This document is intended to serve as personal notes for self-study of *Monte Carlo Methods for Statistical Physics* by M.E.J. Newman and G.T. Barkema.

1 Introduction

1.1 Statistical Mechanics

Many systems of interest are composed of very many simple constituent systems. For example, a liter of oxygen is composed of 10^{22} oxygen molecules. The equations of motion for a single oxygen molecule are relatively simple but the sheer number of oxygen molecules in the larger system render the exact solutions infeasible. Instead, we approach the larger system via probabilistic methods. The goal is to express the state of the larger system as a set of probabilities of being in one state or another.

The typical systems studied in physics for which MCM are useful are systems described by a Hamiltonian H and an associated energy spectrum that is either continuous or given by a discrete set E_0, E_1, E_2, etc. We will also consider systems that interact with a thermal reservoir with which the system can exchange heat. Heat exchanges with the reservoir manifest as a small (negligible) perturbation in the Hamiltonian which pushes the system from one energy state to another.

Modeling the effect of the perturbation on the Hamiltonian is done by defining a *dynamics* for the system. The dynamics of the system may take the form of a set of transition rates. Suppose the system is in a state μ. Then, $R(\mu \to \nu)dt$ is defined to be the probability that the system transitions to a state ν in time dt, where $R(\mu \to \nu)$ is defined to be the *transition rate*. We may also define a set of weights $w_\mu(t)$ which represent that probability that the system will be found in state μ at time t. Using these elements, we can write down a *master equation*:

$$\frac{dw_\mu}{dt} = \sum_\nu [w_\nu(t)R(\nu \to \mu) - w_\mu(t)R(\mu \to \nu)] \tag{1}$$

The first term in the sum is the rate at which the system is transitioning into state μ. The second term is the rate at which the system is transitioning out of state μ.

We may define the expectation value of some observable Q, which takes the value Q_μ in state μ, as

$$\langle Q \rangle = \sum_\mu Q_\mu w_\mu(t)$$

The expectation value of Q can be interpreted two ways:

- It is the mean of the observed value of Q that would be obtained if we measured many identically prepared systems simultaneously.
- It is the time average of the quantity Q. That is, it might be the average observed value of Q after many measurements over time.

1.2 Equilibrium

An *equilibrium* state can be defined as a state in which all of the rates of change dw_μ/dt vanish, and therefore the transition rate into and out of each state μ must be equal. All systems goverened by the master equation must come to equilibrium after some time and so we can define the equilibrium values of the weights, the *equilibrium occupation probabilities*, as

$$p_\mu = \lim_{t\to\infty} w_\mu(t)$$

For a system in equilibrium with thermal reservoir at temperature T, the Boltzmann distribution is

$$p_\mu = \frac{1}{Z}e^{-E_\mu/kT} = \frac{1}{Z}e^{-\beta E_\mu}$$

where Z is the normalization constant, called the *partition function*, and defined as

$$Z = \sum_\mu e^{-\beta E_\mu}$$

Plugging this into our definition of expectation values gives us

$$\langle Q\rangle = \sum_\mu Q_\mu p_\mu(t) = \frac{1}{Z}\sum_\mu Q_\mu e^{-\beta E_\mu}$$

which is the expression for the expectation value of a system in equilibrium. For example, the internal energy U of a system is given by

$$U = \langle E\rangle = \frac{1}{Z}\sum_\mu E_\mu e^{-\beta E_\mu} = -\frac{1}{Z}\frac{\partial Z}{\partial\beta} = -\frac{\partial logZ}{\partial\beta}$$

It is sometimes appropriate in MCM calculations to calculate the partition function, from which many other quantities can be derived. Some examples can be found in Section 1.2 of Newman and Barkema.

1.3 Fluctuations, Correlations, and Responses

It is useful to assess the fluctuations in observable quantities. Doing so allows us to determine how the quantity we are measuring varies over time and therefore how much of an approximation is being made by calculating the expectation value. As an example, the root-mean-square fluctuation in internal energy is given by

$$\sqrt{\langle E^2\rangle - \langle E\rangle^2} = \sqrt{\frac{\partial^2 logZ}{\partial\beta^2}}$$

In the limit of a large system, called the *thermodynamic limit*, it is common for fluctuations to become negligible. For example, RMS fluctuations in internal energy scale as $\sqrt{V}$ in a system of volume V, but the internal energy itself scales

as V so as the system becomes very large, the fluctuations grow increasingly negligible.

Every parameter that can be fixed in a system (e.g. volume, external field) has a conjugate variable (e.g. pressure, magnetization) that is given by a derivative of the free energy. For example,

$$p = -\frac{\partial F}{\partial V}$$

$$M = \frac{\partial F}{\partial B}$$

The Hamiltonian contains terms of the form $-XY$ where Y is a "field" whose value is fixed, and X is the conjugate variable to which it couples. A common trick to calculate the thermal average of a quantity is the following:

- Make up a fictitious field that couples to the quantity of interest.
- Add an appropriate term to the Hamiltonian.
- Set the field to zero after performing the derivative on the free energy.

The *susceptibility* of X to Y is a measurement of the strength of the response of X to variation in Y. Typically, this quantity is denoted by χ and can be defined as

$$\chi = \frac{\partial \langle X \rangle}{\partial Y}$$

The *linear response theorem* tell us that the fluctuations in a variable are proportional to the susceptibility of the variable to its conjugate field.

1.3.1 Correlation Functions

The *disconnected correlation function* $G^{(2)}(i,j)$ is defined as

$$G^{(2)}(i,j) = \langle x_i x_j \rangle$$

This function gives us roughly an idea of how the variables x_i and x_j are correlated. Consider the following scenarios:

- If x_i and x_j move (roughly) together most of the time around the origin zero, then the average of their product will be a positive value.
- If x_i and x_j move (roughly) together but in opposite dirrections most of the time, then the average of their product will be a negative value.
- If x_i and x_j move with no discernible pattern relative to one another, then their product will be sometimes positive and sometimes negative, averaging to zero over a long time.

This definition can be problematic in the case of a *spontaneously broken symmetry state*. This occurs when a variable x develops a non-zero expectation value. If this happens, x_i and x_j may, for instance, always be large positive values leading to a large positive correlation value even if they have nothing to do with one another. To address this problem, we can use the *two-point connected correlation function* $G_c^{(2)}(i,j)$ defined as

$$G_c^{(2)}(i,j) = \langle x_i x_j \rangle - \langle x_i \rangle \langle x_j \rangle$$

$$= \langle (x_i - \langle x_i \rangle) \times (x_j - \langle x_j \rangle) \rangle$$

For reference, the higher-order connected correlation functions for three and four sites can be found in section 1.2.1.

1.4 Example: The Ising Model

The two-dimensional Ising model of a magnet consists of a lattice of sites, where each site contains a magnetic dipole or spin. Each spin is a two-level system denoted by the variable s_i, which can only take on the values ± 1. Exchange interactions are modeled by including terms proportional to $s_i s_j$ in the Hamiltonian. In the simplest case, these terms are identical for each pair of spins (indicating the same interaction strength J all over the lattice). Another simplification is to consider only nearest neighbor interactions. An external magnetic field B can also be introduced. In this case we obtain the following Hamiltonian

$$H = -J \sum_{\langle ij \rangle} s_i s_j - B \sum_i s_i$$

where $\langle ij \rangle$ indicates a sum over nearest neighbors. The minus signs are conventional and fix the signs of the parameters J and B. In the current form:

- A positive J value (*ferromagnetic*) makes the spins want to align to minimize the Hamiltonian.
- A negative J value (*anti-ferromagnetic*) makes the spins want to anti-align.
- A positive B value makes the spins want to orient positively.
- A negative B value makes the spins want to orient negatively - that is, they want to align with the magnetic field.

Each state of the Ising system is an specific "setting" of the spins in the lattice. For N sites in the lattice, there are 2^N states. Earlier we defined the partition function Z as

$$Z = \sum_\mu e^{-\beta E_\mu}$$

where $\{\mu\}$ represents a set of states. So our Ising partition function is given by

$$Z = \sum_{s_1=\pm 1} \sum_{s_2=\pm 1} ... \sum_{s_N=\pm 1} e^{-\beta H} = \sum_{\{s_i\}} e^{-\beta H}$$

From the partition function, we can calculate the internal energy, entropy, free energy, specific heat, and so on (see section 1.2). The mean magnetization $\langle M \rangle$ is given by

$$\langle M \rangle = \left\langle \sum_i s_i \right\rangle$$

and the mean magnetization per spin $\langle m \rangle$ is just

$$\langle m \rangle = \frac{1}{N} \langle M \rangle$$

The magnetic susceptibility is calculated as

$$\frac{\partial \langle M \rangle}{\partial B} = \beta(\langle M^2 \rangle - \langle M \rangle^2)$$

and the magnetic susceptibility per spin is calculated as

$$\chi = \frac{1}{N} \frac{\partial \langle M \rangle}{\partial B}$$

The specific heat is given by

$$c = \frac{k\beta^2}{N}(\langle E^2 \rangle - \langle E \rangle^2)$$

If we introduce a spatially-varying magnetic field by subscripting B, our Hamiltonian becomes

$$H = -J \sum_{\langle ij \rangle} s_i s_j - \sum_i B_i s_i$$

and the mean magnetization per site becomes

$$\langle m_i \rangle = \langle s_i \rangle = \frac{1}{\beta} \frac{\partial \log Z}{\partial B_i}$$

which allows us to write down the connected correlation function

$$G_c^{(2)}(i,j) = \frac{1}{\beta^2} \frac{\partial^2 \log Z}{\partial B_i \partial B_j}$$

1.5 Numerical Methods

There are several computational methods that can be used to solve for quantities of interest in a given model. The most straightforward method involves mapping the model of interest to a lattice of finite size so that the partition function is exactly calculable by computer.

1.5.1 Example: 2D Ising Model

We start with a two-dimensional lattice of 25 spins, arranged in a 5×5 grid. We assume periodic boundary conditions such that spins at one end of the lattice experience interations with spins on the other end. We also start with no externak magnetic field ($B = 0$).

There are 2^{25} possible configurations, or states. However, we can utilize the symmetry of the system to argue that every state has a reflective partner where every spin is flipped. If $B = 0$, this state must have the same energy as its reflective partner. Therefore, we only need to consider one of each pair in the calculation of our partition function and then double the sum.

To calculate the mean magnetization per spin, we can use

$$\langle m \rangle = \frac{1}{N} \langle M \rangle = \frac{1}{N} \left\langle \sum_i s_i \right\rangle$$

and to calculate the specific heat per spin we use

$$C = \frac{\partial U}{\partial T} = -k\beta^2 \frac{\partial U}{\partial \beta} = k\beta^2 \frac{\partial^2 \log Z}{\partial \beta^2}$$

To do: calculate partition function, specific heat and mean magnetization and plot last two. Compare against closed form equations by Onsager. The key conclusion is that the physics is not well modeled on a finite lattice around the critical point ($kT = 2.3J$), but we stil get fairly accurate results away from this point. It is still desirable to study the largest finite-sized system possible to most accurately reason about the properties of the system in the thermodynamic limit. However, brute force calculations scale rapidly in time with the lattice size. We must find other ways to calculate quantities of interest on a large lattice size.

1.5.2 Monte Carlo Methods

To calculate the partition function of a larger lattice for the Ising model, we turn to Monte Carlo simulation. The fundamental idea is to simulate the thermal fluctuations from state to state - in accordance with a set of weights $w_\mu(t)$ as they would be in the real system. We then take a time average of the quantity of interest, which is regarded as an expectation value of that quantity.

For this approach to work, we must find an accurate dynamics for the system that will govern its time evoluation. That is, we must determine the master equation of the system. We do this by

- Choosing a set of rates $R(\mu \to \nu)$ for the transitions.
- Forcing the equilibrium solution for the master equation to be the Boltzmann distribution.
- Use this set of rates to time-evolve the simulated system.

- Estimate the observable quantity in which we're interested.

This approach has the benefit that we do not need to recreate every state to get an accurate solution. We only need to time-evolve through a small fraction of states. However, there are statistical errors that arise from not recreating every state (these can be reduced by lengthening the simulation time). For this reason, it is not wise to calculating the partition function and use this to calculate other values. Instead, we should calculate quantities of interest directly (to reduce the noise of the result).

1.6 Problems

- **1.1** "If a system is in equilibrium with a thermal reservoir at temperature T, the probability of its having a total energy E varies with E in proportion to $e^{-\beta E}$." True of False?

  False. If a system is in equilibrium with a thermal reservoir of temperature T, the probability of its having a total energy E is given by its probability of being in any state with characteristic energy E. This can be written as

  $$P_E = \frac{1}{Z} \sum_{E_\mu = E} e^{-\beta E_\mu}$$

  The partition function factor $1/Z$ is found in all P_E values for any given E. We can also note that, by definition, all E_μ values in the sum of exponentials are simply E. Therefore, if we know the degeneracy of the energy value E (denoted g_E), then the probability of the system having total energy E is given by

  $$P_E = \frac{1}{Z} g_E e^{-\beta E} \propto g_E e^{-\beta E}$$

- **1.2** A certain simple system has only two energy states, with energies E_0 and E_1, and transitions between the two states take place at rates $R(0 \to 1) = R_0 \exp[-\beta(E_1 - E_0)]$ and $R(1 \to 0) = R_0$. Solve the master equation for the probabilities w_0 and w_1 of occupation of the two states as a functions of time with the initial conditions $w_0 = 0$, $w_1 = 1$. Show that as $t \to \infty$ these solutions tend to the Boltzmann probabilities.

  The general form of the master equation is

  $$\frac{dw_\mu}{dt} = \sum_\nu [w_\nu(t) R(\nu \to \mu) - w_\mu(t) R(\mu \to \nu)]$$

  For our two state system, we obtain two coupled equations:

  $$\begin{aligned} \frac{dw_0}{dt} &= w_1(t) R(1 \to 0) - w_0(t) R(0 \to 1) \\ &= w_1(t) R_0 - w_0(t) R_0 \exp[-\beta(E_1 - E_0)] \end{aligned} \tag{2}$$

and

$$\begin{aligned}\frac{dw_1}{dt} &= w_0(t)R(0 \to 1) - w_1(t)R(1 \to 0) \\ &= w_0(t)R_0 \exp[-\beta(E_1 - E_0)] - w_1(t)R_0\end{aligned} \tag{3}$$

which we can write as

$$\begin{bmatrix}\dot{w_0} \\ \dot{w_1}\end{bmatrix} = \begin{bmatrix}-R_0 \exp[-\beta(E_1 - E_0)] & R_0 \\ R_0 \exp[-\beta(E_1 - E_0)] & -R_0\end{bmatrix}\begin{bmatrix}w_0 \\ w_1\end{bmatrix}$$

Solving for $w_0(t)$ and $w_1(t)$ involves solving the eigenvalue problem for the 2×2 matrix above. The characteristic equation is

$$\begin{vmatrix}-R_0 \exp[-\beta(E_1 - E_0)] - \lambda & R_0 \\ R_0 \exp[-\beta(E_1 - E_0)] & -R_0 - \lambda\end{vmatrix} = 0$$

$$(-R_0 \exp[-\beta(E_1 - E_0)] - \lambda)(-R_0 - \lambda) - R_0^2 \exp[-\beta(E_1 - E_0)] = 0$$

$$\lambda(R_0 \exp[-\beta(E_1 - E_0)] + R_0) + \lambda^2 = 0$$

Our two eigenvalues are $\lambda = 0$ and $\lambda = -R_0(\exp[-\beta(E_1 - E_0)] + 1)$ For simplicity, we will temporarily set $\alpha = \exp[-\beta(E_1 - E_0)]$. The associated eigenvectors are

$$C_1 \begin{bmatrix}1 \\ \alpha\end{bmatrix}$$

and

$$C_2 \begin{bmatrix}1 \\ -1\end{bmatrix}$$

Where C_1 and C_2 are arbitrary constants. Our solutions are therefore given by

$$w_0(t) = C_1 + C_2 e^{-R_0(\alpha+1)t}$$
$$w_1(t) = \alpha C_1 - C_2 e^{-R_0(\alpha+1)t}$$

Using the initial condition $w_0(0) = 0$, we discover $C_1 = -C_2$. Then, using the initial condition $w_1(0) = 1$, we find that $C_1 = 1/(\alpha + 1)$. Putting all of this together, we obtain the final solutions

$$w_0(t) = \frac{1 - e^{-R_0(\alpha+1)t}}{\alpha + 1}$$
$$w_1(t) = \frac{\alpha + e^{-R_0(\alpha+1)t}}{\alpha + 1}$$

In the limit $t \to \infty$,

$$\begin{aligned}w_0(t) &= \frac{1}{\alpha + 1} \\ &= \frac{1}{e^{-\beta(E_1 - E_0)} + 1} \\ &= \frac{e^{-\beta E_0}}{e^{-\beta E_1} + e^{-\beta E_0}} \\ &= \frac{e^{-\beta E_0}}{Z}\end{aligned}$$

$$
\begin{aligned}
w_1(t) &= \frac{\alpha}{\alpha+1} \\
&= \frac{e^{-\beta(E_1-E_0)}}{e^{-\beta(E_1-E_0)}+1} \\
&= \frac{e^{-\beta E_1}}{e^{-\beta E_1}+e^{-\beta E_0}} \\
&= \frac{e^{-\beta E_1}}{Z}
\end{aligned}
$$

So, we can see that in the limit $t \to \infty$, we recover the Boltzmann distribution.

- **1.3** A slightly more complex system contains N distinguishable particles, each of which can be in one of two boxes. The particles in the first box have energy $E_0 = 0$ and the particles in the second have energy E_1, and particles are allowed to move back and forward between the boxes under the influence of thermal excitations from a reservoir at temperature T. Find the partition function for this system and then use this result to calculate the internal energy.

  The partition function is given by

$$
Z = \sum_{\mu} e^{-\beta E_\mu}
$$

  where $\{\mu\}$ represents the set of possible states. For a system of N particles, each of which can be in one of two possible states, we have 2^N possible system configurations. Since $E_0 = 0$, E_μ is simply given by $g_1 E_1$ where g_1 is the number of particles in the second box. There is one way to place 0 particles ($E_\mu = 0$) in the second box. There are N ways to place 1 particle ($E_\mu = E_1$) in the second box. There are $N(N-1)/2$ ways to place 2 particles ($E_\mu = 2E_1$) in the second box. Continuing this pattern, we find that the partition function is given by

$$
Z = \sum_{n=0}^{N} \binom{N}{n} e^{-\beta n E_1}
$$

  where the binomial coefficient determines the degeneracy of energy.

  To calculate the internal energy, the expectation value of the energy E, we can use

$$
\begin{aligned}
U &= \frac{1}{Z} \sum_{\mu} E_\mu e^{-\beta E_\mu} \\
&= \frac{1}{Z} \sum_{n=0}^{N} n E_1 \binom{N}{n} e^{-\beta n E_1}
\end{aligned}
$$

Alternatively,

$$\begin{aligned} U &= -\frac{1}{Z}\frac{\partial Z}{\partial \beta} \\ &= -\frac{\partial \log Z}{\partial \beta} \end{aligned}$$

- **1.4** Solve the Ising model in one dimension for the case $B = 0$ as follows. Define a new set of variables σ_i which take values 0 and 1 according to $\sigma_i = \frac{1}{2}(1 - s_i s_{i+1})$ and rewrite the Hamiltonian in terms of these variables for a system of N spins with periodic boundary conditions. Show that the resulting system is equivalent to the one studied in Problem 1.3 in the limit of large N and hence calculate the internal energy as a function of temperature.

  As directed, we define $\sigma_i = \frac{1}{2}(1 - s_i s_{i+1})$. The Hamiltonian of the Ising model in one dimension (with $B = 0$) is given by

$$\begin{aligned} H &= -J \sum_i s_i s_{i+1} \\ &= -J \sum_i s_i s_{i+1} + 1 - 1 \\ &= -J \sum_i -(1 - s_i s_{i+1}) + 1 \\ &= -J \sum_i -2\sigma_i + 1 \end{aligned}$$

  We've now reduced the problem to consideration of a set of variables σ_i, which can only take on the values 0 (if $s_i = s_{i+1}$) and 1 (if $s_i \neq s_{i+1}$). In this way, we've introduced a new binary "spin-like" variable. The partition function of this system is given by

$$\begin{aligned} Z &= \sum_{\{\sigma_i\}} e^{-\beta J \sum_i (2\sigma_i - 1)} \\ &= \sum_{\{\sigma_i\}} e^{\sum_i \beta J} e^{-\beta J \sum_i 2\sigma_i} \end{aligned}$$

  Since each individual σ_i can only take on the values 0 and 1, the quantity $\sum_i 2\sigma_i$ is simply given by $2n$ where n is the number of σ_i variables with the value 1. However, the number of ways of obtaining $2n$ in the exponent is equivalent to the number of ways of choosing n out of N spin-like variables σ_i to be positive. So, we obtain

$$Z = \sum_{n=0}^{N} e^{N\beta J} \binom{N}{n} e^{-\beta J 2n}$$

  Which very similar to the partition function we obtained in the previous problem.

2 Principles of Equilibrium Thermal Monte Carlo Simulation

Key concepts in this section include:

- Importance sampling
- Detailed balance
- Acceptance ratios

2.1 The Estimator

The goal in Monte Carlo simulations is to calculate an expectation value for an observable Q, $\langle Q \rangle$. Ideally, we calculate this value explicitly by averaging over all possible states. But this is not feasible for large systems. The Monte Carlo approach uses knowledge about the system to estimate the average from a randomly selected subset of states (the states are selected randomly, but according to the probability distribution that defines the system).

If we choose M states $\mu_1, \mu_2, ..., \mu_M$, the quantity Q_M is defined as

$$Q_M = \frac{\sum_{i=1}^{M} Q_{\mu_i} p_{\mu_i}^{-1} e^{-\beta E_{\mu_i}}}{\sum_{j=1}^{M} p_{\mu_j}^{-1} e^{-\beta E_{\mu_j}}}$$

This quantity is known as the *estimator* of Q. As the number of states sampled, M, increases Q_M becomes an increasingly accurate estimate for $\langle Q \rangle$. As $M \to \infty$, Q_M becomes $\langle Q \rangle$.

The problem then becomes choosing p_μ values such that the M chosen states to calculate the estimator actually represent the relatively important terms in the expectation value. The technique for choosing these states is known as *importance sampling*.

2.2 Importance Sampling

The goal is to select sample states of the system proportional to the Boltzmann weight of the state. In other words, sampled states are more likely to have a high probability of realization in the simulated system - they are more "relevant" for expectation value calculation. The probability of a particle state μ being chosen should be given by

$$p_\mu = \frac{e^{-\beta E_\mu}}{Z}$$

and therefore our estimator is now given by

$$
\begin{aligned}
Q_M &= \frac{\sum_{i=1}^{M} Q_{\mu_i} p_{\mu_i}^{-1} e^{-\beta E_{\mu_i}}}{\sum_{j=1}^{M} p_{\mu_j}^{-1} e^{-\beta E_{\mu_j}}} \\
&= \frac{\sum_{i=1}^{M} Q_{\mu_i} e^{\beta E_{\mu_i}} e^{-\beta E_{\mu_i}}}{\sum_{j=1}^{M} e^{\beta E_{\mu_j}} e^{-\beta E_{\mu_j}}} \\
&= \frac{1}{M} \sum_{i=1}^{M} Q_{\mu_i}
\end{aligned}
$$

2.2.1 Markov Processes

To actually choose a good set of random states, we rely on *Markov processes* to generate the states appropriately. A Markov processes generates a Markov chain of states. If we start with state μ, the Markov process will generate a state ν according to the transition probability $P(\mu \to \nu)$. This process is repeated over a long enough time such that the states chosen represent the Boltzmann distribution. To eventually come to equilibrium (meaning that our set of states represent the Boltzmann distribution), we require the following:

- Transition probabilities must satisfy $\sum_\nu P(\mu \to \nu) = 1$
- The value of $P(\mu \to \nu)$ for any μ, ν does not vary over time.
- $P(\mu \to \mu)$ may be non-zero.
- *Ergodicity*: it should be possible for the Markov process to reach any state of the system from any other state.
- *Detailed balance*: the Boltzmann probability distribution is obtained, as opposed to another distribution. This can be enforced by applying the condition of detailed balance,

  $$p_\mu P(\mu \to \nu) = p_\nu P(\nu \to \mu)$$

  where p_μ, p_ν are Boltzmann probabilities.

2.3 Acceptance Ratios

It can be difficult to define the set of transition probabilities $P(\mu \to \nu)$ that correctly characterizes the ideal Markov process. There are a set of standard methods (e.g. Metropolis) but we are likely to get far more accurate results by coming up with our own approach, which could be tedious or involve trial-and-error. One way to simplify this process is to note that $P(\mu \to \mu)$ can be non-zero. So, we can start with $P(\mu \to \mu) = 1$ (satisying the detailed balance and normalization conditions), and tweak the other $P(\mu \to \nu)$ values incrementally (with appropriate changes made to $P(\mu \to \mu)$).

Another way to describe this process is to write the transition probabilities in the following way

$$P(\mu \to \nu) = g(\mu \to \nu)A(\mu \to \nu)$$

where

- $g(\mu \to \nu)$ is the *selection probability*, the probability given an initial state μ that our algorithm will generate target state ν.
- $A(\mu \to \nu)$ is the *acceptance ratio*, the probability that we accept and transition to the chosen target state ν.

In choosing acceptance ratios, we should strive to keep them as close to 1 as possible. If all acceptance ratios are low, we are likely to perform very many time steps without actually transitioning to another state. By keeping acceptance ratios high, we place more of a burden on the selection probabilities to simulate the true transition probabilities. The ideal scenario involves absolutely correct selection probabilities and unity acceptance ratios. In reality, we shift some of the burden of correctness off onto the acceptance ratio.

2.3.1 Continuous Time Monte Carlo

The basic idea behind continuous time Monte Carlo algorithms is to vary the length of the time step according to the amount of time we expect the system to reject transitions to other states. A basic continuous time Monte Carlo algorithm is the following:

- Calculate probabilities $P(\mu \to \nu)$ for transitions to all states from current state μ. Choose new state ν according to $P(\mu \to \nu)$ and transition to ν.
- Calculate Δt from $P(\mu \to \nu)$ values. Approximately,

  $$\Delta t \approx \frac{1}{\sum_{\nu \neq \mu} P(\mu \to \nu)}$$

  See section 2.4 for more details.
- Increment time t by Δt to mimic effect of "waiting" for Δt time steps.

2.4 Problems

- **2.1** Derive Equation (2.8) from Equation (1.1).

  The goal is to derive

  $$w_\nu(t+1) = \sum_\mu P(\mu \to \nu) w_\mu(t)$$

  from the master equation

  $$\frac{dw_\mu}{dt} = \sum_\nu [w_\nu(t)R(\nu \to \mu) - w_\mu(t)R(\mu \to \nu)]$$

- **2.2** Consider a system which has just three energy states, with energies $E_0 < E_1 < E_2$. Suppose that the only allowed transitions are ones of the form $\mu \to \nu$, where $\nu = (\mu + 1) \mod 3$. Such a system cannot satisfy detailed balance. Show nonetheless that it is possible to choose the transition probabilities $P(\mu \to \nu)$ so that the Boltzmann distribution is an equilibrium of the dynamics.

3 The Ising Model and the Metropolis Algorithm

Brief recap of the Ising model:

- Model of a magnet in which spins s_i are placed on sites i of a lattice.
- Each spin s_i can take on one of two values, $+1$ or -1.
- If there are N sites in the lattice, there are 2^N possible lattice configurations.
- The Hamiltonian is given by

$$H = -J \sum_{\langle ij \rangle} s_i s_j - B \sum_i s_i$$

  where J is the interaction energy between nearest-neighbor spins and B is an external magnetic field.

3.1 The Metropolis Algorithm

A system that is in thermal equilibrium will only ever have energies within a very narrow range. That is, we do not expect a system in equilibrium to experience a drastic change in energy. For this reason, a simulation of a system in equilibrium may only want to consider transitions to states with energies that are very close to the current energy. One way to achieve this is to enforce *single-spin-flip dynamics*, in which only one spin is flipped during any given transition.

If we enforce single-spin-flip dynamics, the new energy can only differ by $2zJ$, where z is the *lattice coordination number*, the number of neighbors that each site on the lattice has. Single-spin-dynamics also guarantees ergodicity.

The Metropolis algorithm for the Ising model with single-spin-flip dynamics is characterized by a set of acceptance ratios defined as

$$A(\mu \to \nu) = \begin{cases} e^{-\beta(E_\nu - E_\mu)} & if E_\nu - E_\mu > 0 \\ 1 & otherwise. \end{cases} \tag{4}$$

That is, we always accept transitions to equal or lower energy states. If the selected state has a higher energy, we accept it with probability $e^{-\beta(E_\nu - E_\mu)}$.

The selection probabilities $g(\mu \to \nu)$ for each of the possible states (ones where only a single spin has been flipped) are all equal, and the selection probabilities for every other state are zero. For a lattice with N sites, there are only N possible new states so the selection probabilities for those new states are simply

$$g(\mu \to \nu) = \frac{1}{N}$$

The condition of detailed balance then requires

$$\frac{P(\mu \to \nu)}{P(\nu \to \mu)} = \frac{g(\mu \to \nu)A(\mu \to \nu)}{g(\nu \to \mu)A(\nu \to \mu)} = \frac{A(\mu \to \nu)}{A(\nu \to \mu)} = e^{-\beta(E_\nu - E_\mu)}$$

Suppose that for two states μ and ν, μ has the lower energy and ν has the higher energy. In other words, $E_\nu - E_\mu > 0$. From our detailed balance condition, we can see that this means that $A(\nu \to \mu)$ is larger than $A(\mu \to \nu)$. So, we set $A(\nu \to \mu)$ equal to one and, therefore, $A(\mu \to \nu)$ must be given by $e^{-\beta(E_\nu - E_\mu)}$.