

We illustrate how the model works with an example. Assume there are a couple of strings to be aligned. The number of characters in the unknown common ancestor string is usually estimated to be the arithmetical mean of the lengths of the considered strings. In our example, let this be 4. Further assume that for some of the strings there already exists a plausible multiple alignment (obtained by one of the methods described before or proposed by biologists on basis of plausibility). As an example, consider the following multiple alignment.

```

      1      2 3  4
    - C - - G - - A -
    - - - - G C A A A
    - C - C - A - G -
    - T - - T C C T -
    A C G G G C - C -

```

Here, we have decided to interpret columns with a majority of characters over spacing symbols to be derived from one of the conjectured four characters of the unknown ancestor by either match, mutate, or delete events. On the other hand, columns with a majority of spacing symbols are assumed to be the result of insert events. On basis of these decisions we next introduce at all positions of the multiple alignment the correct state that led to the entry.

```

      1      2 3  4
    - M1 - - M2 D3 - M4 -
    - D1 - - M2 M3 I3 M4 I4
    - M1 - I1 D2 M3 - M4 -
    - M1 - - M2 M3 I3 M4 -
    I0 M1 I1 I1 M2 M3 - M4 -

```

Now we estimate transition and emission probabilities of the Hidden Markov model as relative frequencies. For example, there are two transitions from state M_1 to state M_2 , one transition to state D_2 , and one transition to state I_1 . Thus, transition probabilities from state M_1 to other states are estimated as follows: 50% to M_2 , 0% to D_2 , 50% to I_1 . Emission probabilities are estimated similarly, for example, for state M_4 emission probability is estimated for character A as 40%, for C as 20%, for G as 20%, and for T as 20%.

So far, parameters of the model have been fixed on basis of available aligned strings. The general algorithmic problem behind parameter choice is training a model or fitting a model to available data. Parameters are chosen in such a way that the resulting model M has maximum probability of generating the training data. This is called maximization of model likelihood. The required formal concepts to make this precise are the probability that model M generates data D , briefly denoted $P_M(D)$, as well as the likelihood of model M given data D_1, \dots, D_k defined by

$$L(M|D_1, \dots, D_k) = \prod_{i=1}^k P_M(D_k). \quad (2.16)$$