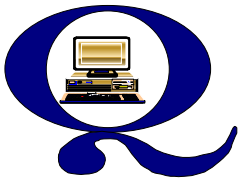


CES Engineering

SIMULATION SYSTEMS

MARKOV CHAIN MONTE CARLO SAMPLING

Prof. Yasser Mostafa Kadah

Motivation

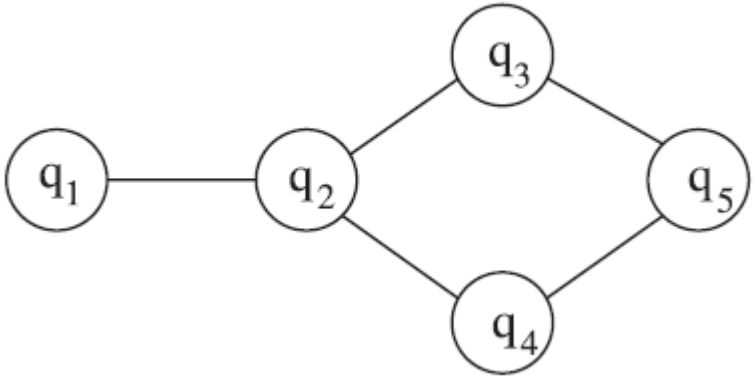
- One of our major concerns in studying simulation methods is figuring out how to sample efficiently from a complicated distribution
  - We may want accurate sampling of a small portion of the probability space
  - If we construct a Markov model of particles moving in a diffuse space and we want accurate sampling of the rate of particle collisions, we may have to simulate the model for an extremely long time to observe any collisions

Metropolis Method

- Metropolis models are very useful for looking at problems in thermodynamics
- It is a technique for using a Markov model to determine the thermodynamic equilibrium of a system of discrete states for which we know potential energies
- Method creates a Markov model whose stationary distribution is the distribution of states at the thermodynamic equilibrium of the system
- Simulating the Markov model then sampling from the states of the system at thermodynamic equilibrium

Metropolis Method

- Suppose we have a system of five states
- If we define some possible ways of moving between states, then we end up with a system we can represent by a graph in which nodes represent states and edges represent allowed transitions between states



Metropolis Method

- Assume that each state q_i has a potential energy E_i
- Thermodynamics tells us that at equilibrium, the stationary probability of being in state q_i (or π_i) is described by a *Boltzmann distribution*

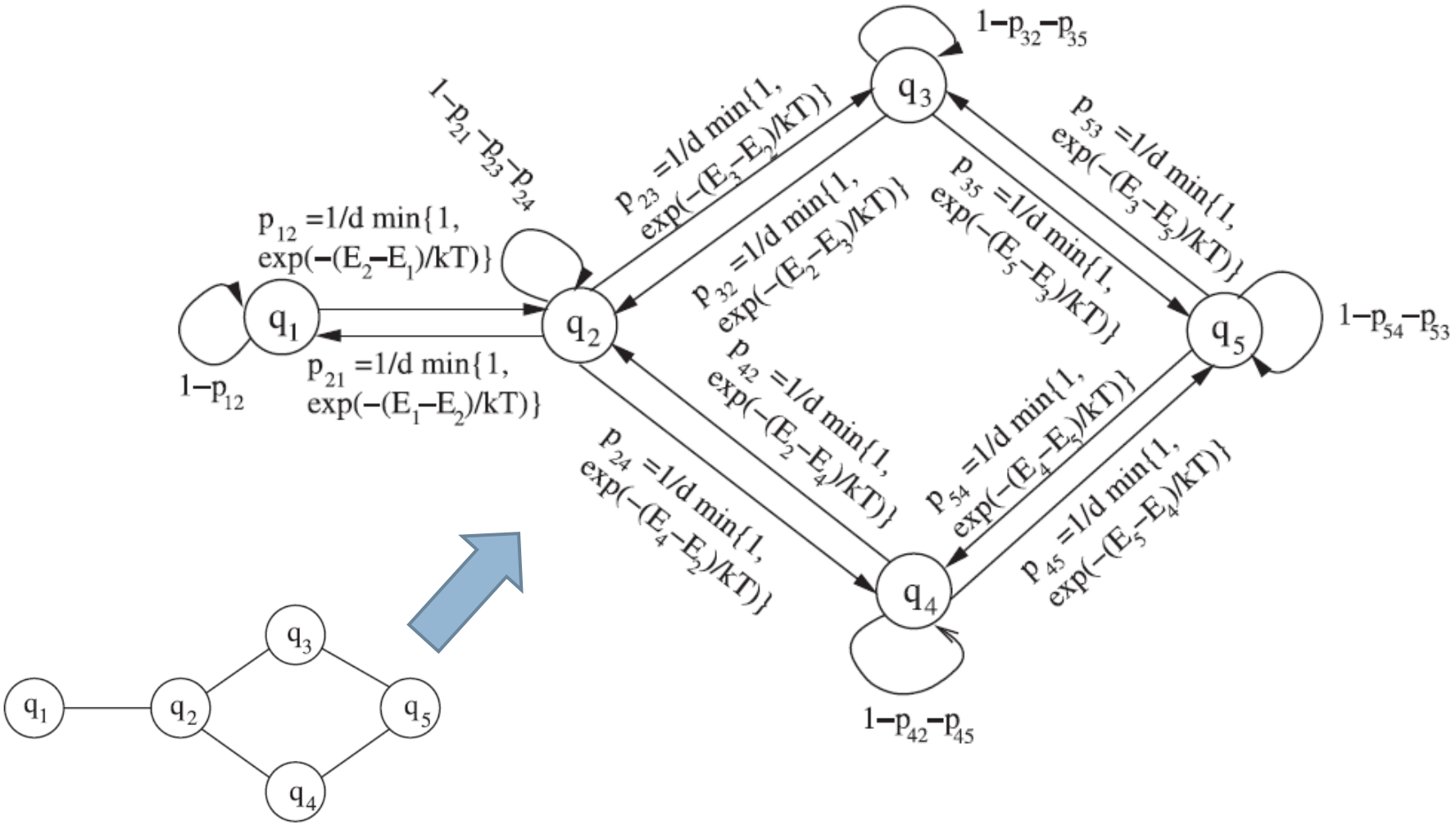
$$\pi_i = \frac{e^{-E_i/kT}}{\sum_{j=1}^{n} e^{-E_j/kT}}$$

  - where k is Boltzmann's constant and T is the absolute temperature

Metropolis Method

- If the number of states is small, then we can calculate this distribution directly
- Metropolis method is helpful when the state set is extremely large and we do not have time to explicitly compute the energy of each state
- The Metropolis method creates a Markov model whose stationary distribution will be the Boltzmann distribution defined by the state energies

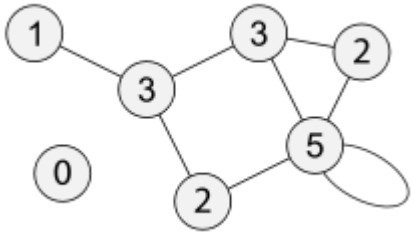
Metropolis Method



Metropolis Method

- Step 1: Pick a random neighbor of qi, which we will call q_j , with probability 1/d , or $q_j=q_i$ with probability $1-d_i/d$ if the degree of node i(d_i) is less than d (maximum degree of any node)
- Step 2: If $E_j \leq E_i$, then move to q_j
- Step 3: If $E_j > E_i$, then with probability $e^{-(Ej-Ei)/kT}$ move to q_j , otherwise stay in q_i

Definition: degree (or valency) of a vertex of a graph is the number of edges incident to the vertex, with loops counted twice



Metropolis Method

- Note that we do not need to know anything about the global structure of the graph or the global energy landscape in order to simulate these transitions.
  - Just need to know the maximum degree of the entire graph, or even an upper bound on it and to have a way to determine the degree of the current node and the energy of each of its neighbors
- If we have a very large state space but spend almost all time at equilibrium in a tiny fraction of the states then the Metropolis method provides a way to sample the equilibrium distribution efficiently without needing to explicitly create the full state graph

Metropolis Method

- We can prove that this model will accurately sample from the Boltzmann distribution.
  - As long as the energies are finite, all of the edges are reversible; if you can go from qi to qj , you can go from q_j to q_i, although generally with different probabilities
- That immediately tells us that the model is ergodic as long as the original graph is connected
  - As long as we wait enough steps, we have a nonzero probability of getting from any node to any other

Metropolis Method

- The Markov models created by the Metropolis method not only have a stationary distribution; they also obey a stronger condition called detailed balance (also known as microreversibility)
  - This property says that given any two states q_i and q_j with transition probabilities p_{ij} and p_{ji} and stationary probabilities π_i and π_j , then,

$$\pi_i p_{ij} = \pi_j p_{ji}.$$

Metropolis Method: Caveats

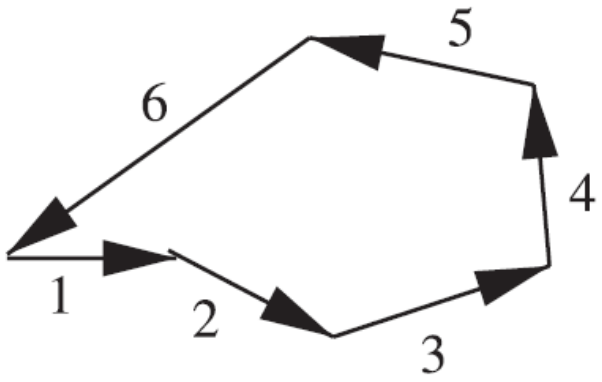
- The Metropolis method is thermodynamically correct, but it is not generally kinetically correct
  - This means that if we have all of the energies right, then the stationary distribution will be the thermodynamic equilibrium distribution, but the pathways between states may not be in any way connected to how the real system will move between states
- Mixing times can be very long, especially if the transitions are poorly chosen
  - Method can easily get stuck in local minima for long periods of time or have difficulty finding the correct trajectories.
  - Difficult to judge whether the method is getting a good sample or is temporarily stuck in one part of the state space

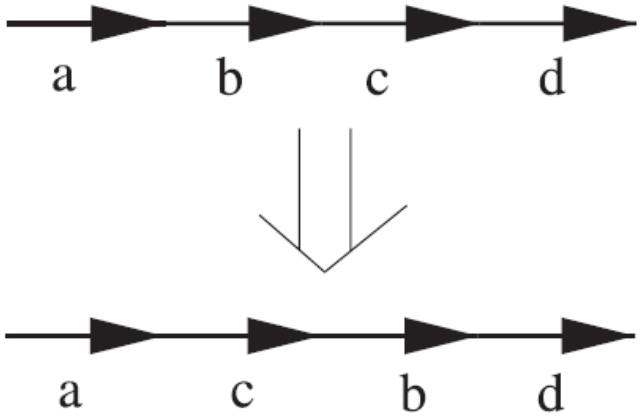
Generalizing the Metropolis Method

- Nothing in the algorithm that requires stationary distribution to be a Boltzmann distribution
  - For any state q_i, we can find transition probabilities to each neighbor state q_j by knowing only the ratio of their stationary probabilities
  - Ignore the energies and work directly with the ratios
    - Step 1: Pick a random neighbor of qi, which we will call q_j , with probability 1/d , or $q_j=q_i$ with probability $1-d_i/d$ if the degree of node i(d_i) is less than d (maximum degree of any node)
    - Step 2: If $\pi_j/\pi_i \geq 1$, then move to q_j
    - Step 3: If $\pi_j/\pi_I < 1$, then with probability π_j/π_i move to q_j , otherwise stay in q_i

Metropolis as an Optimization Method

- Given a discrete optimization problem, we can declare that each possible solution to the problem has an "energy" which is determined by the value of the optimization metric on that solution
- If we define "moves" between possible solutions, then a Metropolis simulation will be expected to move toward low-energy (high-quality) solutions
  - Example: Travelling salesman problem





Metropolis as an Optimization Method

- Simulated Annealing
  - Exactly the same as simulating a Metropolis model, except that we start with high model temperature and we gradually reduce model temperature to reach solution
  - Limit of very high temperature: all transitions are allowed

$$\lim_{T\to\infty} e^{-(E_2-E_1)/kT} = e^0 = 1$$

  - Limit of very low temperature: no transitions are allowed

$$\lim_{T\to 0} e^{-(E_2-E_1)/kT} = e^{-\infty} = 0$$

  - Performance will depend on *cooling schedule*

Gibbs Sampling

- Gibbs sampling is particularly useful for sampling from joint distributions on many variables
- We accomplish this by allowing states of the Markov model to correspond to possible assignments of the full state vector of the jointly distributed variables
- We then allow transitions corresponding to possible changes in a single element of the state vector

Gibbs Sampling

- Assume we are given a joint distribution,

$$Pr\{X_1 = x_1, X_2 = x_2, \ldots, X_n = x_n\} = p(x_1, x_2, \ldots, x_n),$$

  - where R_1;R_2; … ;R_n are the ranges of the respective random variables.

- The state set of the Gibbs sampler is the product of the ranges of the variables:

$$Q = R_1 \times R_2 \times \cdots \times R_n$$

  - That is, there is one state for each possible assignment of values to the random variables

Gibbs Sampling

- There is a transition possible for any change of a single random variable
  - The probability of making any possible transition obeys the following density:

$$Pr\{(x_1, x_2, \ldots, x_i, \ldots, x_n) \rightarrow (x_1, x_2, \ldots, x_i', \ldots, x_n)\}$$

$$= \frac{1}{n} Pr\{X_i = x_i' \mid X_1 = x_1, X_2 = x_2, \ldots, X_{i-1} = x_{i-1}, X_{i+1} = x_{i+1} \ldots, X_n = x_n\}$$

Step 1: Pick a variable Xi uniformly at random.
Step 2: Sample a new value for that one variable from the conditional distribution of that variable, given the current values of all of the other variables:

$$Pr\{X_i = x_i' \mid X_1 = x_1, X_2 = x_2, \ldots, X_{i-1} = x_{i-1}, X_{i+1} = x_{i+1} \ldots, X_n = x_n\}$$

Step 3: Repeat Step 2 to produce a correct sample of the joint distribution of the variables

Importance Sampling

- Metropolis and Gibbs sampling methods provide ways of sampling from distributions that may be hard to express analytically
- Sometimes, though, having a correct sampler is not enough, we also need the sampler to be efficient, in that it gets close to its stationary distribution in a small number of steps
- Importance sampling is a technique we can use when we have a sampler for a given distribution but want to accelerate it

Importance Sampling

- Basic idea is simple: Given a model with state set:

$$Q = \{q_1, \ldots, q_n\}$$

and corresponding stationary distribution:

$$\Pi = \{\pi_1, \ldots, \pi_n\}$$

we construct a new model with the same state set Q but biased stationary distribution,

$$\hat{\Pi} = \{\hat{\pi}_1, \ldots, \hat{\pi}_n\} \qquad \hat{\pi}_i = w_i \pi_i$$

We then sample from distribution $\hat{\Pi}$ but adjust the estimated frequencies for each state qi sampled from $\hat{\Pi}$ by a factor of $1/w_i$

Importance Sampling

- The result is an accurate estimator of π_i, but one in which the variance of the estimates was changed
  - In particular, we generally want to choose the weights w_i such that the variance of the model is reduced, leading to faster estimation of Π
- One common use of importance sampling is to bias a model toward those states that account for the majority of the probability density, thus accelerating estimation of those states

Importance Sampling: Example

- Suppose we have a Monte Carlo model of protein-folding with the following distribution:

$$\Pi = \{\pi_1, \ldots, \pi_n\}$$

- We can generate a modified distribution by attaching a penalty of $e^{k\,ri}$ to each state, where r_i is the radius of gyration of state q_i and k is a scaling constant

$$\hat{\Pi} = \{\pi_1 e^{-kr_1}/Z, \ldots, \pi_n e^{-kr_n}/Z\}$$

Z is a scaling factor used to make the probabilities sum to 1

- Result will be a sampler that is likely to sample compact states more quickly

Importance Sampling: Example

- We know each r_i and can therefore find the vector,

$$\pi_i' = \widehat{\pi}_i \times e^{kr_i} = \pi_i / Z$$

- Solve for Z using the fact that since Π is a probability distribution

$$\sum_i \pi_i = 1 \qquad Z = \left(\sum_i \pi_i' \right)^{-1}$$

- Finally, we can estimate,

$$\pi_i = \pi_i' \times Z$$

Umbrella Sampling

- Special case of importance sampling commonly used in statistical physics problems to give more accurate estimates of frequencies of rare events in a model
- Umbrella sampling proceeds in essentially the same way as general importance sampling, but is biased so that our modified chain spends a disproportionate amount of time in the portion of space we want to estimate accurately
  - For example, reverse weights in previous example to e^{+kr_i}/Z

Umbrella Sampling

- Algorithm similar to that of Importance Sampling but with the new weight functions

1. Scale each Metropolis ratio $\frac{\pi_j}{\pi_i}$ by $e^{k(r_j - r_i)}$ to get a sampler for distribution $\hat{\Pi}$.
2. Scale each estimated $\hat{\pi}_i$ by e^{-kr_i} to get $\pi'_i = \pi_i / Z$.
3. Solve for $Z = (\sum \pi'_i)^{-1}$.
4. Find the unbiased estimates $\pi_i = \pi'_i \times Z$.

Assignments

- Implement the following:
  - Metropolis method
  - Simulated annealing
  - Gibbs sampling
  - Importance sampling
  - Umbrella sampling