Algorithm: Hill-Climbing

Input:
Sequences $t_1, \ldots, t_k$ of lengths $n_1, \ldots, n_k$.

Initialization:
\[
z = 1 \quad \# \text{index of special sequence.}
\]
\[
t^* = t_z, n^* = n_z \quad \# t_1 \text{ is the special sequence, initially.}
\]
\[
\text{for } (j = 2 \text{ to } k) \{
\]
\[
\text{index}[j-1] = j \quad \# \text{index of non-special sequences}
\]
\[
o_j = \text{rand}(1, n_j - w) \quad \# \text{Guess starting offsets}
\]
\[
A'[j-1, 1 \cdots w] = t_j[(o_j+1) \cdots (o_j+w)]
\]
\[
\}
\]
Calculate $P[x,i]$, the propensity matrix of $A'$ with pseudocounts

Search for motif:
Repeat
\[
\{
\]
\[
o^* = \text{argmax}_o \{S(t^*, o)\} \quad \# \text{Select starting offset in } t^*
\]
\[
r = \text{rand}(1, k-1) \quad \# \text{Select new special sequence}
\]
\[
A'[r, 1 \cdots w] = t^*[(o^*+1) \cdots (o^*+w)] \quad \# \text{Replace new special with } t^* \text{ in } A'
\]
\[
y = \text{index}[r]; \quad \text{index}[r] = z; \quad z = y \quad \# \text{store ptr to } t^* \text{ in index}
\]
\[
t^* = t_z; \quad n^* = n_z \quad \# \text{initialize new } t^*
\]
\[
\text{Calculate } P[x,i], \text{ the propensity matrix of } A' \text{ with pseudocounts}
\]
\[
S[x,i] = \log_2 P[x,i]
\]
\[
\}
\]
until($P[\cdot, \cdot]$ stops changing)

Obtain $A$ by adding $t^*[(o^*+1) \cdots (o^*+w)]$ to $A'$
Compute the log odds scoring matrix, $S$, from $A$.

Output:
Local multiple sequence alignment $A$ with scoring matrix $S$.

Algorithm 1: Hill Climbing

The matrices $P$ and $S$ are the propensity and log odds matrices defined in Equations 4.2 and 4.3. Note that $A'$ and $P$ are $(k-1) \times w$ matrices, whereas the output matrices $A$ and $S$ are $k \times w$ matrices. The use of pseudocounts when calculating $P$ and $S$ is recommended to ensure all symbols in the alphabet are represented.
Chapter 4  Modeling motifs: Position Specific Scoring Matrices

Algorithm: Gibbs Sampler
Input:
Sequences $t_1, \ldots, t_k$ of lengths $n_1, \ldots, n_k$.

Initialization:
$z = 1$  # index of special sequence.
$t^*, n^* = n_z$  # $t_1$ is the special sequence, initially.
for ($j = 2$ to $k$) {
  index$[j-1] = j$  # index of non-special sequences
  $o_j = rand(1, n_j - w)$  # Guess starting offsets
  $A'[j-1, 1 \cdots w] = t_j[(o_j+1) \cdots (o_j+w)]$
}

Calculate $P[x,i]$, the propensity matrix of $A'$ with pseudocounts

Search for motif:
Repeat
{
  for ($o = 0$ to $(n^*-w)$)
  
  pdf$[o] = \frac{\prod_{j=1}^{w} P[t^*[o+j], j]}{\sum_{i=0}^{n^*-w} \prod_{j=1}^{w} P[t^*[i+j], j]}$
  
  With probability pdf$[o]$, $o^* = o$  # Select starting offset in $t^*$
  $r = rand(1, k-1)$  # Select new special sequence
  $A'[r, 1 \cdots w] = t^*[(o^*+1) \cdots (o^*+w)]$  # Replace new special with $t^*$ in $A'$
  $y = index[r]$; $index[r] = z$; $z = y$  # store ptr to $t^*$ in index
  $t^* = t_z$; $n^* = n_z$  # initialize new $t^*$
  Calculate $P[x,i]$, the propensity matrix of $A'$ with pseudocounts
}
until($P[\cdot, \cdot]$ stops changing)

Obtain $A$ by adding $t^*[(o^*+1) \cdots (o^*+w)]$ to $A'$
Compute the log odds scoring matrix, $S$, from $A$.

Output:
Local multiple sequence alignment $A$ with scoring matrix $S$.

Algorithm 2: Gibbs Sampler