时序模式

2021/12/13

目录

时序模式

- 就餐饮企业而言，经常会碰到这样的问题：

  由于餐饮行业是生产和销售同时进行的，因此销售预测对于餐饮企业十分必要。如何基于菜品历史销售数据，做好餐饮销售预测？以便减少菜品脱销现象和避免因备料不足而造成的生产延误，从而减少菜品生产等待时间，提供给客户更优质的服务，同时可以减少安全库存量，做到生产准时制，降低物流成本。

- 餐饮销售预测可以看作是基于时间序列的短期数据预测，预测对象为具体菜品销售量。

时序模式

- 常用按时间顺序排列的一组随机变量 $X_1, X_2, \cdots, X_t$ 来表示一个随机事件的时间序列，简记为 $\{X_t\}$ ；用 $x_1, x_2, \cdots, x_n$ 或 $\{x_t, t = 1, 2, \cdots, n\}$ 表示该随机序列的 n 个有序观察值，称之为序列长度为 n 的观察值序列。
- 本节应用时间序列分析的目的就是给定一个已被观测了的时间序列，预测该序列的未来值。

时序模式——时间序列模型

- 常用的时间序列模型见下表：

模型名称	描述
平滑法	平滑法常用于趋势分析和预测，利用修匀技术，削弱短期随机波动对序列的影响，使序列平滑化。根据所用平滑技术的不同，可具体分为移动平均法和指数平滑法。
趋势拟合法	趋势拟合法把时间作为自变量，相应的序列观察值作为因变量，建立回归模型。根据序列的特征，可具体分为线性拟合和曲线拟合。
组合模型	时间序列的变化主要受到长期趋势（ T ）、季节变动（ S ）、周期变动（ C ）和不规则变动（ ε ）这四个因素的影响。根据序列的特点，可以构建加法模型和乘法模型。
组合模型	加法模型： $x_t = T_t + S_t + C_t + \varepsilon_t$
	乘法模型： $x_t = T_t \times S_t \times C_t \times \varepsilon_t$

时序模式——时间序列模型

- 常用的时间序列模型见下表：

模型名称	描述
AR模型	$x_t = \phi_0 + \phi_1 x_{t-1} + \phi_2 x_{t-2} + \cdots + \phi_p x_{t-p} + \varepsilon_t$ 以前 p 期的序列值 $x_{t-1}, x_{t-2}, \cdots, x_{t-p}$ 为自变量、随机变量 X_t 的取值 x_t 为因变量建立线性回归模型。
MA模型	$x_t = \mu + \varepsilon_t - \theta_1 \varepsilon_{t-1} - \theta_2 \varepsilon_{t-2} - \cdots - \theta_q \varepsilon_{t-q}$ 随机变量 X_t 的取值 x_t 与以前各期的序列值无关，建立与前 q 期的随机扰动 $\varepsilon_{t-1}, \varepsilon_{t-2}, \cdots, \varepsilon_{t-q}$ 的线性回归模型。
ARMA模型	$x_t = \phi_0 + \phi_1 x_{t-1} + \phi_2 x_{t-2} + \cdots + \phi_p x_{t-p} + \varepsilon_t - \theta_1 \varepsilon_{t-1} - \theta_2 \varepsilon_{t-2} - \cdots - \theta_q \varepsilon_{t-q}$ 随机变量 X_t 的取值 x_t 不仅与以前 p 期的序列值有关，还与前 q 期的随机扰动有关。

时序模式——时间序列模型

- 常用的时间序列模型见下表：

模型名称	描述
ARIMA模型	许多非平稳序列差分后会显示出平稳序列的性质，称这个非平稳序列为差分平稳序列。对差分平稳序列可以使用ARIMA模型进行拟合。
ARCH模型	ARCH模型能准确地模拟时间序列变量的波动性的变化，适用于序列具有异方差性并且异方差函数短期自相关。
GARCH模型及其衍生模型	GARCH模型称为广义ARCH模型，是ARCH模型的拓展。相比于ARCH模型，GARCH模型及其衍生模型更能反映实际数据中的长期记忆性、信息的非对称性等性质。

- 本节将重点介绍AR模型、MA模型、ARMA模型和ARIMA模型。

时序模式——时间序列的预处理

- 拿到一个观察值序列后，首先要对它的**纯随机性**和**平稳性**进行检验，这两个重要的检验称为序列的预处理。根据检验结果可以将序列分为不同的类型，对不同类型的序列会采取不同的分析方法。
- **对于纯随机序列**，又叫白噪声序列，就意味着序列的各项之间没有任何相关关系，序列在进行完全无序的随机波动，可以终止对该序列的分析。
- **对于平稳非白噪声序列**，它的均值和方差是常数，现已有一套非常成熟的平稳序列的建模方法。通常是建立一个线性模型来拟合该序列的发展，借此提取该序列的有用信息。ARMA模型是最常用的平稳序列拟合模型；
- **对于非平稳序列**，由于它的均值和方差不稳定，处理方法一般是将其转变为平稳序列，这样就可以应用有关平稳时间序列的分析方法，如建立ARMA模型来进行相应得研究。如果一个时间序列经差分运算后具有平稳性，成该序列为差分平稳序列，可以使用ARIMA模型进行分析。

1. 平稳性检验

如果时间序列 $\{X_t, t \in T\}$ 在某一常数附近波动且波动范围有限，即有常数均值和常数方差，并且相距 k 期的序列变量之间的影响程度是一样的，则称 $\{X_t, t \in T\}$ 为平稳序列。

对序列的平稳性的检验有两种检验方法，一种是根据时序图和自相关图的特征做出判断的图检验，该方法操作简单、应用广泛，缺点是带有主观性；另一种是构造检验统计量进行的方法，目前最常用的方法是单位根检验。

时序模式——时间序列的预处理

1. 平稳性检验

- 时序图检验

根据平稳时间序列的均值和方差都为常数的性质，平稳序列的时序图显示该序列值始终在一个常数附近随机波动，而且波动的范围有界；如果有明显的趋势性或者周期性那它通常不是平稳序列。

- 自相关图检验

平稳序列具有短期相关性，这个性质表明对平稳序列而言通常只有近期的序列值对现时值得影响比较明显，间隔越远的过去值对现时值得影响越小。

随着延迟期数 k 的增加，平稳序列的自相关系数 ρ_k （延迟 k 期）会比较快的衰减趋向于零，并在零附近随机波动，而非平稳序列的自相关系数衰减的速度比较慢，这就是利用自相关图进行平稳性检验的标准。

时序模式——时间序列的预处理

1. 平稳性检验

- 单位根检验

最广泛使用的方法是ADF test：原假设是时间序列有单位根并且是非平稳的。如果ADF test的P-value小于显著水平（0.05），就可以拒绝原假设。

时序模式——时间序列的预处理

2. 纯随机性检验

- 如果一个序列式纯随机序列，那么它的序列值之间应该没有任何关系，即满足 $\gamma(k)=0, k\neq 0$ ，这是一种理论上才会出现的理想状态，实际上纯随机序列的样本自相关系数不会绝对为零，但是很接近零，并在零附近随机波动。
- 纯随机性检验也称白噪声检验，一般是构造检验统计量来检验序列的纯随机性，常用的检验统计量有 Q 统计量、LB 统计量，由样本各延迟期数的自相关系数可以计算得到检验统计量，然后计算出对应的 p 值，如果 p 值显著大于显著性水平 α ，则表示该序列不能拒绝纯随机的原假设，可以停止对该序列的分析。

时序模式——平稳时间序列分析

- ARMA模型的全称是自回归移动平均模型，它是目前最常用的拟合平稳序列的模型。
- ARMA模型又可以细分为AR模型、MA模型和ARMA模型三大类。都可以看作是多元线性回归模型。
- 下面将分别介绍AR模型、MA模型和ARMA模型三大模型。

自相关函数（ACF）描述了该序列的当前值与其过去的值之间的相关程度，包括直接和间接的相关性信息。

$$\rho_k = \frac{corr(X_{t+k}, X_t)}{corr(X_t, X_t)}$$

k是间隔的阶数

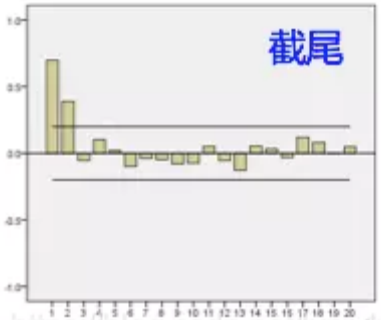
偏自相关函数（PACF） 是部分自相关函数或者偏自相关函数。它不是找到像ACF这样的滞后与当前的相关性，而是找到残差（在去除了之前的滞后已经解释的影响之后仍然存在）与下一个滞后值的相关性。

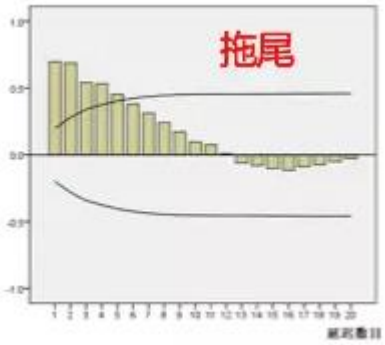
$$X_t = \alpha_0 + \alpha_1 X_{t-1} + \alpha_2 X_{t-2} + \alpha_3 X_{t-3}$$

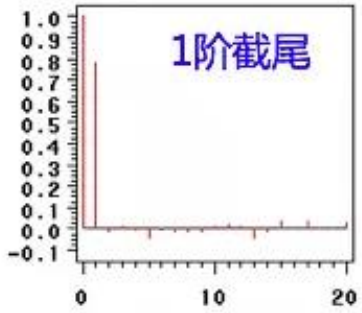
α_3为第三阶偏自相关系数，也就是排除了中间变量X_{t-1}、 X_{t-2}

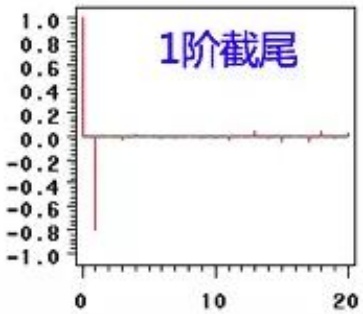
ACF和PACF图的特性

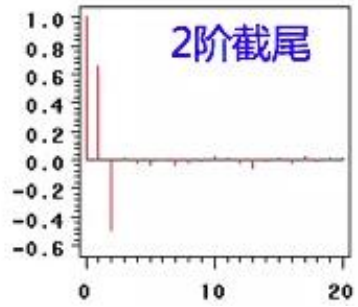
- **截尾**：在大于某个常数k后快速趋于0为k阶截尾
- **拖尾**：始终有非零取值，不会在k大于某个常数后就恒等于零(或在0附近随机波动)

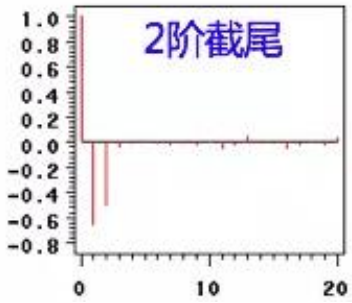


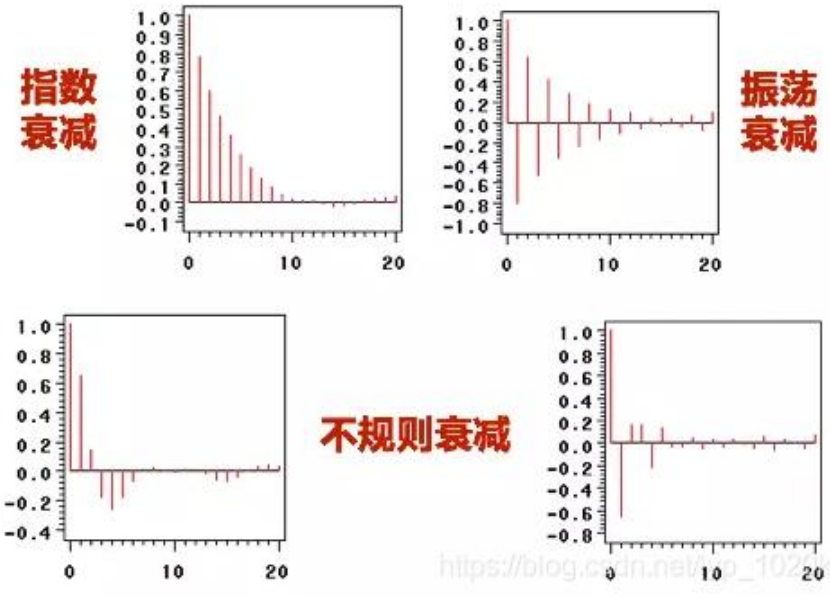












时序模式——平稳时间序列分析

平稳时间序列的ARMA模型建模步骤。

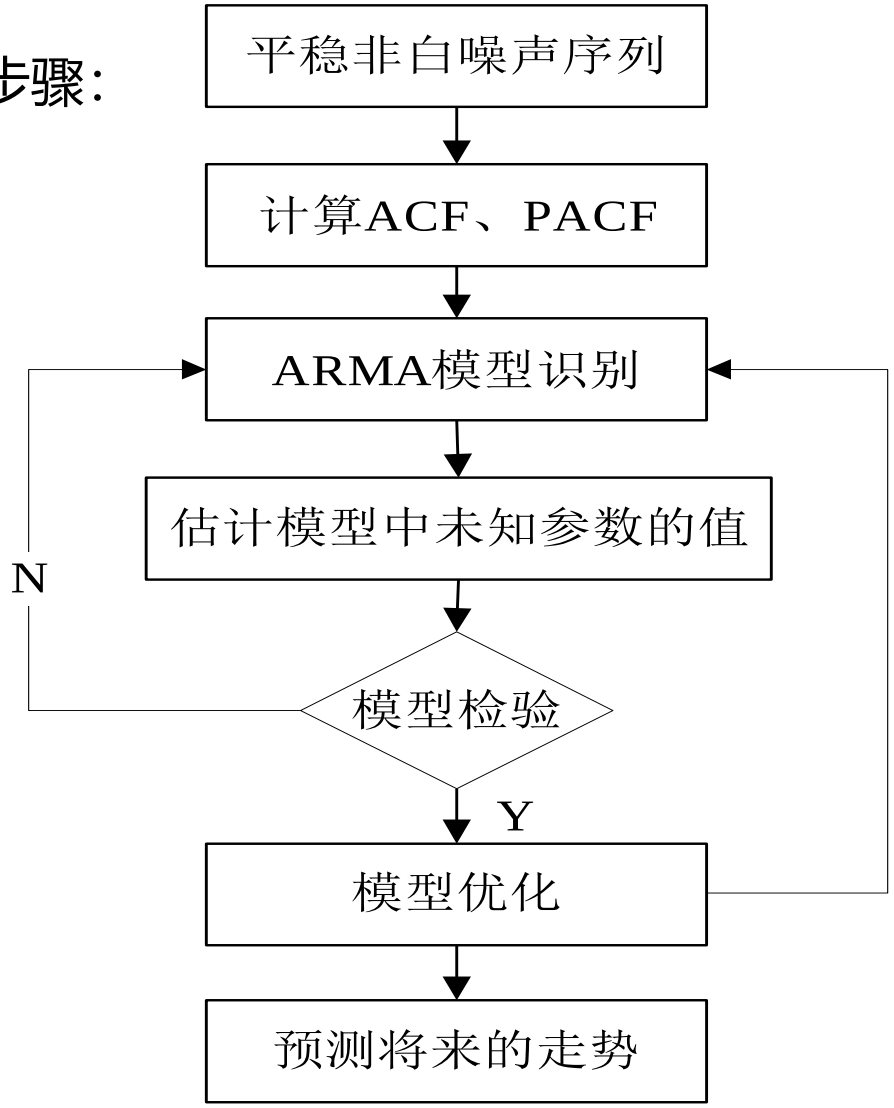
- 某个时间序列经过预处理，被判定为**平稳非白噪声序列**，就可以利用ARMA模型进行建模。
- 由AR模型、MA模型和ARMA模型的**自相关系数和偏自相关系数**的性质，选择出合适的模型。
- AR、MA和ARMA模型自相关系数和偏自相关系数的性质如下：

模型	自相关系数（ACF）	偏自相关系数（PACF）
$AR(p)$	拖尾	阶截尾
$MA(q)$	阶截尾	拖尾
$ARMA(p,q)$	拖尾	拖尾

时序模式——平稳时间序列分析

平稳时间序列的ARMA模型建模步骤。

- 平稳时间序列建模步骤：



时序模式——平稳时间序列分析

平稳时间序列的ARMA模型建模步骤。

- 平稳时间序列建模步骤：

1. 计算ACF和PACF

   先计算非平稳白噪声序列的自相关系数（ACF）和偏自相关系数（PACF）

2. ARMA模型识别

   由AR模型、MA模型和ARMA模型的自相关系数和偏自相关系数的性质，选择出合适的模型。

3. 模型中参数的估计
4. 模型检验
5. 模型优化
6. 模型应用

时序模式——平稳时间序列分析

1. AR模型

具有如下结构的模型称为 p 阶自回归模型，简记为 $AR(p)$ ：

$$x_t = \phi_0 + \phi_1 x_{t-1} + \phi_2 x_{t-2} + \cdots + \phi_p x_{t-p} + \varepsilon_t$$

即在 t 时刻的随机变量 X_t 的取值 x_t 是前 p 期 $x_{t-1}, x_{t-2}, \cdots, x_{t-p}$ 的多元线性回归，认为 x_t 主要是受过去 p 期的序列值的影响。误差项是当期的随机干扰 ε_t ，为零均值白噪声序列。

平稳 $AR(p)$ 模型的性质见下表：

统计量	性质
均值	常数均值
方差	常数方差
自相关系数（ACF）	拖尾
偏自相关系数（PACF）	阶截尾

时序模式——平稳时间序列分析

3. ARMA模型

具有如下结构的模型称为自回归移动平均模型，简记为 $ARMA(p,q)$ ：

$$x_t = \phi_0 + \phi_1 x_{t-1} + \phi_2 x_{t-2} + \cdots + \phi_p x_{t-p} + \varepsilon_t - \theta_1 \varepsilon_{t-1} - \theta_2 \varepsilon_{t-2} - \cdots - \theta_q \varepsilon_{t-q}$$

即在 t 时刻的随机变量 X_t 的取值 x_t 是前 p 期 $x_{t-1}, x_{t-2}, \cdots, x_{t-p}$ 和前 q 期 $\varepsilon_{t-1}, \varepsilon_{t-2}, \cdots, \varepsilon_{t-q}$ 的多元线性函数，误差项是当期的随机干扰 ε_t ，为零均值白噪声序列。认为 x_t 主要是受过去 p 期的序列值和过去 q 期的误差项的共同影响。

平稳 $MA(p,q)$ 的性质见下表：

时序模式——平稳时间序列分析

3. ARMA模型

平稳 $ARMA(p,q)$ 的性质见下表：

统计量	性质
均值	常数均值
方差	常数方差
自相关系数（ACF）	拖尾
偏自相关系数（PACF）	拖尾

时序模式——非平稳时间序列分析

- 前面介绍了对平稳时间序列的分析方法。实际上，在自然界中绝大部分序列都是非平稳的。因而对非平稳序列的分析更普遍、更重要，创造出来的分析方法也更多。
- 对非平稳时间序列的分析方法可以分为确定性因素分解的时序分析和随机时序分析两大类 。
- 确定性因素分解的方法把所有序列的变化都归结为四个因素（长期趋势、季节变动、循环变动和随机波动）的综合影响。可以建立加法模型和乘法模型等。
- 根据时间序列的不同特点，随机时序分析可以建立的模型有ARIMA模型、残差自回归模型、季节模型、异方差模型等。
- 本节重点介绍ARIMA模型对非平稳时间序列进行建模。

时序模式——非平稳时间序列分析

ARIMA模型

- 1阶差分

  相距1期的两个序 列值之间的减法运算称为 1阶差分运算；

- k步差分

  相距 k期的两个序列值之间的减法运算称为 k 步差分运算。

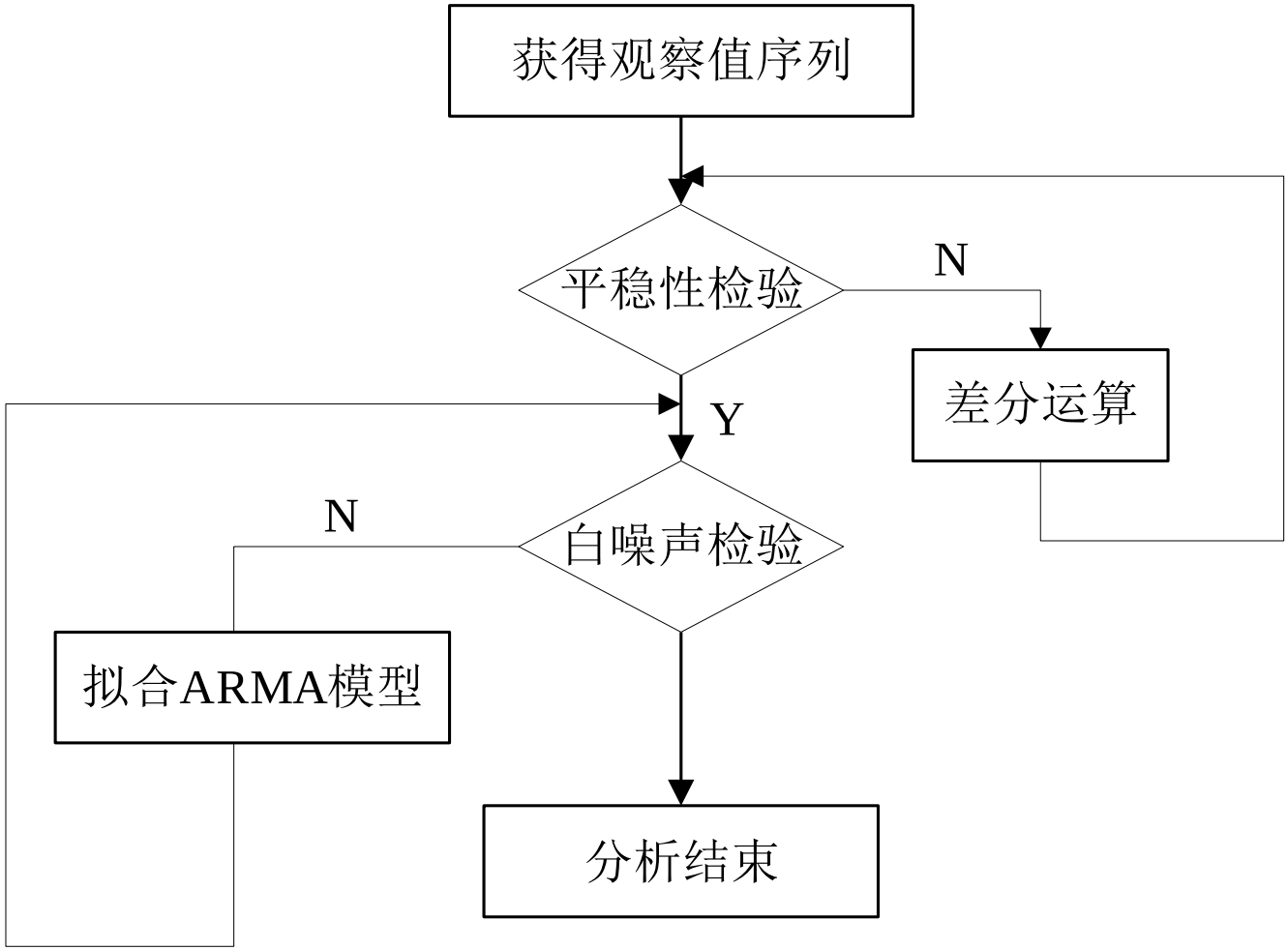
- 差分运算具有强大的确定性信息提取能力，许多非平稳序列差分后会显示出平稳序列的性质，这时称这个非平稳序列为差分平稳序列。

- 对差分平稳序列可以使用ARMA模型进行拟合。

- ARIMA模型的实质就是差分运算与ARMA模型的组合，掌握了ARMA模型的建模方法和步骤以后，对序列建立ARIMA模型是比较简单的。

ARIMA模型

- 差分平稳时间序列的ARIMA模型建模步骤如下：



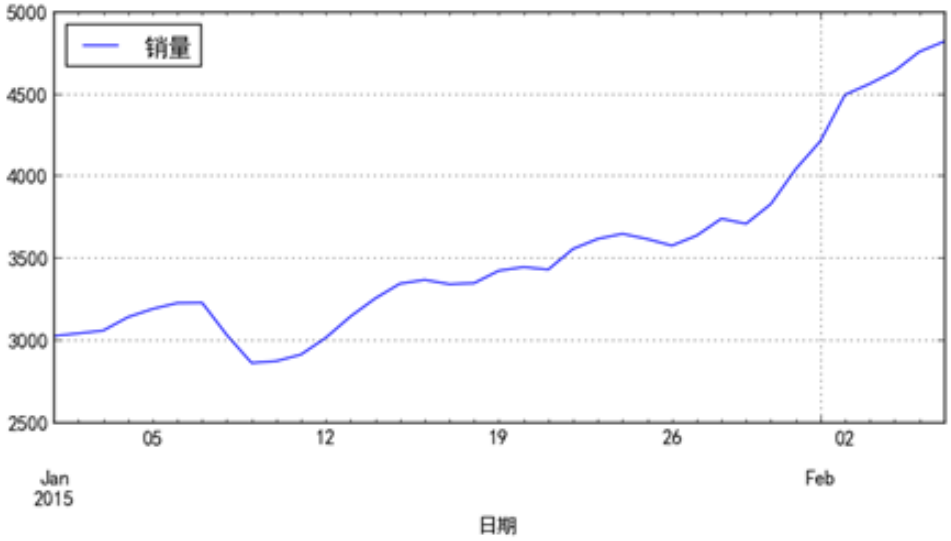
时序模式——案例实现

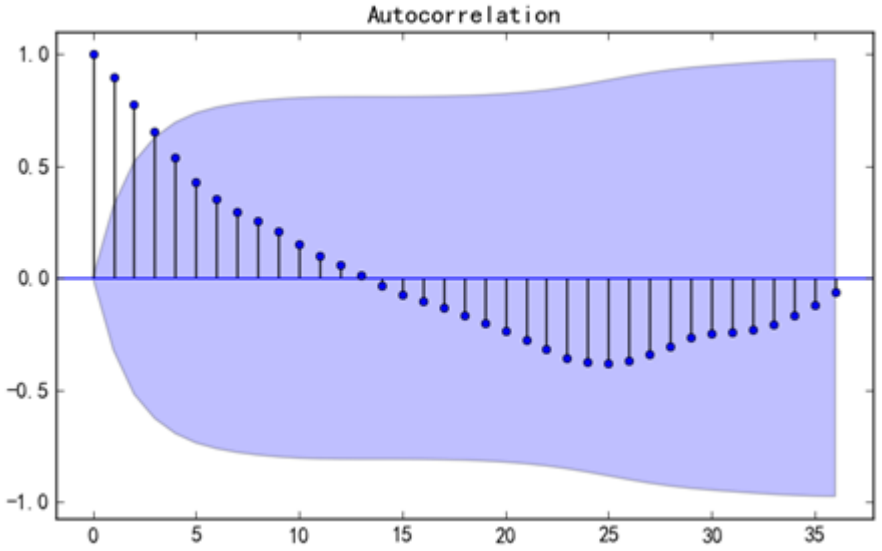
下面应用以上的理论知识，对2015/1/1到2015/2/6某餐厅的销售数据进行建模。部分数据如下表：

日期	销量
2015/1/1	3023
2015/1/2	3039
2015/1/3	3056
2015/1/4	3138
2015/1/5	3188
2015/1/6	3224
2015/1/7	3226
2015/1/8	3029
2015/1/9	2859
2015/1/10	2870
…	…

时序模式——案例实现

1. 序列的平稳性检验--图检验





- 左上的时序图显示该序列具有明显的单调递增趋势，可以判断为是非平稳序列；
- 右上的自相关图显示自相关系数长期大于零，说明序列间具有很强的长期相关性；
- 结论：销售序列是非平稳序列

时序模式——案例实现

1. 序列的平稳性检验--单位根检验

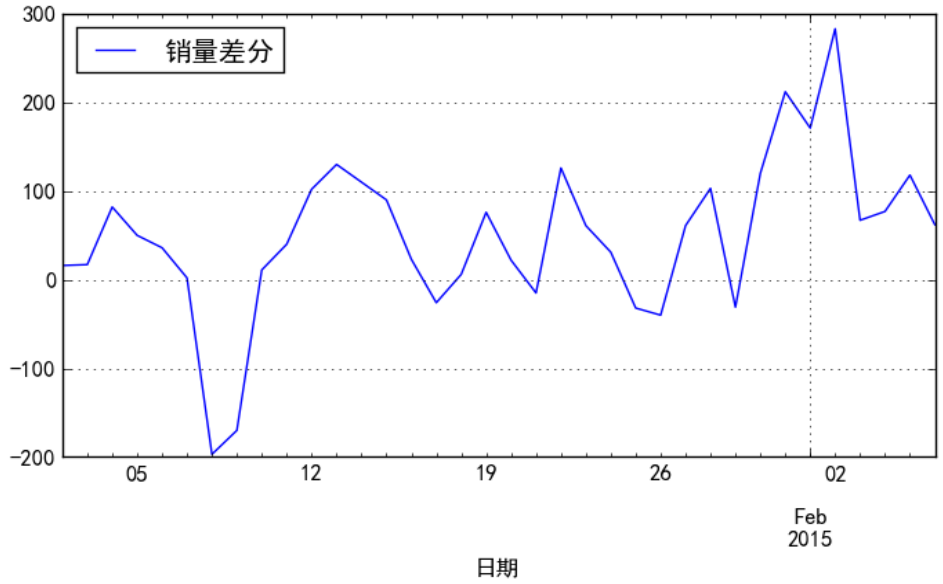
adf	cValue			p值
	1%	5%	10%	
1.8138	-3.7112	-2.9812	-2.6301	0.9984

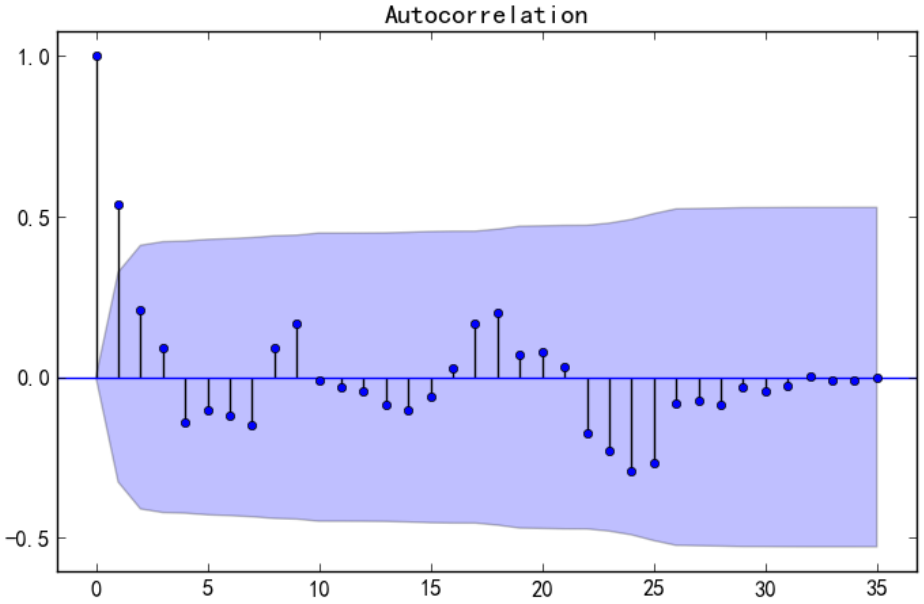
- 单位根检验统计量对应的 p 值显著大于0.05。
- 最终将该序列判断为非平稳序列（非平稳序列一定不是白噪声序列）。

时序模式——案例实现

2. 对原始序列进行一阶差分，并进行平稳性和白噪声检验

2.1 对一阶差分后的序列做平稳性检验





adf	cValue			p值
	1%	5%	10%	
-3.1561	-3.6327	-2.9485	-2.6130	0.0227

时序模式——案例实现

2. 对原始序列进行一阶差分，并进行平稳性和白噪声检验

2.1 对一阶差分后的序列做平稳性检验

- 结果显示，一阶差分之后的序列的时序图在均值附近比较平稳的波动、自相关图有很强的短期相关性、单位根检验p值小于0.05；
- 所以一阶差分之后的序列是平稳序列。

时序模式——案例实现

2. 对原始序列进行一阶差分，并进行平稳性和白噪声检验

2.2 对一阶差分后的序列做白噪声检验

stat	p值
11.304	0.007734

- 输出的p值远小于0.05，所以一阶差分之后的序列是平稳非白噪声序列。

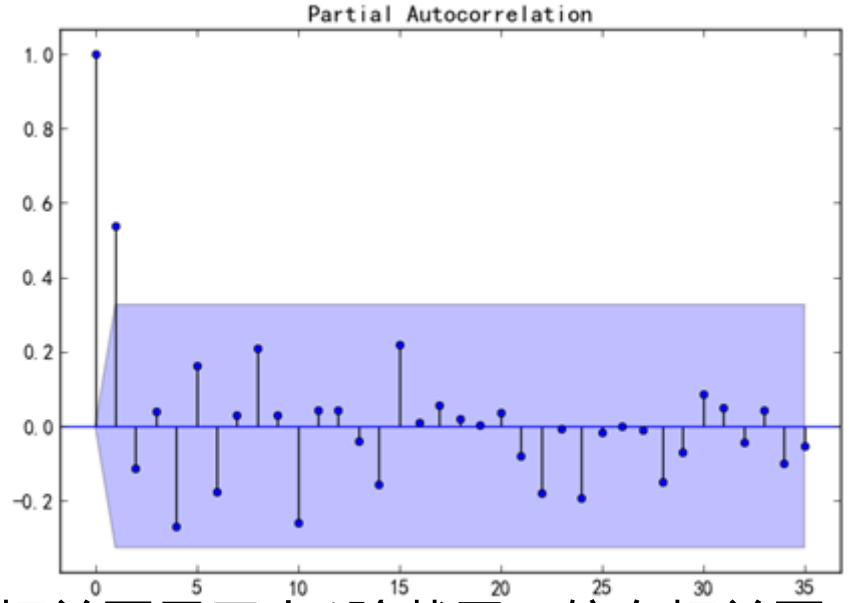
时序模式——案例实现

3. 对一阶差分之后的平稳非白噪声序列拟合ARMA模型

3.1 模型定阶

模型定阶就是确定p和q。

第一种是人为识别的方法：根据ARMA模型识别原则进行模型定阶。



- 一阶差分后自相关图显示出1阶截尾，偏自相关图（如上）显示出拖尾性，所以可以考虑用 MA（1）模型拟合1阶差分后的序列，即对原始序列建立ARIMA（0，1，1）模型。

时序模式——案例实现

3. 对一阶差分之后的平稳非白噪声序列拟合ARMA模型

3.1 模型定阶

第二种方法：相对最优模型识别。

计算ARMA（p，q）当 p 和 q 均小于等于 5 的所有组合的 BIC 信息量，取其中 BIC 信息量达到最小的模型阶数。

- 计算完成BIC矩阵是：

432.068472	422.510082	426.088911	426.595507
423.628276	426.073601	NaN	NaN
426.774824	427.395787	430.709154	NaN
430.317524	NaN	NaN	436.478109

时序模式——案例实现

3. 对一阶差分之后的平稳非白噪声序列拟合ARMA模型

3.1 模型定阶

第二种方法：相对最优模型识别。

p值为1、q值为 0 时最小 BIC 值为:422.510082。

- 可以用AR（1）模型拟合一阶差分后的序列，即对原始序列建立ARIMA（1，1，0）模型。
- 虽然两种方法建立的模型是不一样的，但是可以检验两个模型均通过了检验。
- 实际上对原始序列建立ARIMA（1，1，1）模型也是通过检验的。
- 说明了模型具有非唯一性，进行模型选择优化是有必要的。

时序模式——案例实现

3. 对一阶差分之后的平稳非白噪声序列拟合ARMA模型

3.2 模型检验

p值为：0.627016，大于0.05，残差为白噪声序列，模型通过检验。

3.2 参数估计及检验

Parameter	Coef.	Std. Err.	t
const	49.956	20.139	2.4806
ma.L1.D.销量	0.671	0.1648	4.0712

时序模式——案例实现

4. ARIMA模型预测

应用ARIMA（0，1，1）对2015/1/1到2015/2/6某餐厅的销售数据做为期5天的预测，结果如下：

2015/2/7	2015/2/8	2015/2/9	2015/2/10	2015/2/11
4874.0	4923.9	4973.9	5023.8	5073.8

需要说明的是，利用模型向前预测的时期越长，预测误差将会越来越大，这是时间预测只能进行短期预测的典型特点 。

时序模式——Python主要时序模式算法

Python实现时序模式的主要库是StatsModels（当然，如果Pandas能做的，就可以利用Pandas先做），算法主要是ARIMA模型，在使用该模型进行建模时，需要进行一系列判别操作，主要包含平稳性检验、白噪声检验、是否差分、AIC和BIC指标值、模型定阶，最后再做预测。与其相关的函数如下表所示。

时序模式——Python主要时序模式算法

函数名	函数功能	所属工具箱
acf()	计算自相关系数	statsmodels.tsa.stattools
plot_acf()	画自相关系数图	statsmodels.graphics.tsaplots
pacf()	计算偏相关系数	statsmodels.tsa.stattools
plot_pacf()	画偏相关系数图	statsmodels.graphics.tsaplots
adfuller()	对观测值序列进行单位根检验	statsmodels.tsa.stattools
diff()	对观测值序列进行差分计算	Pandas对象自带的方法
ARIMA()	创建一个ARIMA时序模型	statsmodels.tsa.arima_model
summary()或summaty2	给出一份ARIMA模型的报告	ARIMA模型对象自带的方法
aic/bic/hqic	计算ARIMA模型的AIC/BIC/HQIC指标值	ARIMA模型对象自带的变量

时序模式——Python主要时序模式算法

函数名	函数功能	所属工具箱
forecast()	应用构建的时序模型进行预测	ARIMA模型对象自带的方法
acorr_ljungbox()	Ljung-Box检验，检验是否为白噪声	statsmodels.stats.diagnostic

1. acf()

功能：计算自相关系数

使用格式：

autocorr = acf(data, unbiased=False, nlags=40, qstat=False, fft=False, alpha=None)

输入参数data为观测值序列（即为时间序列，可以是DataFrame或Series），返回参数autocorr为观测值序列自相关函数。其余为可选参数，如qstat=True时同时返回Q统计量和对应p值。

时序模式——Python主要时序模式算法

2. plot_acf()

功能：画自相关系数图

使用格式：

p = plot_acf(data)

返回一个Matplotlib对象，可以用.show()方法显示图像。

3.pacf() / plot_pacf()

功能：计算偏相关系数/画偏相关系数图

使用格式：使用跟acf() / plot_acf()类似，不再赘述。

时序模式——Python主要时序模式算法

4. adfuller()

功能：计对观测值序列进行单位根检验（ADF test)

使用格式：

h = adffuller(Series, maxlag=None, regression='c', autolag='AIC', store=False, regresults=False)

输入参数Series为一维观测值序列，返回值依次为adf、pvalue、usedlag、nobs、critical values、icbest、regresults、resstore。

5.diff()

功能：对观测值序列进行差分计算

使用格式：

D.diff()

D为Pandas的DataFrame或Series。

时序模式——Python主要时序模式算法

6. arima

功能：设置时序模式的建模参数，创建ARIMA时序模型。

使用格式：

```
arima = ARIMA(data, (p,1,q)).fit()
```

data参数为输入的时间序列，p、q为对应的阶，d为差分次数。

7.summary() /summary2()

功能：生成已有模型的报告

使用格式：

```
arima.summary() / arima.summary2()
```

其中arima为已经建立好的ARIMA模型，返回一份格式化的模型报告，包含模型的系数、标准误差、p值、AIC、BIC等详细指标。

时序模式——Python主要时序模式算法

8. aic/bic/hqic

功能：计算ARIMA模型的AIC、BIC、HQIC指标值

使用格式：

arima.aic/arima.bic/arima.hqic

其中arima为已经建立好的ARIMA模型，返回值是Model时序模型得到的AIC、BIC、HQIC指标值。

9.forecast()

功能：用得到的时序模型进行预测

使用格式：

a,b,c = arima.forecast(num)

输入参数num为要预测的天数，arima为已经建立好的ARIMA模型。a为返回的预测值，b为预测的误差，c为预测置信区间。

时序模式——Python主要时序模式算法

10. acorr_ljungbox()

功能：检测是否为白噪声序列

使用格式：

acorr_ljungbox(data, lags=1)

输入参数data为时间序列数据，lags为滞后数，返回统计量和p值。

Thank You!