**一，预备研究**

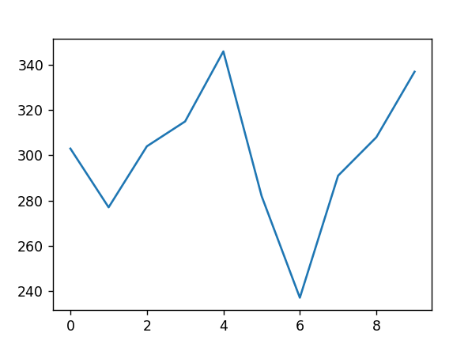
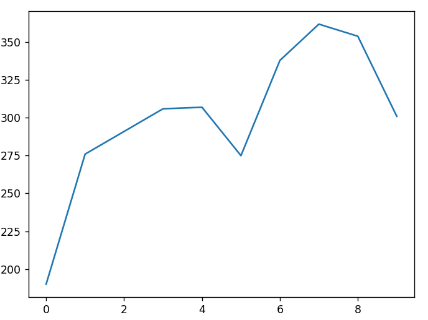
**1.1，相邻序列尾数相关性**

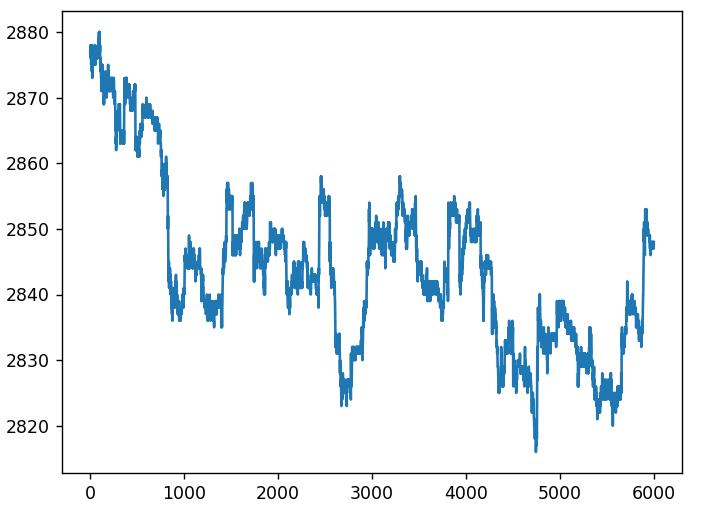
对玉米20230317.2258之前的45747根1min\_close\_bar临近1分钟线的尾数进行统计

tail（close[-3000:]）左

tail（close[-6000:-3000]）右

close[-6000:]下





猜想，顺序相关可能会有更高的相关性，下面讨论临近序列的尾数相关性关系。

***研究1：***

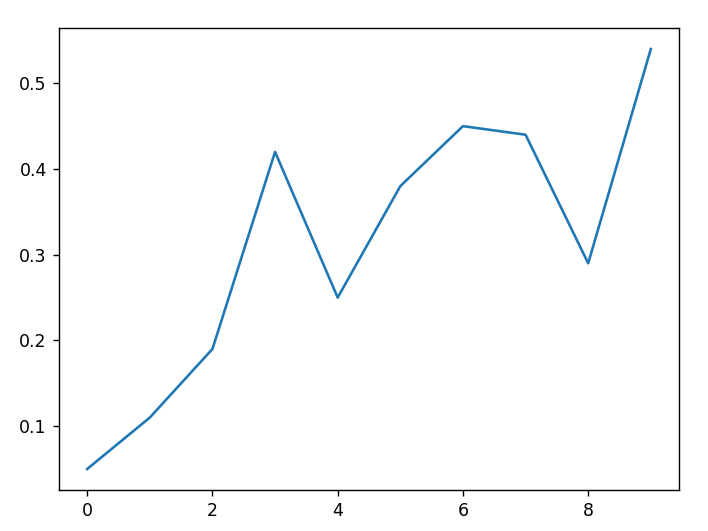
下图为step=1k到10k为单位的相邻，尾数的相关性。

各个周期内都呈现正相关性并不明显且波动较大，均值为

[0.05, 0.11, 0.19, 0.42, 0.25, 0.38, 0.45, 0.44, 0.29, 0.54]，

整体随step递增

在4k、7k、10k取峰值。

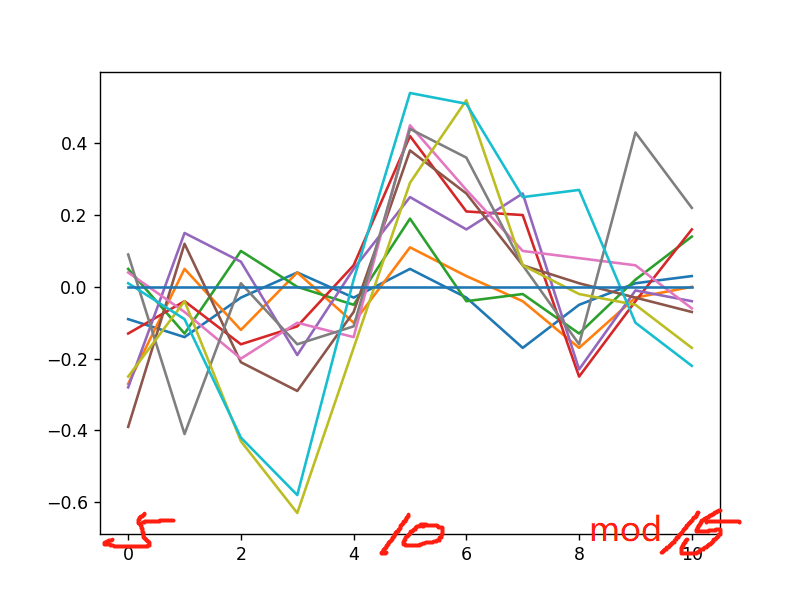


***研究2：***

下图为step=1k to 10k，mod=5 to 15的相邻，尾数的相关性。

各个step内相关性均值整体在mod=10时最高，在mod=8时整体最小。

在step=4k，mod=5 to 15的相邻，在mod=10最大0.42



***研究3：***

连续顺序的相关性，与1000次随机顺序的相关性

step=1k,mod=5时，连续顺序的相关性：-0.0923；1000次随机顺序的相关性：0.0054

step=1k,mod=6时，连续顺序的相关性：-0.1388；1000次随机顺序的相关性：-0.0093

step=1k,mod=7时，连续顺序的相关性：-0.0325；1000次随机顺序的相关性：-0.0106

step=1k,mod=8时，连续顺序的相关性：0.0408；1000次随机顺序的相关性：-0.0100

step=1k,mod=9时，连续顺序的相关性：-0.0271；1000次随机顺序的相关性：0.0050

step=1k,mod=10时，连续顺序的相关性：0.0522；1000次随机顺序的相关性：0.0279

step=1k,mod=11时，连续顺序的相关性：-0.028；1000次随机顺序的相关性：0.0267

step=1k,mod=12时，连续顺序的相关性：-0.1691；1000次随机顺序的相关性：0.0058

step=1k,mod=13时，连续顺序的相关性：-0.0502；1000次随机顺序的相关性：0.0065

step=1k,mod=14时，连续顺序的相关性：0.0058；1000次随机顺序的相关性：0.0021

step=1k,mod=15时，连续顺序的相关性：0.0285；1000次随机顺序的相关性：0.0015

首先，可以观察到，大多数，都是连续的abs大于随机的abs，体现了近期的相关性的特殊性，和无序相关性的普遍性。且mod<10时，无序相关性收敛到0。

然而，我们其中注意到mod=11时较为特殊，无序相关性0.0267大于近期相关性-0.028，这说明在mod11时，远期的相关性远大于近期的相关性。实验：

step=0.5k,mod=11时，连续顺序的相关性：0.0203；1000次随机顺序的相关性：0.0142

step=1k,mod=11时，连续顺序的相关性：-0.028；1000次随机顺序的相关性：0.0294

step=2k,mod=11时，连续顺序的相关性：0.0295；1000次随机顺序的相关性：0.0551

step=3k,mod=11时，连续顺序的相关性：-0.0419；1000次随机顺序的相关性：0.0898

step=4k,mod=11时，连续顺序的相关性：0.2118；1000次随机顺序的相关性：0.1453

step=5k,mod=11时，连续顺序的相关性：0.1608；1000次随机顺序的相关性：0.0954

step=6k,mod=11时，连续顺序的相关性：0.262；1000次随机顺序的相关性：0.1800

step=7k,mod=11时，连续顺序的相关性：0.2723；1000次随机顺序的相关性：0.1578

step=8k,mod=11时，连续顺序的相关性：0.3612；1000次随机顺序的相关性：0.1533

step=9k,mod=11时，连续顺序的相关性：0.5246；1000次随机顺序的相关性：0.4182

step=10k,mod=11时，连续顺序的相关性：0.5144；1000次随机顺序的相关性：0.3545

而：

step=1k,mod=10时，连续顺序的相关性：0.0522；1000次随机顺序的相关性：0.0292

step=2k,mod=10时，连续顺序的相关性：0.1134；1000次随机顺序的相关性：0.0823

step=3k,mod=10时，连续顺序的相关性：0.1881；1000次随机顺序的相关性：0.0980

step=4k,mod=10时，连续顺序的相关性：0.4197；1000次随机顺序的相关性：0.1233

step=5k,mod=10时，连续顺序的相关性：0.2498；1000次随机顺序的相关性：0.1595

step=6k,mod=10时，连续顺序的相关性：0.3759；1000次随机顺序的相关性：0.2199

step=7k,mod=10时，连续顺序的相关性：0.4481；1000次随机顺序的相关性：0.3754

step=8k,mod=10时，连续顺序的相关性：0.4433；1000次随机顺序的相关性：0.2677

step=9k,mod=10时，连续顺序的相关性：0.2852；1000次随机顺序的相关性：0.2684

step=10k,mod=10时，连续顺序的相关性：0.5394；1000次随机顺序的相关性：0.4181

可知：各个step的mod=11的无序相关性均大于0，而顺序相关性整体随step递增。一个猜想性的解释，该现象说明mod11或者临近mod10的mod值代表了一定的近期行情的倾斜特征，因而其顺序相关性与step正相关。进一步研究mod9,mod8等略小的价格周期，会发现相类似的显现，说明了与mod10临近的价格周期的形成与行情的倾斜特征联系很大。可以作为斜向网格建立的基础。

随机相关性与长周期相关性为正值，是因为其与mod10耦合造成的。而各周期的mod10的顺序相关性都一定程度强于随机结果，因而进一步验证了mod10的价格离散值主要周期的统治性。

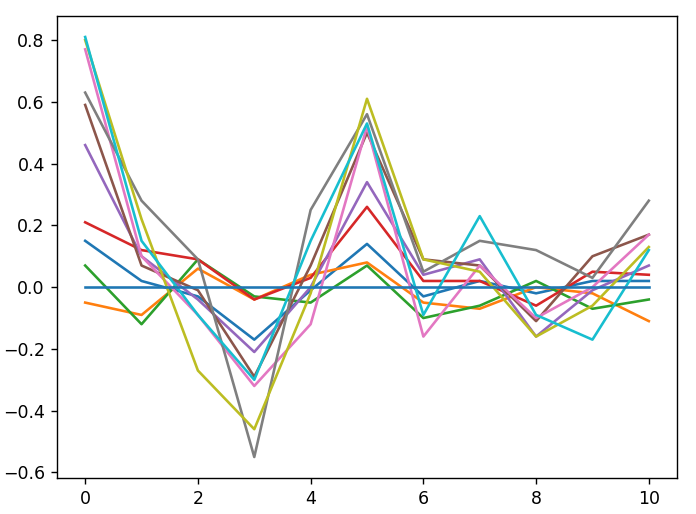
***研究4：***

对于rb的初步研究，选取同期rb\_close\_1min数据，做：

等同于研究一的结果：

[0.15, -0.05, 0.07, 0.21, 0.46, 0.59, 0.77, 0.63, 0.8, 0.81]

等同于研究二的结果：



初步说明对于rb，mod5比mod10更为核心。

**结论：**

1，进一步验证了mod10时普遍的价格聚集周期。**对于mod10临近的周期的确有更相关的趋势。**

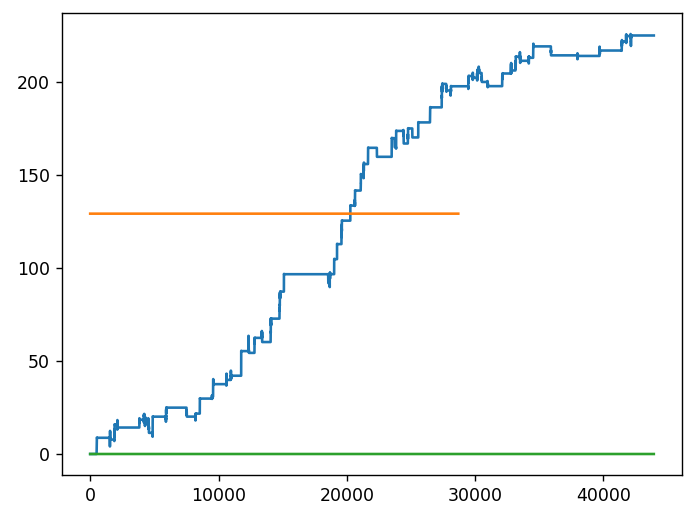
2，在任何相邻时间周期上，mod10都是正相关的，但是强弱不同；整体上时间周期越长，平均的相关性越大。

3，发现mod=11时的特殊现象，猜想：**与mod10临近的价格周期的形成与行情的倾斜特征联系很大**，并据此可以作为斜向网格建立的基础。

**1.2，傅里叶分析**

由于上述研究，存在价格区间的周期性，和倾斜周期性，因此进行傅里叶分析，数据源同上。

利用对滚动价格图形的二维傅里叶变换建立特征，导入随机森林回测结果如下。



**1.3，日线思路**

思路：在日线中寻找规律，在生成标签化特征代入到分钟线内放大利润。

记录以下日线特征：

**1.4，用PCA方法升级二维傅里叶变换至二维卷积核**

***原因分析：***

1，2D傅里叶变换

优点：可以从各种角度抓住周期，方便价差形成。

缺点：不一定反应数据内在规律，以及影响涨跌的主要特征

2，PCA核方法

优点：抓住数据主要规律

缺点：无法与数据标签建立联系

3，LDA核方法

优点：与数据标签建立联系

缺点：标签离散化，而非相关值。

***注：****有PCA或LDA的回归方法，然而只是将PCA的结果进行回归的操作。*

**推荐：先2D傅里叶变换（或者小波变换），再做LDA，最后随机森林。**