

# Задачи оценивания значимости выравнивания при помощи скрытых марковских моделей

Власенко Даниил Владимирович  
Научный руководитель: к.ф.-м.н. Коробейников А.И.

Санкт-Петербургский государственный университет  
Кафедра "Статистического моделирования"

Санкт-Петербург  
Декабрь 2021

# Выравнивание последовательностей

## Определение

*Выравнивание последовательностей — размещение двух или более последовательностей друг под другом таким образом, чтобы было легче увидеть их схожие участки.*

A	C	E	A	A	F	A	E	
C	E	A	F	D	C	E		
A	C	E	A	A	F	A	—	E
—	C	E	A	—	F	D	C	E

## Определение

*Значимость выравнивания — действительное число  $s$ , отражающее сходство последовательностей.*

- достаточно ли высокая значимость, чтобы считать последовательность не шумом, или шум мог добиться такой значимости.
- достаточно ли низкая значимость, чтобы считать последовательность шумом, или не шум мог получить такую значимость.

## Определение

*Ложноположительная вероятность значимости  $s$  — это вероятность того, что шум получит значимость равную или выше  $s$ .*

## Определение

Пусть  $X_n$  и  $Y_n$  дискретные стохастические процессы,  $n \geq 1$ . Пара  $(X_n, Y_n)$  называется скрытой марковской моделью, если

- $X_n$  — марковский процесс, поведение которого напрямую не наблюдается ("скрытый");
- $P(Y_n = y_n | X_1 = x_1, \dots, X_n = x_n) = P(Y_n | X_n = x_n)$  для любого  $n \geq 1$ , где  $x_1, \dots, x_n$  — значения, принимаемые процессом  $X_n$  (**состояния модели**),  $y_n$  — значение, принимаемое процессом  $Y_n$  (**наблюдаемый символ модели**).

