Имеем 2 паттерна, которые хотим соединить. Как-то зафиксировали их начала и концы. Далее, имеем распределение расстояний на которых второй паттерн встречается после первого. Каждая точка в этом распределении говорит, на каком расстоянии находится конец j-го паттерна от начала i-го. У каждой реализации паттерна(тут мы пришли к дискретизации и называем конкретные моменты времени, где паттерн имеет место быть!) существует свое правдоподобие(α_i, β_j).

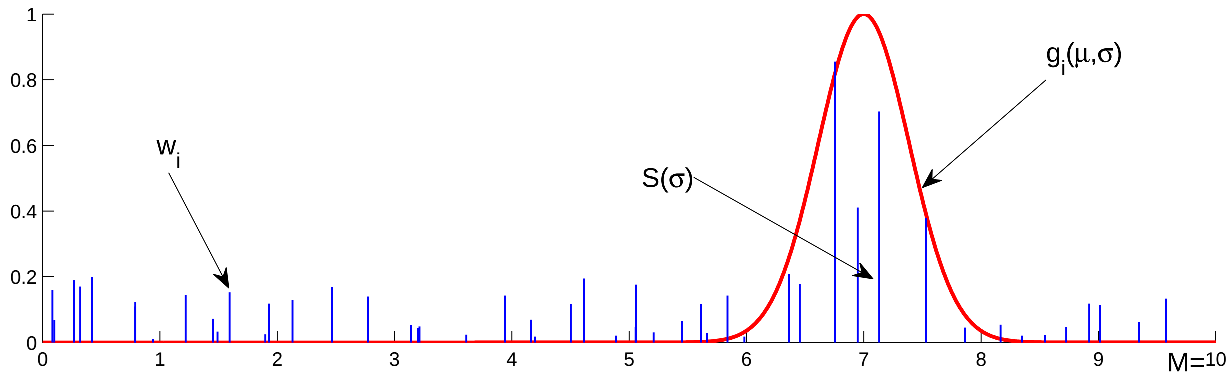
Рассмотрим следующую сумму k:

$$\begin{aligned} k &= \sum_{i=1}^{N} w_i g_i(\mu, \sigma), \\ w_i &= \alpha_i \beta_i, \\ g_i(\mu, \sigma) &= \frac{1}{\sqrt{2\pi}\sigma} exp\left(-\frac{(x_i - \mu)^2}{\sigma^2}\right) \end{aligned}$$

Так же, пусть:

$$\begin{aligned} S(\sigma) &= \sigma\sqrt{2\pi}, \text{ площадь под гауссианой;} \\ N &— \text{ количество точек в распределении;} \\ N_{eff} &= \frac{\sum_{i=1}^{N} w_i}{\max_i w_i}; \\ M &— \text{ размер окна} \end{aligned}$$

В общем примерно так:

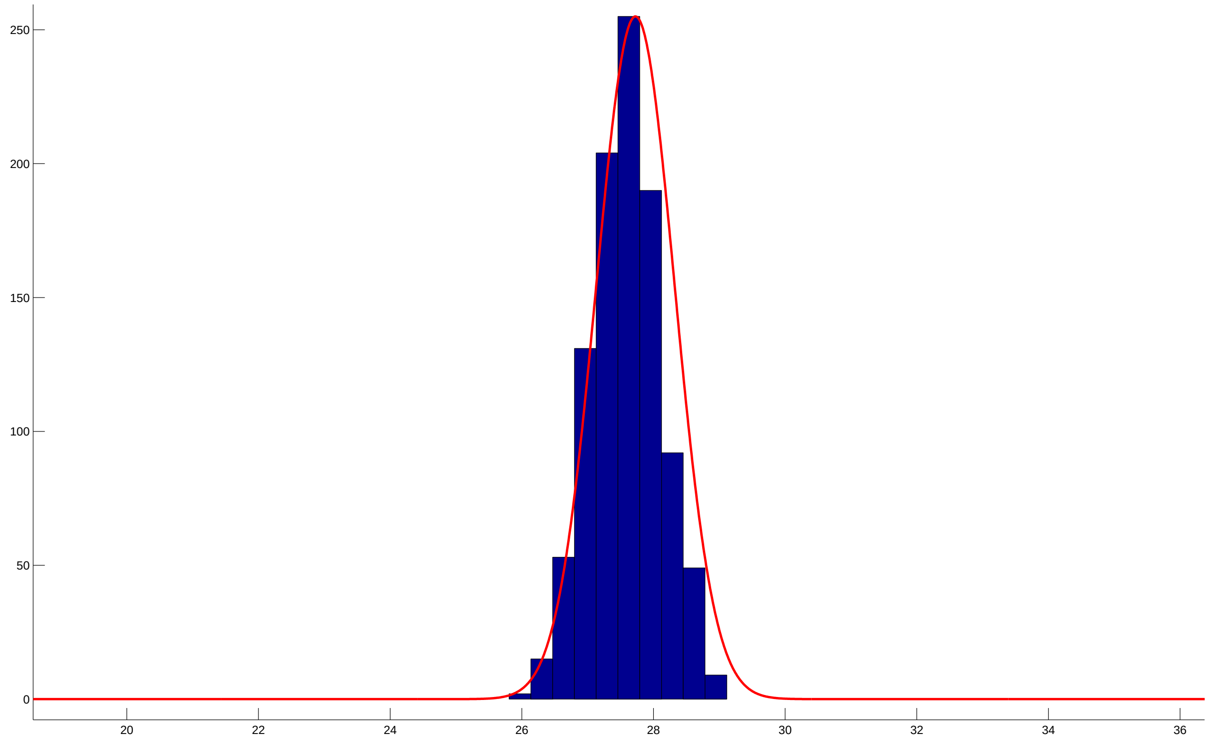


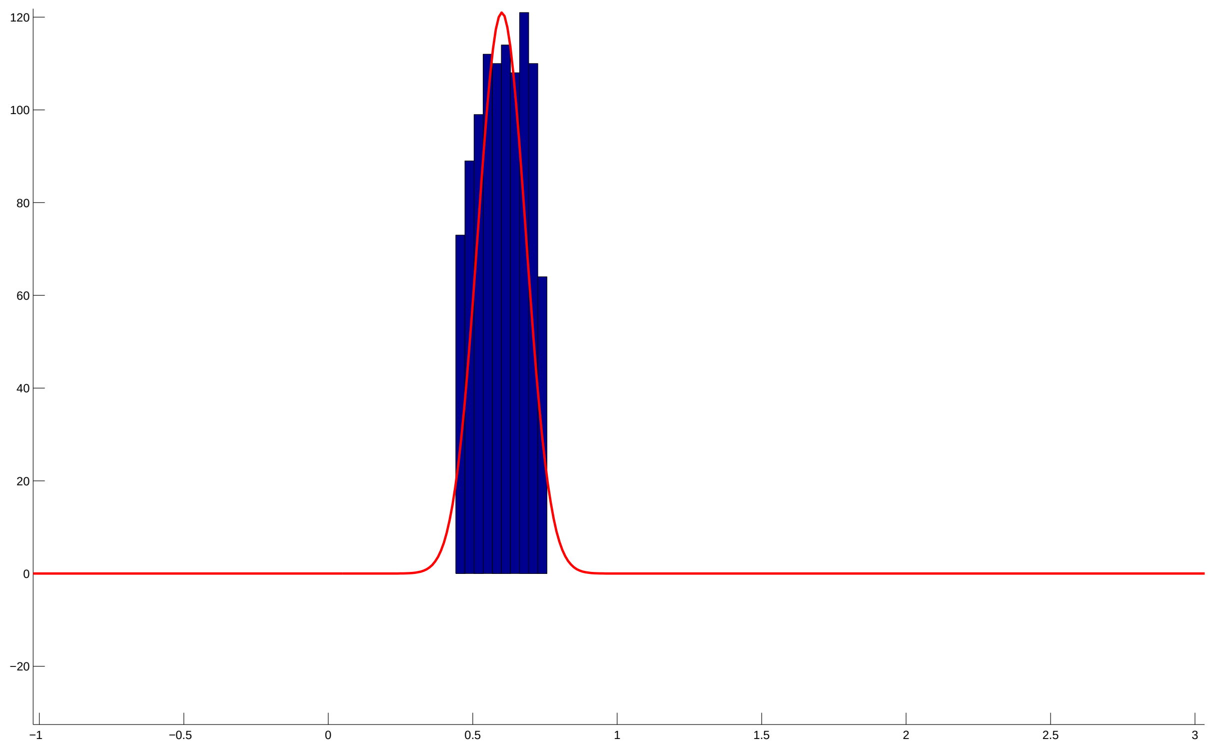
Если принять гипотезу о равномерном распределении расстояний(нулевая гипотеза – в данных нету никаких закономерностей, то случайная величина k будет иметь, по ЦПТ, при $N \to \infty$ следующее распределение:

$$k \sim N\left(\frac{S}{M}N, \frac{(M-S)S}{M^2}N\right), \text{— если } w_i = 1$$

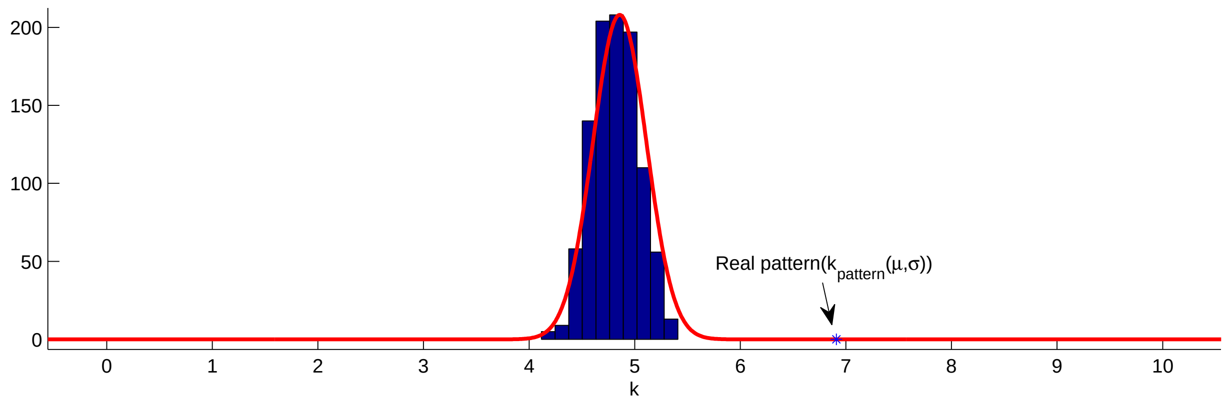
$$k \sim N\left(\frac{S}{M}N\ mean(w_i), \frac{(M-S)S}{M^2}Ndisp(w_i)\right), \text{— пытаемся учесть } w_i \quad (1)$$

Вот что получается на практике, если w_i брать из равномерного распределения(более мы ничего не можем о нем сказать). Надо отметить, что при $N < 50$ мат ожидание бывает слегка сдвинуто. Следующие графики представлены для значений N 1000 и 20 соответственно.





Теперь нужно на реальных данных вычислить значение величины k и посмотреть на сколько оно было вероятно, исходя из формулы (1). Чем меньше вероятность данного события в случайных денных, тем больше оно похоже на паттерн. Идея, в общем, такая...



Таким образом, мы должны минимизировать:

$$\begin{cases} N_{\left(\frac{S}{M}N\ mean(w_i),\frac{(M-S)S}{M^2}Ndisp(w_i)\right)}[k_{pattern}(\mu,\sigma)] \rightarrow \min\limits_{\mu,\sigma} \\ k_{pattern}(\mu,\sigma) > \dfrac{S}{M}N\ mean(w_i) \end{cases}$$

Здесь второе условие использовано для того что бы показать, что нам нужен именно правый «хвост» гауссианы.