時間序列分析

第二章：時間序列的預處理

主講老師：江愷瑤

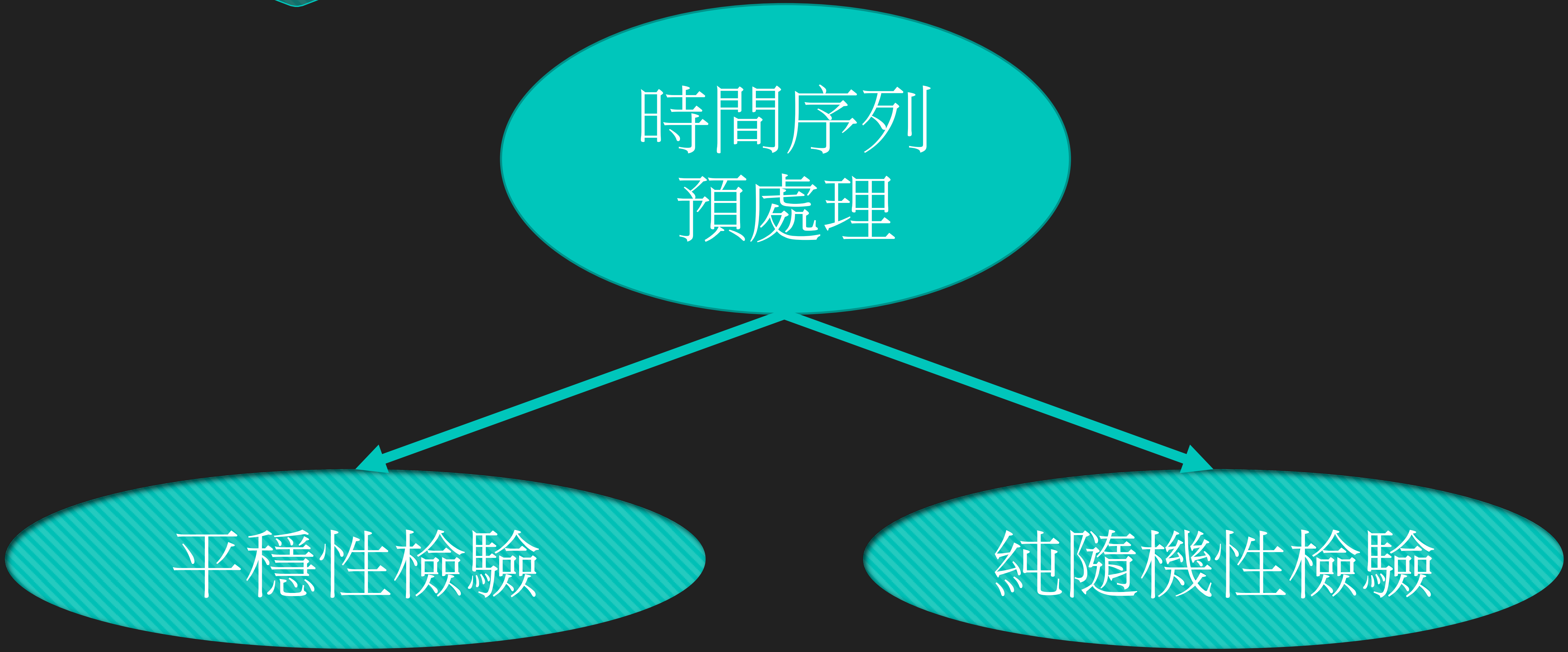
第二章：時間序列的預處理

第2講：純隨機性檢驗

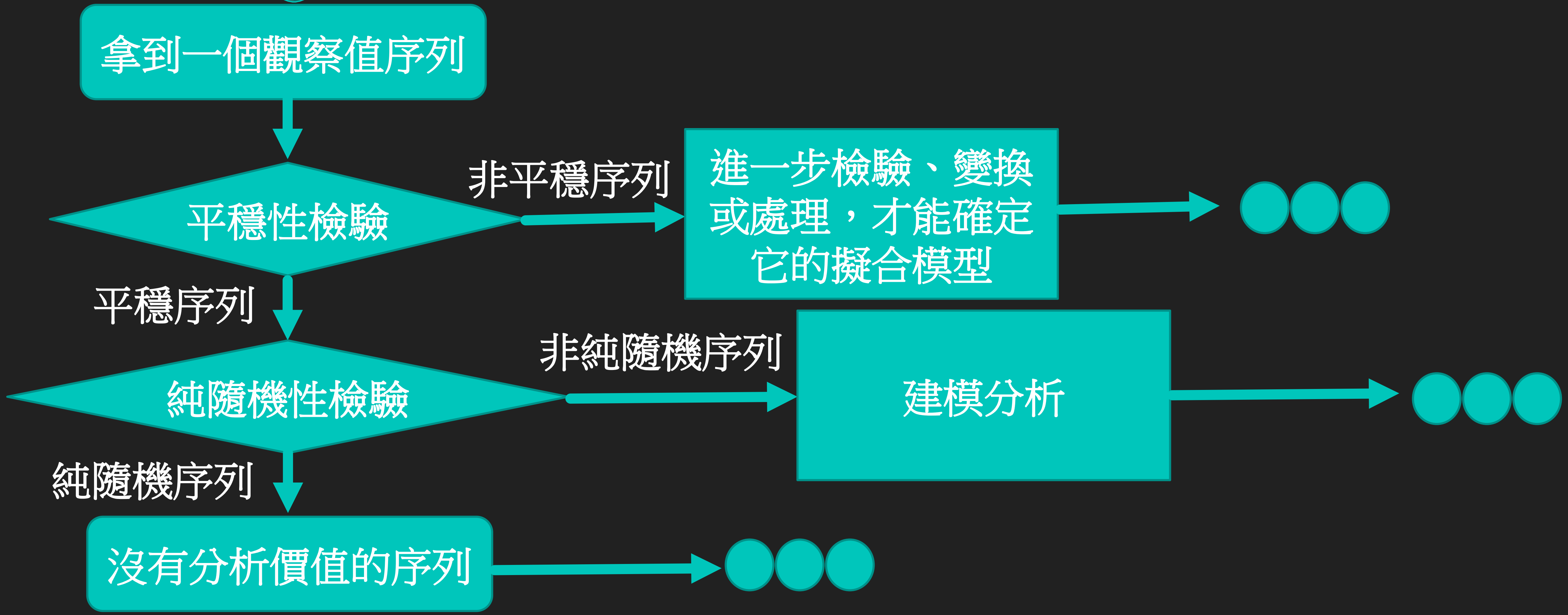
目錄

- 純隨機序列定義
- 白噪聲序列性質
- 純隨機性檢驗

時間序列的預處理



時間序列的預處理


拿到一個觀察值序列
平穩性檢驗
非平穩序列
進一步檢驗、變換或處理，才能確定它的擬合模型
平穩序列
純隨機性檢驗
非純隨機序列
建模分析
純隨機序列
沒有分析價值的序列

純隨機性檢驗意義

- 只有那些序列值之間具有密切的相關關系，歷史數據對未來發展有一定影響的序列，才值得去建模挖掘歷史信息中的有效信息，用來預測未來發展。

- 如果序列值之間沒有任何相關性，即序列是一個沒有記憶的序列，過去行為對將來沒有影響。這種純隨機序列是沒有任何分析價值的。

純隨機序列定義

純隨機序列定義

- 如何時間序列$\{X_t\}$滿足以下性質：

① 任取$t \in T$，有$E(X_t) = \mu$

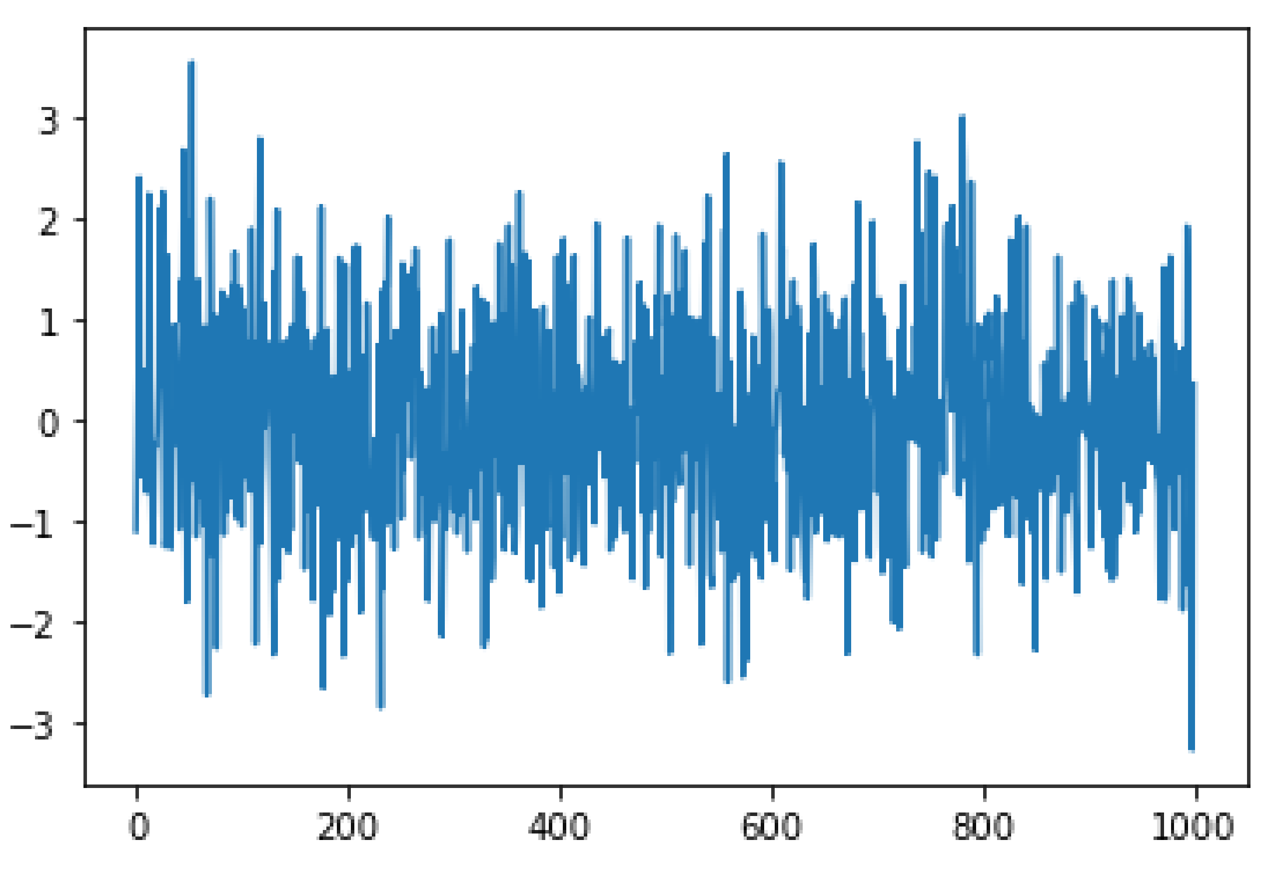
② 任取$t, s \in T$，有

$$\gamma(t, s) = \begin{cases} \sigma^2, & t = s \\ 0, & t \neq s \end{cases}$$

- 則稱序列$\{X_t\}$為純隨機序列，也稱為白噪聲(white noise)序列，記為$X_t \sim WN(\mu, \sigma^2)$。

純隨機序列定義

- 人們最初發現白光具有這種特性，因而也稱為白噪聲序列。
- 白噪聲序列一定是平穩序列，是最簡單的平穩序列。



白噪聲序列的性質

白噪聲序列的性質

- 白噪聲序列雖然簡單，但在時間序列分析中所起的作用非常大。

① 純隨機性

② 方差齊性

白噪聲序列的性質——純隨機性

- 白噪聲序列具有以下性質：

$$\gamma(k) = 0, \forall k \neq 0$$

- 這說明序列各項之間沒有任何相關關系，**序列在進行完全無序的隨機波動**。此序列對應的隨機事件沒有包含任何值得提取的有用信息。

白噪聲序列的性質——純隨機性

- 如果序列之間出現某種顯著相關關系：

$$\gamma(k) \neq 0, \exists k \neq 0$$

- 序列不是純隨機序列，**序列間隔k期的序列值之間存在一定程度的相互影響關系，即相關信息**。接下來的分析目的就是把這種相關信息提取出來。

- 當相關信息被提取出來後，剩下的殘差序列就應該呈現出純隨機性的性質。這可以用來判斷相關信息是否充分提取。

白噪聲序列的性質——方差齊性

- 方差齊性是指序列中每個變量的方差都相等，即

$$D(X_t) = \gamma(0) = \sigma^2$$

- 如果序列不滿足方差齊性，則稱序列具有異方差性質。
- 馬爾可夫定理：只有方差齊性時，用最小二乘法估計得到的未知參數估計值才是准確的、有效的；否則，估計值就不是方差最小線性無偏估計，擬合模型的預測精度會受影響。

- 檢驗擬合模型的殘差是否方差齊性，若不是，則模型還未充分提取相關信息，後期考慮用異方差模型來擬合。（非平穩序列的隨機分析）

純隨機性檢驗

純隨機性檢驗

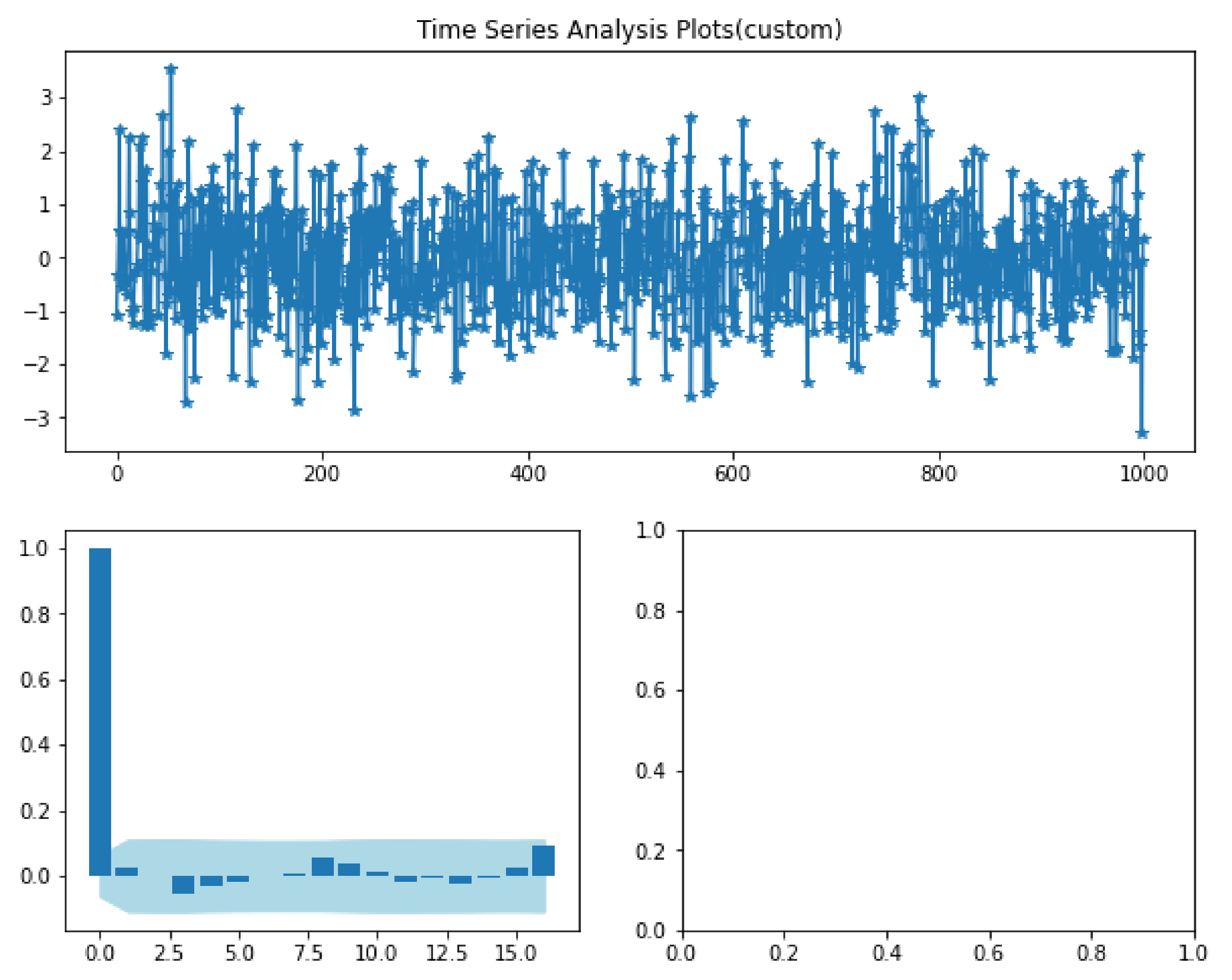
- 純隨機性檢驗也稱爲白噪聲檢驗，是專門用來檢驗序列是否爲純隨機序列的一種方法。
- 純隨機性的序列的序列值之間沒有任何相關關系，即

$$\gamma(k) = 0, \forall k \neq 0$$

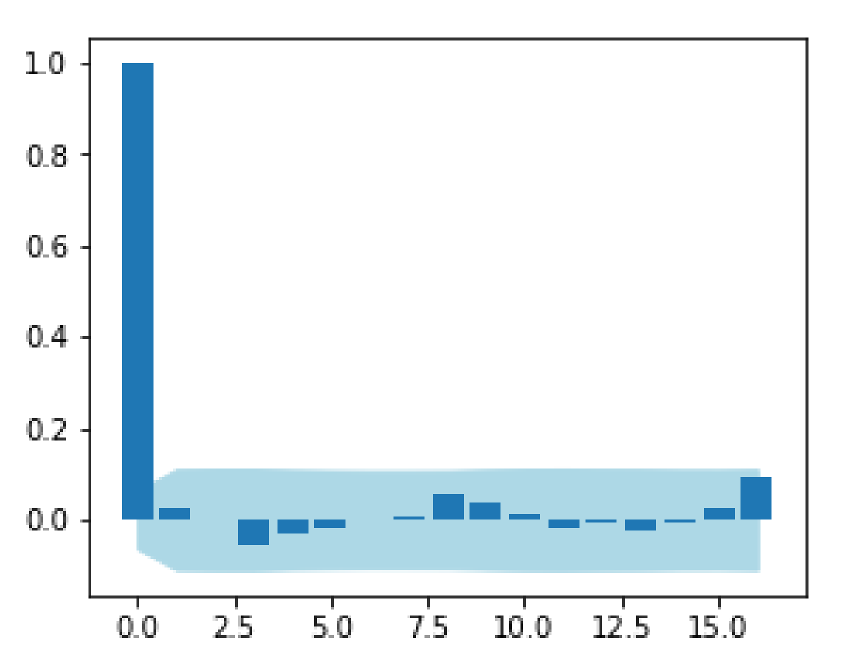
- 實際上，由于觀察值序列的有限性，純隨機序列的樣本自相關係數不會絕對爲0。

純隨機性檢驗

- 右圖是白噪聲序列
  - 時序圖
  - 樣本自相關圖
- 沒有一個樣本自相關系數嚴格等於零。它們都很小，在零附近以一個小幅度做隨機波動。



純隨機性檢驗



- 考慮樣本自相關系數統計性質，從統計學意義上來判斷序列的性質。
- **Barlett定理**
- 如果一時間序列是純隨機的，得到一個觀察期數為n的觀察序列$\{x_t, t = 1,2,\ldots,n\}$，那麼該序列的延遲非零期的樣本**自相關系數將近似服從均值為零、方差為序列觀察期數倒數的正態分布**，即

$$\hat{\rho}_k \dot{\sim} N\left(0, \frac{1}{n}\right), \forall k \neq 0$$

- 其中n為序列觀察期數。

純隨機性檢驗——假設檢驗

- 序列之間差異性是絕對的，而相關性是偶然的，於是假設條件為

$$H_0: \rho_1 = \rho_2 = \cdots = \rho_m = 0, \forall m \geq 1$$

H_1: 至少存在某個$\rho_k \neq 0, \forall m \geq 1, k \leq m$

原假設：延遲期數小於或等於m的序列值之間相互獨立

備擇假設：延遲期數小於或等於m的序列值之間有相關性

純隨機性檢驗——檢驗統計量

- **Q統計量（Box and Pierce）**

$$Q = n\sum_{k=1}^{m}\hat{\rho}_k^2 \dot{\sim} \chi^2(m)$$

n為序列觀察期數
m為指定延遲期數

- **Q統計量近似服從自由度為m的卡方分布**

- 當Q統計量大於$\chi^2_{1-\alpha}(m)$分位點，或統計量的P值小於α時，則可以以$1-\alpha$的置信水平拒絕原假設，認為序列不是白噪聲序列
- 否則，接受原假設，認為序列為純隨機序列。

純隨機性檢驗——檢驗統計量

- **LB統計量（Box and Ljung）**彌補Q統計量在小樣本場合不太精確

$$LB = n(n+2)\sum_{k=1}^{m}\left(\frac{\hat{\rho}_k^2}{n-k}\right) \dot{\sim} \chi^2(m)$$

n為序列觀察期數
m為指定延遲期數

- LB統計量是Q統計量的修正，習慣把它們統稱為Q統計量，分別記為Q_{BP}和Q_{LB}統計量。在各種檢驗場合普遍采用的Q統計量是指LB統計量。

純隨機性檢驗——卡方分布

- 卡方分布表
- https://www.medcalc.org/manual/chi-square-table.php

	P										
DF	**0.995**	**0.975**	**0.20**	**0.10**	**0.05**	**0.025**	**0.02**	**0.01**	**0.005**	**0.002**	**0.001**
1	0.0000393	0.000982	1.642	2.706	3.841	5.024	5.412	6.635	7.879	9.550	10.828
2	0.0100	0.0506	3.219	4.605	5.991	7.378	7.824	9.210	10.597	12.429	13.816
3	0.0717	0.216	4.642	6.251	7.815	9.348	9.837	11.345	12.838	14.796	16.266
4	0.207	0.484	5.989	7.779	9.488	11.143	11.668	13.277	14.860	16.924	18.467
5	0.412	0.831	7.289	9.236	11.070	12.833	13.388	15.086	16.750	18.907	20.515
6	0.676	1.237	8.558	10.645	12.592	14.449	15.033	16.812	18.548	20.791	22.458
7	0.989	1.690	9.803	12.017	14.067	16.013	16.622	18.475	20.278	22.601	24.322
8	1.344	2.180	11.030	13.362	15.507	17.535	18.168	20.090	21.955	24.352	26.124
9	1.735	2.700	12.242	14.684	16.919	19.023	19.679	21.666	23.589	26.056	27.877
10	2.156	3.247	13.442	15.987	18.307	20.483	21.161	23.209	25.188	27.722	29.588
11	2.603	3.816	14.631	17.275	19.675	21.920	22.618	24.725	26.757	29.354	31.264
12	3.074	4.404	15.812	18.549	21.026	23.337	24.054	26.217	28.300	30.957	32.909
13	3.565	5.009	16.985	19.812	22.362	24.736	25.472	27.688	29.819	32.535	34.528

純隨機性檢驗——例子

- 白噪聲序列的$\hat{\rho}_k$
- 計算LB
- 查卡方分布表

延遲	LB統計量檢驗	
	LB統計量	P值
6	2.38	0.854
12	5.38	0.909

- $\alpha = 0.05$

```python
rho=np.array([1,-0.001,-0.037,-0.006,0.012,-0.025,-0.014,0.009,-0.010,-0.027,-0.0
```

```python
print(rho)
```

```
[ 1.    -0.001 -0.037 -0.006  0.012 -0.025 -0.014  0.009 -0.01   -0.027
 -0.025 -0.014  0.035]
```

```python
LB=np.zeros(13)
n=1000
for i in range(1,12+1):
    for k in range(1,i+1):
        LB[i]=LB[i]+rho[k]**2/(n-k)
LB=LB*n*(n+2)

lag=np.arange(13)
myd=pd.DataFrame(np.c_[lag,LB],columns=['lags','LB'])

print(myd)
```

```
    lags        LB
0    0.0  0.000000
1    1.0  0.001003
2    2.0  1.375490
3    3.0  1.411671
4    4.0  1.556538
5    5.0  2.185935
6    6.0  2.383512
7    7.0  2.465247
8    8.0  2.566255
9    9.0  3.303346
10  10.0  3.935922
11  11.0  4.134499
12  12.0  5.376857
```

whitenoise_LBtest.ipynb

- P值顯著大於顯著性水平α，不能拒絕純隨機的原假設。

純隨機性檢驗——例子

- statsmodels中的lb_test

```
In [18]: def purerandtest(y):
             a,b=lb_test(y,lags=None,boxpierce=False)
             LB_purerand=pd.DataFrame(np.c_[a,b],columns=['LB','Pvalue'])
             LB_purerand['lags']=range(1,len(a)+1)
             print('----time series: LB pure randomness test----')
             print(LB_purerand)
```

```
In [19]: purerandtest(myl)
```

```
----time series: LB pure randomness test----
           LB    Pvalue  lags
0    0.789049  0.374388     1
1    2.689701  0.260579     2
2    3.854109  0.277655     3
3    4.473036  0.345757     4
4    4.562313  0.471590     5
5    4.670572  0.586704     6
6    6.211752  0.515253     7
7    6.216927  0.622947     8
8    7.929293  0.541287     9
9   12.707674  0.240478    10
10  16.923687  0.110153    11
11  17.332170  0.137524    12
12  17.368041  0.183021    13
13  18.395957  0.189337    14
```

whitenoise.ipynb

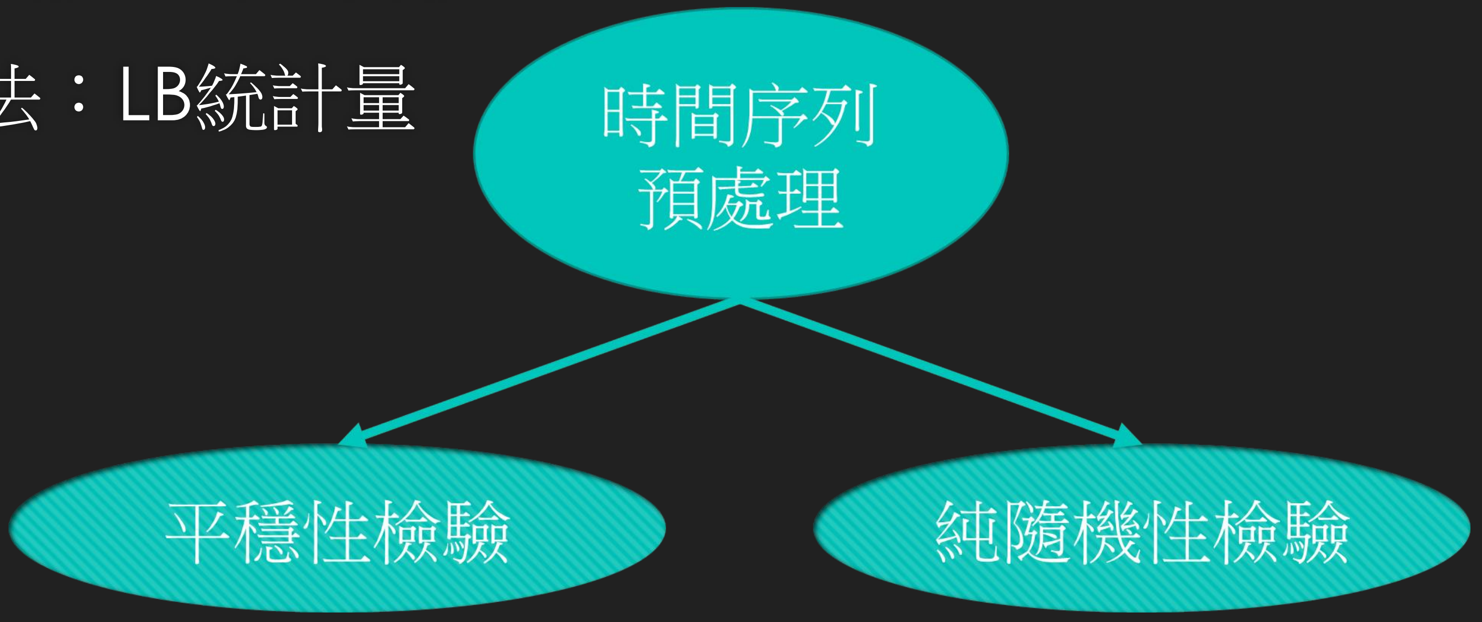
純隨機性檢驗——例子

- **思考：**本題爲何只作了短期延遲的無自相關性檢驗，就能判定爲白噪聲序列？
- 首先本題是平穩序列，其次平穩序列一般具有短期相關性。即若序列有顯著相關性，通常只存在延遲時間較短的序列值間。例子：股票， GDP 等 。
- 一個平穩序列短期延遲的序列值間無顯著相關性，則長期延遲間一般更不存在。
- 一個平穩序列存在短期相關性，則該序列一定不是白噪聲序列，沒必要進行長期延遲檢驗。

總結

本節內容

- 了解純隨機性檢驗意義
- 掌握純隨機序列定義
- 白噪聲序列性質：純隨機性、方差齊性
- 掌握純隨機性檢驗的方法：LB統計量



作業

作業2-2A

作業

- 讀取附錄1.4.csv中的1949-1998年北京市每年最高氣溫序列。
  ① 自己編寫函數計算延遲k相關系數$\hat{\rho}_k$
  ② 自己編寫函數計算LB統計量
  ③ 查表，找出延遲6期、延遲12期的LB統計量的P值
  ④ 利用statsdmodels的lb_test取得延遲1-12期的LB統計量和P值

- 判斷1949-1998年北京市每年最高氣溫序列是否白噪聲序列（$alpha = 0.05$）

作業2-2B

作業

- 對data1.5.csv中的1950-1998年北京市城鄉居民定期儲蓄所占比例序列，進行平穩性與純隨機性進行檢驗。（利用statsmodels的函數）

① 畫出時序圖

② 進行自相關圖檢驗

③ 進行純隨機性檢驗（延遲6期和12期，$alpha = 0.05$）

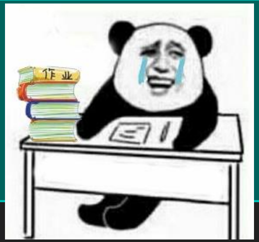
④ 用文字描述你的結果。

作業2-2C

作業

- table2.6.csv是1969年1月到1973年9月在芝加哥海德公園內每28天發生的搶包案件數。
- 判斷該序列$\{x_t\}$的平穩性和純隨機性。
- 對該序列進行運算$y_t = x_t - x_{t-1}$，判斷$\{y_t\}$平穩性和純隨機性。

作業



作業

提交文件：word文件(參考HW2-1.docx)和ipynb文件

截止日期：2021年9月24日23:59

提交地方：tronclass的hw_2.1&2.2

課上作業

課上作業2-2A

- 考慮序列{1,2,3,…,19,20}

1. 判斷序列是否平穩
2. 計算該序列的平均值μ
3. 計算該序列的方差σ^2
4. 計算該序列的樣本自相關係數$\hat{\rho}_k (k = 1,2,\ldots,6)$

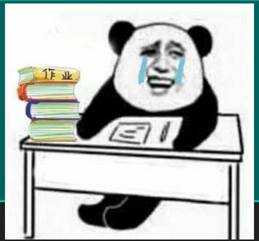
課上作業2-2B

- 若序列長度為100，前12個樣本自相關系數如下：

k	1	2	3	4	5	6
ρ_k	0.02	0.05	0.10	-0.02	0.05	0.01
k	7	8	9	10	11	12
ρ_k	0.12	-0.06	0.08	-0.05	0.02	-0.05

1. 該序列能否視為純隨機序列？

課上作業



作業

小組作業：1-2人一組

提交文件：紙

截止時間：本次課下課前