

ARIMA

- AR、MA、ARMA、ARIMA、SARIMA组成区别

- AR(p):包含 p 个自回归参数
- MA(q):包含 q 个移动平均参数
- ARMA(p,q):参数个数 $n=p+q$
 - 组成: AR+MA
- ARIMA
 - 组成: AR+I+MA (自回归+差分+移动平均)
- SARIMA
 - ARIMA参数条件中多一个S (季节性参数)

- 使用ARMA数据条件: 需要是平稳非白噪声序列

- ARIMA比ARMA多一个差分过程(I)

因为当序列为非平稳时, 需要使用差分将非平稳序列转化为平稳序列

- 那么如何判断数据是否平稳?

- ADF检验 (单位根检验)

最普适, 用的最多的方法; 该讲不含ADF检验统计学原理

- 基本逻辑: ADF检验就是判断序列是否存在单位根: 如果序列平稳, 就不存在单位根; 否则, 就会存在单位根 (基于假设检验)

- H_0 假设: 存在单位根

如果得到的显著性检验统计量小于三个置信度 (10%, 5%, 1%), 则对应有 (90%, 95, 99%) 的把握来拒绝原假设

- 差分 (I) : 转化非平稳序列为平稳序列 (消除趋势性和季节性)

参考: [时间序列模型\(四\): ARIMA模型 - 知乎](#)

- 阶数 (order) 与滞后 (lag)

- 阶数: 进行多少次差分才能使序列平稳

使用最多不会超过2阶差分 (一次差分就短一位序列)

- 滞后: 时间序列数据点之间的时间差

- 对月度数据的时间序列, y_{t-1} 就代表了 y_t 的一个月的滞后

- 滞后差分 (多步差分)

- 一般的一阶差分是滞后为1的差分 (lag-1differences), 即相邻观测值相减
 - 如果是相隔1个值的两个观测值相减, 即滞后为2
 - 相比于高阶差分, 多步差分使用频率大于高阶差分

- 非季节性差分 (阶数d) : 消除趋势。

阶差分可以剔除趋势性影响

- 季节性差分（步数D和周期S）：消除周期性波动。

步差分可以剔除季节性影响

- 如何决定几阶差分？几步差分？后会使数据变得平稳

- 确定非季节性差分阶数d

- 以ADF检验为指标，对差分后数据进行ADF检验，通过检验即数据变得平稳

- 确定季节性差分步数D

数据存在季节性（如月度、季度），需在SARIMA模型中引入季节性差分

- 月度数据每年重复波动,则: $S=12, D=1$, 即 $y_t' = y_t - y_{t-12}$

s是季节周期

- 具体操作:

- 先进行非季节性差分（阶差分）（确定d），消除趋势。
- 若存在季节性，再进行季节性差分（步差分）（确定D）。

- ARIMA和SARIMA

- ARIMA：数据无季节性，仅需处理趋势和随机性（参数：p,d,q）

d-差分阶数，p阶截尾（pacf）、q阶截尾（acf）

- SARIMA：数据有季节性，需额外季节性参数（P,D,Q,S）

D—差分步数；S—周期

- 利用ACF图和PACF图来判断使用哪个模型

- acf和pacf

- 什么是acf和pacf

- ACF 描述了一个观测值和滞后值之间的自相关，包括直接和间接的相关性

假设你有一个时间序列，比如每天的股票价格。ACF的作用是回答：“今天的价格和昨天、前天、大前天的

价格有多大的总体关联性？”

这个“关联性”包括了所有中间日期的间接影响。

数学原理：

- 定义：ACF(k) = 今天 (y_t) 和k天前 (y_{t-k}) 的相关系数，**不控制中间日期** ($y_{t-1}, y_{t-2}, \dots, y_{t-k+1}$) 的影响。

- 公式：

$$ACF(k) = \frac{Cov(y_t, y_{t-k})}{\sqrt{Var(y_t) \cdot Var(y_{t-k})}}$$

其中，Cov是协方差，Var是方差。

- PACF：残差与下一个滞后值的相关性

通俗理解：

PACF的问题是：“如果排除昨天、前天...到k-1天前的所有影响，今天的价格和k天前的价格还有多少直接关联？”

它像一把刀，剔除中间变量的干扰，只保留最纯粹的直接关系。

数学原理：

• 定义：PACF(k) = 在控制 $y_{t-1}, y_{t-2}, \dots, y_{t-k+1}$ 的条件下， y_t 和 y_{t-k} 的相关系数。

• 计算方法（线性回归视角）：

1. 用 y_t 对 $y_{t-1}, y_{t-2}, \dots, y_{t-k+1}$ 做回归，得到残差 e_t （剔除中间滞后项的影响）。
2. 用 y_{t-k} 对 $y_{t-1}, y_{t-2}, \dots, y_{t-k+1}$ 做回归，得到残差 e_{t-k} 。
3. PACF(k) = e_t 和 e_{t-k} 的相关系数。

残差的相关系数

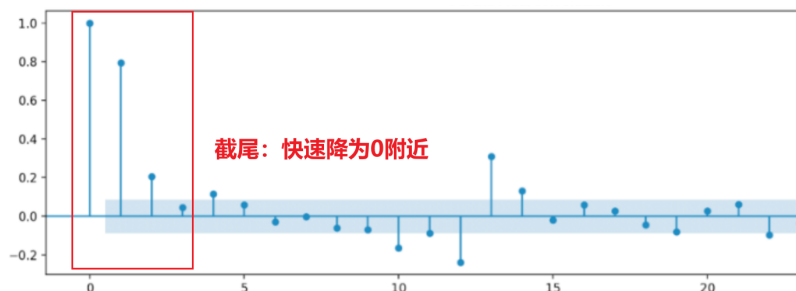
公式简化：

$$\text{PACF}(k) = \text{Corr}(y_t - \hat{y}_t, y_{t-k} - \hat{y}_{t-k})$$

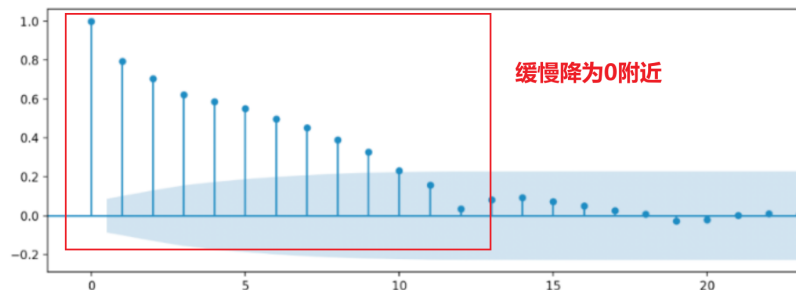
其中， \hat{y}_t 是 y_t 用 $y_{t-1}, \dots, y_{t-k+1}$ 预测的值， \hat{y}_{t-k} 同理。

• 如何看acf和pacf图

• 截尾



• 拖尾



• 拖尾的情况



• 判定：

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

• ARIMA：差分后ACF和PACF拖尾

• 举例：

• 假设时间序列是 y_1, y_2, y_3, y_4

• ACF(2)：计算 y_4 和 y_2 的相关系数，不排除 y_3 的影响。

• PACF(2)：先让 y_4 和 y_2 各自减去 y_3 的预测值得到残差，再计算残差的相关系数。

• 利用ACF图和PACF图识别AR/MA模型的阶数

仅适用于纯AR或纯MA模型，ARMA不适用使用图判断

- AR(p)/MA(q)阶截尾：k阶位置之后，突然降至0，p=k/q=k
突然下降
- 拖尾：k阶位置开始，落入两倍标准差之内，即为k阶拖尾
逐渐衰减过程

- 混合模型的阶数判断

在ARIMA/ARMA中参数个数 $n=p+q$

- AIC/BIC信息准则

参考：[时间序列学习（5）：ARMA模型定阶（AIC、BIC准则、Ljung-Box检验）_ljungbox检验-CSDN博客](#)

- AIC

k-参数个数；L-模型参数（p，q）的似然函数；n-数据个数

$$AIC = 2k - 2 \ln(L)$$

- BIC

k-参数个数；L-模型参数（p，q）的似然函数；n-数据个数

$$BIC = k \ln(n) - 2 \ln(L)$$

- 区别

- 若目标是预测，优先使用AIC；若目标是解释数据生成机制/大样本，优先使用BIC。

三、AIC与BIC的关键区别

特征	AIC	BIC
目标	预测准确性（最小化预测误差）	模型真实性（选择真实模型）
惩罚项	$2k$	$k \ln(n)$ 适用于大样本
样本量影响	不受样本量直接影响	样本量越大，惩罚越重
适用场景	小样本或预测导向	大样本或理论模型选择
模型倾向	可能选择更复杂模型	更倾向简单模型

- ARIMA（p,d,q）

- 确定差分阶数d

- 使用ADF平稳性检验，检验一阶差分/二阶差分是否平稳

- 确定AR(p)和MA(q)

-

- **ACF截尾** → **MA(q)**: 最后一个显著峰值的滞后阶数为 q 。
- **PACF截尾** → **AR(p)**: 最后一个显著峰值的滞后阶数为 p 。

- 定阶：使用AIC/BIC比较不同(p,q)组合，选择AIC/BIC最小的组合作为模型参数

- SARIMA(p,d,q)(P,D,Q,S)

- 确定非季节性参数(p,d,q)

1. 确定非季节性参数 (p, d, q)

- **与非季节ARIMA相同**：通过差分和ACF/PACF判断 d, p, q 。

- 确定季节性参数(P,D,Q,S)

2. 确定季节性参数 (P, D, Q)

- 步骤：

1. 季节性差分 D ：

- 观察周期性波动（如每年重复），进行一阶季节性差分 ($D = 1$)。
- 公式： $y'_t = y_t - y_{t-S}$ （如 $S = 12$ 对应月度数据）。
- 检验平稳性（ADF/KPSS）。

2. 季节性ACF/PACF分析：

- 分析差分后序列在周期滞后 $(S, 2S, \dots)$ 处的ACF和PACF。
- **季节性ACF截尾** → **季节MA(Q)**：最后一个显著季节性峰值的滞后为 Q 。
- **季节性PACF截尾** → **季节AR(P)**：最后一个显著季节性峰值的滞后为 P 。

- d、D、S固定，使用AIC/BIC确定(p,q)(P,Q)组合

- ARIMA代码

含时间序列分解、ADF检验、基于acf/pacf模型选择、p, q最大候选次数的确定、基于AIC/BIC最优参数组合的选择、代码逻辑解析、模型检验、最终预测结果分析