Олег Дмитриев, ВБиБ ВШЭ
ДЗ 2

Задание 1. Выведите рекуррентную формулу количества всех возможных выравниваний последовательностей длины n и m пользуясь разбиением всех выравниваний на непересекающиеся блоки.

Ответ: Пусть у нас есть две последовательности a и b длины m и n. Тогда обозначим число всех возможных выравниваний через $f(m,n)$. Мы можем разбить все выравнивания на непересекающиеся блоки в зависимости от того как заканчивается выравнивание:

$$\begin{pmatrix} \dots & a_m \\ \dots & b_n \end{pmatrix} \text{ или } \begin{pmatrix} \dots & \_ \\ \dots & b_n \end{pmatrix} \text{ или } \begin{pmatrix} \dots & a_m \\ \dots & \_ \end{pmatrix}.$$

Стоит также сказать, что следующие выравнивания будем считать не тождественными:

$$\begin{pmatrix} \dots & \mathrm{a_m} & \_ & \dots \\ \dots & \_ & \mathrm{b_1} & \dots \end{pmatrix}, \quad \begin{pmatrix} \dots & \_ & \mathrm{a_1} & \dots \\ \dots & \mathrm{b_n} & \_ & \dots \end{pmatrix}$$

С точки зрения биоинформатики мне кажется это правильным. Тогда можно получить рекуррентное выражение:

$$f(m,n) = f(m-1,m-1) + f(m,n-1) + f(m-1,n)$$

с начальными условиями:

$$f(0,k) = f(k,0) = f(0,0) = 1$$

Задание 2. Получите точную формулу, основываясь на начальные условия и рекуррентную формулу.

Ответ: Я не знал как вывести из рекуррентного соотношения, поэтому вывел из комбинаторных соображений. Пусть $m > n$, тогда верно:

$$f(m,n) = \sum_{i=m}^{m+n} \binom{i}{i-m} \binom{m}{n-(i-m)} = \sum_{j=0}^{n} \binom{j+m}{m} \binom{m}{n-j}$$

где i - длина полученного выравнивания; $\binom{i}{i-m}$ - количество различных выборов мест для гэпов в первой строке; $\binom{m}{n-(i-m)}$ - количество различных выборов мест для гэпов во второй строке с учетом того, что гэпам в первой строке соответствуют символы, а не гэпы.

Задание 3. Воспользуйтесь приближением Стирлинга чтобы получить приближенную формулу количества выравниваний.

Ответ: Формула Стирлинга для числа сочетаний:

$$\binom{n}{k} \approx \frac{\left(\frac{n}{e}\right)^n}{\left(\frac{n}{e}\right)^{n-k} e^{-k} k!} = \frac{n^k}{k!}$$

Тогда:

$$f(m,n) \approx \sum_{j=0}^{n} \frac{(j+m)^m}{m!} \frac{m^{n-j}}{(n-j)!}$$