02-712 Week 10

Biological Modeling and Simulation

Aidan Jan

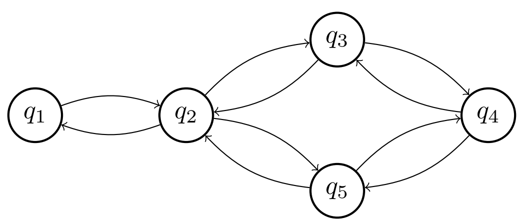
October 28, 2025

Ergolic Graphs

An Ergolic graph is one where every state is reachable from every other state with a non-zero probability. Suppose we have an ergodic graph with states:

$$Q = (q_1, q_2, q_3, q_4, q_5)$$

which includes transitions in both directions between: $(q_1, q_2), (q_2, q_3), (q_3, q_4), (q_4, q_5), (q_2, q_5)$. Note that q_2 has the highest degree in the graph (degree 3).



We will use a model called the metropolis model on this graph:

1. Given a state q_i, pick a neighbor q_j uniformly at random with probability $\frac{1}{d}$ (or q_i again with probability $1 - \frac{d_i}{d}$ if $d_i < d$)
2. If $E_j < E_i$, move to q_i.
3. If $E_j > E_i$, move with probability $e^{-(E_j - E_i)/kT}$
4. Go to step 1.

Remember that E represents the "energy" of a state.
What the description above means is that, let a probability be on every edge. All transitions add up to less than or equal to 1. If less than 1, then the probability of a self loop is the remaining fraction.

Detailed Balance

Detailed balance means that $\pi_i p_{ij} = \pi_j p_{ji}$. Essentially, the probability of being in i and moving to j is equal to the probability of being in j and moving to i. This property is present in *every* metropolis model.

This reason for this is that every metropolis model satisfies the Kolmogarov Criterion. Basically, imagine a set of nodes in a cycle. If moving from one node to another, where there is a *decrease* in energy, then we would move with a probability of $\frac{1}{d}$. If there was an *increase* of energy, then we would move with a probability of $\frac{1}{d}e^{-\Delta_i/kT}$.

The total probability of going around the entire cycle would be

$$\left(\frac{1}{d}\right)^{k+1} e^{-\frac{1}{kT}\sum \Delta_i}, \quad \text{where } i \text{ represents entries where energy increases}$$

Now, if we go around the cycle the other way, then all the signs flip. The number of $\frac{1}{d}$ factors stay the same, but the deltas in the exponential term will be all those that did not appear originally. However, since a manhattan model has the property that the probability of $\pi_i p_{ij} = \pi_j p_{ji}$ as described earlier, these two exponentials must be equal. Therefore, every manhattan model must exhibit detailed balance.

Weighted Sequence Sampling

Suppose we have a DNA sequence, and we want to calculate dinucleotide frequency bias. For example, consider the target sequence GG, and we have the following sequences:

- $ACAGTAC$ - 1
- $ACGGTAC$ - 2
- $ACGGTGG$ - 4
- $ACGGGAC$ - 4

A weight is the number of GG's present in the target sequence, assuming

We can model this with a metropolis model.

- Let the state set $Q = \{A, C, G, T\}^n$.
- Transitions: $NN \cdots (n) \cdots NN \rightarrow NN \cdots (N') \cdots NN$
- Given q_i:
  1. Pick base j uniformly from 1 to n.
  2. Pick new nucleotide uniformly at random with probability $\frac{1}{3}$
  3. Compare number of GG. If we have $\ldots GAG \cdots \rightarrow \ldots GGG \ldots$, then $\Delta GG = 2$. (Moving backwards would be -2)
  4. If $\Delta_{GG} > 0$, accept the change.
  5. If $\Delta_{GG} < 0$, accept with probability $2^{\Delta_{GG}}$

Metropolis for Optimization

Suppose we have an objective function $I(c)$. We can calculate a score $e^{-I(c)/z}$ for every state, and use a Metropolis model to pick the highest probability state out of the entire model by tracking the amount of time spent on each state.

Gibbs Sampling

$$p(x_1, x_2, \ldots, x_n) = \Pr\{X_1 = x_1, X_2 = x_2, \ldots, X_n = x_n\}, \quad X_1 \in R_1$$

[FILL]

Algorithm:

1. Pick x_j uniformly at random from $1 \ldots n$.
2. Sample x_j from the conditional density $\Pr\{x'_j | x_1, \ldots, x_j, x_{j+1}, \ldots, x_n\}$
3. Repeat (go to step 1)

Example: Ising Model

Suppose we have a row of magnets that are close to each other. Each magnet either points up or down. Same repels, opposites attract.

Let the magnets be $X_1, \ldots, X_n$, where $X_i \in \{+1, -1\}$. To use the Gibbs sampling model,

1. Pick $i \leftarrow \lceil u[0, n] \rceil$
2. Sample of X_i

$$\Pr\{X_i = -1 | X_j \neq X_i\} = \left(e^{((-1)x_{i-1} + (-1)x_{i+1})g/kT}\right) / \left(e^{((-1)X_{i-1} + (-1)X_{i+1})g/kT} \cdot e^{(X_{i-1} + X_{i+1})g/kT}\right)$$

$$\Pr\{X_i = +1 | X_j \neq X_i\} = \left(e^{(x_{i-1} + x_{i+1})g/kT}\right) / \left(e^{((-1)X_{i-1} + (-1)X_{i+1})g/kT} \cdot e^{(X_{i-1} + X_{i+1})g/kT}\right)$$

3. (note that the denominators of both are the same)

Importance Sampling

Given $Q = \{q_1, \ldots, q_n\}$ and $\Pi = \{\pi_1, \ldots, \pi_n\}$

1. Make $\hat{\Pi} = \{\hat{\pi_1}, \ldots, \hat{\pi_n}\} = \hat{\Pi}_i = \frac{\pi_i w_i}{\sum \pi_i w_i}$
2. Sample from $\hat{\Pi}$
3. Scale estimates $\hat{\pi_1}, \ldots, \hat{\pi_n}$ by w_i
4. (w_i represents weights.)

Umbrella Sampling