1 背景介绍

1.1 线性复杂度

对于一个给定的二元序列$s=s_0s_1\cdots s_{N-1}$，假如有一种线性关系能够由前k个元素将后续所有的元素都推导出来，即存在一组二元域上的系数$\lambda_1,\lambda_2,\cdots,\lambda_k$，使得

$$s_i=\lambda_1 s_{i-1}+\lambda_2 s_{i-2}+\cdots+\lambda_k s_{i-k} \qquad i=k+1,k+2,\cdots,N-1$$

成立。称成员最少的这样一组系数的个数k为序列s的线性复杂度。且规定全零序列的线性复杂度为0.

1.2 线性复杂度的性质

对于长度为N的序列$s=s_0s_1\cdots s_{N-1}$，记$L_n(s)$为其前n位的线性复杂度，即其子序列$s_0s_1\cdots s_{n-1}$的线性复杂度。

性质 1.2.1 $0\leq L_N(s)\leq N$

性质 1.2.2 *在已知$L_n(s)$的前提下，如果$n+1-L_n(s)>L_n(s)$时，L_{n+1}可能的等于$L_n(s)$或$n+1-L_n(s)$，取这两个值的可能性各占一半；否则L_{n+1}只能等于$L_n(s)$。*

1.3 线性复杂度图谱

对于给定的序列$s=s_0s_1\cdots s_{N-1}$，先假定$L_0(s)=0$。将点

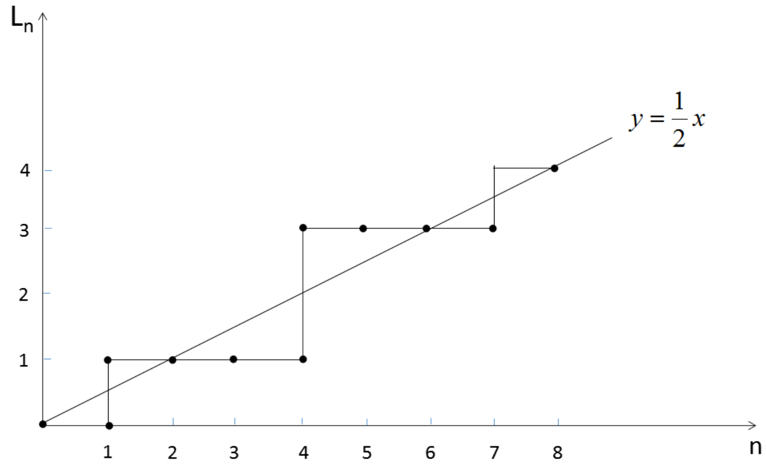
$$(0,L_0(s)),(1,L_0(s)),(1,L_1(s)),\cdots,(N,L_{N-1}(s)),(N,L_N(s))$$

依次连接得到的图叫做序列s的线性复杂度图谱。因为一般的线性复杂度图谱都死围绕直线$y=\frac{1}{2}x$上下波动的，我们一般也把直线$y=\frac{1}{2}x$画出以作参照。

一个典型的线性复杂度图谱如下所示。

定义 1.3.1 *对于某一长度为N的序列s，定义其**k阶离差和**（k=1,2,3,.....）为*

$$D^k(s)=\sum_{i=1}^{N}\mid L_i(s)-\frac{i}{2}\mid^k$$



2 问题描述

对于长度为$2W$的序列s，假如该序列的每一位都是随机取值（即以$\frac{1}{2}$的概率取1，以$\frac{1}{2}$的概率取0，并且不同位之间的取值相互独立），则其k阶离差和只能取有限的值，因而其k阶离差和的概率分布是一个离散分布。我们的目标是计算出$W = 250$即长度为500的随机序列的3阶离差和的概率分布。我们记该随机变量为D_{2W}^k，其概率分布为F_{2W}^k。因为D_{2W}^k只取有限的一些值，因此可将F_{2W}^k表示为

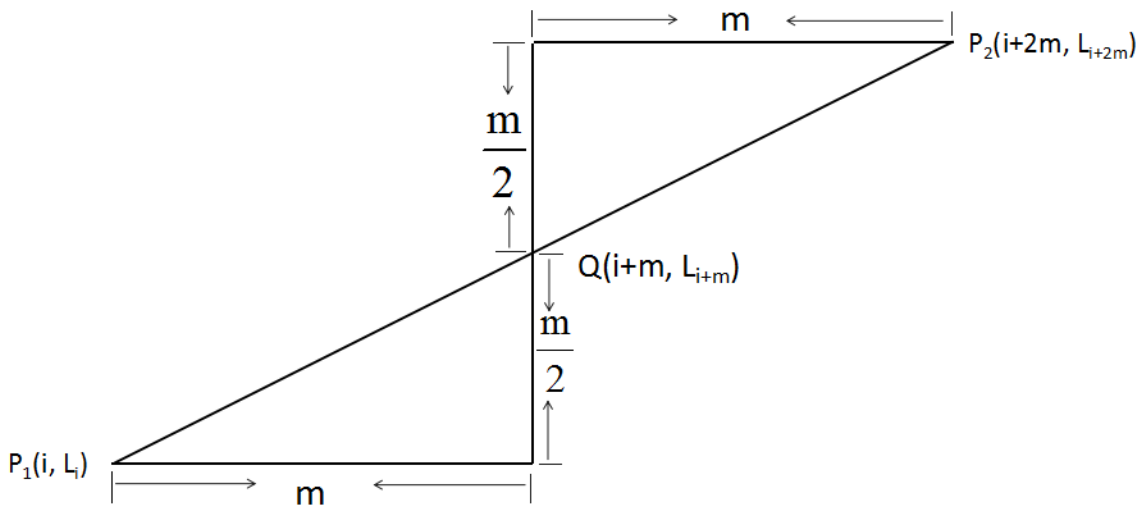
$$F_{2W}^k = \{[a_1, \Pr(a_1)], [a_2, \Pr(a_2)], \cdots\}$$

由线性复杂度的性质可知，当某点$(n, L_n(s))$在直线$y = \frac{1}{2}x$之下或就在该直线上时，下一点的$L_{n+1}(s)$的值均以$\frac{1}{2}$的概率取$L_n(s)$或$n+1-L_n(s)$，在图形上的直观意义是下一点等可能地保持水平或者跳到直线$y = \frac{1}{2}x$上面去；而当该点在直线$y = \frac{1}{2}x$上方时，下一点必然维持水平。

2.1 解决算法

对于求长度为$2M$的随机序列的k阶离差和的概率分布，我们采用的是一种递归的思想。对此首先来介绍一个概念：全等三角形对。

在一个线性复杂度图谱和参照直线$y = \frac{1}{2}x$组成的图中，我们经常能找到一对或者多对全等的直角三角形，如图所示称这样的一对全等三角形为全等三角对，以下简称三角对。三角对中的一个三角形的水平宽度为m，将



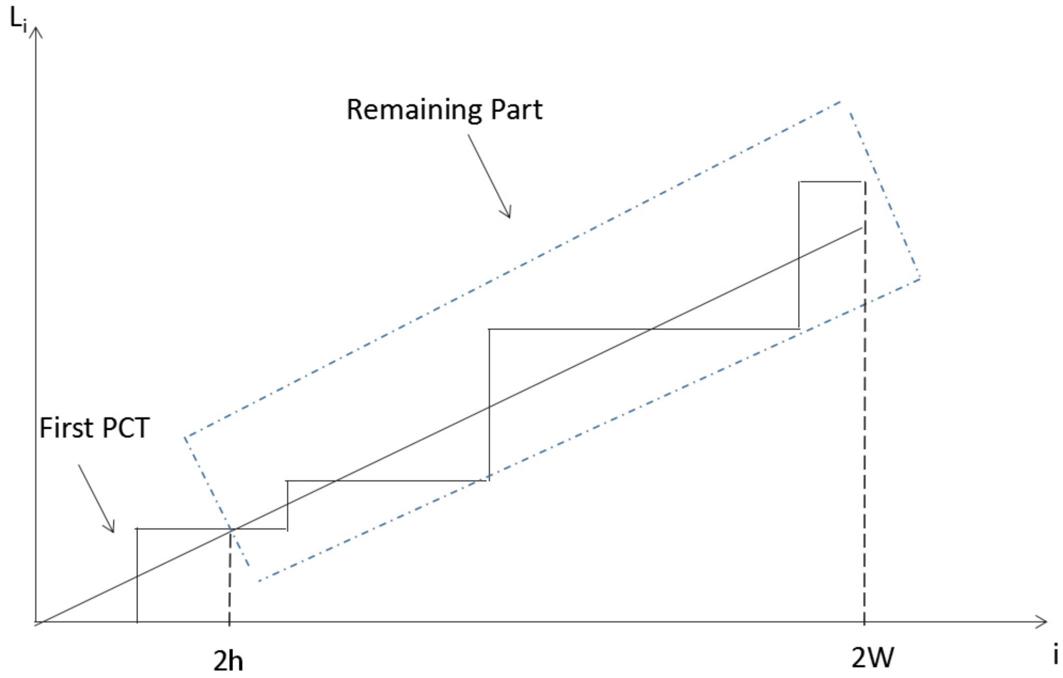
此三角对记为T_m。

如此一来，假如我们已知$F_0, F_2, F_4, \cdots, F_{2W-2}$，欲求$F_{2W}$，则如下图所示，我们可以将整个线性复杂度图谱分为一个从原点开始的三角对T_h和剩余的部分，显然这里h可取$1, 2, \cdots, W$。而且，T_h出现的概率很容易计算，易知

$$
\begin{aligned}
\Pr(T_h) =& \frac{1}{2^h} \\
D^k(T_h) =& \sum_{i=1}^{2h} \mid L_i(s) - \frac{i}{2} \mid^k \\
=& 2\sum_{i=1}^{h} \mid L_i(s) - \frac{i}{2} \mid^k - \mid L_h(s) - \frac{h}{2} \mid^k
\end{aligned}
$$

而剩余部分的水平宽度将小于$2W$，因此对于第一个三角对为T_h的情况，其概率分布可以看成以出现T_h为条件的条件概率分布。为方便表述，我们先引入运算符号$\circ$:

$$
[x_1, y_1] \circ [x_2, y_2] = [x_1 + x_2, y_1 \times y_2]
$$
$$
\{[x_1, y_1], [x_2, y_2], \cdots\} \circ [x, y] = \{[x_1, y_1] \circ [x, y], [x_2, y_2] \circ [x, y], \cdots\}
$$



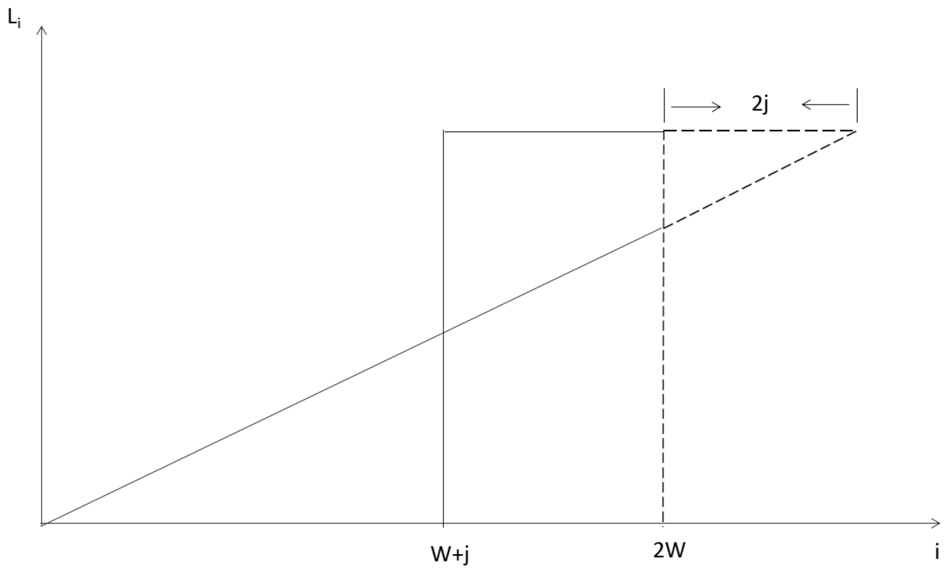
于是，对于第一个三角形是T_h的情况，这一部分的概率分布为

$$F_{2W-2h} \circ [D^k(T_h), \Pr(T_h)]$$

但是还有一种情况没考虑进来，就是整个线性复杂度图谱根本就连一个完整的三角对都没有，换句话说，整个线性复杂度图谱就是个残缺的三角对，正如下图所示： 同样，为了方便表述我们用T_h'来表示上图所示的不完整三角对，这表明其下底边长为h，易知j可取$W+1, W+2, \cdots, 2W$。易知

$$\begin{aligned}
\Pr(T_h') =& \frac{1}{2^h} \\
D^k(T_h') =& \sum_{i=1}^{2W} \mid L_i(s) - \frac{i}{2} \mid^k \\
=& 2\sum_{i=1}^{h} \mid L_i(s) - \frac{i}{2} \mid^k - \sum_{i=1}^{2h-2W-1} \mid L_i(s) - \frac{i}{2} \mid^k
\end{aligned}$$

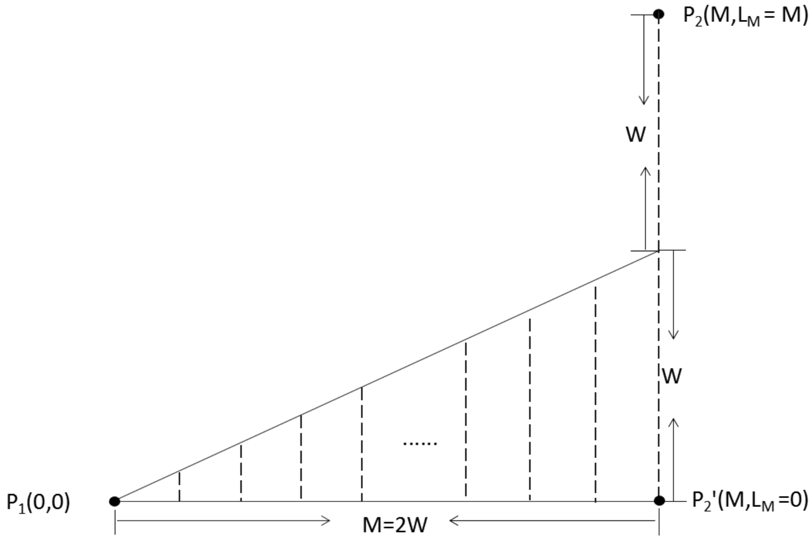
这里需要特别说明的是上面的计算概率公式只对$h = W+1, W+2, \cdots, 2W-1$成立，因为对于$T_{2W}'$是有两种情况与之对应的，其实也就是最后一点跳与



不跳这两种情况，如下图所示： 也就是说$\Pr(T'_{2W}) = 2\frac{1}{2^{2W}} = \frac{1}{2^{2W-1}}$。

至此，我们可得得到计算F_{2W}的递归公式

$$F_{2W} = (\bigcup_{h=1}^{W} F_{2W-2h} \circ [D^k(T_h), \Pr(T_h)]) \bigcup (\bigcup_{h=W+1}^{2W} [\Pr(T'_h), D^k(T'_h)])$$


P2(M,LM = M)
W
W
......
P1(0,0)
P2'(M,LM =0)
M=2W