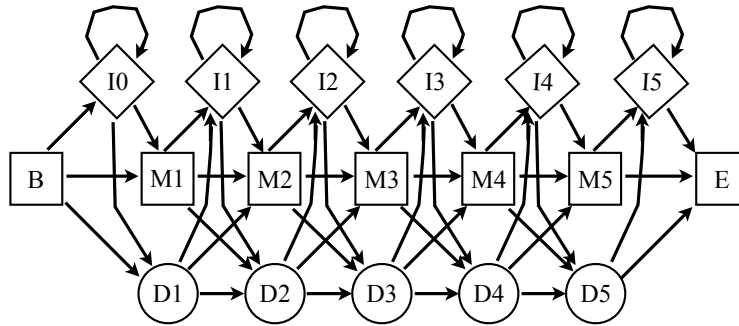


Math 5233: Profile HMMs; due April 15th

Consider a profile HMM as shown below. The consensus sequence is QWERT; each match state emits the corresponding consensus letter with probability $\frac{19}{20}$, and any other amino acid letter with probability $\frac{1}{380}$. The insertion states I_k emit each amino acid with probability $\frac{1}{20}$. The $B \rightarrow M1$ and $M_i \rightarrow M_j$ transitions occur with probability $\frac{7}{8}$. The $M_i \rightarrow I_i$ and $M_i \rightarrow D(i+1)$ transitions occur with probability $\frac{1}{16}$. The $I_i \rightarrow I_i$ self-transitions occur with probability $\frac{1}{4}$. The $I_i \rightarrow M(i+1)$ transitions occur with probability $\frac{11}{16}$, while $I_i \rightarrow D(i+1)$ transitions occur with probability $\frac{1}{16}$. Finally, the $D_i \rightarrow D(i+1)$ and $D_i \rightarrow I_i$ transitions occur with probability $\frac{1}{16}$, while $D_i \rightarrow M(i+1)$ occurs with probability $\frac{7}{8}$. Note that the sum of all probabilities leaving a state should sum to 1.



- (1) What is the \log_2 likelihood of the most probable path of the consensus sequence in this model?

- (2) What is the alignment generated by the sequence QIWET, and what is its log likelihood?

- (3) How do the log likelihoods of the sequences AQWERT, AAQWERT, AAAQWERT, etc, compare to that of the consensus?