Hidden Markov Models

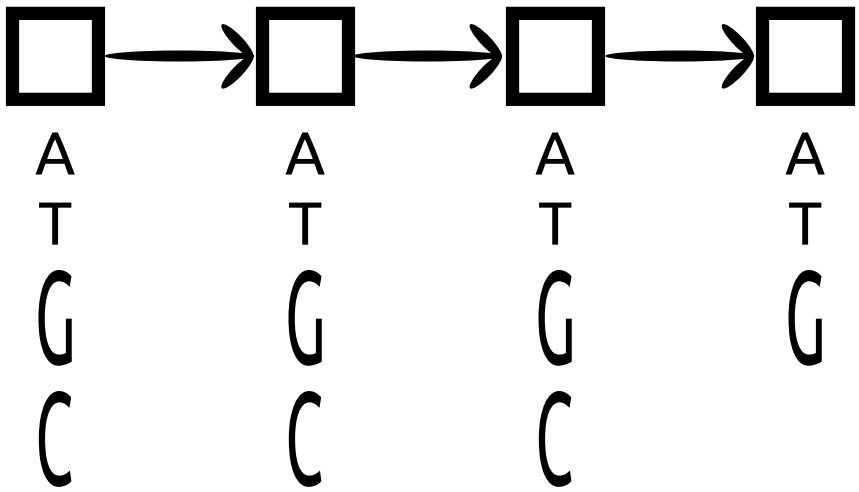
Mark Voorhies

4/1/2011

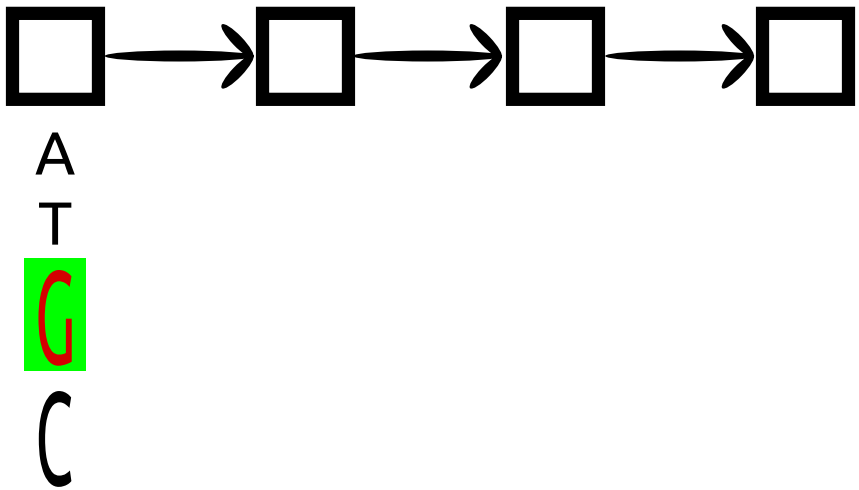
Review

- Multiple alignments and trees
- Profiles
- Our treatment of gaps is still *ad hoc*

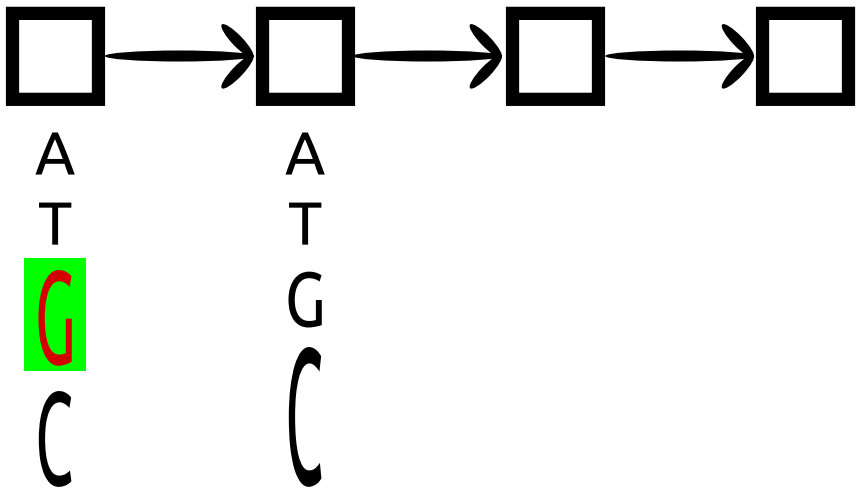
0^{th} order Markov Model



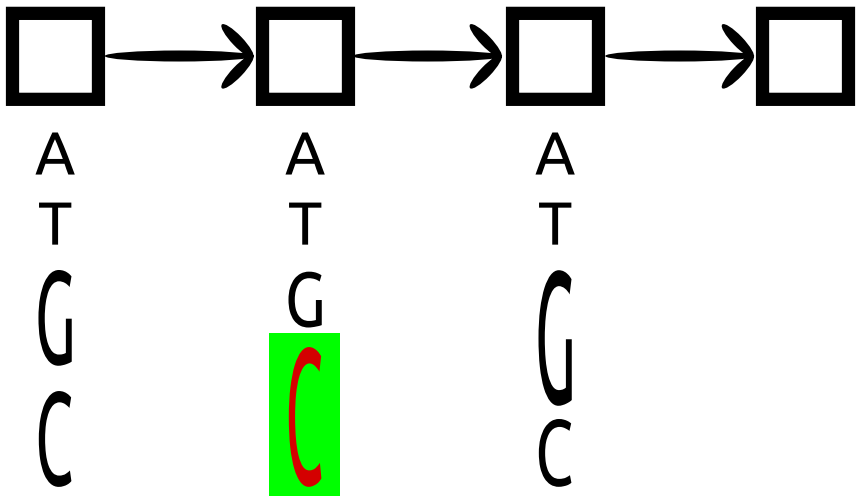
1^{st} order Markov Model



1^{st} order Markov Model



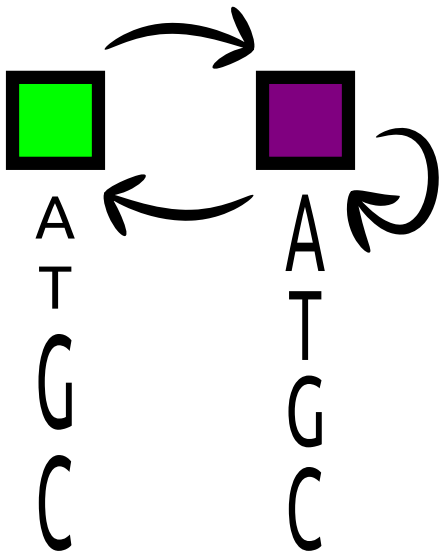
1^{st} order Markov Model



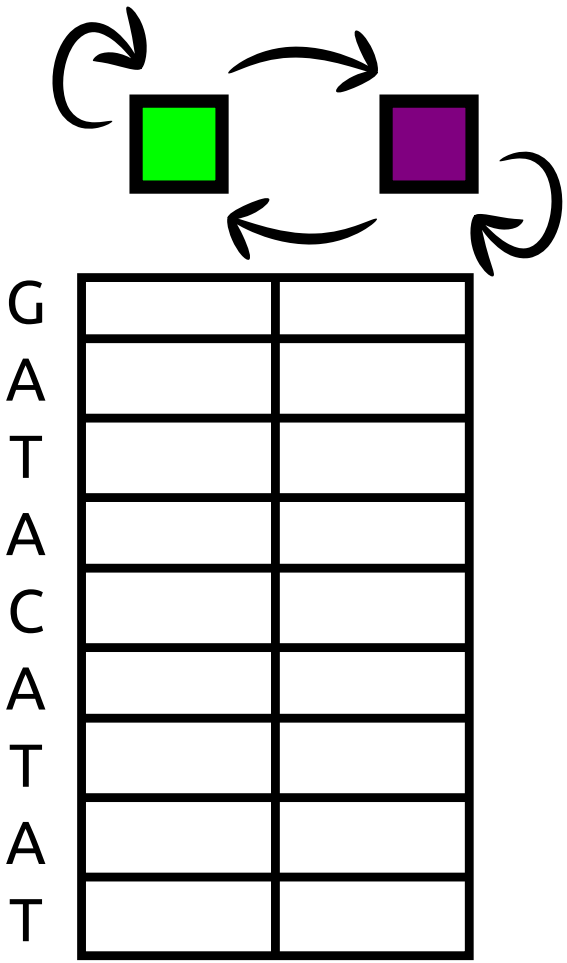
What are Markov Models good for?

- Background sequence composition
- Spam

Hidden Markov Model

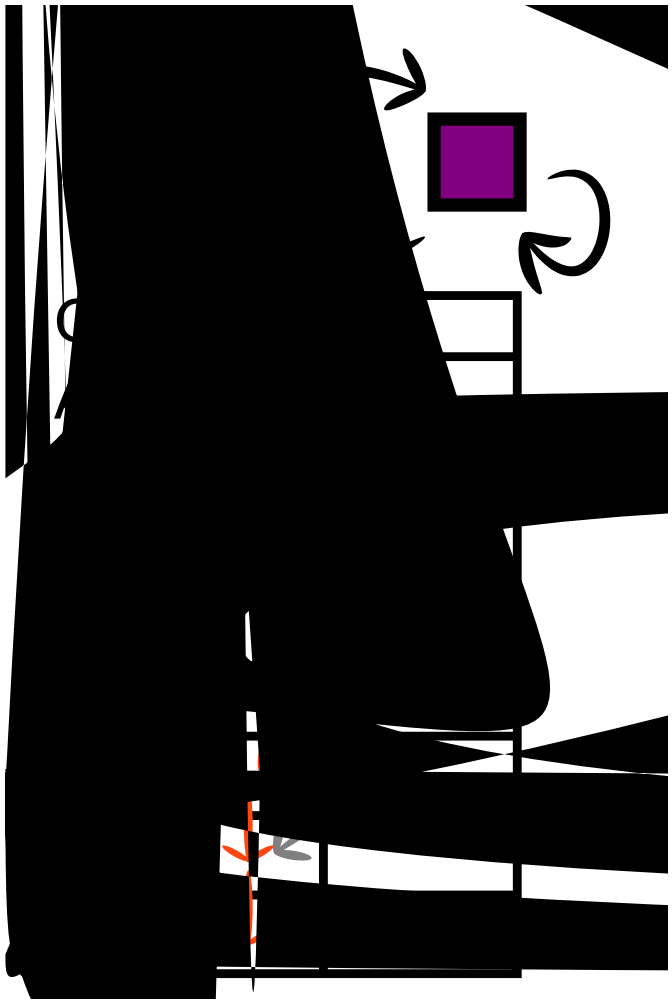


The Viterbi algorithm: Alignment



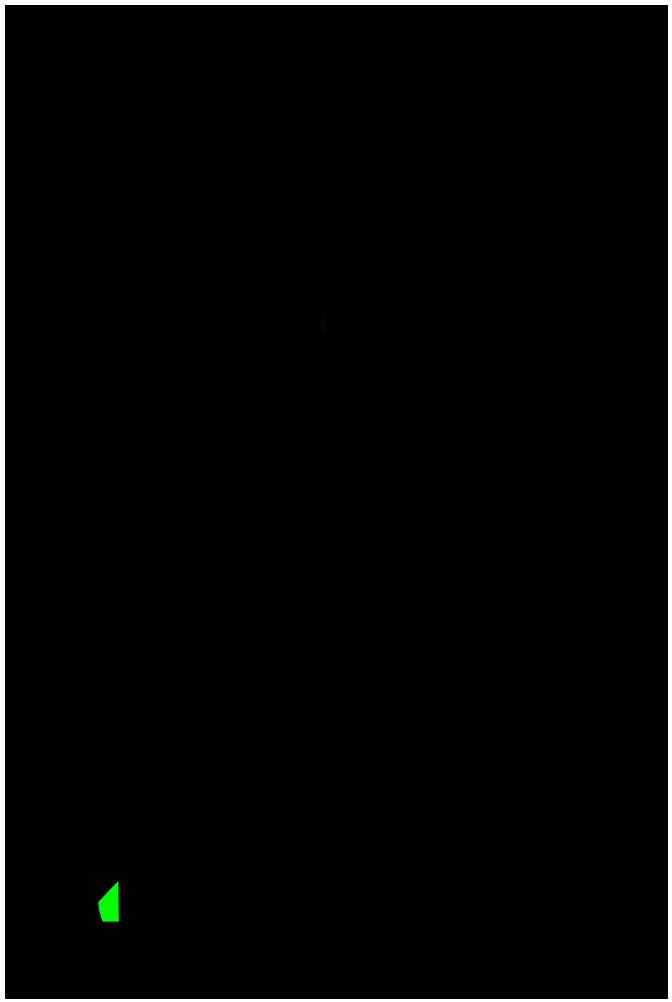
The Viterbi algorithm: Alignment

The Forward algorithm: Net probability



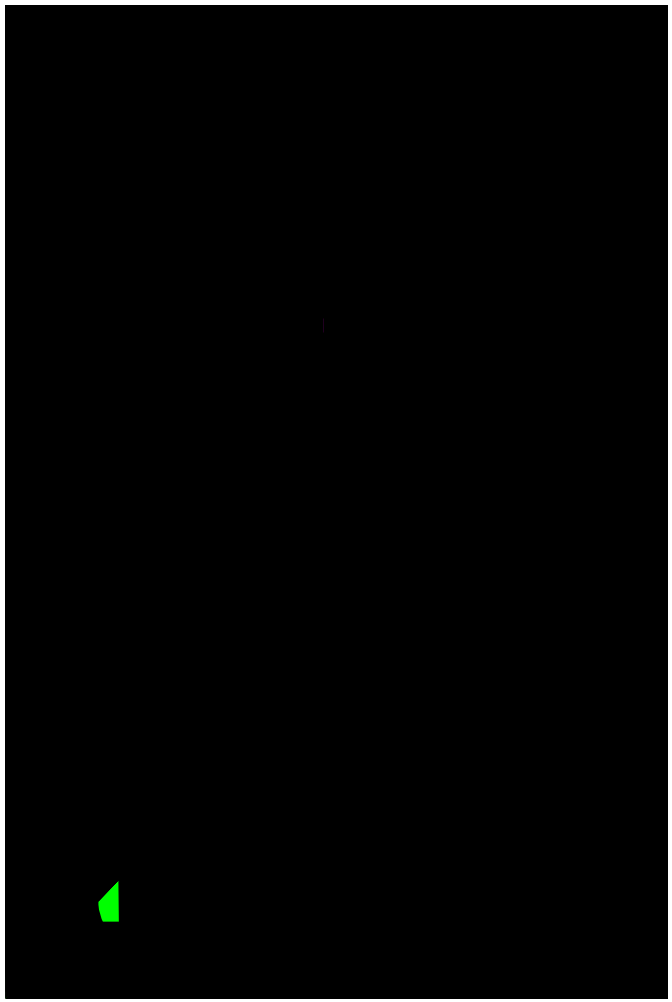
- Probability-weighted sum over all possible paths
- Simple modification of Viterbi (although *summing* probabilities means we have to be more careful about rounding error)
- Result is the probability that the observed sequence is explained by the model
- In practice, this probability is compared to that of a null model (*e.g.*, random genomic sequence)

Training an HMM



- If we have a set of sequences with known hidden states (*e.g.*, from experiment), then we can calculate the emission and transition probabilities directly
- Otherwise, they can be iteratively fit to a set of unlabeled sequences that are known to be true matches to the model

Training an HMM



- If we have a set of sequences with known hidden states (*e.g.*, from experiment), then we can calculate the emission and transition probabilities directly
- Otherwise, they can be iteratively fit to a set of unlabeled sequences that are known to be true matches to the model
- The most common fitting procedure is the Baum-Welch algorithm, a special case of expectation maximization (EM)

Profile Alignments: Plan 7

(Image from Sean Eddy, PLoS Comp. Biol. 4:e1000069)

Rigging Plan 7 for Multi-Hit Alignment

(Image from Sean Eddy, PLoS Comp. Biol. 4:e1000069)