Lecture Outline: Molecular Evolution (part 1)

1. **Features of Molecular Evolution**
   (a) Possible multiple changes on edges
   (b) Transition/transversion bias
   (c) Non-uniform base composition
   (d) Rate variation across sites
   (e) Dependence among sites
   (f) Codon position
   (g) Protein structure

2. **Continuous-time Markov Chains**
   (a) Probabilistic framework
   - Essentially, all models are wrong, but some are useful.

     George Box
   - A probabilistic framework provides a platform for formal statistical inference
   - Examining goodness of fit can lead to model refinement and a better understanding of the actual biological process
   - Model refinement is a continuing area of research
   - Most common models of molecular evolution treat sites as independent
   - These common models just need to describe the substitutions among four bases at a single site over time.

   (b) Markov property
   - Use the notation $X(t)$ to represent the base at time t.
   - Formal statement:

   $$\mathsf{P}\{X(s+t)=j \mid X(s)=i, X(u)=x(u) \text{ for } u<s\} = \mathsf{P}\{X(s+t)=j \mid X(s)=i\}$$

   - Informal understanding: given the present, the past is independent of the future
   - If the expression does not depend on the time s, the Markov process is called *homogeneous*.

3. **Rate Matrix**
   (a) Positive off-diagonal rates of transition
   (b) Negative total on the diagonal
   (c) Row sums are zero
   (d) Example

   $$Q = \{q_{ij}\} = \begin{pmatrix} -1.1 & 0.3 & 0.6 & 0.2 \\ 0.2 & -1.1 & 0.3 & 0.6 \\ 0.4 & 0.3 & -0.9 & 0.2 \\ 0.2 & 0.9 & 0.3 & -1.4 \end{pmatrix}$$

4. **Alarm Clock Description**
   (a) Exponential distribution
   - Only continuous-time distribution with *memoryless property* needed for the Markov property.
   - Single parameter λ is called the *rate*.
   - Density is $f(t) = \lambda \mathrm{e}^{-\lambda t}, \quad \text{for } t \geq 0.$
   - Density satisfies $\int_0^\infty f(t)\mathrm{d}t = 1.$

- Cumulative distribution function is $\mathsf{P}\{T \le t\} = F(t) = \int_0^t f(s) \mathrm{d}s = 1 - \mathrm{e}^{-\lambda t}$.
- Tail probability (probability of no event in time t) is $\mathrm{e}^{-\lambda t}$.
- Mean is $1/\lambda$.

(b) Exponential time to next event

- If the current state is i, the time to the next event is exponentially distributed with rate $-q_{ii}$ defined to be q_i.

(c) Probability of the specific transition

- Given a transition occurs from state i, the probability that the transition is to state j is proportional to q_{ij}, namely $q_{ij} / \sum_{k \ne i} q_{ik}$.

5. **Path Probability Density Calculation**

(a) Example: Begin at A, change to G at time 0.3, change to C at time 0.8, and then no more changes before time $t = 1$.

$$\mathsf{P}\{\text{path}\} = \mathsf{P}\{\text{begin at A}\} \times \left(1.1\mathrm{e}^{-(1.1)(0.3)} \cdot \frac{0.6}{1.1}\right) \times \left(0.9\mathrm{e}^{-(0.9)(0.5)} \cdot \frac{0.3}{0.9}\right) \times \left(\mathrm{e}^{-(1.1)(0.2)}\right)$$

6. **Transition Probabilities**

(a) Matrix multiplication

- Compute AB where A is an $m \times n$ matrix and B is an $n \times p$ matrix. (Note that the number of columns in A must match the number of rows in B.)
- The ij element of the matrix AB is the dot product of the ith row if A and the jth row of B.

$$AB_{ij} = \sum_{k=1}^{n} a_{ik} b_{kj}$$

- Example:

$$A = \begin{pmatrix} -1 & 0.4 & 0.6 \\ 0.8 & -2 & 1.2 \\ 0 & 1 & -1 \end{pmatrix} B = \begin{pmatrix} -1 & 0 & 1 \\ 1 & -2 & 1 \\ 0 & 0.5 & -0.5 \end{pmatrix} AB = \begin{pmatrix} 1.4 & -0.5 & -0.9 \\ -2.8 & 4.6 & -1.8 \\ 1 & -2.5 & 1.5 \end{pmatrix}$$

(b) Matrix exponentiation

- For a square matrix A, the matrix exponential is defined to be

$$\mathrm{e}^A = \sum_{k=0}^{\infty} \frac{A^k}{k!} = I + A + \frac{A^2}{2} + \frac{A^3}{6} + \cdots$$

(c) Transition matrix

- For a continuous time Markov chain, the *transition matrix* whose ij element is the probability of being in state j at time t given the process begins in state i at time 0 is $P(t) = \mathrm{e}^{Qt}$.
- A probability transition matrix has non-negative values and each row sums to one.
- Each row contains the probabilities from a probability distribution on the possible states of the Markov process.
- Examples:

$$P(0.1) = \begin{pmatrix} 0.897 & 0.029 & 0.055 & 0.019 \\ 0.019 & 0.899 & 0.029 & 0.053 \\ 0.037 & 0.029 & 0.916 & 0.019 \\ 0.019 & 0.080 & 0.029 & 0.872 \end{pmatrix} \qquad P(0.5) = \begin{pmatrix} 0.605 & 0.118 & 0.199 & 0.079 \\ 0.079 & 0.629 & 0.118 & 0.174 \\ 0.132 & 0.118 & 0.671 & 0.079 \\ 0.079 & 0.261 & 0.118 & 0.542 \end{pmatrix}$$

$$P(1) = \begin{pmatrix} 0.407 & 0.190 & 0.276 & 0.126 \\ 0.126 & 0.464 & 0.190 & 0.219 \\ 0.184 & 0.190 & 0.500 & 0.126 \\ 0.126 & 0.329 & 0.190 & 0.355 \end{pmatrix} \qquad P(10) = \begin{pmatrix} 0.200 & 0.300 & 0.300 & 0.200 \\ 0.200 & 0.300 & 0.300 & 0.200 \\ 0.200 & 0.300 & 0.300 & 0.200 \\ 0.200 & 0.300 & 0.300 & 0.200 \end{pmatrix}$$

7. **Stationary Distribution**

   (a) Long-run average

   - Well behaved continuous-time Markov chains have a *stationary distribution*, often designated π (not the constant close to 3.14 related to circles).
   - When the time t is large enough, the probability $P_{ij}(t)$ will be close to π_j for each i. (See $P(10)$ from earlier.)
   - The stationary distribution can be thought of as a long-run average— over a long time, the proportion of time the state spends in state i converges to π_i.

   (b) Multiplication property

   - The stationary distribution is an eigenvector of Q^T, the transpose of Q, associated with the eigen value 0.
   - This means that $\pi^T Q = 0^T$.
   - It also follows that $\pi^T P(t) = \pi^T$ for any time t. (If you begin in the stationary distribution, you remain in the stationary distribution.)

   (c) Usual parameterization of rate matrix

   - The matrix $Q = \{q_{ij}\}$ is typically parameterized as $q_{ij} = r_{ij}\pi_j/\mu$ for $i \neq j$ which guarantees that π will be the stationary distribution when $r_{ij} = r_{ji}$.

8. **Scaling**

   (a) Expected number of substitutions per unit time

   - The expected number of substitutions per unit time is the average rate of substitution which is a weighted average of the rates for each state weighted by their stationary distribution.

$$\mu = \sum_i \pi_i q_i$$

   - If the matrix Q is reparameterized so that all elements are divided by μ, then the unit of measurement becomes one substitution.

9. **Time-reversibility**

   (a) Conceptual understanding

   - A continuous-time Markov chain is *time-reversible* if the probability of a sequence of events is the same going forward as it is going backwards.
   - Look at example from earlier.

   (b) Time-reversibility condition

   - The matrix Q is the matrix for a time-reversible Markov chain when $\pi_i q_{ij} = \pi_j q_{ji}$ for all i and j. That is the overall rate of substitutions from i to j equals the overall rate of substitutions from j to i for every pair of states i and j.