

Exercise sheet 5: Probalign

For the following exercises on Probalign, we use an affine gap penalty with $g(k) = \alpha + \beta k = -0.5 - 0.25k$, there temperature $T = 1$ and the similarity function $\sigma(x_i, y_j)$:

$$\sigma(x_i, y_j) = \begin{matrix} & \begin{matrix} A & C & G & T \end{matrix} \\ \begin{matrix} A \\ C \\ G \\ T \end{matrix} & \begin{pmatrix} 2 & -1 & -1 & -1 \\ -1 & 2 & -1 & -1 \\ -1 & -1 & 2 & -1 \\ -1 & -1 & -1 & 2 \end{pmatrix} \end{matrix}$$

Exercise 1

Question 1A Compute the Boltzmann-weighted score for the following alignments:

$$\begin{array}{ll} \text{(a)} & \begin{array}{l} \text{x: --AGCGG} \\ \quad ||:|| \\ \text{y: ACAGGGG} \end{array} \\ \text{(b)} & \begin{array}{l} \text{x: AGCGG-----} \\ \quad : \\ \text{y: -----ACAGGGG} \end{array} \end{array}$$

Hint 1 : Formulae

$$S(a) = \sum_{x_i \sim y_j \in a} \sigma(x_i, y_j) + \sum \text{gap penalties} e^{\frac{S(a)}{T}} = \left(\prod_{x_i \sim y_j \in a} e^{\frac{\sigma(x_i, y_j)}{T}} \right) \times e^{\frac{\sum \text{gap penalties}}{T}}$$

Hint 2 For each alignment you only need to calculate e^x once.

Hint 3: Calculations

$$\begin{array}{ll} \text{(a)} & e^{\sigma(A,A)} \times e^{3\sigma(G,G)} \times e^{\sigma(C,G)} \times e^{g(2)} = e^2 \times e^6 \times e^{-1} \times e^{-0.5+(-0.25 \times 2)} = e^6 \\ \text{(b)} & e^{\sigma(G,A)} \times e^{g(4)} \times e^{g(6)} = e^{-1} \times e^{0.5+(-0.25 \times 4)} \times e^{0.5+(-0.25 \times 6)} = e^{-4.5} \end{array}$$

Solution

$$\begin{array}{ll} \text{(a)} & e^6 = 403.43 \\ \text{(b)} & e^{-4.5} = 0.011 \end{array}$$

Exercise 2

Question 2A Derive the recursion formula for $Z_{i,j}^I$. Allow insertions after deletions and vice versa.

Solution

$$Z_{i,j}^I = Z_{i,j-1}^I \times e^{\frac{\beta}{T}} + Z_{i,j-1}^M \times e^{\frac{g(1)}{T}} + Z_{i,j-1}^D \times e^{\frac{g(1)}{T}}$$

Question 2B Compute the partition function $Z(T)$ by dynamic programming for the sequences $\mathbf{x}=\text{ACC}$ and $\mathbf{y}=\text{AC}$. Allow insertions after deletions and vice versa. In order to simplify the computations, you can round to two digits after the decimal point.

Hint 1: Formulae Initialization:

$$\begin{aligned} Z_{i,0}^M &= Z_{0,j}^M = 0, Z_{0,0}^M = 1 \\ Z_{i,0}^I &= 0 \\ Z_{0,j}^D &= 0 \end{aligned}$$

Recursion:

$$\begin{aligned} Z_{i,j}^M &= Z_{i-1,j-1} \times e^{\frac{\sigma(x_i, y_j)}{T}} \\ Z_{i,j}^I &= Z_{i,j-1}^I \times e^{\frac{\beta}{T}} + Z_{i,j-1}^M \times e^{\frac{g(1)}{T}} + Z_{i,j-1}^D \times e^{\frac{g(1)}{T}} \\ Z_{i,j}^D &= Z_{i-1,j}^D \times e^{\frac{\beta}{T}} + Z_{i-1,j}^M \times e^{\frac{g(1)}{T}} + Z_{i-1,j}^I \times e^{\frac{g(1)}{T}} \\ Z_{i,j} &= Z_{i,j}^M + Z_{i,j}^I + Z_{i,j}^D \end{aligned}$$

Solution

| $Z^{\{M\}}$ | - | A | C |
|-------------|---|------|-------|
| - | 1 | 0.00 | 0.00 |
| A | 0 | 7.39 | 0.17 |
| C | 0 | 0.17 | 57.90 |
| C | 0 | 0.14 | 30.42 |

| $Z^{\{I\}}$ | - | A | C |
|-------------|---|------|------|
| - | 0 | 0.47 | 0.37 |
| A | 0 | 0.22 | 3.77 |
| C | 0 | 0.17 | 2.00 |
| C | 0 | 0.14 | 1.63 |

| $Z^{\wedge}\{D\}$ | - | A | C |
|-------------------|------|------|-------|
| - | 0.00 | 0.00 | 0.00 |
| A | 0.47 | 0.22 | 0.17 |
| C | 0.37 | 3.68 | 2.00 |
| C | 0.29 | 3.10 | 29.85 |

| Z | - | A | C |
|---|------|------|-------|
| - | 1.00 | 0.47 | 0.37 |
| A | 0.47 | 7.84 | 4.12 |
| C | 0.37 | 4.12 | 61.89 |
| C | 0.29 | 3.37 | 61.90 |

Exercise 3

The partition function of the reverse sequences $x^* = CA$ and $y^* = CCA$ is given in the matrix Z^* :

| $Z^{\wedge}\{*\}$ | - | A | C |
|-------------------|------|------|-------|
| - | 1.00 | 0.47 | 0.37 |
| A | 0.47 | 7.84 | 4.12 |
| C | 0.37 | 7.43 | 8.45 |
| C | 0.29 | 4.94 | 61.90 |

Question 3A Find a mapping from matrix $Z_{k,l}^*$ to $Z'_{i,j}$. Which position in matrix Z^* corresponds to which position in matrix Z' ?

Solution $Z'_{i,j}$ is the partition function of the alignment $x_j \dots x_{|x|}$ with $y_i \dots y_{|y|}$.

$Z_{k,l}^*$ is the partition function of the alignment $x_{|x|} \dots x_{|x|-k+1}$ with $y_{|y|} \dots y_{|y|-l+1}$.

$$j = |x| - l + 1 \Leftrightarrow l = |x| - j + 1$$

$$i = |y| - k + 1 \Leftrightarrow k = |y| - i + 1$$

| $Z^{\wedge}\{*\}$ | - | C | A |
|-------------------|-------|-------|-------|
| - | (0,0) | (0,1) | (0,2) |
| A | (1,0) | (1,1) | (1,2) |
| C | (2,0) | (2,1) | (2,2) |
| C | (3,0) | (3,1) | (3,2) |

| $Z \setminus \{ \}$ | - | A | C | -1 |
|---------------------|-------|-------|-------|-------|
| - | (0,0) | (0,1) | (0,2) | (0,3) |
| A | (1,0) | (1,1) | (1,2) | (1,3) |
| C | (2,0) | (2,1) | (2,2) | (2,3) |
| C | (3,0) | (3,1) | (3,2) | (3,3) |
| - | (4,0) | (4,1) | (4,2) | (4,3) |

Question 3B Use Z, Z^* and the mapping from Z^* to Z' to compute the probability of the alignment edges (1, 1), (2, 2), (3, 1) and (3, 2) between x and y .

Hint 1: Formulae

$$P(x_i \sim y_j | x, y) = \frac{Z_{i-1, j-1} \times e^{\frac{\sigma(x_i, y_j)}{T}} \times Z'_{i+1, j+1}}{Z(T)} Z_{i, j}^M = Z_{i-1, j-1} \times e^{\frac{\sigma(x_i, y_j)}{T}}$$

Hint 2 Mapped positions:

$$\begin{aligned} Z'_{2,2} &\iff Z_{2,1}^* \\ Z'_{3,3} &\iff Z_{1,0}^* \\ Z'_{4,2} &\iff Z_{0,1}^* \\ Z'_{4,3} &\iff Z_{0,0}^* \end{aligned}$$

Solution Alignment edge (1, 1):

$$P(x_1 \sim y_1 | x, y) = \frac{Z_{0,0} \times e^{\frac{\sigma(x_1, y_1)}{T}} \times Z'_{2,2}}{Z(T)} = \frac{Z_{0,0} \times e^{\frac{\sigma(x_1, y_1)}{T}} \times Z_{2,1}^*}{Z(T)} = \frac{Z_{1,1}^M \times Z_{2,1}^*}{Z(T)} = \frac{7.39 \times 7.43}{61.90} = 0.89$$

Alignment edge (2, 2):

$$P(x_2 \sim y_2 | x, y) = \frac{Z_{1,1} \times e^{\frac{\sigma(x_2, y_2)}{T}} \times Z'_{3,3}}{Z(T)} = \frac{Z_{1,1} \times e^{\frac{\sigma(x_2, y_2)}{T}} \times Z_{1,0}^*}{Z(T)} = \frac{Z_{2,2}^M \times Z_{1,0}^*}{Z(T)} = \frac{57.90 \times 0.47}{61.90} = 0.44$$

Alignment edge (3, 1):

$$P(x_3 \sim y_1 | x, y) = \frac{Z_{3,1}^M \times Z'_{4,2}}{Z(T)} = \frac{Z_{3,1}^M \times Z_{0,1}^*}{Z(T)} = \frac{0.14 \times 0.47}{61.90} = 0.001$$

Alignment edge (3, 2):

$$P(x_3 \sim y_2 | x, y) = \frac{Z_{3,2}^M \times Z'_{4,3}}{Z(T)} = \frac{Z_{3,1}^M \times Z_{0,0}^*}{Z(T)} = \frac{30.42 \times 1}{61.90} = 0.49$$