

Algorithms for Computational Biology

HMM Notation

If you are studying from the Durbin, et. al. *Biological Sequence Analysis* book, you may find yourself a little lost in the notation, as it is different from Michael's.

Equations 3.4, 3.5 and 3.7 in Durbin show the major equations involved in the Viterbi recursion:

$$\begin{aligned}
 a_{kl} &= P(\pi_i = l | \pi_{i-1} = k) \\
 e_k(b) &= P(x_i = b | \pi_i = k) \\
 v_l(i+1) &= e_l(x_{i+1}) \max_k (v_k(i) a_{kl})
 \end{aligned}$$

Michael's equivalents are

$$\begin{aligned}
 a_{ji} &= Pr(q_t = s_i | q_{t-1} = s_j) \\
 b_i(A_k) &= Pr(o_t = A_k | q_t = s_i) \\
 \delta_t(i) &= b_i(o_t) \max_{j=1}^N a_{ji} \delta_{t-1}(j)
 \end{aligned}$$

This shows a lot of the differences in the variables and subscripts used.

Durbin	MB	Description
b	A_k	character in the input alphabet (for DNA, A,C,G,T,...)
i	t	current position
l, k	i, j	state indices
(none)	N	number of states in the HMM
L	L	length of the input sequence
(none)	s_i	possible state in an HMM (for $i = 1..N$)
e_k	b_i	emission function for a certain state (for $k, i = 1..N$)
x_i	o_t	observation or part of the input sequence (for $i, t = 1..L$)
v	δ	probability of the maximum path (Viterbi value)
π_i	q_t	a state in a state path through the sequence (for $i, t = 1..L$)