HMM miscellanea:

By choosing the appropriate HMM architecture, Markov chains with orders greater than 1 can be incorporated into the HMM framework.

Example:

We have been assuming that GC-rich and AT-rich states are organized along a sequence according to a first order Markov chain.

In this case, the value of y_{i+1} (0 if AT-rich at site $i+1$ and 1 if GC-rich at site $i+1$) depended only on the value of y_i.

In other words, $\Pr(y_{i+1} \mid y_1, y_2, \ldots, y_i) = \Pr(y_{i+1} \mid y_i)$.

Now let's assume that we have a second order Markov chain. In this case, the value of y_{i+1} depends on the values of y_i and y_{i-1}.

Assume:

$$\begin{aligned}
\Pr\left(y_{i+1}=0 \mid y_i=0, y_{i-1}=0\right) &= 0.9 \\
\Pr\left(y_{i+1}=0 \mid y_i=0, y_{i-1}=1\right) &= 0.8 \\
\Pr\left(y_{i+1}=0 \mid y_i=1, y_{i-1}=0\right) &= 0.7 \\
\Pr\left(y_{i+1}=0 \mid y_i=1, y_{i-1}=1\right) &= 0.2
\end{aligned}$$

Define a new type of state w_i where $w_i = 2y_i + y_{i-1}$. We can define a first order Markov chain in terms of w_i the probabilities of w_{i+1} given the value of w_i are:

	w_{i+1}			
	0	1	2	3
w_i				
0	0.9	0.0	0.1	0.0
1	0.8	0.0	0.2	0.0
2	0.0	0.7	0.0	0.3
3	0.0	0.2	0.0	0.8

Back to 1st order chain for GC-rich and AT-rich states ...

Assume $\Pr(y_{i+1} = 1 \mid y_i = 1) = p$.

So, $\Pr(y_{i+1} = 0 \mid y_i = 1) = 1 - p$.

Probability that a GC-island has length L is: $p^{L-1}(1-p)$.
This is a geometric distribution.

We can modify the HMM so that GC-islands must be at least some minimum length M.

As an example, let's change how the value of y_i is determined.

We will still have $y_i = 0$ when site i is an AT-rich state.

Now, we will have $y_i = k$ when site i is a GC-rich state and where site y_{i-k} is the nearest AT-rich site to site i in the 5' direction along the sequence.

For $1 \leq k < M$, we can set

$$\Pr\left(y_{i+1} = k+1 \mid y_i = k\right) = 1$$

For $k \geq M$, we will have two probabilities that may be nonzero when $y_i = k$. These possibly nonzero probabilities are $\Pr\left(y_{i+1} = k+1 \mid y_i = k\right)$ and $\Pr\left(y_{i+1} = 0 \mid y_i = k\right)$.

Also, $\Pr\left(y_{i+1} = 1 \mid y_i = 0\right)$ and $\Pr\left(y_{i+1} = 0 \mid y_i = 0\right)$ may be nonzero.

Other more complicated length probability distributions are also possible by modifying the HMM architecture (e.g., Nick Goldman, David Jones, and I have exploited this possibility to mimic length distributions of α-helices, β-strands, turns, and coils in our studies of protein evolution).

Sometimes, it is worthwhile to add an explicit "BEGIN" and "END" hidden Markov state to a model.

The treatment of these "BEGIN" and "END" states is straightforward.

HMM parameter estimation when the path y is known:

Emission probabilities can be estimated by the number of times a residue type was emitted divided by the number of times the residue type could have been emitted.

Transition probabilities can be estimated by the number of times a transition was taken divided by the number of times that it could have been taken.

HMM parameter estimation with the path y is unknown:

Parameter estimation gets more complicated. There are many kinds of numerical optimization routines that can be adopted in conjunction with HMM's but these will not be discussed here.

The one type of routine that we briefly mention is the Baum-Welch algorithm, a variant of the expectation-maximization (EM) algorithm.

If we do not observe the hidden states of an HMM, we can still make inferences regarding them.

Step 1: Make rough guess about HMM parameter values to initialize Baum-Welch algorithm

Step 2: Calculate log-likelihood.

Step 3: Given the parameter values, calculate expected number of times a residue type was emitted by a specific (e.g., GC-rich) hidden Markov state. Calculate expected number of times residue type could have been emitted by that specific hidden Markov state. Divide former by latter for new estimates of emission probabilities.

For example, expected number of times that an "A" could have been emitted by a GC-rich state is:

$$\sum_{i=1}^{N} \Pr\left(y_i = 1 \mid x\right)$$

Calculate expected number of times a transition was taken. Calculate expected number of times a transition could have been taken. Divide former by latter to obtain new estimates of transition probabilities.

Step 4: Go to Step 2 but stop after step 2 if log-likelihood only changes by a tiny amount from previous cycle.

Correction: Page 66 in my version of the Durbin et al. text (page 65 in other versions) reads "... it can be argued that when the primary use of the HMM is to produce decodings via Viterbi alignments, then it is good to train using them." *The text is misleading here – such an argument would be flawed*