

1) Не получается задавать т-паттерн с фиксированным первым элементом.

В случае двухэлементного паттерна полная неразбериха. Первый символ вообще почти никогда не учитывается. Я считаю, что первому символу нужно давать какую-то номинальную дисперсию.

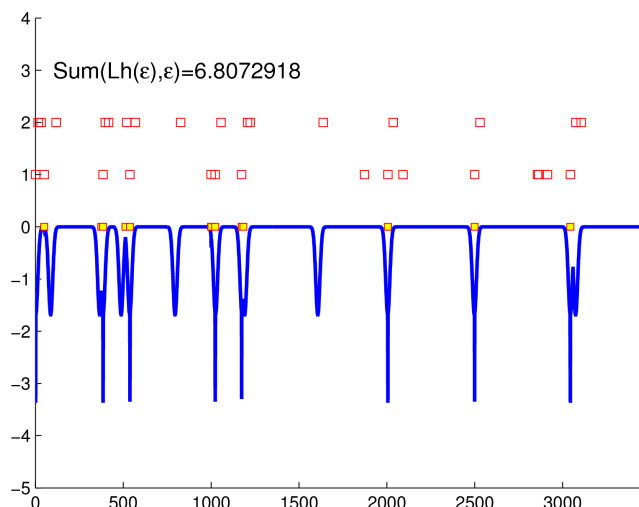


Illustration 1: Фиксирован 1-й элемент. Мало шума. Паттерн есть.

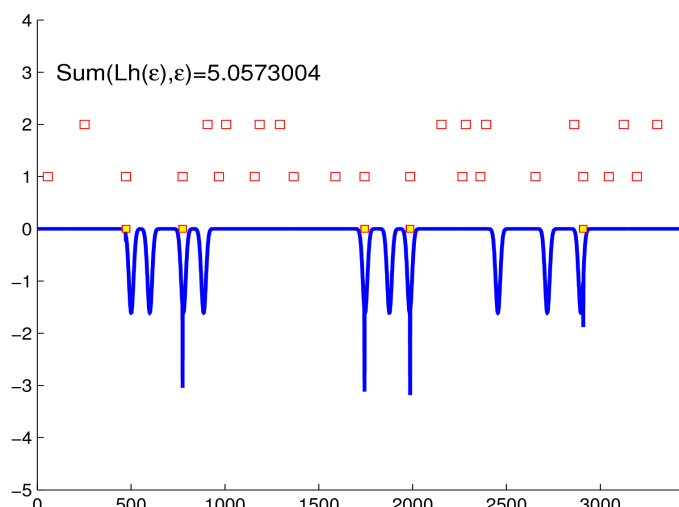


Illustration 2: Фиксирован 1-й элемент. Мало шума. Паттерна нету.

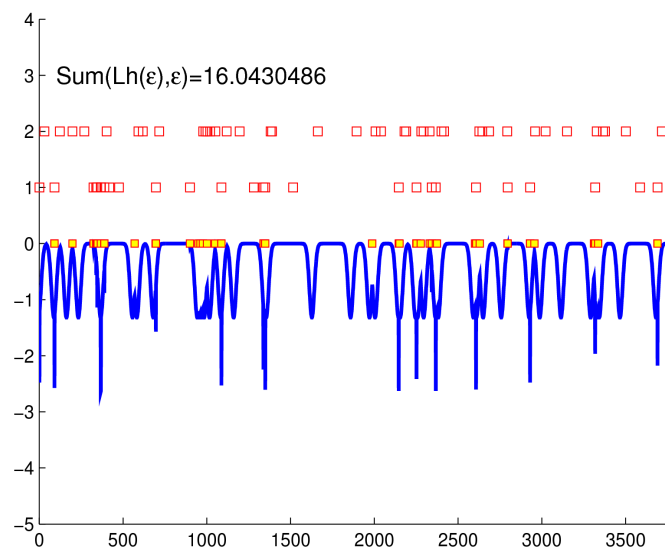


Illustration 3: Фиксирован 1-й элемент. Много шума. Паттерн есть.

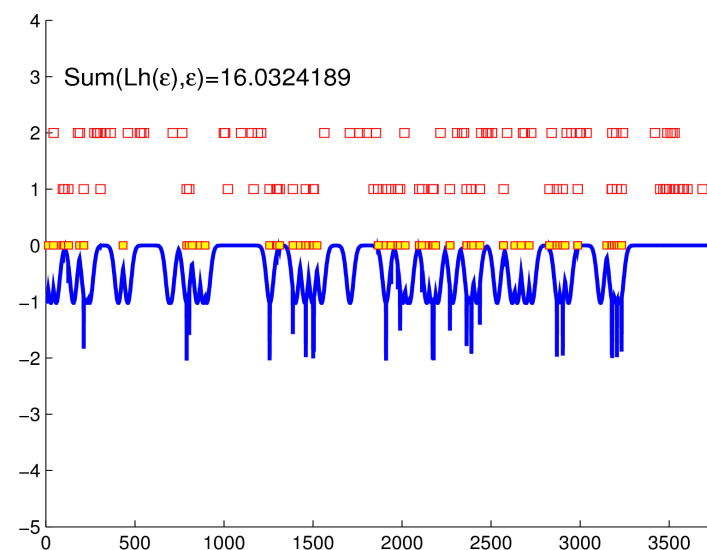


Illustration 4: Фиксирован 1-й элемент. Много шума. Паттерна нету.

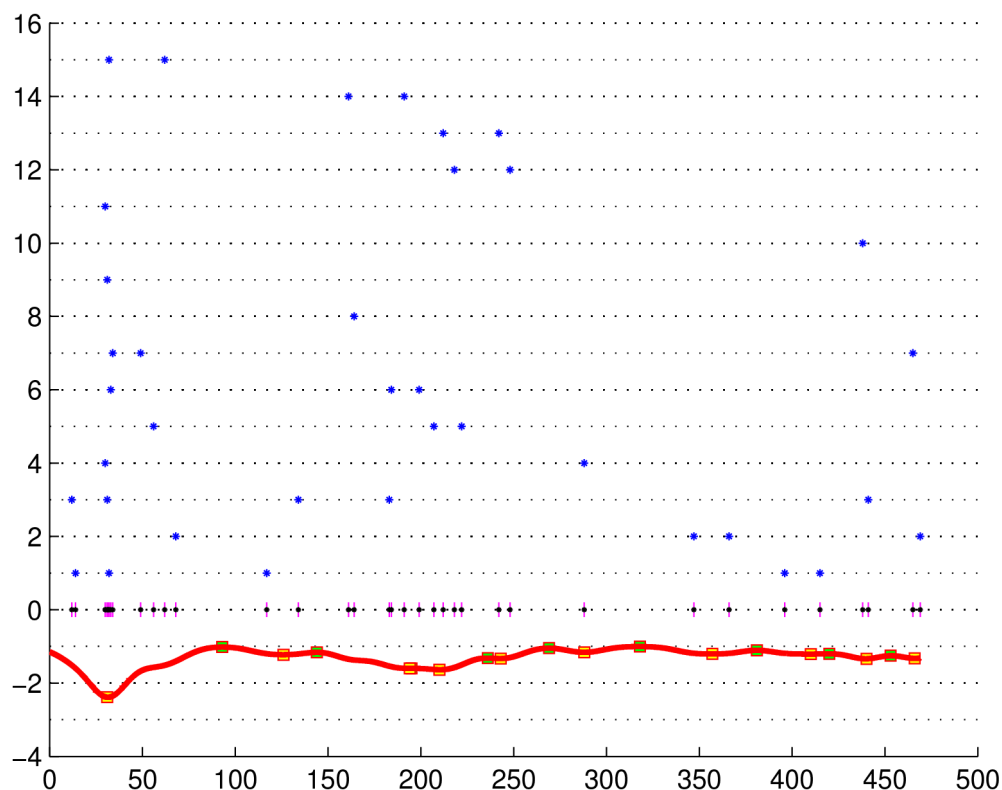
2) Когда мы полагаем, что символ именно пропущен, а не просто далеко находится.

Сложная задача оптимизации, в общем случае перебор 2^n и это N_i раз. При подсчете правдоподобия паттерна в каждой точке непонятно, как лучше считать.

3) При определении паттернов.

Например, мы хотим посмотреть связь между паттернами ab и cd, при условии, что существует паттерн abcd. Тогда можно:

1. Смотреть распределения расстояний между концами ab и началами cd. Получится что-то типа этого:



но, нам придется учитывать само количество появлений паттерна, от чего мы в принципе хотим уйти. Тут есть свои сложности, но в принципе это будет работать.

2. Можно смотреть смещения между функциями правдоподобия.

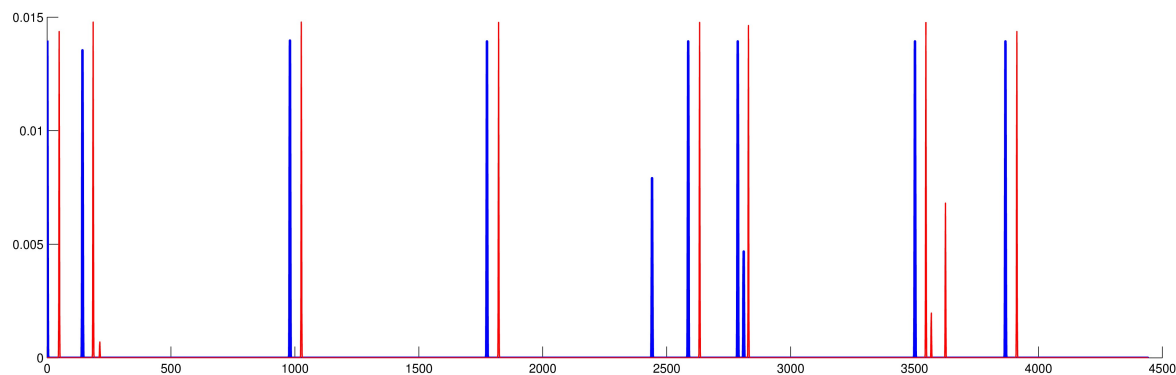
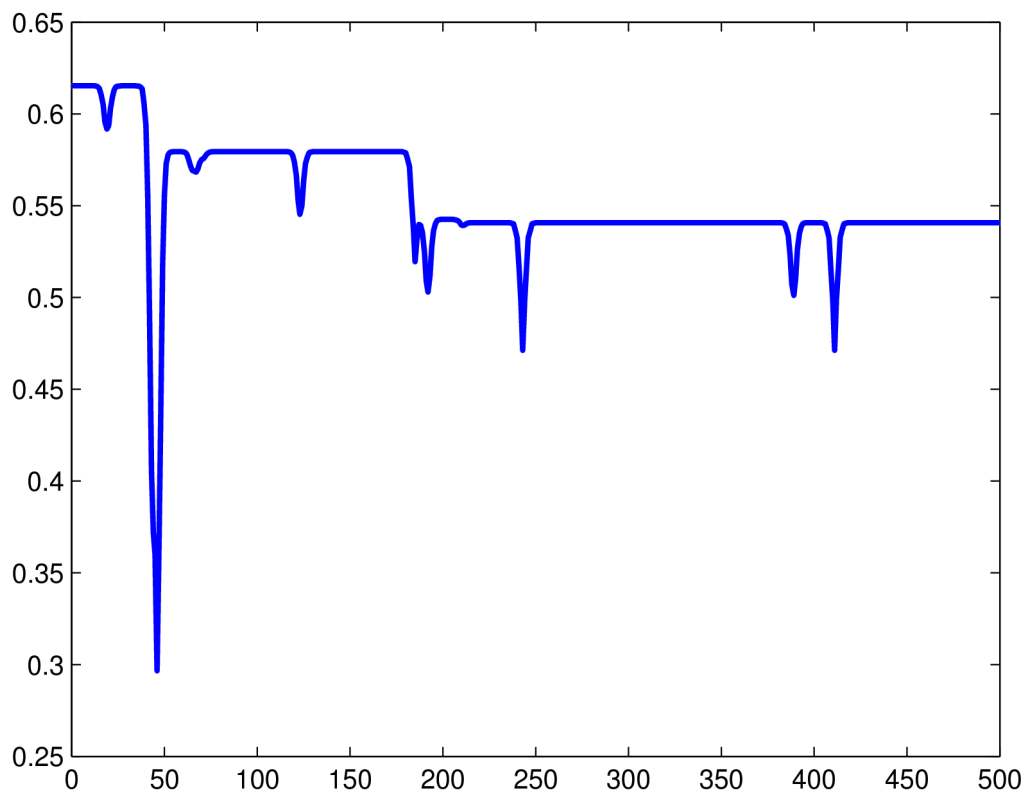


Illustration 5: Синий - правдоподобие первого паттерна. Красный - второго.

Тогда, если построить график зависимости величины «непохожести» (интеграл от модуля разности) функций в зависимости от смещения, то там где функции больше всего похожи, тем расстоянием эти паттерны и связаны.



Здесь минимум функции достигается в точке 46. Однако, это расстояние, связывающее *начало* паттерна ab с началом паттерна cd. В принципе, ничего не мешает считать правдоподобие первого паттерна относительно его конца..

Так же не очень ясно как считать дисперсию, с которой будут связаны паттерны ab и cd.

Как эффективно искать смещение?

3. Скорее всего надо нормировать правдоподобие – делить на кол-во элементов