*In this discussion, I assume the existence of the count matrix and the transition matrix have been discussed.*

**A. Calculating Free Energy from a Transition Matrix**

The free energy of a state relates directly to the stationary probability of being in state when the system is in equilibrium, where

(1)

where is the left eigenvector of the row-stochastic transition matrix with eigenvalue which relates to the equilibrium state. We assume that is ergodic; that is, every state can be reached from every other state, and thus there exists only one eigenvalue of unity and hence is unique.1 The stationary probability is related to the Gibbs’ free energy by

(2)

where is the Boltzmann constant and is temperature in Kelvin. Then the free energy of state relative to some reference state is.

()

However, from equation (1), it is clear that

()

and it will be convenient to therefore calculate as

()

**B. Uncertainty of Transition Probabilities**

It is relatively straightforward to consider the likelihood of a certain transition count matrix given the knowledge of the transition probability matrix , . However, in order to calculate the uncertainty in transition probability values given an experimentally determined count matrix, is required for a distribution of transition matrices . A detailed discussion of the Bayesian statistics used in these calculations is given in Ref 2.

We will from here onwards assume a uniform prior distribution given by Dirichlet parameters

()

and thus a resultant count matrix given by the sum of observed transition counts and prior transition counts

()

**C. Eigenvector Uncertainty**

For an -dimensional transition matrix , the eigenvector given by gives a system of linear equations. The computation involved in determining the uncertainty in each value of from the uncertainty in is NP-hard.2,3

We propose two methods of estimating the uncertainty. The first, a closed-form distribution, relies on a first-order Taylor polynomial and a multivariate normal distribution, as proposed by Singhal et al4. The second employs a brute-force approach, sampling transition matrices from a Dirichlet distribution based on the count matrix to create a distribution of eigenvectors , whose standard deviation can thus be calculated.

**Method 1 – Multivariate normal approximation & Taylor series approximation**

In Ref 4, Meyer et al. detail a method for obtaining the sensitivity associated with eigenvalues and eigenvectors of a matrix subject to perturbations. The sensitivity is given as follows

()

where is the identity matrix. To find , the perturbation in is simply a unit perturbation in the position . Thus we find

()

Singhal et al. then detail the process required to obtain the variance in a single element of the stationary probability vector from these sensitivities5. Let

()

and

()

Consider shorthand for denoting the th row of , shorthand for denoting the th row of . Then the variance is given by

()

This method provides a closed-form calculation for the variance of is , which is potentially slower than the next method for large matrices but notably does not involve a sampling constant which is generally much larger than the dimension of the matrix.

**Method 2 – Brute force approach**

The improvement in efficiency gained from the above method may not be worth the error introduced by its key assumptions. Here we detail a method which requires no assumptions but whose running time is slightly slower (dependent on the number of samples.)

From previously established results on transition matrices5, the distribution of the transition probabilities from the th state is a Dirichlet distribution with parameters .

We sampled this Dirichlet distribution times and the left eigenvectors of each sample are taken. The standard deviation of the collection of eigenvectors is then computed.

This method relies on no assumptions and produces uncertainty data in time, where is the time taken to sample from a normal distribution, given that sampling from the Dirichlet distribution takes time5 and eigendecomposition takes time.6

**D. Propagation of uncertainty to free energy surface**

The uncertainty in free energy is propagated by a simple calculation from the eigenvector , where

()

**E. Calculation of most likely folding path**

From any one configuration, the most likely transition is that which leads to the adjacent configuration of lowest free energy. This leads to the discovery of free energy basins, from which the most likely informative transition is that which leads to the configuration of lowest energy which is adjacent to any configuration within the basin. Thus an algorithm was devised to determine the most likely folding path using a graph to model the free energy surface.

* Choose a starting node, in this case the extended state
* Repeat until the folded state is reached:
  + Move to the adjacent configuration of lowest free energy
  + If this configuration has been visited before, we have reached a cycle
    - Combine all configurations in the cycle so that every configuration adjacent to the cycle is considered adjacent to all configurations within the cycle, and configurations within the cycle are no longer adjacent to each other (this avoids becoming ‘stuck’ within a local free energy minimum.)

As this algorithm only passes over each vertex once in the initial search, and at maximum must pass over one more vertex for each one searched in the creation of cycles, the total running time of the algorithm is .