left[-N\left(\frac{m^2}{2} - \beta h m\right)\right]. \tag{27}$$

This has a maximum at $m = m^\star$ where

$$m^\star = \beta h. \tag{28}$$

Note that the maximum arises because $P(m)$ is a product of $\exp(-Nm^2/2)$, which *decreases* rapidly as m increases, and $\exp(N\beta h m)$ which *increases* rapidly as m increases, see the figure.
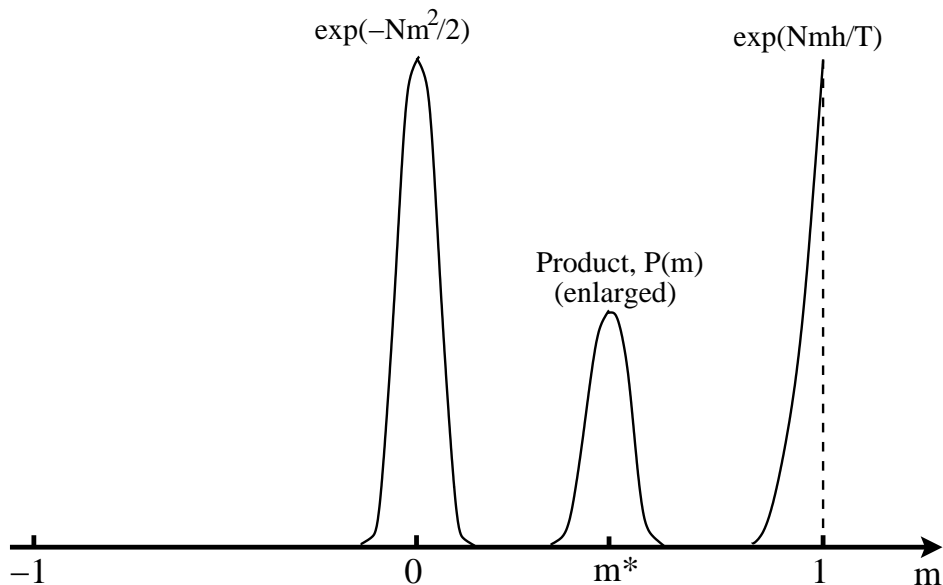
$P(m)$ is *sharply peaked* at $m^\star$ as we can see by expanding Eq. (27) up to second order in $\delta m = m - m^\star$:

$$P(m) = P(m^\star) \exp(-N\delta m^2/2), \tag{29}$$

This shows that the width of $P(m)$ is of order $1/\sqrt{N}$, and so all the states that a large system visits (in equilibrium) in a field h have magnetization very close to $m^\star$.

To summarize we have seen for our simple model, the Ising model in which the energy comes entirely from the external field, that

1. At $T = \infty$, which corresponds to equal weighting of all states, a large system only visits states with m close to zero, the left hand peak in the above figure.

2. At finite T, a large system only visits states which have m close to $m^\star = \beta h$, the middle peak in the above figure.

It turns out that analogous results hold much more generally. In particular, at infinite temperature the states visited by the system will have energy close to a certain value (corresponding to the energy where most of the states lie) but at a finite temperature the system will visit states with energy in the vicinity of a *different* value.

III. MONTE CARLO TECHNIQUES

For additional information on this section see *A Guide to Monte Carlo Simulations in Statistical Physics* by David Landau and Kurt Binder, or the first 4 pages of the article on Monte Carlo on my web site `http://bartok.ucsc.edu/peter/converge.pdf`.

The example in the previous section is very simple. Each spin is independent of other spins. One can make the model a much richer model for magnetism by including interactions between nearby sites. We suppose that the spins are on a lattice, which could be simple cubic lattice in three-dimensions, a square lattice in two-dimensions, or a chain in one-dimension, and the energy depends on the relative orientation of neighboring pairs, i.e.

$$\mathcal{H} = -J \sum_{\langle i,j \rangle} S_i S_j - h \sum_i S_i, \tag{30}$$

where J is the interaction strength and the sum $\langle i, j \rangle$ is over nearest neighbor pairs. If $J > 0$ the spin prefer to align parallel at low temperature, and such a system is said to be "ferromagnetic". If $J < 0$ neighboring spins prefer to point in opposite directions at low temperature, and such systems are said to be "antiferromagnetic".

In general, the problem is now highly non-trivial. Only in one-dimension is there a fairly simple exact solution. In two dimensions, some quantities like the energy and magnetization in zero field can be calculated with monumental effort, while in in three dimensions there are no exact solutions.

In two and three dimensions there is a transition at a finite temperature T_c where the spins align spontaneously even in the absence of an external field h. If $J > 0$ the spins align parallel and give rise to a state with a macroscopic magnetization. This is then a very simple model for the magnetism of iron. The magnetization per spin m increases as T goes below T_c and, in this model, tends to unity as $T \to 0$. This is shown in the figure.
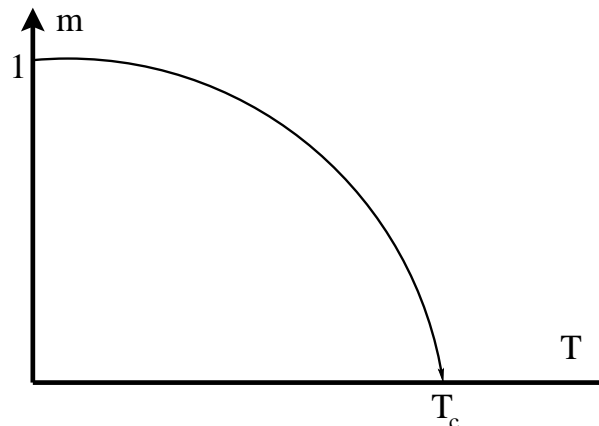
In one dimension there is no transition at a finite T, though at $T = 0$ all the spins are aligned so one often talks about a transition at zero temperature.

Since there are no exact solutions in three dimensions, it is useful to be able to simulate models like the Ising model numerically. In particular we would like to calculate averages as in Eq. (4). As we've said, except for very small N, there are too many states to enumerate them all, so we need to do some sort of "random" sampling. However, we can't do truly random sampling, which we used earlier in the class for Monte Carlo integration, because, as discussed in the previous section, the vast majority of states generated do not have the correct energy for the (finite) temperature being considered.

The usual technique is to generate states *with the Boltzmann probability distribution itself*. Assuming for the moment that we can do this, then our estimate for $\langle A \rangle$ in Eq. (4) is

$$\langle A \rangle \simeq \frac{1}{M_{\text{meas}}} \sum_{\ell=1}^{M_{\text{meas}}} A_\ell, \tag{31}$$

where M_{meas} is the number of "measurements" done on quantity A. Note that we give equal weight to all the generated states and the Boltzmann factor does not appear *explicitly*, though it is present *implicitly* because the states are generated with the Boltzmann probability.

In order to generate states with the Boltzmann weight we use an iterative procedure. Starting from an initial state "0", which may *not* be a state with the correct energy for the temperature being considered, we generate states 1, 2, 3 etc. such that eventually, states *are* generated with the Boltzmann distribution for that temperature. It is convenient to think of this sequence as representing the state of the system at successive "times" t.

Suppose the system is in state ℓ with energy E_ℓ at time t. What will the state be at time $t+1$? The procedure is to choose a new trial state m, with energy E_m, which typically differs from ℓ by flipping a single spin. The state at time $t+1$ is then taken to be m with a certain probability $w_{\ell \to m}$ (discussed below), and otherwise the system stays in the old state ℓ. We start the system off in some initial state ℓ_0 say,

$$P_\ell(0) = \delta_{\ell, \ell_0} \ , \tag{32}$$

and require that, at long times, the $P_\ell(t)$ equal the equilibrium, Boltzmann, distribution P_ℓ^{eq}:

$$\lim_{t \to \infty} P_\ell(t) = P_\ell^{\text{eq}}. \tag{33}$$

Let's see how, with an appropriate choice of the w, this method will eventually generate states with a Boltzmann distribution. We consider the equation, known as the *Master Equation*, which tells us how the probability for the system to be in the different states changes with time. If $P_\ell(t)$ is the probability that the system is in state ℓ at time t, then the master equation is

$$P_\ell(t+1) - P_\ell(t) = \sum_{m \neq \ell} \left(P_m(t) w_{m \to \ell} - P_\ell(t) w_{\ell \to m} \right). \tag{34}$$

The first term on the RHS describes transitions into state l from state m at time $t+1$, which increases P_l, and the second describes transitions out of l into m.

Clearly a *necessary* condition for Eq. (33) to be valid is that if the distribution is in equilibrium at time t then it remains the equilibrium distribution at time $t+1$. From Eq. (34), this requires

$$\sum_{m \neq \ell} \left(P_m^{\text{eq}}(t) w_{m \to \ell} - P_\ell^{\text{eq}}(t) w_{\ell \to m} \right) = 0. \tag{35}$$

For convenience, we usually require that each term in the sum vanish separately, which leads to the important *detailed balance condition*

$$\frac{w_{\ell \to m}}{w_{m \to \ell}} = \frac{P_m^{\text{eq}}}{P_\ell^{\text{eq}}} = \exp(-\beta(E_m - E_\ell)). \tag{36}$$

The detailed balance condition does not determine the w uniquely, only the ratio of w for a transition to that for the inverse transition. A common choice which satisfies detailed balance is the *Metropolis* updating probability:

$$w_{\ell \to m} = \begin{cases} \exp(-\beta \Delta E) & \text{if } \Delta E > 0 \\ 1 & \text{otherwise,} \quad \text{(Metropolis probability)} \end{cases} \tag{37}$$

where $\Delta E = E_m - E_l$ is the energy change in going from state ℓ to state m. It is easy to see that Eq. (36) is satisfied because if the change in energy is positive for one of the transitions it must be negative for the inverse transition. If $\Delta E = 0$ precisely then $w_{\ell \to m} = w_{m \to \ell} = 1 = \exp(-\beta \Delta E)$ so things still work out. Physically the Metropolis updating rule means that we always accept the move if the system gains energy ($\Delta E < 0$), but, if $\Delta E > 0$, we only accept the move with a probability which is less than unity and which decreases as ΔE increases. Intuitively this clearly means there is a greater likelihood of the system being in states of lower energy.

Another common choice is the "heatbath" probability, where, irrespective of the sign of ΔE, one flips the spin with probability

$$w_{\ell \to m} = \frac{1}{1 + \exp(\beta \Delta E)} \qquad \text{(Heat bath probability)}. \tag{38}$$

This satsifies detailed balance because $f(x) = \exp(-x)f(-x)$ with $f(x) = (\exp(x) + 1)^{-1}$.

If the detailed balance condition is satisfied then, once the system has reached equilibrium, it will stay in equilibrium. However, it is also necessary to show that the system will *converge* to equilibrium starting from some arbitrary initial distribution like Eq. (32). Most of the proofs of this are quite mathematical. However, a simple proof is given in the article on my web site quoted above. Convergence occurs for an algorithm satsifying detailed balance provided, in addition, it is "ergodic", which means that starting from any state one can get to any other state in a finite number of moves. Physically ergodicity means that the system won't get trapped in just part of the configuration space. Here we will just assume that convergence occurs if the detailed balance condition is satisfied.

We can now describe how to implement a Monte Carlo simulation for the Ising model:

1. The basic unit of Monte Carlo time is the "sweep" which involves going through the lattice (usually sequentially though sometimes a random sequence is used) and flipping each spin in turn with the probability in Eq. (37) or Eq (38). To implement this, compute the energy ΔE to flip the spin at site i, generate a random number r with a uniform distribution between 0 and 1, and flip S_i if

$$r < \exp(-\beta \Delta E) \qquad \text{(Metropolis)} \tag{39}$$
$$r(1 + \exp(\beta \Delta E)) < 1 \qquad \text{(Heat bath)}. \tag{40}$$

   Note that ΔE only involves the orientation of spins which are *neighbors* of site i.

2. Since the system is not initially in equilibrium, perform M_{drop} initial sweeps during which no measurements are made, in order to equilibrate it.

3. Then perform M_{run} sweeps during which measurements are carried out. One might do a measurement after every sweep, in which case the number of measurements M_{meas} is equal to M_{run}, or possibly only after every few sweeps, in which case $M_{\text{meas}} < M_{\text{run}}$. Averages are then given by Eq. (31).

This type of sampling, in which we don't choose points completely at random but choose them predominantly in the important region of configuration space, is called *importance sampling*. It was introduced into statistical physics by Metropolis et al. (1953).

As usual we need to estimate error bars. This is not as simple in our earlier class discussion of error bars in Monte Carlo integrals. There we used the fact that the different data was *statistically independent*. (Remember that two random variables X and Y are statistically independent if $\langle X\,Y \rangle = \langle X \rangle \langle Y \rangle$.) In that case, if we have M estimates $X_\alpha,\ \alpha = 1, 2, \cdots, M$, the final estimate X_{est} for the average of X and its error bar ϵ are given by

$$X_{\text{est}} = \frac{1}{M} \sum_{\alpha=1}^{M} X_\alpha, \tag{41}$$

$$\epsilon = \frac{\sigma}{\sqrt{M-1}}, \tag{42}$$

where

$$\sigma^2 = \frac{1}{M} \left(\sum_{\alpha=1}^{M} X_\alpha^2 \right) - X_{\text{est}}^2, \tag{43}$$

is the variance of the M values X_α. However, the relationship (42) between σ and the error ϵ is only true for statistically independent data. For our Monte Carlo simulations there will be a correlation between $X(t_0)$ and $X(t_0 + t)$ up some time τ called the *relaxation time*. (We measure time in units of a sweep from now on.) Roughly speaking, we should get ϵ from σ in Eq. (42) by dividing by $\sqrt{M/\tau}$ rather than by $\sqrt{M}$ (forgetting about the usually unimportant factor of -1).

In order to estimate the error bar from the simulation we need to have statistically independent data. This is usually obtained by "binning". We divide our M_{meas} measurements for X say into n groups or "bins". The number of sweeps corresponding to each bin is M_{run}/n. If this is much larger than τ there won't be much correlation between

averages obtained from different bins. (Only a small amount due to the data at the end of one bin being correlated with that at the beginning of the next bin.) Hence we estimate the error bar from Eqs. (41)–(43) in which X_α refers to the average from a bin, and $M = n$, the number of bins.

Knowing how many sweeps to include in a bin requires an estimate of τ. For the purposes of the homework assignment for Physics 242 students, this is not essential and any plausible assumption will do. However, for professional work one could compute

$$C(t_0, t_0 + t) = \langle X(t_0)X_(t_0 + t)\rangle - X_{\text{est}}^2, \tag{44}$$

where the average here is taken over different values of the starting time, t_0, in a given run. The "time-dependent correlation function" $C(t_0, t_0 + t)$ will tend to zero for $t > \tau$.

An alternative approach to getting the error bar is to do n completely independent runs, starting from different initial states ℓ_0 and with different seeds for the random numbers. The data from the different runs will certainly be uncorrelated. A small disadvantage with this approach is that equilibration, i.e. the initial M_{drop} sweeps, has to be done separately for each run, which is a bit wasteful.