Revision of Poisson Process

Ramon Fallon

March 3, 2011

Fundamental Concepts of Computational Evolution

- Given Data: Sequence Alignments
- Markov Chain ruled by a Poisson Process
- Instantaneous Substitution Rate Matrix, `Q`
- Obtaining the Transition Probability Matrix, `P (t)`

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

- Models defined by parameters attached to Substitution Matrix
- Representation of Trees
- Likelihood and its calculation on a tree

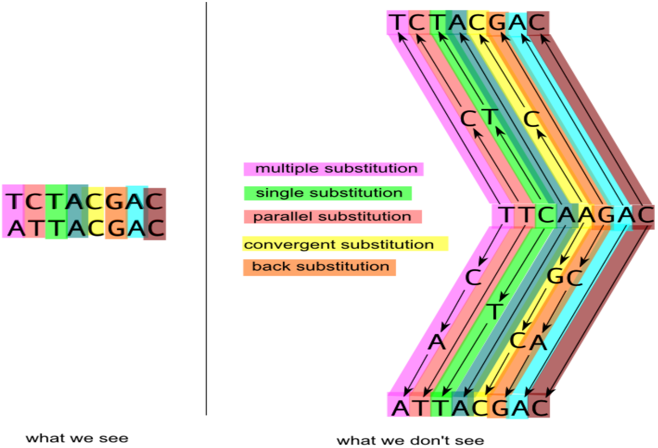
Sources

Scott V. Edwards.
Natural Selection and Phylogenetic Analysis.
PNAS, 106(22):8799–8800, June 2009.

J. Felsenstein.
Evolutionary Trees from DNA Sequences: A Maximum Likelihood Approach.
Journal of Molecular Evolution, 17:368–376, 1981.

Ziheng Yang.
Computational Molecular Evolution.
Oxford University Press, 2006.

Alignments: what we don't see [Yang(2006)]

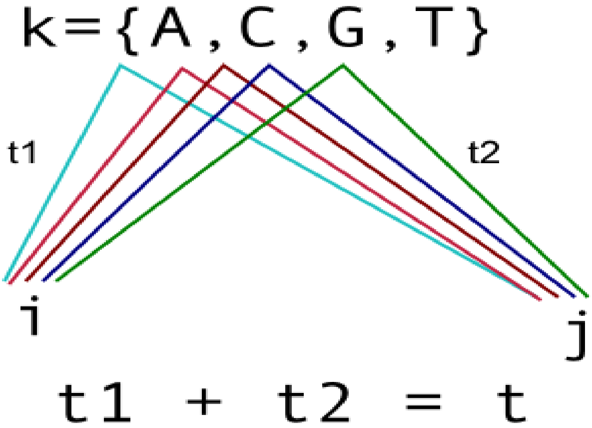


- **multiple hits**: we deal with this via Chapman-Kolmogorov.
- **convergent substitution**, natural selection at work. See [Edwards(2009)].

Chapman-Kolmogorov Equations

Referred to as **Pulley Principle** by [Felsenstein(1981)].

$$p_{ij}(t) = \sum_k p_{ik}(t_1)p_{kj}(t_2)$$



Dealing with multiple hits

We get the Poisson distribution by approximating the binomial:

$$\Pr(K = k) = b(k; n, p) = \binom{n}{k} p^k (1-p)^{n-k}$$

Look at example of $K = 0$, with n large and p small so we way substitute $\lambda = np$:

$$b(0; n, p) = (1-p)^n = \left(1 - \frac{\lambda}{n}\right)^n$$

Taking logs and using the Taylor expansion for $log(1-x)$:

$$\log b(0; n, p) = n \log\left(1 - \frac{\lambda}{n}\right) = -\lambda - \frac{\lambda^2}{2n} - \dots$$

Truncating to first term, and redoing for $K = 0, 1, 2, \dots, k \dots$:

$$b(0; n, p) \approx e^{-\lambda}; b(1; n, p) \approx \lambda e^{-\lambda}; b(k; n, p) \approx \frac{\lambda^k}{k!} e^{-\lambda};$$