research summary week11

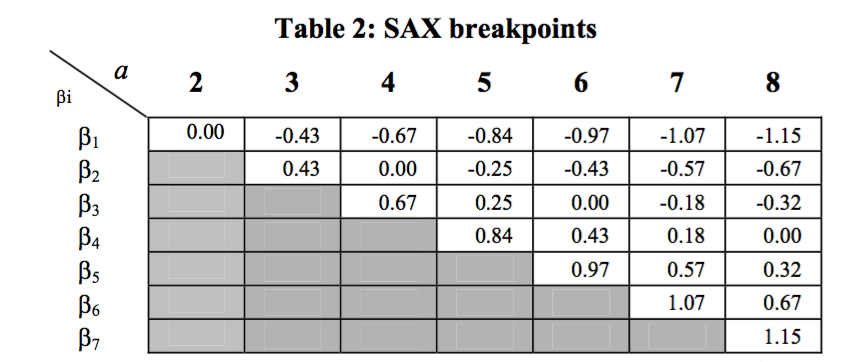
SAX的距离计算、lower bound和紧度

# SAX 与iSAX

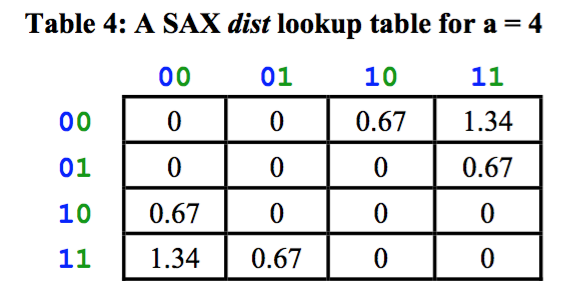
Sax采用Gaussian curve来划分，首先会将PAA归一化，均值为0，标准差为1.

We normalize each time series to have a mean of zero and a standard deviation of one before converting it to the PAA.

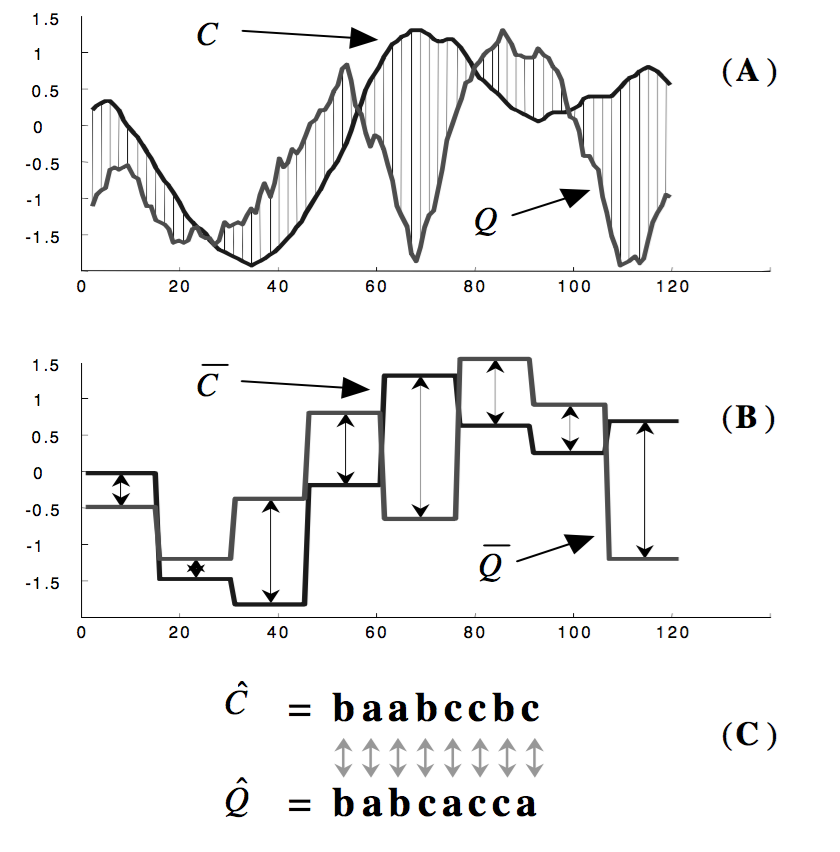
根据概率分布密度，整个区间会被等分为等概率的a份：



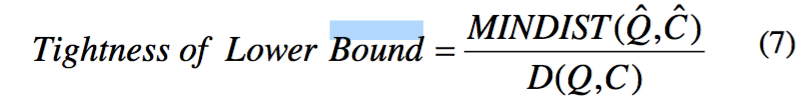
由于建立索引的需要，在iSAX中采取了a=2n的划分方式，由上图可以看到，其实a=4的划分完全是a=8划分的子集。



距离度量采用尽可能小的估计，对于a=4，最远可能相差0.67\*2这么多，相邻的两块之间距离算作0，保证lower bound。



紧界估计：tightness of lower bound:



SAX的划分的精度是可以控制的。同时非常适合于index，因为他是自顶向下细分的；suffix不需要？



当使用的bit位增多时，lower bound tightness会变得很好。

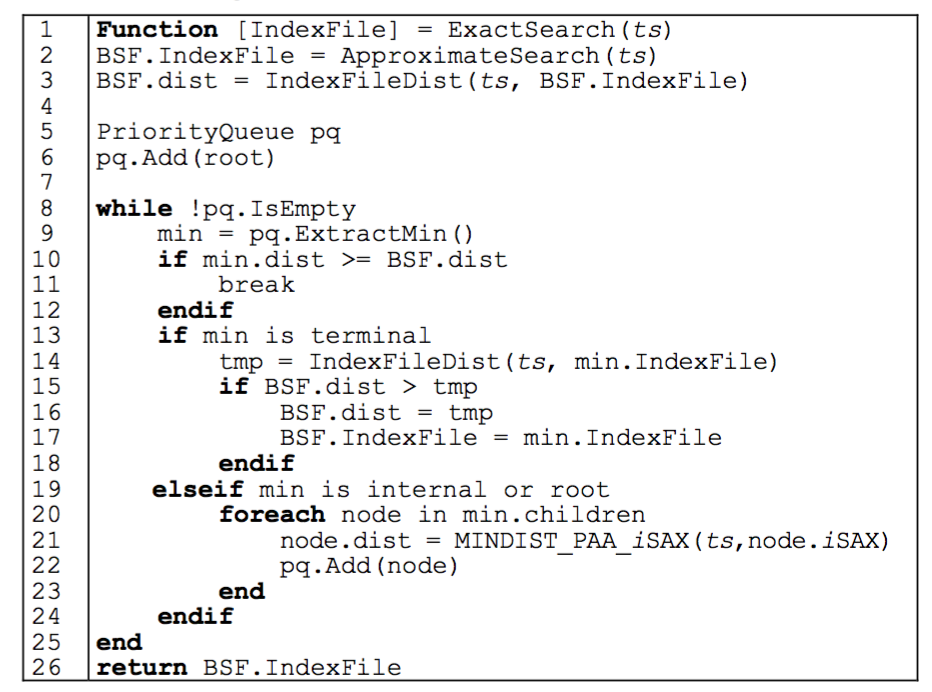


ultra fast approximate query:

The method of approximation is derived from the intuition that two similar time series are often represented by the same *i*SAX word. Given this assumption, the approximate result is obtained by attempting to find a terminal node in the index with the same *i*SAX representation as the query.

由于iSAX具有二叉树性质，没有overlap，因此直接将待查询的片段去索引中寻找。如果有则返回，没有则回到最后一次分裂的节点再向下寻找。

Exact search

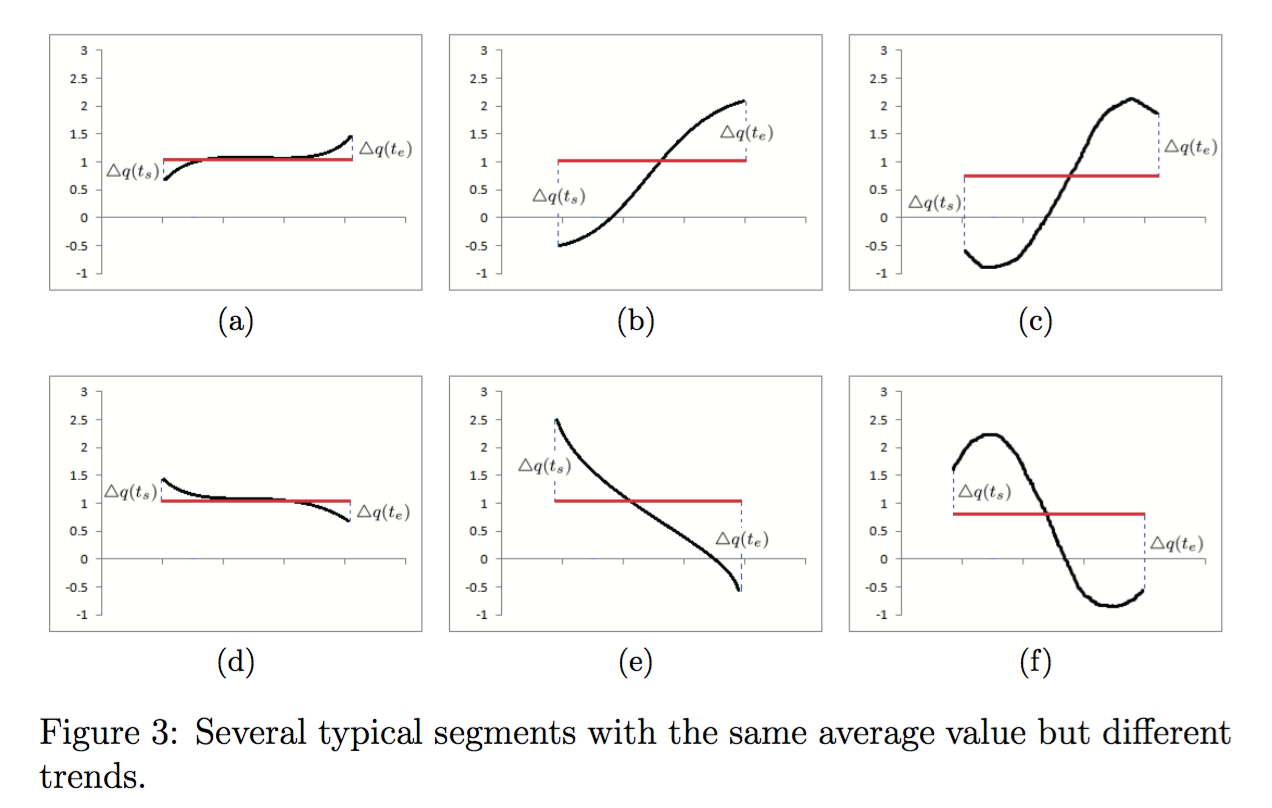


遍历式查找；

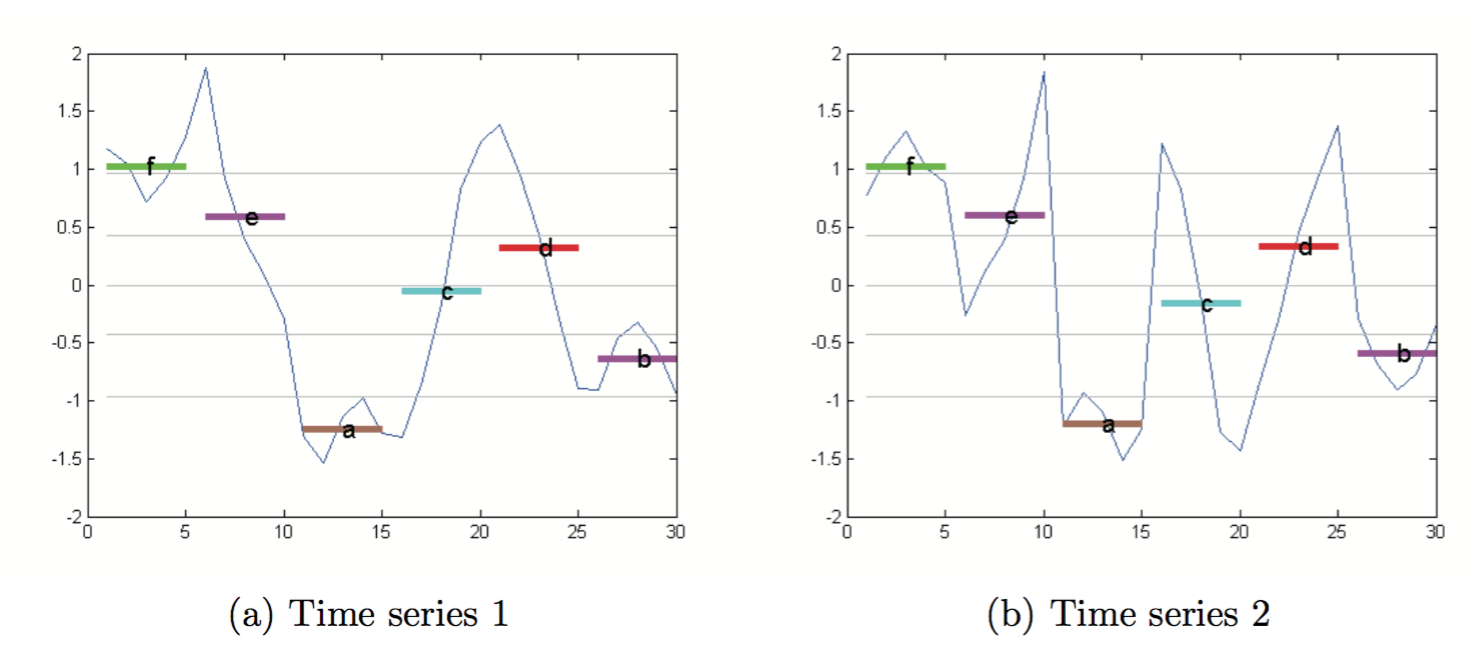
# 3.SAX-TD

An Improvement of Symbolic Aggregate Approximation Distance Measure for Time Series

(a) level and slight up, (b) obvious up, (c) down, up and then down, (d) level and slight down, (e) obvious down, (f) up, down and then up.



original SAX kas a major limitation:



argue: the real time series seems more complete and disorder.

Three

# 4. Extended SAX

Extended SAX: Extension of Symbolic Aggregate Approximation for Financial Time Series Data Representation