In each project so far, we have come across a common numerical technique along with its application to a specific area in physics.

In this project, we will deal with a set of techniques useful in situations where we are interested in some sort of averages over a very large "phase space" or space of possible parameters

- The particular problem we will examine is that of phase transitions in a simple classical model of magnetism in statistical mechanics.
- It is useful to review some probability theory as a prelude.

Analyzing Big Data in Physics:

- Problems where the phase space is very large:
  - Classical gas of molecules
  - Classical Ising magnet
  - Stock markets
  - Traffic
- Caution! Analyzing stock markets with statistical mechanics ideas should be done with a lot of care. It is hard to account for the behavior of humans!

Basics of Probability:

Binomial Distribution:

Suppose you have a coin which flips to heads with a probability, p

- Suppose you flip the coin n times.
- The probability of there being precisely k flips to heads is ${}^{n}C_{k}p^{k}(1-p)^{(n-k)}$

Poisson Distribution

- Suppose we know that the average number of patients arriving at an emergency room between 10 and 11 pm is 5.
- What is the probability that 4 patients show up on a given night?
  - For any given person, the probability of showing up to an emergency room is some small number, p
  - The number of people who would visit a particular emergency room is, say N.
  - The answer is then: $\binom{N}{k} p^k (1-p)^{n-k}$ where k is 4.
  - However, we know neither N, nor p, but we know $Np = 5$.
  - It turns out that the Answer is well approximated by : $e^{-5} \frac{5^4}{4!}$

Poisson Limit Theorem:

- Let p_n be a sequence of real numbers in $[0, 1]$ such that the sequence np_n converges to a finite limit λ. Then:
  $\lim_{n\to\infty} \binom{n}{k} p_n^k (1 - p_n)^{n-k} = e^{-\lambda} \frac{\lambda^k}{k!}$

Proof:

$$\begin{aligned}
&\binom{n}{k} p^k (1-p)^{n-k} \\
&\simeq \lim_{n\to\infty} \frac{n(n-1)(n-2)\dots(n-k+1)}{k!} \left(\frac{\lambda}{n}\right)^k \left(1-\frac{\lambda}{n}\right)^{n-k} \\
&= \lim_{n\to\infty} \frac{n^k + O\left(n^{k-1}\right)}{k!} \frac{\lambda^k}{n^k} \left(1-\frac{\lambda}{n}\right)^{n-k} \\
&= \lim_{n\to\infty} \frac{\lambda^k}{k!} \left(1-\frac{\lambda}{n}\right)^{n-k}
\end{aligned}$$

- Since $\lim_{n\to\infty} \left(1-\frac{\lambda}{n}\right)^n = e^{-\lambda}$

and $\lim_{n\to\infty} \left(1-\frac{\lambda}{n}\right)^{-k} = 1$

This leaves $\binom{n}{k} p^k (1-p)^{n-k} \simeq \frac{\lambda^k e^{-\lambda}}{k!}$

A motivating example of a large-ish phase space

Suppose we have a random number generator at our disposal and we want to get an estimate of π using this random number generator. How do we do this?

- We can generate pairs of integers (x, y) where $-1 \leq x, y \leq 1$ and count the fraction of such pairs such that $x^x + y^2 < 1$. As we produce more and more data, this fraction should start to approach $\pi/4$. Here we see an example of a problem, where we desire to sample a phase space (the interior of a square) in a random way and eight point is weighted equally in our final calculation.
- In statistical physics, we want to sample over phase space with a Boltzmann weight. How can we do that using a numerical algorithm?

It is useful to introduce some terminology:

- A stochastic process is a time dependent random process and is described by a set of time-dependent random variables, usually prescribed at a set of discrete time points. A simple example is the position of a gas molecule undergoing collisions. For some more details, see the Appendix.
- A Markov chain is a particular type of stochastic process where they time dependent variable is a state

Markov Chains

– We describe a Markov chain as follows: We have a set of states, $S = s_1, s_2, \ldots, s_r$. The process starts in one of these states and moves successively from one state to another. Each move is called a step. If the chain is currently in state s_i, then it moves to state s_j at the next step with a probability denoted by p_{ij} , and this probability does not depend upon which states the chain was in before the current.

Markov Chains

– The probabilities p_{ij} are called transition probabilities. The process can remain in the state it is in, and this occurs with probability p_{ii}. An initial probability distribution, defined on \$S, specifies the starting state. Usually this is done by specifying a particular state as the starting state. A picturesque description of a Markov chain (due to R. Howard) is of a frog jumping on a set of lily pads. The frog starts on one of the pads and then jumps from lily pad to lily pad with the appropriate transition probabilities.

Markov Chains II

– Example 11.1 (From Kemeny, Snell, and Thompson) The Land of Oz is blessed by many things, but not by good weather. They never have two nice days in a row. If they have a nice day, they are just as likely to have snow as rain the next day. If they have snow or rain, they have an even chance of having the same the next day. If there is change from snow or rain, only half of the time is this a change to a nice day. With this information we form a Markov chain as follows. We take as states the kinds of weather R, N, and S. From the above information we determine the transition probabilities. These are most conveniently represented in a square array as

$$\mathbf{P} = \begin{array}{c} \\ \mathrm{R} \\ \mathrm{N} \\ \mathrm{S} \end{array} \begin{array}{c} \begin{array}{ccc} \mathrm{R} & \mathrm{N} & \mathrm{S} \end{array} \\ \begin{pmatrix} 1/2 & 1/4 & 1/4 \\ 1/2 & 0 & 1/2 \\ 1/4 & 1/4 & 1/2 \end{pmatrix} \end{array} .$$

Markov Chains III

- The entries in the first row of the matrix P in Example 11.1 represent the probabilities for the various kinds of weather following a rainy day. Similarly, the entries in the second and third rows represent the probabilities for the various kinds of weather following nice and snowy days, respectively. Such a square array is called the matrix of transition probabilities, or the transition matrix .
- The key property of these Markov chains is the following: For some systems (including classical magnets) with a probability distribution of states, Markov chains, (i.e., appropriate transition matrices) may be chosen such that the long time distribution of states in the chain is the same as the probability distribution of states in the system (thermal Boltzmann distribution of states for the classical magnet). More precise statements about when such a behavior happens can be found in the Appendix.

Application in Stat Mech:

- You have a system that has a certain set of states $1, 2, \ldots N$ each of which has an associated probability π_n, but you don't necessarily know the π_n. You do, however, know the relative probabilities π_i/π_j. More precisely, π_i can be computed, but is harder to compute than simply π_i/π_j.
- You want to generate a sample $\sigma_1, \sigma_2, \ldots, \sigma_M$ with the property that as $N \to \infty$, the number of times we get state i divided by the number of samples approaches its associated probability (number of times state i appears)$/M \to \pi_n$
- We try to find a Markov chain that does this.
- We notice that the Metropolis algorithm generates precisely such a Markov chain, so we run the Metropolis algorithm to generate the sample.

Statistical Mechanics:

– Consider a system with a finite number of states $1, 2, \ldots, N$ with associated probabilities $\pi_1, \pi_2, \ldots, \pi_n$. This means that

$$\sum_{n=1}^{N} \pi_n = 1, \qquad \pi_n \geq 0. \tag{1}$$

An observable O of such a system is a physical quantity that has a certain value o_n if the system occupies state n.

Statistical Average:

– The average value of an observable is defined as

$$\langle O \rangle = \sum_{n=1}^{N} \pi_n o_n. \tag{2}$$

For example, for a system in equilibrium with a heat bath, the the states could be states of definite energy, the probabilities would be the probability of finding the system in one of these states, and the energy itself would be an observable – each state has a certain associated energy value.

Computing Statistical Averages:

- It can often be difficult to compute averages of observables for physical systems for various reasons. For example, the system might have an enormous number of states causing naive numerical evaluation of averages to be difficult or impossible. In such cases, a common workaround is to generate a representative random sample $\sigma_1, \sigma_2, \ldots, \sigma_M$ of states and compute the sample average

$$\langle O \rangle_M = \frac{1}{M} \sum_{n=1}^{N} o_{\sigma_n} \tag{3}$$

Computing Stat Avgs

- As M becomes large, the sample average will converge to the average provided the samples are chosen appropriately;

$$\langle O \rangle_M \to \langle O \rangle, \qquad M \to \infty. \tag{4}$$

- A powerful method for generating random samples and thus allowing one to compute sample averages is to use a Markov chain to generate a sequence of random samples. This approach is called Markov Chain Monte Carlo (MCMC). In a computational setting, the Markov chain can be generated according to some algorithm. Perhaps the most well-known, simplest, and most elegant of all of these is the so-called **Metropolis-Hastings algorithm**:

Metropolis Hastings algorithm:

- Initiate the sequence of samples in a state σ_1 of your choosing, and successively generate states in the following way:
- If σ is the current state, propose a new state σ' according to a conditional probability distribution q, namely $q(\sigma'|\sigma)$ is the probability of proposing σ' given the current state is σ.
- Accept the proposed state, namely add it as the next state in the sample sequence, with probability

$$A(\sigma'|\sigma) = \min\left(1, \frac{\pi_{\sigma'}}{\pi_\sigma}\frac{q(\sigma|\sigma')}{q(\sigma'|\sigma)}\right) \tag{5}$$

Metropolis Hastings algorithm:

- If the state is not accepted, in other words if it is rejected, add the current state σ to your sequence instead as the next state.
- Repeat these steps for a large number of samples.
- The proposal distribution is often chosen to be symmetric

$$q(\sigma|\sigma') = q(\sigma'|\sigma). \tag{6}$$

} Choosing this sort of proposal distribution simplifies the acceptance probability to

$$A(\sigma'|\sigma) = \min\left(1, \frac{\pi_{\sigma'}}{\pi_\sigma}\right), \tag{7}$$

and in this case the algorithm is often called simply the Metropolis algorithm.

Ising Model:

- The 2D Ising model is a simple model of a ferromagnetic material with a phase transition.
- The model consists of a 2D lattice $L \times L$ of spins $s_i \in \{-1, +1\}$. Each spin interacts only with its nearest neighbours. The energy is expressed as

$$E(\{s_i\}) = -\sum_{\langle i,j \rangle} s_i s_j - H \sum_i s_i, \tag{8}$$

where $\{s_i\}$ is notation for the entire configuration of spins, H is the external magnetic field, and $\langle i, j \rangle$ implies a summation over all nearest-neighbour pairs.

Ising Model:

– We have normalized energy with J, spin with $\hbar/2$, and magnetic field with J/μ, where J is the exchange energy and μ is the atomic magnetic moment. When the system is in contact with a heat bath at temperature T the equilibrium probability density is the Boltzmann distribution

–

$$\rho(\{s_i\}) = Z^{-1} \exp(-E(\{s_i\})/T), \tag{9}$$

where the partition function Z is the sum of exponential factors $\exp(-E/T)$ over all possible configurations.

Phase Transitions

– Below the critical temperature

$$T_c = \frac{2}{\log(1+\sqrt{2})} \approx 2.2692 \tag{10}$$

the spins are perferentially aligned in a given direction; above T_c the spins have no mean orientation for $H = 0$.

Appendix

Stochastic Process

In probability theory and related fields, a stochastic or random process is a mathematical object usually defined as a collection of random variables. Historically, the random variables were associated with or indexed by a set of numbers, usually viewed as points in time, giving the interpretation of a stochastic process representing numerical values of some system randomly changing over time, such as the growth of a bacterial population, an electrical current fluctuating due to thermal noise, or the movement of a gas molecule

Note: A stochastic process is a time dependent process depending on randomness. From a mathematical point of view, a stochastic process $Y_X(t)$ is a random variable Y , which is a function of another random variable X and time $t \geq 0$, i.e. $Y_X(t) = f(X, t)$.)

Markov Chains

– We describe a Markov chain as follows: We have a set of states, $S = s_1, s_2, \ldots, s_r$. The process starts in one of these states and moves successively from one state to another. Each move is called a step. If the chain is currently in state s_i, then it moves to state s_j at the next step with a probability denoted by p_{ij} , and this probability does not depend upon which states the chain was in before the current.

Markov Chains

– The probabilities p_{ij} are called transition probabilities. The process can remain in the state it is in, and this occurs with probability p_{ii}. An initial probability distribution, defined on S, specifies the starting state. Usually this is done by specifying a particular state as the starting state. A picturesque description of a Markov chain (due to R. Howard) is of a frog jumping on a set of lily pads. The frog starts on one of the pads and then jumps from lily pad to lily pad with the appropriate transition probabilities.

Markov Chains II

- Example (From Kemeny, Snell, and Thompson): The Land of Oz is blessed by many things, but not by good weather. They never have two nice days in a row. If they have a nice day, they are just as likely to have snow as rain the next day. If they have snow or rain, they have an even chance of having the same the next day. If there is change from snow or rain, only half of the time is this a change to a nice day. With this information we form a Markov chain as follows. We take as states the kinds of weather R, N, and S. From the above information we determine the transition probabilities. These are most conveniently represented in a square array as

$$\mathbf{P} = \begin{array}{c} \\ \mathrm{R} \\ \mathrm{N} \\ \mathrm{S} \end{array} \begin{array}{c} \begin{array}{ccc} \mathrm{R} & \mathrm{N} & \mathrm{S} \end{array} \\ \begin{pmatrix} 1/2 & 1/4 & 1/4 \\ 1/2 & 0 & 1/2 \\ 1/4 & 1/4 & 1/2 \end{pmatrix} \end{array} .$$

Markov Chains III

- The entries in the first row of the matrix P in the above Example represent the probabilities for the various kinds of weather following a rainy day. Similarly, the entries in the second and third rows represent the probabilities for the various kinds of weather following nice and snowy days, respectively. Such a square array is called the matrix of transition probabilities, or the transition matrix .
- **Theorem** Let P be the transition matrix of a Markov chain. The ijth entry $p(n)$ of the matrix P^n gives the probability that the Markov chain, starting in state s_i, will be in state $s_j$after n steps.

Markov Chains IV

- We now consider the long-term behavior of a Markov chain when it starts in a state chosen by a probability distribution on the set of states, which we will call a probability vector. A probability vector with r components is a row vector whose entries are non-negative and sum to 1. If u is a probability vector which represents the initial state of a Markov chain, then we think of the i th component of u as representing the probability that the chain starts in state s_i.
- Let P be the transition matrix of a Markov chain, and let u be the probability vector which represents the starting distribution. Then the probability that the chain is in state s_i after n steps is the i th entry in the vector $u(n) = uP^n$.

Absorbing Chains

- Definition: A state s_i of a Markov chain is called absorbing if it is impossible to leave it (i.e., $p_{ii} = 1$). A Markov chain is absorbing if it has at least one absorbing state, and if from every state it is possible to go to an absorbing state (not necessarily in one step).
- Definition: In an absorbing Markov chain, a state which is not absorbing is called transient.

Ergodic Chains

- A second important kind of Markov chain we shall study in detail is an ergodic Markov chain, defined as follows.
- Definition: A Markov chain is called an *ergodic* chain if it is possible to go from every state to every state (not necessarily in one move). In many books, ergodic Markov chains are called irreducible.
- Definition: A Markov chain is called a *regular* chain if some power of the transition matrix has only positive elements. P In other words, for some n, it is possible to go from any state to any state in exactly n steps. It is clear from this definition that every regular chain is ergodic. On the other hand, an ergodic chain is not necessarily regular, as the following examples show.

Regular Markov Chains

– Any transition matrix that has no zeros determines a regular Markov chain. However, it is possible for a regular Markov chain to have a transition matrix that has zeros. The transition matrix of the Land of Oz example of Section 11.1 has $P_{NN} = 0$ but the second power P^2 has no zeros, so this is a regular Markov chain.

Fundamental Limit Theorem

– **Theorem**: Let P be the transition matrix for a regular chain. Then, as $n \to \infty$, the powers P^n approach a limiting matrix W with all rows the same vector w. The vector w is a strictly positive probability vector (i.e., the components are all positive and they sum to one).

Regular Markov Chains

- **Theorem** Let P be a regular transition matrix, let $W = \lim_{n\to\infty} P_n$, let w be the common row of W, and let c be the column vector all of whose components are 1. Then
- (a) $wP = w$, and any row vector v such that $vP = v$ is a constant multiple of w.
- (b) $Pc = c$, and any column vector x such that $Px = x$ is a multiple of c.