Molecular Evolution

Bret Larget

Departments of Botany and of Statistics University of Wisconsin—Madison

September 15, 2011

Molecular Evolution

Features of Molecular Evolution

- Possible multiple changes on edges
- 2 Transition/transversion bias
- One of the second se
- Rate variation across sites
- Opendence among sites
- **o** Codon position
- Protein structure

1 / 13

A Famous Quote About Models

Essentially, all models are wrong, but some are useful.

George Box

Molecular Evolution

Molecular Evolution

Models

3 / 13

Probability Models

- 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.

The Markov Property

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

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

= P {X(s+t) = j | 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*.

Molecular Evolution

Probabilistic framework

Continuous-time Markov Chains

5 / 13

Rate Matrix

- Positive off-diagonal rates of transition
- Negative total on the diagonal
- Row sums are zero
- Example

$$Q = \{q_{ij}\} = \left(egin{array}{ccccc} -1.1 & 0.3 & 0.6 & 0.2 \ 0.2 & -1.1 & 0.3 & 0.6 \ 0.4 & 0.3 & -0.9 & 0.2 \ 0.2 & 0.9 & 0.3 & -1.4 \end{array}
ight)$$

Alarm Clock Description

- If the current state is *i*, the time to the next event is exponentially distributed with rate -q_{ii} defined to be q_i.
- Given a transition occurs from state *i*, the probability that the transition is to state *j* is proportional to q_{ij} , namely $q_{ij} / \sum_{k \neq i} q_{ik}$.

Molecular Evolution

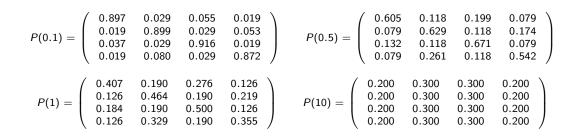
Probabilistic framework

Continuous-time Markov Chains

7 / 13

Transition Probabilities

- 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 $P(t) = e^{Qt}$.
- 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.



```
Molecular Evolution
```

Probabilistic framework

Continuous-time Markov Chains

9 / 13

The Stationary Distribution

- 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_i 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.

Parameterization

• 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 $r_{ij} = r_{ji}$.

Molecular Evolution

Probabilistic framework

Continuous-time Markov Chains

11 / 13

Scaling

• 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.

Time-reversibility

 The matrix Q is the matrix for a time-reversible Markov chain when π_iq_{ij} = π_jq_{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.

Molecular Evolution

Probabilistic framework

Continuous-time Markov Chains

13 / 13