NJUAI 南京大学 人工智能学院
SCHOOL OF ARTIFICIAL INTELLIGENCE, NANJING UNIVERSITY
南京大學 NANJING UNIVERSITY 1902

05 平稳和非平稳时间序列分析

线性过程、ARMA模型及其扩展

Wold分解定理与线性时不变LTI系统

- Wold分解定理：离散**平稳**时间序列$\{X_t\}$分解

$$Y_t = V_t + \xi_t = \sum_{j=1}^{\infty} \phi_j Y_{t-j} + \sum_{j=0}^{\infty} \psi_j \varepsilon_{t-j}$$

$\{V_t\}$为确定性平稳序列，$\{\xi_t\}$ 为随机性平稳序列

- $Y_0 = V_0 + \xi_0 = \psi_0 \varepsilon_0 = \varepsilon_0$
- $Y_1 = V_1 + \xi_1 = \phi_1 Y_0 + \psi_0 \varepsilon_1 + \psi_1 \varepsilon_0$
- …

LTI系统的性质可用于时间序列

- 差分方程描述

$$\sum_{i=0}^{N} a_i Y_{n-i} = \sum_{j=0}^{M} b_j \varepsilon_{n-j}$$

- 卷积描述$Y = \varepsilon * h$

$$Y_t = \sum_{k=-\infty}^{\infty} h_k \varepsilon_{t-k}$$

使用差分方程判断系统的稳定性（给定极小的输入后输出有界）

线性过程

- **线性过程**：写为以下形式的随机过程

$$Y_t - \mu = \sum_{i=0}^{\infty} \psi_i \varepsilon_{t-i}$$

μ是随机过程的均值，一般假设均值$\mu = 0$

- $\{\varepsilon_t\}$为白噪声，可进行缩放使$\psi_0 = 1$
- 序列可用白噪声的现行加权得到，因此白噪声过程也称为触发系统的震荡序列（a *series of shocks* that drive the system）

原始线性过程

线性过程$Y_t = \sum_{i=0}^{\infty} \psi_i \varepsilon_{t-i}$，$\{\varepsilon_t\}$为白噪声，可进行缩放使$\psi_0 = 1$，即

$$\begin{aligned} Y_t &= \psi_0 \varepsilon_t + \psi_1 \varepsilon_{t-1} + \psi_2 \varepsilon_{t-2} + \cdots \\ &= \varepsilon_t + \psi_1 \varepsilon_{t-1} + \psi_2 \varepsilon_{t-2} + \cdots \\ &= \varepsilon_t + \psi_1 \mathbf{B} \varepsilon_t + \psi_2 \mathbf{B}^2 \varepsilon_t + \cdots \\ &= (1 + \psi_1 \mathbf{B} + \psi_2 \mathbf{B}^2 + \cdots) \varepsilon_t = \psi(\mathbf{B}) \varepsilon_t \end{aligned}$$

基于延迟算子表示

$$\psi(\mathbf{B}) = 1 + \sum_{i=1}^{\infty} \psi_i \mathbf{B}^i = \sum_{i=0}^{\infty} \psi_i \mathbf{B}^i$$

假设$\psi_0 = 1$。$\psi(\mathbf{B})$一般称为（关联ε_t和Y_t的）传递函数（**transfer function**）；当把$\mathbf{B}$作为数值（可以取复数数值）带入时也可视为ψ的生成函数（**generating function**）

ψ/π加权

- 一个简单的线性过程

$$Y_t = \varepsilon_t - \theta\varepsilon_{t-1} = (1-\theta\mathbf{B})\varepsilon_t$$

其中$\psi_1 = -\theta, \psi_j = 0\ (j>1)$，因此

$$(1-\theta\mathbf{B})^{-1}Y_t = \varepsilon_t$$

- 若$|\theta|<1$，则

$$(1-\theta\mathbf{B})^{-1}Y_t$$
$$= \left(1+\theta\mathbf{B}+\theta^2\mathbf{B}^2+\theta^3\mathbf{B}^3+\cdots\right)Y_t$$
$$= \varepsilon_t$$

则

$$Y_t + \theta Y_{t-1} + \theta^2 Y_{t-2} + \cdots = \varepsilon_t$$

- 可定义$\pi_j = \theta^j$

ψ加权

记

$$\psi(\mathbf{B}) = 1 + \sum_{i=1}^{\infty}\psi_i\mathbf{B}^i$$

则

$$Y_t = \psi(\mathbf{B})\varepsilon_t$$

π加权

记

$$\pi(\mathbf{B}) = 1 + \sum_{i=1}^{\infty}\pi_i\mathbf{B}^i$$

则

$$\pi(\mathbf{B})Y_t = \varepsilon_t$$

$$\pi(\mathbf{B})\psi(\mathbf{B}) = 1$$

线性过程的性质

- 自协方差函数

$$\gamma(k) = \mathrm{Cov}\left[\sum_{i=0}^{\infty} \psi_i \varepsilon_{t-i}, \sum_{i=0}^{\infty} \psi_i \varepsilon_{t+k-i}\right] = \sigma_\varepsilon^2 \sum_{i=0}^{\infty} \psi_i \psi_{i+k}$$

- 【平稳性（Stationarity）】若当$|z| > 1$时

$$\psi(z) = \sum_{i=0}^{\infty} \psi_i z^{-i} = \psi_0 z^0 + \psi_1 (z^{-1}) + \psi_2 (z^{-1})^2 + \cdots$$

  收敛，则线性过程$Y_t = \psi(\mathbf{B})\varepsilon_t$平稳

- 【可逆性（Invertibility）】若当$|z| > 1$时

$$\pi(z) = \sum_{i=0}^{\infty} \pi_i z^{-i}$$

  收敛，则线性过程$\pi(\mathbf{B})Y_t = \varepsilon_t$可逆

线性过程的平稳/收敛性

- 线性过程

$$Y_t = \sum_{i=0}^{\infty} \psi_i \varepsilon_{t-i}$$

由于$E[|\varepsilon_t|] \le \sigma$，若有

$$\sum_{i=-\infty}^{\infty} |\psi_i| < \infty$$

则

$$E|Y_t| \le \sum_{j=-\infty}^{\infty} \left(|\psi_j| E|\varepsilon_{t-j}|\right)$$
$$\le \left(\sum_{j=-\infty}^{\infty} |\psi_j|\right)\sigma < \infty$$

- 已知

$$\gamma(k) = \mathrm{Cov}\left[\sum_{i=0}^{\infty} \psi_i \varepsilon_{t-i}, \sum_{i=0}^{\infty} \psi_i \varepsilon_{t+k-i}\right]$$
$$= \sigma_\varepsilon^2 \sum_{i=0}^{\infty} \psi_i \psi_{i+k}$$

则方差

$$\sigma_Y^2 = \gamma_{YY}(0) = \sigma_\varepsilon^2 \sum_{i=0}^{\infty} \psi_i^2$$

- 平稳性条件来源于方差的**有界性**，要求$\sum_{i=0}^{\infty} \psi_i^2$有界，相当于要求$\psi_i$绝对可和

$$\sum_{i=0}^{\infty} |\psi_i| < \infty \Leftrightarrow \sum_{i=0}^{\infty} \psi_i z^{-i} < \infty, |z| > 1$$

平稳性条件的几种表示方式

- 线性过程：写为以下形式的随机过程

$$Y_t = \sum_{i=0}^{\infty} \psi_i \varepsilon_{t-i}$$

- $\{\varepsilon_t\}$为白噪声，可进行缩放使$\psi_0 = 1$

方差

$$\sigma_Y^2 = \gamma_{YY}(0) = \sigma_\varepsilon^2 \sum_{i=0}^{\infty} \psi_i^2$$

平稳性和序列ψ_i的性质有关

绝对可和

$$\sum_{i=0}^{\infty} |\psi_i| < \infty$$

平方可和

$$\sum_{i=0}^{\infty} \psi_i^2 < \infty$$

生成函数

类比差分方程判断

$$\sum_{i=0}^{\infty} \psi_i z^{-i} < \infty, |z| > 1$$

线性过程的可逆性

- 针对 $Y_t = \varepsilon_t - \theta\varepsilon_{t-1} = (1-\theta\mathbf{B})\varepsilon_t$ 有

$$\varepsilon_t = (1-\theta\mathbf{B})^{-1}Y_t = \left(1+\theta\mathbf{B}+\theta^2\mathbf{B}^2+\cdots+\theta^{\boldsymbol{k}}\mathbf{B}^{\boldsymbol{k}}\right)\left(1-\theta^{k+1}\mathbf{B}^{\boldsymbol{k+1}}\right)^{-1}Y_t$$

$$Y_t = -\theta Y_{t-1} - \theta^2 Y_{t-2} - \cdots - \theta^k Y_{t-k} + \varepsilon_t - \theta^{k+1}\varepsilon_{t-k-1}$$

- 若$|\theta| < 1$，且令$k \to \infty$，可定义$\pi_j = \theta^j$

$$(1-\theta\mathbf{B})^{-1}Y_t = \left(1+\theta\mathbf{B}+\theta^2\mathbf{B}^2+\theta^{\mathbf{3}}\mathbf{B}^{\mathbf{3}}+\cdots\right)Y_t = \varepsilon_t$$

$$Y_t = -\theta Y_{t-1} - \theta^2 Y_{t-2} - \cdots + \varepsilon_t$$

距离当前时刻越远， Y_{t-j}权重越小

- 若$|\theta| \geq 1$，则当前Y_t依赖于$\{Y_{t-1},\ldots,Y_{t-k}\}$，且$k$越大权重$\theta^k$越大
- 因此，要求$|\theta| < 1$并定义此种情形为“可逆”，这一条件等价于：

$$\sum_{j=0}^{\infty}|\theta|^j \equiv \sum_{j=0}^{\infty}|\pi_j| < \infty$$

特殊的线性过程

- 线性过程：写为以下形式的随机过程

$$Y_t = \sum_{i=0}^{\infty} \psi_i \varepsilon_{t-i}$$

- 设$\boldsymbol{\psi}(\boldsymbol{z}) = \boldsymbol{\theta}(\boldsymbol{z})/\boldsymbol{\phi}(\boldsymbol{z})$, $\theta(z)$, $\phi(z)$为多项式

AR过程

- Y_t表示为

$$Y_t + \phi_1 Y_{t-1} + \cdots + \phi_p Y_{t-p} = \varepsilon_t$$

$\{\varepsilon_t\}$为白噪声

- 从$p+1$开始π权重为0
- 使用延迟算子表示为

$$\left(1 + \phi_1 \mathbf{B} + \cdots + \phi_p \mathbf{B}^p\right) Y_t = \phi(\mathbf{B}) Y_t = \varepsilon_t$$

$$\boldsymbol{\psi}(\boldsymbol{z}) = \boldsymbol{\theta}(\boldsymbol{z})/\boldsymbol{\phi}(\boldsymbol{z})$$

MA过程

- Y_t表示为

$$Y_t = \varepsilon_t + \theta_1 \varepsilon_{t-1} + \cdots + \theta_q \varepsilon_{t-q}$$

$\{\varepsilon_t\}$为白噪声

- 从$q+1$开始ψ权重为0
- 使用延迟算子表示为

$$Y_t = \left(1 + \theta_1 \mathbf{B} + \cdots + \theta_q \boldsymbol{B}^q\right) \varepsilon_t = \theta(\boldsymbol{B}) \varepsilon_t$$

概要

1. AR模型

2. MA模型

3. AR模型思想的应用

4. ARMA模型：
(平稳时间序列模型)

5. ARIMA模型：
(非平稳时间序列模型)

6. 模型预测

AR模型的定义

- p阶**自回归**模型，简记为$AR(p)$

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

注意AR模型的表示方法和系数

  - $\phi_p \neq 0$
  - $E(\varepsilon_t) = 0$，$V(\varepsilon_t) = \sigma_\varepsilon^2$，$E(\varepsilon_t \varepsilon_s) = 0, s \neq t$
  - $EY_s \varepsilon_t = 0, \forall s < t$

- 特别当 $\phi_0 = 0$时，称为中心化$AR(p)$模型

$$\mu = \frac{\phi_0}{1 - \phi_1 - \cdots - \phi_p}$$

- 令 $Y'_t = Y_t - \mu$，称 $\{Y'_t\}$是$\{Y_t\}$的**中心化**序列

- **延迟算子**类似于一个时间指针，当前序列值乘以一个延迟算子，就相当于把当前序列值的时间向过去拨了一个时刻

- 记**B**为延迟算子，有

$$\begin{aligned} Y_{t-1} &= \mathbf{B} Y_t \\ Y_{t-2} &= \mathbf{B}^2 Y_t \\ &\vdots \\ Y_{t-p} &= \mathbf{B}^p Y_t, \forall p \geq 1 \end{aligned}$$

- 所以$AR(p)$模型的**简写**形式如下导出

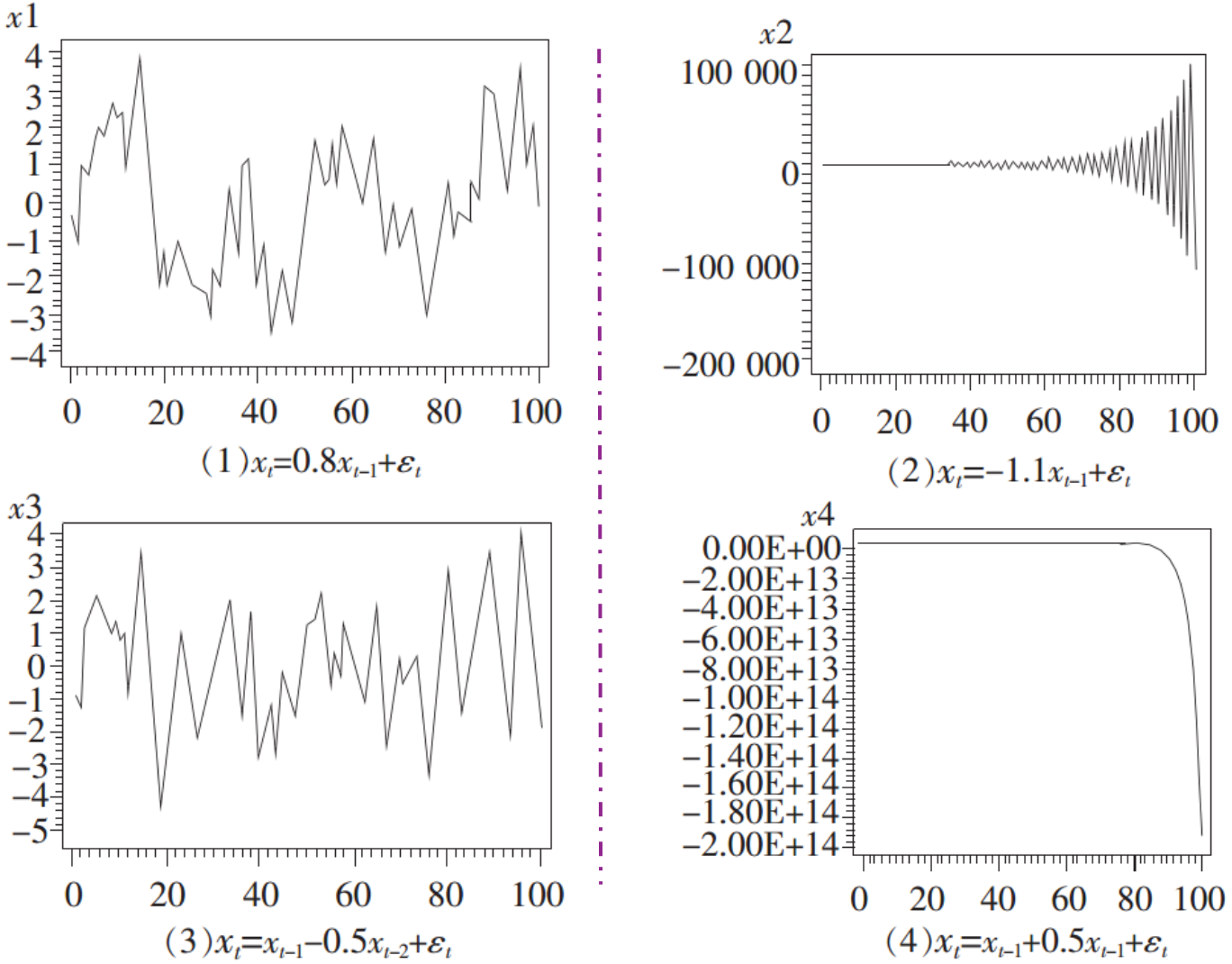
$$\begin{aligned} & Y_t - \phi_1 Y_{t-1} - \cdots - \phi_p Y_{t-p} = \varepsilon_t \\ \Leftrightarrow\ & (1 - \phi_1 \mathbf{B} - \cdots - \phi_p \mathbf{B}^p) Y_t = \varepsilon_t \\ \Leftrightarrow\ & \Phi(\mathbf{B}) Y_t = \varepsilon_t \end{aligned}$$

AR模型平稳性判别

- 要拟合一个平稳序列的发展，用来拟合的模型显然也应该是平稳的。AR模型是常用的平稳序列的拟合模型之一，但并非所有的 AR 模型都是平稳的。

- 判别方法
  - 特征根判别法（基于差分方程）
  - 平稳域判别法

平稳特征



非平稳特征

AR模型平稳的特征根判别

- $AR(p)$模型：

$$Y_t = \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \cdots + \phi_p Y_{t-p} + \varepsilon_t$$

- 转化为

$$Y_t - \phi_1 Y_{t-1} - \phi_2 Y_{t-2} - \cdots - \phi_p Y_{t-p} = \varepsilon_t$$

- p阶自回归序列平稳，要求p个非零特征根都在**单位圆内**，即

$$\lambda^p - \phi_1 \lambda^{p-1} - \cdots - \phi_p = 0$$

$$\Rightarrow |\lambda_i| < 1 \quad , i = 1,2,\cdots,p$$

- $AR(p)$模型：

$$\left(1 - \phi_1 \mathbf{B} - \phi_2 \mathbf{B}^2 - \cdots - \phi_p \mathbf{B}^p\right) Y_t = \varepsilon_t$$

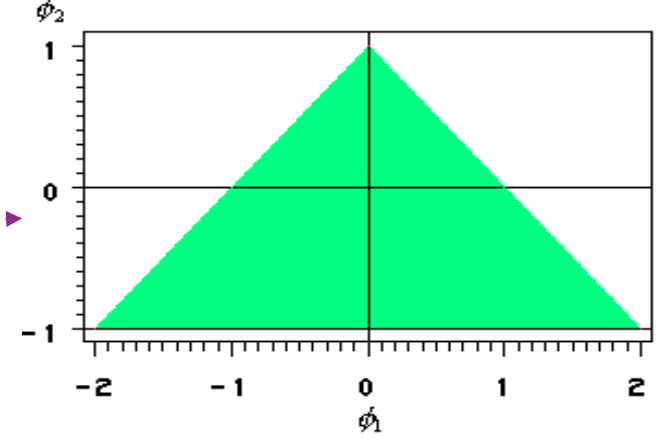
- p阶自回归序列平稳的条件是**自回归系数多项式**的p个根都在**单位圆外**

$$\Phi(z) = 1 - \phi_1 z - \phi_2 z^2 - \cdots - \phi_p z^p = 0$$

$$\Rightarrow z_i = \frac{1}{\lambda_i}, i = 1,2,\cdots,p$$

- 等价于要求$\phi(z^{-1}) = 0$的根在单位圆内

平稳域判别



- $AR(1)$模型

$$Y_t = \phi Y_{t-1} + \varepsilon_t$$

- 特征根

$$\lambda = \phi$$

- 平稳域

$$\{\phi | -1 < \phi < 1\}$$

- $AR(2)$模型

$$Y_t = \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \varepsilon_t$$

- 特征根

$$\lambda_1 = \frac{\phi_1 + \sqrt{\phi_1^2 + 4\phi_2}}{2}, \qquad \lambda_2 = \frac{\phi_1 - \sqrt{\phi_1^2 + 4\phi_2}}{2}$$

- 平稳域

(1) $|\phi_2| = |\lambda_1 \lambda_2| < 1$

(2) $\phi_2 + \phi_1 = -\lambda_1\lambda_2 + \lambda_1 + \lambda_2 = 1 - (1 - \lambda_1)(1 - \lambda_2) < 1$

(3) $\phi_2 - \phi_1 = -\lambda_1\lambda_2 - \lambda_1 - \lambda_2 = 1 - (1 + \lambda_1)(1 + \lambda_2) < 1$

$\Rightarrow \{\phi_1, \phi_2 \big| |\phi_2| < 1,\ \phi_2 \pm \phi_1 < 1\}$

用特征根和平稳域判别法检验 AR 模型的平稳性

- 考察如下四个模型的平稳性

(1) $Y_t = 0.8Y_{t-1} + \varepsilon_t$

(2) $Y_t = -1.1Y_{t-1} + \varepsilon_t$

模型	1	2
特征根判别	$\lambda_1 = 0.8$	$\lambda_1 = -1.1$
平稳域判别	$\phi = 0.8$	$\phi = -1.1$
结论	平稳	非平稳

平稳AR模型的统计性质：均值

- 如果$AR(p)$模型满足**平稳性**条件，则有

$$EY_t = E(\phi_0 + \phi_1 Y_{t-1} + \cdots + \phi_p Y_{t-p} + \varepsilon_t)$$

- 根据平稳序列均值为常数，且$\{\varepsilon_t\}$为白噪声序列，有

$$EY_t = \mu, E(\varepsilon_t) = 0, \forall t \in T$$

- 推导出

$$\mu = \frac{\phi_0}{1 - \phi_1 - \cdots - \phi_p}$$

- 方差如何计算？两边同时取方差？

平稳AR模型的统计性质：方差

- 要得到平稳 $AR(p)$模型的方差, 需要借助Green函数的帮助

- Green函数的定义：假设$\{Y_t\}$为任意p阶的**平稳** AR 模型, 那么一定存在一个**常数**序列

$$\{G_j, j = 1,2,\cdots\}$$

使得$\{Y_t\}$可以等价表达为**纯随机序列**$\{\varepsilon_t|\varepsilon_t \sim N(0,\sigma^2)\}$的线性组合, 即

$$Y_t = \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \cdots + \phi_p Y_{t-p} + \varepsilon_t$$
$$= G_0\varepsilon_t + G_1\varepsilon_{t-1} + G_2\varepsilon_{t-2}\cdots$$

这个常数序列$\{G_j, j = 1,2,\cdots\}$就称为 Green函数

- 基于Green函数，可以求出$AR(p)$模型的方差为

$$V(Y_t) = (G_0^2 + G_1^2 + G_2^2 + \cdots)\sigma_\varepsilon^2$$

线性过程中从ψ加权转换成π加权

Green函数的计算

- 求平稳 $AR(1)$ 模型

$$Y_t = \phi_1 Y_{t-1} + \varepsilon_t$$

的Green函数的递推公式, 并基于Green函数求解 $AR(1)$ 模型的方差。

- 平稳$AR(1)$模型的Green函数递推公式为

$$G_j = \begin{cases} 1, & j = 0 \\ \phi_1^j, & j = 1,2,\cdots \end{cases}$$

- 平稳$AR(1)$模型的方差为

$$V(Y_t) = \sum_{j=0}^{\infty} G_j^2\, V(\varepsilon_t) = \sum_{j=0}^{\infty} {\phi_1}^{2j}\, \sigma_\varepsilon^2 = \frac{\sigma_\varepsilon^2}{1-\phi_1^2}$$

ϕ/G加权

- 基于AR思想的随机过程

$$Y_t - \phi_1 Y_{t-1} - \cdots - \phi_p Y_{t-p} = \varepsilon_t$$

ϕ加权

- 记

$$\phi(\mathbf{B}) = \left(1 + \phi_1 \mathbf{B} + \cdots + \phi_p \mathbf{B}^p\right)$$

则

$$\phi(\mathbf{B}) Y_t = \varepsilon_t$$

$$\phi(\mathbf{B}) G(\mathbf{B}) = 1$$

G加权

- 记

$$G_j = \begin{cases} G_0 = 1 \\ \sum_{k=1}^{j} \phi'_k G_{j-k}, \quad j = 1,2,\cdots \end{cases}$$

则

$$Y_t = G(\mathbf{B}) \varepsilon_t$$

平稳AR模型的统计性质：协方差/自相关函数

- 在平稳$AR(p)$模型两边同乘Y_{t-k}，再求期望

$$E(Y_tY_{t-k}) = \phi_1 E(Y_{t-1}Y_{t-k}) + \cdots + \phi_p E\left(Y_{t-p}Y_{t-k}\right) + E(\varepsilon_t Y_{t-k}), \forall k \geq 1$$

- 根据

$$E(\varepsilon_t Y_{t-k}) = 0$$

- 得**协方差函数**的**递推公式**

$$\gamma_k = \phi_1\gamma_{k-1} + \phi_2\gamma_{k-2} + \cdots + \phi_p\gamma_{k-p}, \forall k \geq 1$$

- 自相关函数的定义

$$\rho_k = \frac{\gamma_k}{\gamma_0}$$

- 平稳$AR(p)$模型的**自相关函数**递推公式

$$\rho_k = \phi_1\rho_{k-1} + \phi_2\rho_{k-2} + \cdots + \phi_p\rho_{k-p}$$

协方差、自相关函数的计算

- 平稳$AR(1)$模型

$$Y_t = \phi_1 Y_{t-1} + \varepsilon_t$$

- **方差**为

$$\gamma_0 = \frac{\sigma_\varepsilon^2}{1-\phi_1^2}$$

- 协方差递推公式

$$\gamma_k = \phi_1 \gamma_{k-1} = \phi_1^k \gamma_0$$

- 协方差函数公式为

$$\gamma_k = \phi_1^k \frac{\sigma_\varepsilon^2}{1-\phi_1^2}, \forall k \geq 1$$

- 自相关函数

$$\rho_k = \phi_1^k, k \geq 0$$

协方差、自相关函数的计算

- 平稳$AR(2)$模型

$$Y_t = \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \varepsilon_t$$

- 平稳$AR(2)$模型的协方差函数递推公式为

$$\begin{cases} \gamma_0 = \dfrac{1-\phi_2}{(1+\phi_2)(1-\phi_1-\phi_2)(1+\phi_1-\phi_2)}\sigma_\varepsilon^2 \\ \gamma_1 = \dfrac{\phi_1\gamma_0}{1-\phi_2} \\ \gamma_k = \phi_1\gamma_{k-1} + \phi_2\gamma_{k-2}, \quad k \geq 2 \end{cases}$$

- 自相关函数

$$\rho_k = \begin{cases} 1, & k = 0 \\ \dfrac{\phi_1}{1-\phi_2}, & k = 1 \\ \phi_1\rho_{k-1} + \phi_2\rho_{k-2}, & k \geq 2 \end{cases}$$

AR模型自相关函数的性质

- AR模型的自相关函数的表达式实际上是一个**齐次差分方程**，它的**通解**形式为

$$\rho_k = \sum_{i=1}^{p} c_i \lambda_i^k \text{，} \ |\lambda_i| < 1, \text{且} c_1, \cdots, c_p \text{为任意实数}$$

- 根据自相关函数的通解形式，可以判断AR模型的自相关函数具有如下特征
  - 呈指数衰减

$$|\lambda_i| < 1 \Rightarrow \rho_k = \sum_{i=1}^{p} c_i \lambda_i^k \to 0$$

  - 拖尾性

$$c_1, \cdots, c_p \text{为任意实数} \Rightarrow \rho_k = \sum_{i=1}^{p} c_i \lambda_i^k \text{不会恒等于零}, \forall k > \text{某个常数}$$

自相关函数的性质

- 考察如下AR模型的自相关图

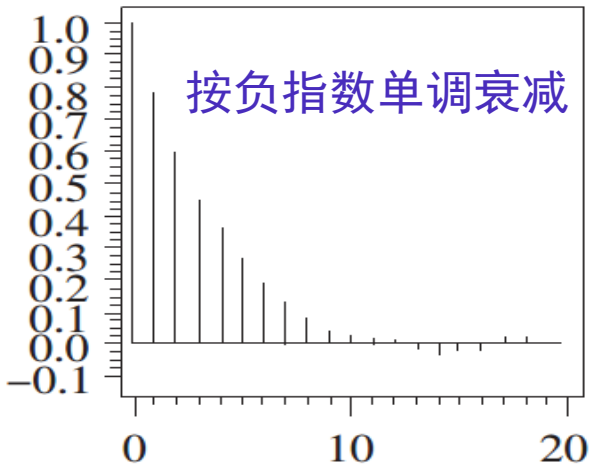
(1) $Y_t = 0.8Y_{t-1} + \varepsilon_t$

(2) $Y_t = -0.8Y_{t-1} + \varepsilon_t$

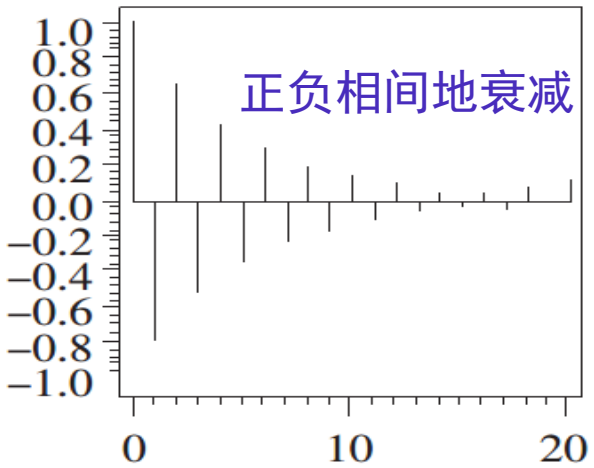
(3) $Y_t = Y_{t-1} - 0.5Y_{t-2} + \varepsilon_t$

(4) $Y_t = -Y_{t-1} - 0.5Y_{t-2} + \varepsilon_t$

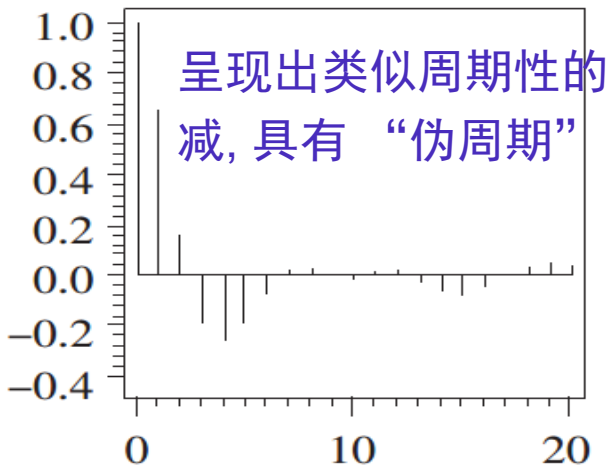
这四个平稳 AR 模型，不论特征根是实根还是复根，是正根还是负根，它们的自相关函数都呈现出**拖尾**性和呈**指数衰减**到零值附近的性质。



(1) $x_t=0.8x_{t-1}+\varepsilon_t$

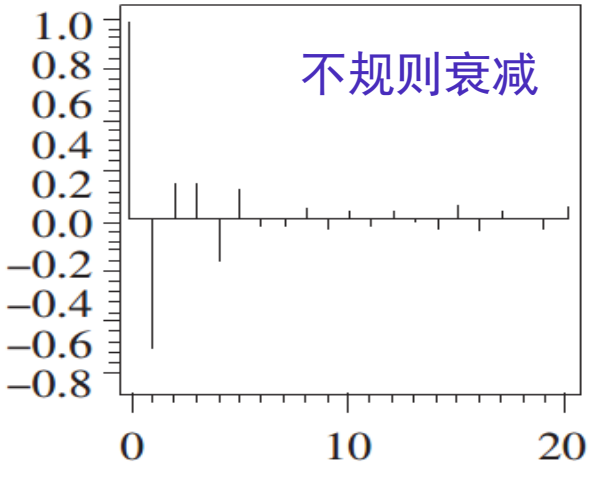


(2) $x_t=-0.8x_{t-1}+\varepsilon_t$



呈现出类似周期性的余弦衰减，具有“伪周期”特征

(3) $x_t=x_{t-1}-0.5x_{t-2}+\varepsilon_t$



(4) $x_t=-x_{t-1}-0.5x_{t-2}+\varepsilon_t$

平稳AR模型的统计性质：偏自相关函数

- 对于平稳$AR(p)$序列，所谓滞后k**偏自相关函数**就是指在**给定**中间$k-1$个随机变量$Y_{t-1}, Y_{t-2}, \cdots, Y_{t-k+1}$的条件下，$Y_{t-k}$对$Y_t$影响的相关度量：

$$\rho_{Y_t, Y_{t-k} | Y_{t-1}, \cdots, Y_{t-k+1}}$$

$$= \frac{E[(Y_t - \hat{E}Y_t)(Y_{t-k} - \hat{E}Y_{t-k})]}{E[(Y_{t-k} - \hat{E}Y_{t-k})^2]}$$

$$\hat{E}Y_t = E[Y_t \mid Y_{t-1}, \cdots, Y_{t-k+1}],$$

$$\hat{E}Y_{t-k} = E[Y_{t-k} \mid Y_{t-1}, \cdots, Y_{t-k+1}]$$

- 用过去k期的序列值对Y_t做k阶**自回归**拟合：

$$Y_t = \phi_{k1}Y_{t-1} + \phi_{k2}Y_{t-2} + \cdots \phi_{k(k-1)}Y_{t-k+1} + \phi_{kk}Y_{t-k} + \varepsilon_t$$

$$\hat{E}Y_t = \phi_{k1}Y_{t-1} + \phi_{k2}Y_{t-2} + \cdots \phi_{k(k-1)}Y_{t-k+1}$$

$$+\phi_{kk}\hat{E}(Y_{t-k}) + E\left(\varepsilon_t \middle| Y_{t-1}, \cdots, Y_{t-k+1}\right)$$

$$= \phi_{k1}Y_{t-1} + \phi_{k2}Y_{t-2} + \cdots \phi_{k(k-1)}Y_{t-k+1} + \phi_{kk}\hat{E}(Y_{t-k})$$

$$\therefore E\left[\left(Y_t - \hat{E}Y_t\right)\left(Y_{t-k} - \hat{E}Y_{t-k}\right)\right]$$

$$= \phi_{kk}E[(Y_{t-k} - \hat{E}Y_{t-k})^2]$$

$$\Rightarrow \rho_{Y_t, Y_{t-k} | Y_{t-1}, \cdots, Y_{t-k+1}}$$

$$= \frac{E[(Y_t - \hat{E}Y_t)(Y_{t-k} - \hat{E}Y_{k-t})]}{E[(Y_{t-k} - \hat{E}Y_{t-k})^2]} = \phi_{kk}$$

基于Yule-Walker方程组计算偏自相关函数

- 在方程

$$Y_t = \phi_{k1}Y_{t-1} + \phi_{k2}Y_{t-2} + \cdots \phi_{k(k-1)}Y_{t-k+1} + \phi_{kk}Y_{t-k}$$

等号两边同时乘以$Y_{t-l}, \forall l > 0$，等号两边求期望再除以方差，得

$$\rho_l = \phi_{k1}\rho_{l-1} + \phi_{k2}\rho_{l-2} + \cdots + \phi_{kk}\rho_{l-k}, \forall l > 0$$

- 取前k个方程构成的方程组即Yule-Walker方程组

$$\begin{cases} \rho_1 = \phi_{k1}\rho_0 + \phi_{k2}\rho_1 + \cdots + \phi_{kk}\rho_{k-1} \\ \rho_2 = \phi_{k1}\rho_1 + \phi_{k2}\rho_0 + \cdots + \phi_{kk}\rho_{k-2} \\ \cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots \\ \rho_k = \phi_{k1}\rho_{k-1} + \phi_{k2}\rho_{k-2} + \cdots + \phi_{kk}\rho_0 \end{cases}$$

- 解Yule-Walker方程组可得参数$(\phi_{k1}, \phi_{k2}, \cdots, \phi_{kk})'$的解，最后一个参数的解即为延迟$k$偏自相关函数

AR(1)模型偏自相关函数的计算

- $AR(1)$模型

$$Y_t = \phi_1 Y_{t-1} + \varepsilon_t$$

- Yule-Walker方程

$$\rho_1 = \phi_{11}\rho_0 \Rightarrow \phi_{11} = \rho_1 = \phi_1$$

- 偏自相关函数的解

$$\phi_{kk} = \begin{cases} \phi_1, & k = 1 \\ 0, & k \geq 2 \end{cases}$$

AR(2)模型偏自相关函数的计算

- $k=1$时，Yule−Walker方程为

$$\phi_{11}=\rho_1$$

对$AR(2)$模型又有

$$\rho_1=\phi_1\rho_0+\phi_2\rho_1$$

$$\Rightarrow \phi_{11}=\rho_1=\frac{\phi_1}{1-\phi_2}$$

- $k=2$时，Yule−Walker方程为

$$\begin{cases}\rho_1=\phi_{21}\rho_0+\phi_{22}\rho_1\\ \rho_2=\phi_{21}\rho_1+\phi_{22}\rho_0\end{cases}$$

对$AR(2)$模型又有

$$\begin{cases}\rho_1=\phi_1+\phi_2\rho_1\\ \rho_2=\phi_1\rho_1+\phi_2\end{cases}$$

$$\Rightarrow \phi_{22}=\phi_2$$

$AR(2)$模型的偏相关系数为$\phi_{kk}=\begin{cases}\frac{\phi_1}{1-\phi_2},k=1\\ \phi_2,k=2\\ 0,k\geq 3\end{cases}$

基于矩阵结构计算偏自相关函数

- Yule-Walker方程写成矩阵形式

$$\begin{pmatrix} 1 & \rho_1 & \cdots & \rho_{k-1} \\ \rho_1 & 1 & \cdots & \rho_{k-2} \\ \vdots & \vdots & \cdots & \vdots \\ \rho_{k-1} & \rho_{k-2} & \cdots & 1 \end{pmatrix} \begin{pmatrix} \phi_{k1} \\ \phi_{k2} \\ \vdots \\ \phi_{kk} \end{pmatrix} = \begin{pmatrix} \rho_1 \\ \rho_2 \\ \vdots \\ \rho_k \end{pmatrix}$$

- 根据Cramer法则

$$\phi_{kk} = \frac{D_k}{D}$$

其中

$$D = \begin{vmatrix} 1 & \rho_1 & \cdots & \rho_{k-1} \\ \rho_1 & 1 & \cdots & \rho_{k-2} \\ \vdots & \vdots & \vdots & \vdots \\ \rho_{k-1} & \rho_{k-2} & \cdots & 1 \end{vmatrix}, D_k = \begin{vmatrix} 1 & \rho_1 & \cdots & \rho_1 \\ \rho_1 & 1 & \cdots & \rho_2 \\ \vdots & \vdots & \vdots & \vdots \\ \rho_{k-1} & \rho_{k-2} & \cdots & \rho_k \end{vmatrix}$$

AR模型偏自相关函数的性质

- AR模型偏自相关函数p阶截尾，是指

$$\phi_{kk} = 0, k > p$$

要证明这一点, 实际上只要能证明当$k > p$时, $D_k = 0$即可。

- 记$\xi_i = \begin{pmatrix} \rho_{i-1} \\ \rho_{i-2} \\ \vdots \\ \rho_{i-k} \end{pmatrix}$，$i = 1,2,\cdots,k$，$\eta = \begin{pmatrix} \rho_1 \\ \rho_2 \\ \vdots \\ \rho_k \end{pmatrix}$，对于$AR(p)$模型有：$\eta = \phi_1\xi_1 + \phi_2\xi_2 + \cdots + \phi_p\xi_p$

$$\Rightarrow D_k = \begin{vmatrix} 1 & \rho_1 & \cdots\rho_{p-1}\cdots\rho_1 \\ \rho_1 & 1 & \cdots\rho_{p-2}\cdots\rho_2 \\ \vdots & \vdots & \vdots \quad \vdots \quad \vdots \quad \vdots \\ \rho_{k-1} & \rho_{k-2} & \cdots\rho_{k-p}\cdots\rho_k \end{vmatrix} = 0，\ k > p \Rightarrow \phi_{kk} = \frac{D_k}{D} = 0，\ k > p$$

偏自相关函数

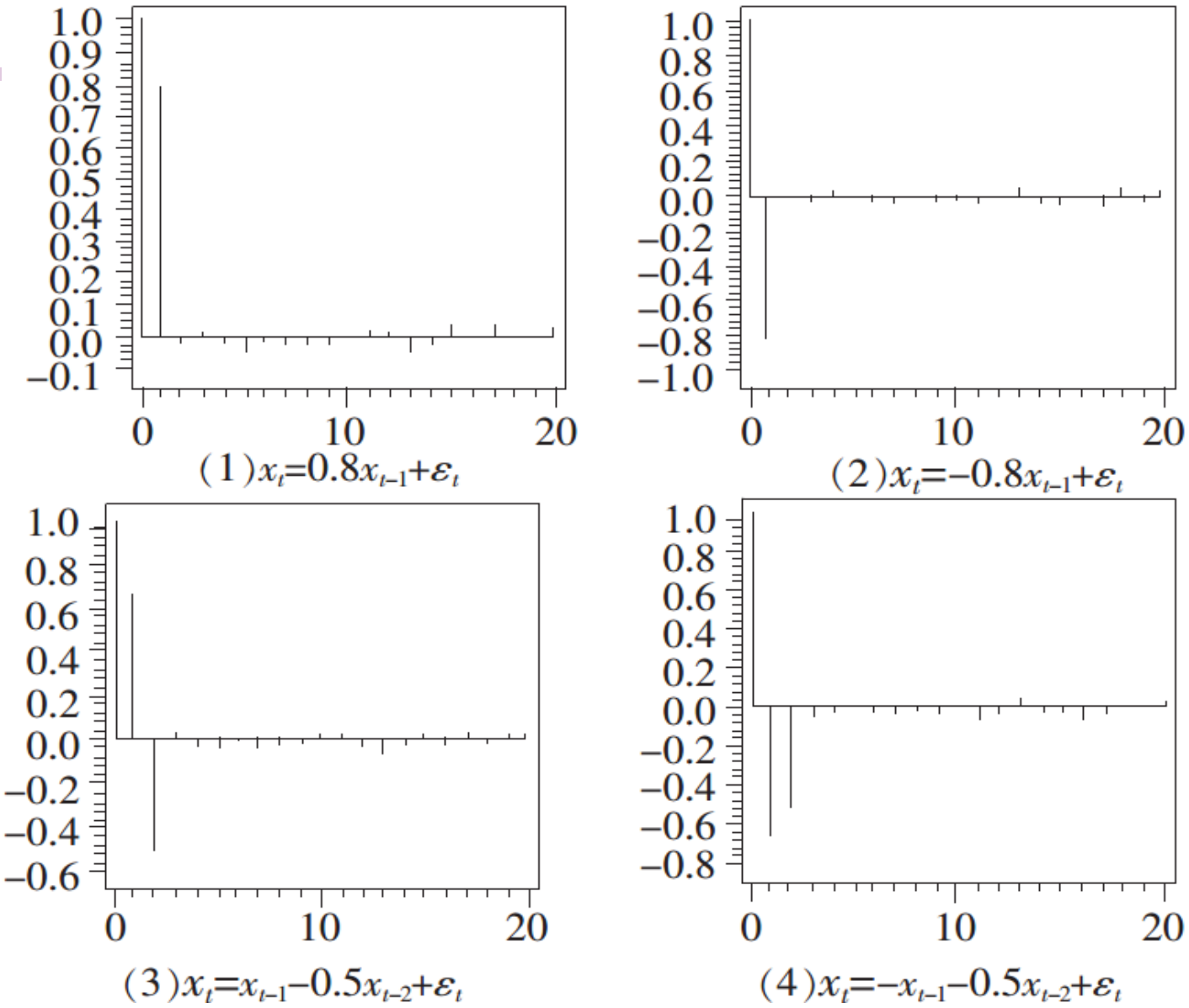
- 求如下AR模型的偏自相关函数

- (1)$Y_t = 0.8Y_{t-1} + \varepsilon_t$
- (2)$Y_t = -0.8Y_{t-1} + \varepsilon_t$
- (3)$Y_t = Y_{t-1} - 0.5Y_{t-2} + \varepsilon_t$
- (4)$Y_t = -Y_{t-1} - 0.5Y_{t-2} + \varepsilon_t$

(1) $x_t = 0.8x_{t-1} + \varepsilon_t$	$\phi_{kk} = \begin{cases} 0.8, & k=1 \\ 0, & k \geqslant 2 \end{cases}$
(2) $x_t = -0.8x_{t-1} + \varepsilon_t$	$\phi_{kk} = \begin{cases} -0.8, & k=1 \\ 0, & k \geqslant 2 \end{cases}$
(3) $x_t = x_{t-1} - 0.5x_{t-2} + \varepsilon_t$	$\phi_{kk} = \begin{cases} \frac{2}{3}, & k=1 \\ -0.5, & k=2 \\ 0, & k \geqslant 3 \end{cases}$
(4) $x_t = -x_{t-1} - 0.5x_{t-2} + \varepsilon_t$	$\phi_{kk} = \begin{cases} -\frac{2}{3}, & k=1 \\ -0.5, & k=2 \\ 0, & k \geqslant 3 \end{cases}$

偏自相关函数

- 求如下AR模型的偏自相关函数
- (1)$Y_t = 0.8Y_{t-1} + \varepsilon_t$
- (2)$Y_t = -0.8Y_{t-1} + \varepsilon_t$
- (3)$Y_t = Y_{t-1} - 0.5Y_{t-2} + \varepsilon_t$
- (4)$Y_t = -Y_{t-1} - 0.5Y_{t-2} + \varepsilon_t$



概要

1. AR模型

2. MA模型

3. AR模型思想的应用

4. ARMA模型：
(平稳时间序列模型)

5. ARIMA模型：
(非平稳时间序列模型)

6. 模型预测

MA模型的定义

- 具有如下结构的模型称为q阶**移动平均（Moving Average）**模型，简记为$MA(q)$

$$\begin{cases} Y_t = \mu + \varepsilon_t - \theta_1\varepsilon_{t-1} - \theta_2\varepsilon_{t-2} - \cdots - \theta_q\varepsilon_{t-q} \\ \theta_q \neq 0 \\ E(\varepsilon_t) = 0,\ Var(\varepsilon_t) = \sigma_\varepsilon^2, E(\varepsilon_t\varepsilon_s) = 0, s \neq t \end{cases}$$

- 特别当$\mu = 0$时，称为中心化$MA(q)$模型

- 基于延迟算子，$MA(q)$模型又可以简记为

$$Y_t = \Theta(\mathbf{B})\varepsilon_t, \qquad \Theta(\mathbf{B}) = 1 - \theta_1\mathbf{B} - \theta_2\mathbf{B}^2 - \cdots - \theta_q\mathbf{B}^q$$

其中$\Theta(\mathbf{B})$称为q阶移动平均系数多项式

MA模型的统计性质

- 常数均值

$$EY_t = E\left(\mu + \varepsilon_t - \theta_1\varepsilon_{t-1} - \theta_2\varepsilon_{t-2} - \cdots - \theta_q\varepsilon_{t-q}\right) = \mu$$

- 常数方差

$$\begin{aligned} Var(Y_t) &= Var(\mu + \varepsilon_t - \theta_1\varepsilon_{t-1} - \theta_2\varepsilon_{t-2} - \cdots - \theta_q\varepsilon_{t-q}) \\ &= (1 + \theta_1^2 + \cdots + \theta_q^2)\sigma_\varepsilon^2 \end{aligned}$$

- 自协方差函数与自相关函数**q阶截尾**

$$\gamma_k = \begin{cases} (1 + \theta_1^2 + \cdots + \theta_q^2)\sigma_\varepsilon^2, & k = 0 \\ (-\theta_k + \sum_{i=1}^{q-k} \theta_i\theta_{k+i})\sigma_\varepsilon^2, & 1 \le k \le q \\ 0, & k > q \end{cases}, \qquad \rho_k = \begin{cases} 1, & k = 0 \\ \dfrac{-\theta_k + \sum_{i=1}^{q-k} \theta_i\theta_{k+i}}{1 + \theta_1^2 + \cdots + \theta_q^2}, & 1 \le k \le q \\ 0, & k > q \end{cases}$$

MA模型的自相关函数

- MA(1)模型

$$Y_t = \varepsilon_t - \theta_1 \varepsilon_{t-1}$$

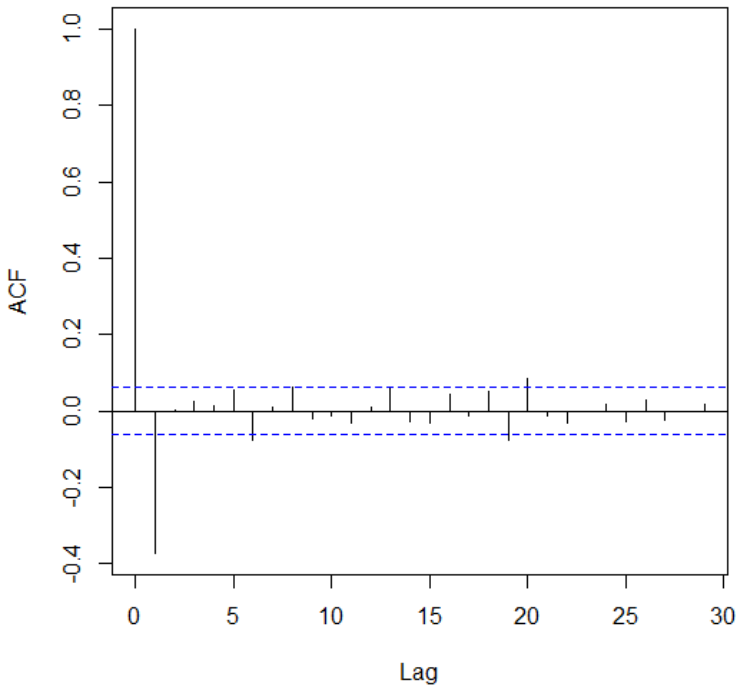
- 自相关函数

$$\rho_k = \begin{cases} 1 & , k = 0 \\ \dfrac{-\theta_1}{1+\theta_1^2} & , k = 1 \\ 0 & , k \geq 2 \end{cases}$$

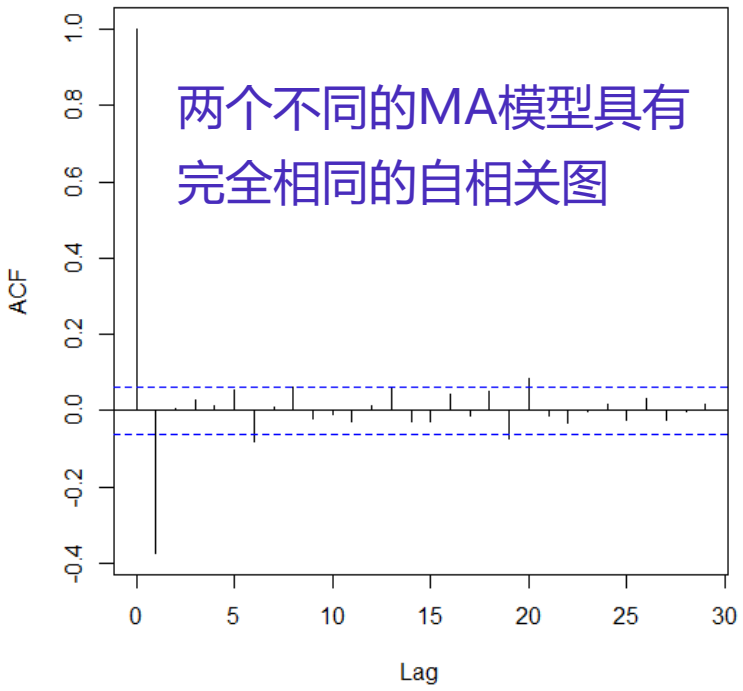
两个不同的MA模型自相关函数也正好相等

$$\rho_k = \begin{cases} -0.4, k = 1 \\ 0, k \geq 2 \end{cases}$$

（1）$Y_t = \varepsilon_t - 2\varepsilon_{t-1}$



（2）$Y_t = \varepsilon_t - 0.5\varepsilon_{t-1}$



两个不同的MA模型具有完全相同的自相关图

样本自相关图显示出$MA(1)$ 模型自相关函数一阶截尾

MA模型的自相关函数

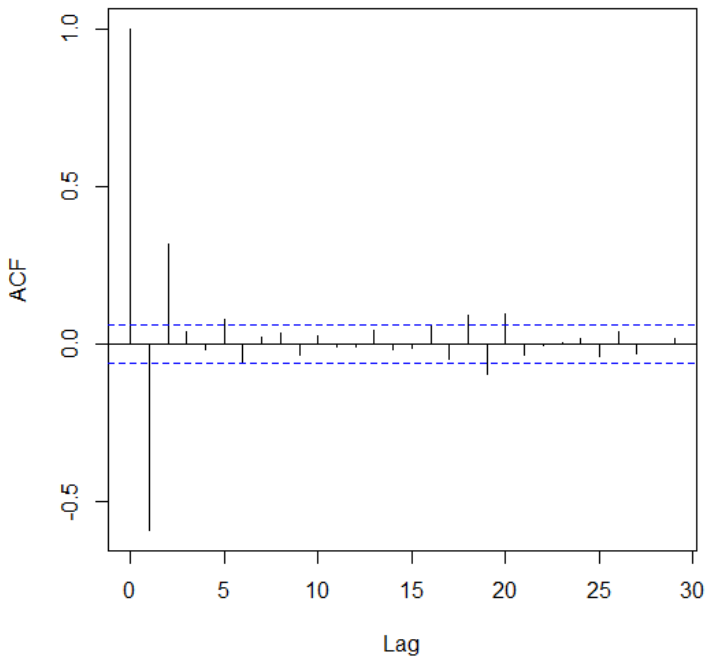
- MA(2)模型

$$Y_t = \varepsilon_t - \theta_1 \varepsilon_{t-1} - \theta_2 \varepsilon_{t-2}$$

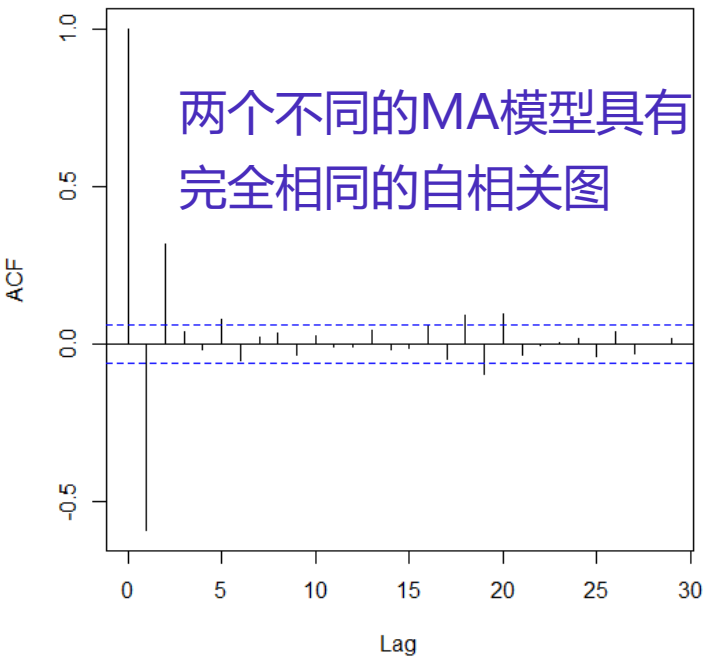
- 自相关函数

$$\rho_k = \begin{cases} 1 & ,k = 0 \\ \dfrac{-\theta_1 + \theta_1\theta_2}{1 + \theta_1^2 + \theta_2^2} & ,k = 1 \\ \dfrac{-\theta_2}{1 + \theta_1^2 + \theta_2^2} & ,k = 2 \\ 0 & ,k \geq 3 \end{cases}$$

（3）$Y_t = \varepsilon_t - \frac{4}{5}\varepsilon_{t-1} + \frac{16}{25}\varepsilon_{t-2}$



（4）$Y_t = \varepsilon_t - \frac{5}{4}\varepsilon_{t-1} + \frac{25}{16}\varepsilon_{t-2}$



两个不同的MA模型具有完全相同的自相关图

样本自相关图清晰显示出 $MA(2)$模型自相关函数二阶截尾

可逆MA模型

模型1：$Y_t = \varepsilon_t - \theta\varepsilon_{t-1}$

模型2：$Y_t = \varepsilon_t - \frac{1}{\theta}\varepsilon_{t-1}$

$$\frac{Y_t}{1-\theta\mathbf{B}} = \varepsilon_t$$

$$\rho = \frac{-\theta}{1+\theta^2}$$

$$\frac{Y_t}{1-\frac{1}{\theta}\mathbf{B}} = \varepsilon_t$$

$|\theta| < 1$, 模型1可逆

$|\theta| > 1$, 模型2可逆

可逆MA模型定义：若一个MA模型能够表示成为收敛的AR模型形式，那么该MA模型称为可逆MA模型

MA模型的可逆条件

- $MA(q)$ 模型的**可逆**概念和 $AR(p)$ 模型的**平稳**概念是**对偶**概念。
- $MA(q)$模型的可逆条件是该模型特征方程

$$\lambda^q - \theta_1\lambda^{q-1} - \theta_2\lambda^{q-2} - \cdots - \theta_q = 0$$

的q个非零特征根都在单位圆内

$$|\lambda_i| < 1, \quad i = 1,2,\cdots,q$$

或移动平滑系数多项式$\Theta(u) = 0$的根都在单位圆外

$$u_i = \left|\frac{1}{\lambda_i}\right| > 1, i = 1,2,\cdots,q$$

低阶MA模型系数可逆域

- 根据MA模型的结构，求出特征方程的特征根，根据特征根都在单位圆内的约束条件，可以求出满足可逆条件的系数取值空间，这就是MA模型的系数**可逆域**。
- MA模型的系数可逆域与AR模型的平稳域具有对偶关系
- $MA(1)$模型的系数可逆域

$$\{\theta | -1 < \theta < 1\}$$

- $MA(2)$模型的系数可逆域

$$\{\theta_1, \theta_2 \big| |\theta_2| < 1，且\theta_2 \pm \theta_1 < 1\}$$

MA模型的偏自相关函数

- 对于一个可逆$MA(q)$模型，可以等价写成$AR(\infty)$模型形式

$$I(\mathbf{B})Y_t = \varepsilon_t$$

$$I_j = \begin{cases} 1, & j \geq 0 \\ \sum_{k=1}^{l} \theta_k' I_{j-k}, & j \geq 1 \end{cases}$$

- $AR(p)$模型偏自相关函数p阶截尾，所以可逆 $MA(q)$ 模型偏自相关函数 ∞ 阶截尾，即具有偏自相关函数**拖尾属性。**

- $MA(1)$模型：

$$Y_t = \varepsilon_t - \theta_1 \varepsilon_{t-1}$$

- 根据偏自相关函数的定义，延迟k阶偏自相关函数是如下方程组的最后一个系数

$$\rho_j = \phi_{k1}\rho_{j-1} + \phi_{k2}\rho_{j-2} + \cdots + \phi_{k(k-1)}\rho_{j-k+1} + \phi_{kk}\rho_{j-k},$$

- 对$j = 1,2,\cdots,k$依次求方程，可得$MA(1)$模型任意k阶偏自相关函数的通解为

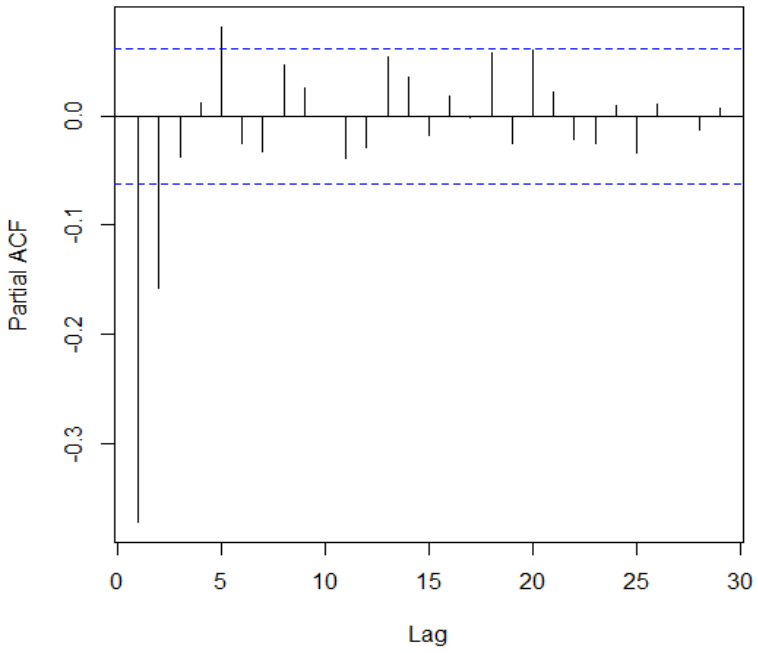
$$\phi_{kk} = \frac{-\theta_1^k}{\sum_{j=0}^{k} \theta_1^{2j}}, k \geq 1$$

MA模型的偏自相关函数

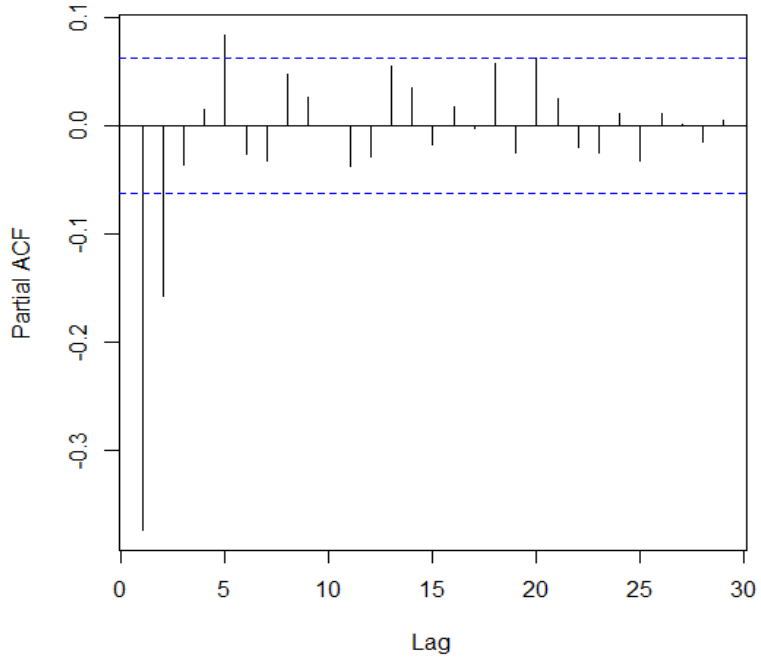
- 考察MA模型偏自相关函数的拖尾性

- (1) $Y_t = \varepsilon_t - 2\varepsilon_{t-1}$
- (2) $Y_t = \varepsilon_t - 0.5\varepsilon_{t-1}$
- (3) $Y_t = \varepsilon_t - \frac{4}{5}\varepsilon_{t-1} + \frac{16}{25}\varepsilon_{t-2}$
- (4) $Y_t = \varepsilon_t - \frac{5}{4}\varepsilon_{t-1} + \frac{25}{16}\varepsilon_{t-2}$

(1) $Y_t = \varepsilon_t - 2\varepsilon_{t-1}$



(2) $Y_t = \varepsilon_t - 0.5\varepsilon_{t-1}$



MA模型的偏自相关函数

- 考察MA模型偏自相关函数的拖尾性

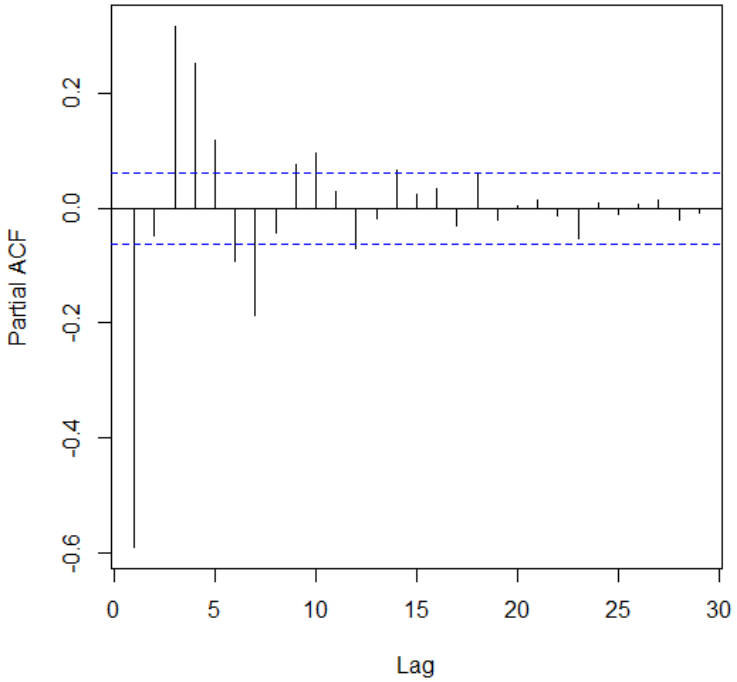
- (1) $Y_t = \varepsilon_t - 2\varepsilon_{t-1}$

- (2) $Y_t = \varepsilon_t - 0.5\varepsilon_{t-1}$

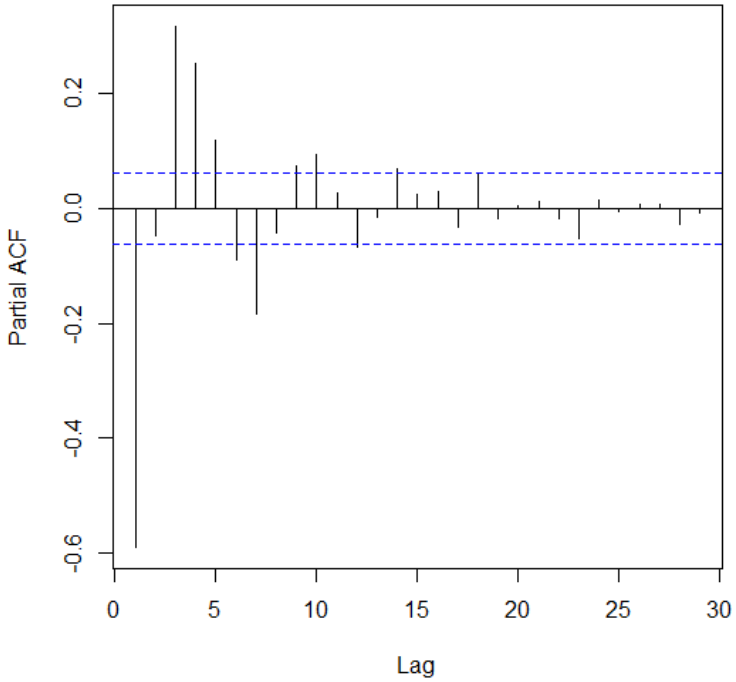
- (3) $Y_t = \varepsilon_t - \frac{4}{5}\varepsilon_{t-1} + \frac{16}{25}\varepsilon_{t-2}$

- (4) $Y_t = \varepsilon_t - \frac{5}{4}\varepsilon_{t-1} + \frac{25}{16}\varepsilon_{t-2}$

(3) $Y_t = \varepsilon_t - \frac{4}{5}\varepsilon_{t-1} + \frac{16}{25}\varepsilon_{t-2}$



(4) $Y_t = \varepsilon_t - \frac{5}{4}\varepsilon_{t-1} + \frac{25}{16}\varepsilon_{t-2}$



概要

1. AR模型

2. MA模型

3. AR模型思想的应用

4. ARMA模型：
(平稳时间序列模型)

5. ARIMA模型：
(非平稳时间序列模型)

6. 模型预测

概要

3. AR模型思想的应用 → 概率估计
(通用思路)

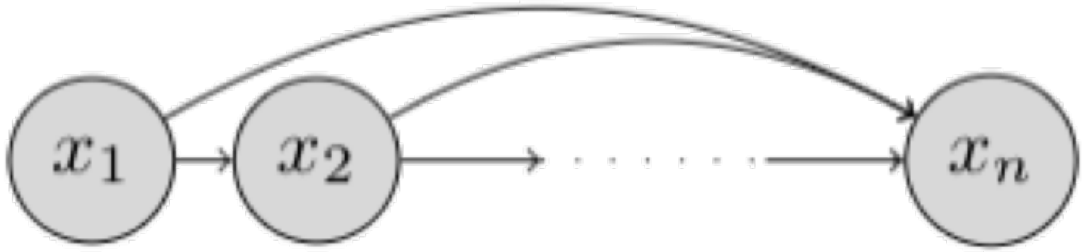
- 语音模型
- 语言模型
- 图像模型
- ……

基于AR的概率分布构建

- 针对数据集$\mathcal{D}$，包含观察值$\{x_i\}$，关于某个随机变量$\mathbf{x}$的分布表示为

$$p(\mathbf{x}) = \prod_{i=1}^{n} p(x_i \mid x_1, x_2, \dots, x_{i-1}) = \prod_{i=1}^{n} p(x_i \mid \mathbf{x}_{<i})$$

其中$\mathbf{x}_{<i} = [x_1, x_2, \dots, x_{i-1}]$表示所有时间在$i$之前的观察值组成的向量



- 注意
  - AR模型对概率的表示（重构）是精确的
  - AR模型的复杂度随n快速增长

https://deepgenerativemodels.github.io/notes/autoregressive/

AR生成模型

- AR用于生成式模型（Generative model），$p(x_i \mid \mathbf{x}_{<i})$能够进行数据抽样
- 若数据$x_i \in \{0,1\}$，则可以设置

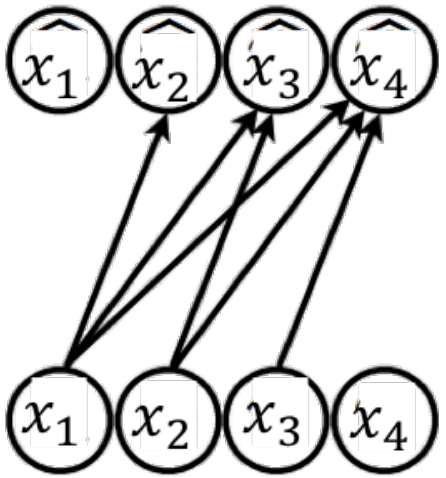
$$p_{\theta_i}(x_i \mid \mathbf{x}_{<i}) = \mathrm{Bern}\big(f_i(x_1, x_2, \dots, x_{i-1})\big)$$

其中θ_i是参数，$f_i: \{0,1\}^{i-1} \to [0,1]$；$f_i(x_1, x_2, \dots, x_{i-1})$设置了用于生成$x_i$的伯努利分布的均值，模型参数总量为$\sum_i |\theta_i|$

- Fully-Visible Sigmoid Belief Network (FVSBN)

$$f_i(x_1, x_2, \dots, x_{i-1}) = \sigma\left(\alpha_0^{(i)} + \alpha_1^{(i)} x_1 + \cdots + \alpha_{i-1}^{(i)} x_{i-1}\right)$$

$\theta_i = \left\{\alpha_0^{(i)}, \alpha_1^{(i)}, \dots, \alpha_{i-1}^{(i)}\right\}$为模型参数，参数总量为$\sum_i^n i = \mathcal{O}(n^2)$



FVSBN

Brendan J. Frey, Geoffrey E. Hinton, Peter Dayan. Does the Wake-sleep Algorithm Produce Good Density Estimators? NIPS 1995.

NADE：基于AR的生成模型

- 使用MLP构造f:

$$f_i(x_1, x_2, \dots, x_{i-1}) = \sigma\left(\boldsymbol{\alpha}^{(i)}\mathbf{h}_i + b_i\right), \qquad \mathbf{h}_i = \sigma(A_i\mathbf{x}_{<\mathbf{i}} + \mathbf{c}_i)$$

参数$\theta_i = \left\{A_i \in \mathbb{R}^{d\times(i-1)}, \mathbf{c}_i \in \mathbb{R}^d, \boldsymbol{\alpha}^{(i)} \in \mathbb{R}^d, b_i \in \mathbb{R}\right\}$，参数量为$\mathcal{O}(n^2 d)$

- Neural *Autoregressive* Density Estimator (NADE)，不同f_i参数**共享**

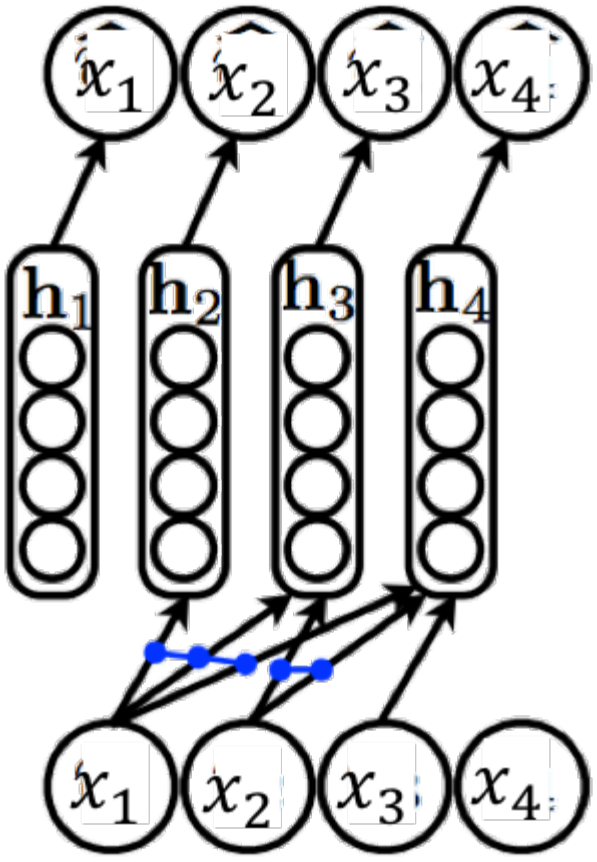
$$f_i(x_1, x_2, \dots, x_{i-1}) = \sigma\left(\boldsymbol{\alpha}^{(i)}\mathbf{h}_i + b_i\right), \qquad \mathbf{h}_i = \sigma\left(W_{\cdot,<i}\mathbf{x}_{<\mathbf{i}} + \mathbf{c}\right)$$

参数$\theta = \left\{W \in \mathbb{R}^{d\times n}, \mathbf{c} \in \mathbb{R}^d, \left\{\boldsymbol{\alpha}^{(i)} \in \mathbb{R}^d\right\}_{i=1}^n, \left\{b_i \in \mathbb{R}\right\}_{i=1}^n\right\}$，参数量为$\mathcal{O}(nd)$

- 隐层变量可使用**递归**方法计算

$$\mathbf{h}_i = \sigma(\mathbf{a}_i), \qquad \mathbf{a}_{i+1} = \mathbf{a}_i + W[\cdot, i]x_i$$

其中$\mathbf{a}_1 = \mathbf{c}$，复杂度为$\mathcal{O}(nd)$



NADE

Hugo Larochelle, Iain Murray. The Neural Autoregressive Distribution Estimator. AISTATS 2011.

NADE：基于AR的生成模型

- 条件概率生成

$$p\left(x_{o_d} = 1 \mid \boldsymbol{x}_{o_{<d}}\right) = \sigma\left(\boldsymbol{V}_{o_d,}\boldsymbol{h}_d + b_{o_d}\right)$$

$$\boldsymbol{h}_d = \sigma\left(\boldsymbol{W}_{\cdot,o_{<d}}\boldsymbol{x}_{o_d} + \boldsymbol{c}\right)$$

- 通过递归方法计算概率
  - 初始化

$$\boldsymbol{h}_1 = \sigma(\boldsymbol{a}_1),\ \text{where } \boldsymbol{a}_1 = \boldsymbol{c}$$

  - 迭代

$$\boldsymbol{h}_d = \sigma(\boldsymbol{a}_d),\ \text{for } d \in \{2, \dots, D\}$$

where $\boldsymbol{a}_d = \boldsymbol{W}_{\cdot,o_{<d}}\boldsymbol{x}_{o_{<d}} + \boldsymbol{c} = \boldsymbol{W}_{\cdot,o_{d-1}}x_{o_{d-1}} + \boldsymbol{a}_{d-1}$

- 最大似然

$$\frac{1}{N}\sum_{n=1}^{N} -\log p\left(\boldsymbol{x}^{(n)}\right) = \frac{1}{N}\sum_{n=1}^{N}\sum_{d=1}^{D} -\log p\left(x_{o_d}^{(n)} \mid \boldsymbol{x}_{o_{<d}}^{(n)}\right)$$

Algorithm 1 Computation of $p(\boldsymbol{x})$ and learning gradients for NADE.

Input: training observation vector $\boldsymbol{x}$ and ordering o of the input dimensions.
Output: $p(\boldsymbol{x})$ and gradients of $-\log p(\boldsymbol{x})$ on parameters.

Computing $p(\boldsymbol{x})$
$\boldsymbol{a}_1 \leftarrow \boldsymbol{c}$
$p(\boldsymbol{x}) \leftarrow 1$
for d from 1 to D **do**
  $\boldsymbol{h}_d \leftarrow \text{sigm}\left(\boldsymbol{a}_d\right)$
  $p(x_{o_d} = 1 \mid \boldsymbol{x}_{o_{<d}}) \leftarrow \text{sigm}\left(\boldsymbol{V}_{o_d,\cdot}\boldsymbol{h}_d + b_{o_d}\right)$
  $p(\boldsymbol{x}) \leftarrow p(\boldsymbol{x})\left(p(x_{o_d} = 1 \mid \boldsymbol{x}_{o_{<d}})^{x_{o_d}} + \left(1 - p(x_{o_d} = 1 \mid \boldsymbol{x}_{o_{<d}})\right)^{1-x_{o_d}}\right)$
  $\boldsymbol{a}_{d+1} \leftarrow \boldsymbol{a}_d + \boldsymbol{W}_{\cdot,o_d}x_{o_d}$
end for

Computing gradients of $-\log p(\boldsymbol{x})$
$\delta\boldsymbol{a}_D \leftarrow 0$
$\delta\boldsymbol{c} \leftarrow 0$
for d from D to 1 **do**
  $\delta b_{o_d} \leftarrow \left(p(x_{o_d} = 1 \mid \boldsymbol{x}_{o_{<d}}) - x_{o_d}\right)$
  $\delta\boldsymbol{V}_{o_d,\cdot} \leftarrow \left(p(x_{o_d} = 1 \mid \boldsymbol{x}_{o_{<d}}) - x_{o_d}\right)\boldsymbol{h}_d^\top$
  $\delta\boldsymbol{h}_d \leftarrow \left(p(x_{o_d} = 1 \mid \boldsymbol{x}_{o_{<d}}) - x_{o_d}\right)\boldsymbol{V}_{o_d,\cdot}^\top$
  $\delta\boldsymbol{c} \leftarrow \delta\boldsymbol{c} + \delta\boldsymbol{h}_d \odot \boldsymbol{h}_d \odot (1 - \boldsymbol{h}_d)$
  $\delta\boldsymbol{W}_{\cdot,o_d} \leftarrow \delta\boldsymbol{a}_d x_{o_d}$
  $\delta\boldsymbol{a}_{d-1} \leftarrow \delta\boldsymbol{a}_d + \delta\boldsymbol{h}_d \odot \boldsymbol{h}_d \odot (1 - \boldsymbol{h}_d)$
end for
return $p(\boldsymbol{x}), \delta\boldsymbol{b}, \delta\boldsymbol{V}, \delta\boldsymbol{c}, \delta\boldsymbol{W}$

NADE的变种

- RNADE: Real-Valued NADE
  - 假设$f_i(x_1, x_2, \dots, x_{i-1})$为高斯混合模型，估计$\{\mu_{i,1}, \mu_{i,2}, \dots, \mu_{i,K}, \Sigma_{i,1}, \Sigma_{i,2}, \dots, \Sigma_{i,k}\}$

$$p\left(x_{o_d} \mid \boldsymbol{x}_{o_{<d}}\right) = \sum_{c=1}^{C} \pi_{o_d,c} \mathcal{N}\left(x_{o_d}; \mu_{o_d,c}, \sigma_{o_d,c}^2\right)$$

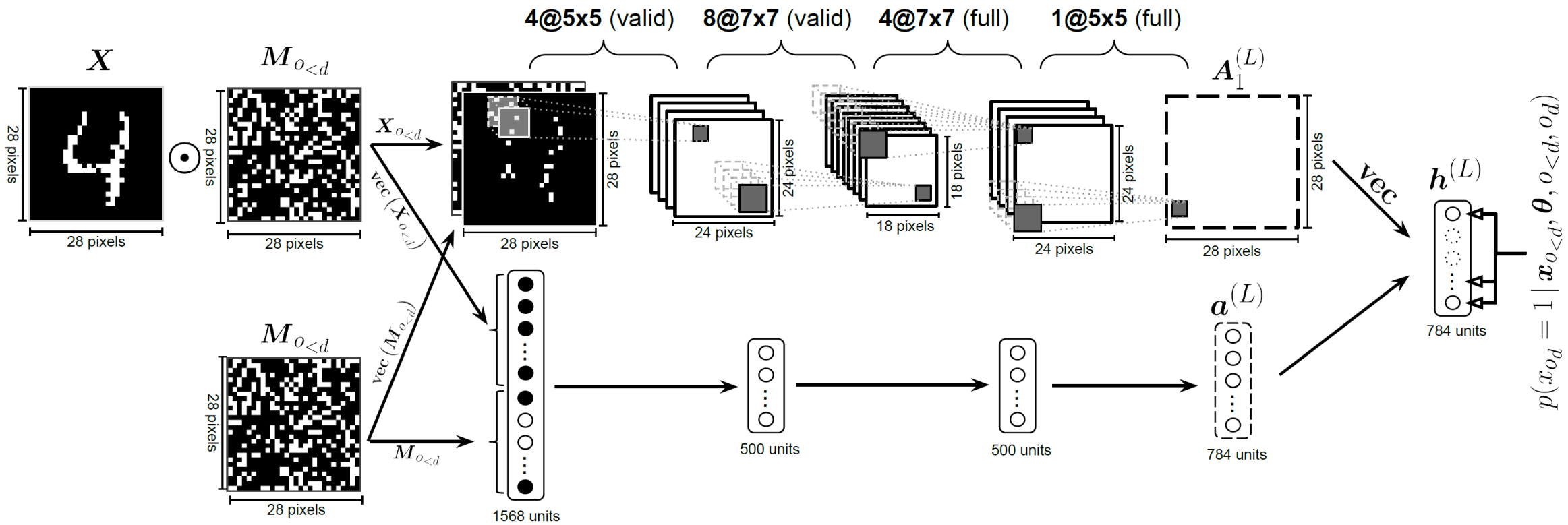
- EoNADE: Ensemble of NADE
  - 针对不同顺序$\{o^{(k)}\}_{k=1}^{K}$进行估计，再集成

$$\frac{1}{K} \sum_{k=1}^{K} p\left(\boldsymbol{x} \mid \theta, o^{(k)}\right)$$

Benigno Uria et al. Neural Autoregressive Distribution Estimation. JMLR. 17: 205:1-205:37 (2016)

NADE的变种

- ConvNADE：NADE with ConvNet



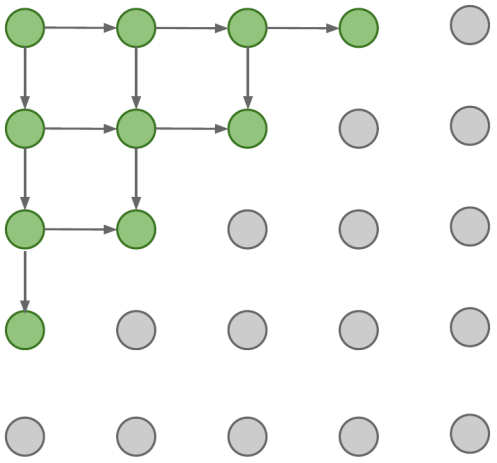
Benigno Uria et al. Neural Autoregressive Distribution Estimation. JMLR. 17: 205:1-205:37 (2016)

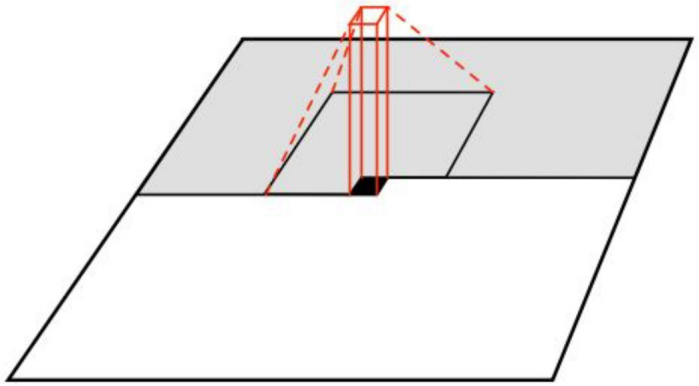
PixelRNN和PixelCNN

- PixelRNN
  - 从左上角生成$n \times n$图片（序列化预测后续内容）

$$p(\mathbf{x}) = \prod_{i=1}^{n^2} p(x_i \mid x_1, \ldots, x_{i-1})$$

  - 使用RNN等模型构建条件概率
  - 序列化生成图片非常慢
- PixelCNN
  - 使用CNN针对部分上下文区域建模条件概率
  - 一定程度加快训练，但生成图片仍非常慢





Aäron van den Oord, Nal Kalchbrenner, Koray Kavukcuoglu. Pixel Recurrent Neural Networks. ICML 2016.

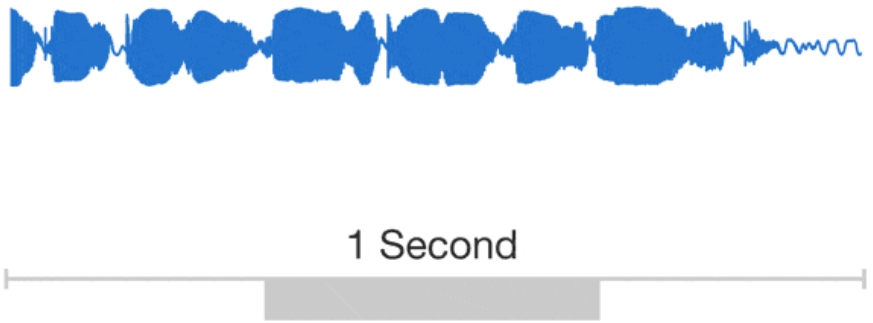
Aäron van den Oord et al. Conditional Image Generation with PixelCNN Decoders. NIPS 2016.

WaveNet

- 生成语音数据（具有极高分辨率）

- 语音序列$\{x_1, \dots, x_T\}$概率的分解

$$p(\mathbf{x}) = \prod_{t=1}^{T} p(x_t \mid x_1, \dots, x_{t-1})$$



使用类似PixelCNN的方式，用CNN构建条件概率，关键在于如何进行卷积

Aäron van den Oord et al. WaveNet: A Generative Model for Raw Audio. SSW 2016.

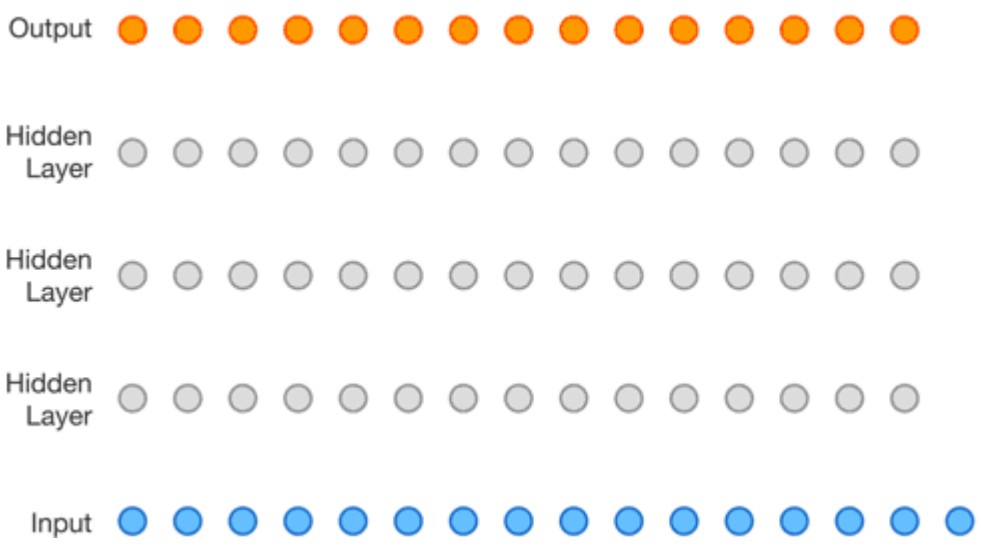
WaveNet的Casual Convolution

- 使用"因果"卷积（Causal/Masked Convolution）构建条件概率

$$p(x_{t+1} \mid x_1, \dots, x_t)$$

条件概率的生成和未来数值$x_{t+1}, \dots, x_{t+T}$无关

- 使用Dilated Convolution提升效率，作用于更粗的粒度（similar to pooling or strided convolutions, but here the output has the **same** size as the input）



Dilated causal convolutions for dilations 1, 2, 4, and 8.

Word2Vec

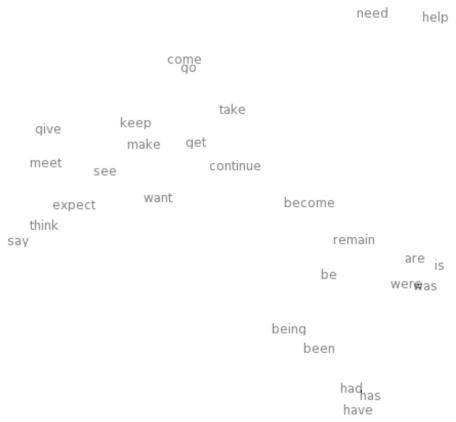
- 主要思路：通过上下文确定单次的语义表示。

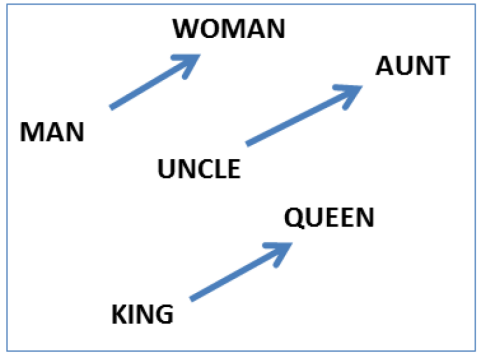
Distributional semantics: A word's meaning is given by the words that freque

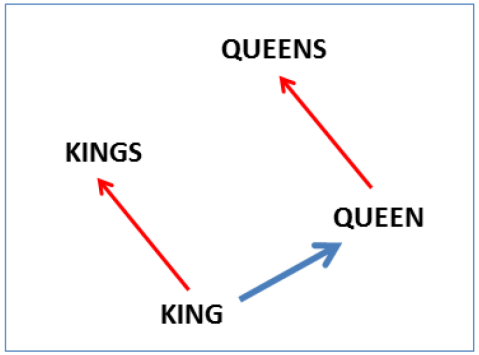
…government debt problems turning into	***banking***	*crises as happened in 2009…*
…saying that Europe needs unified	***banking***	*regulation to replace the hodgepodge…*
…India has just given its	***banking***	*system a shot in the arm…*

"banking"的语义可由上下文推断

$$banking = \begin{pmatrix} 0.286 \\ 0.792 \\ -0.177 \\ -0.107 \\ 0.109 \\ -0.542 \\ 0.349 \\ 0.271 \end{pmatrix}$$



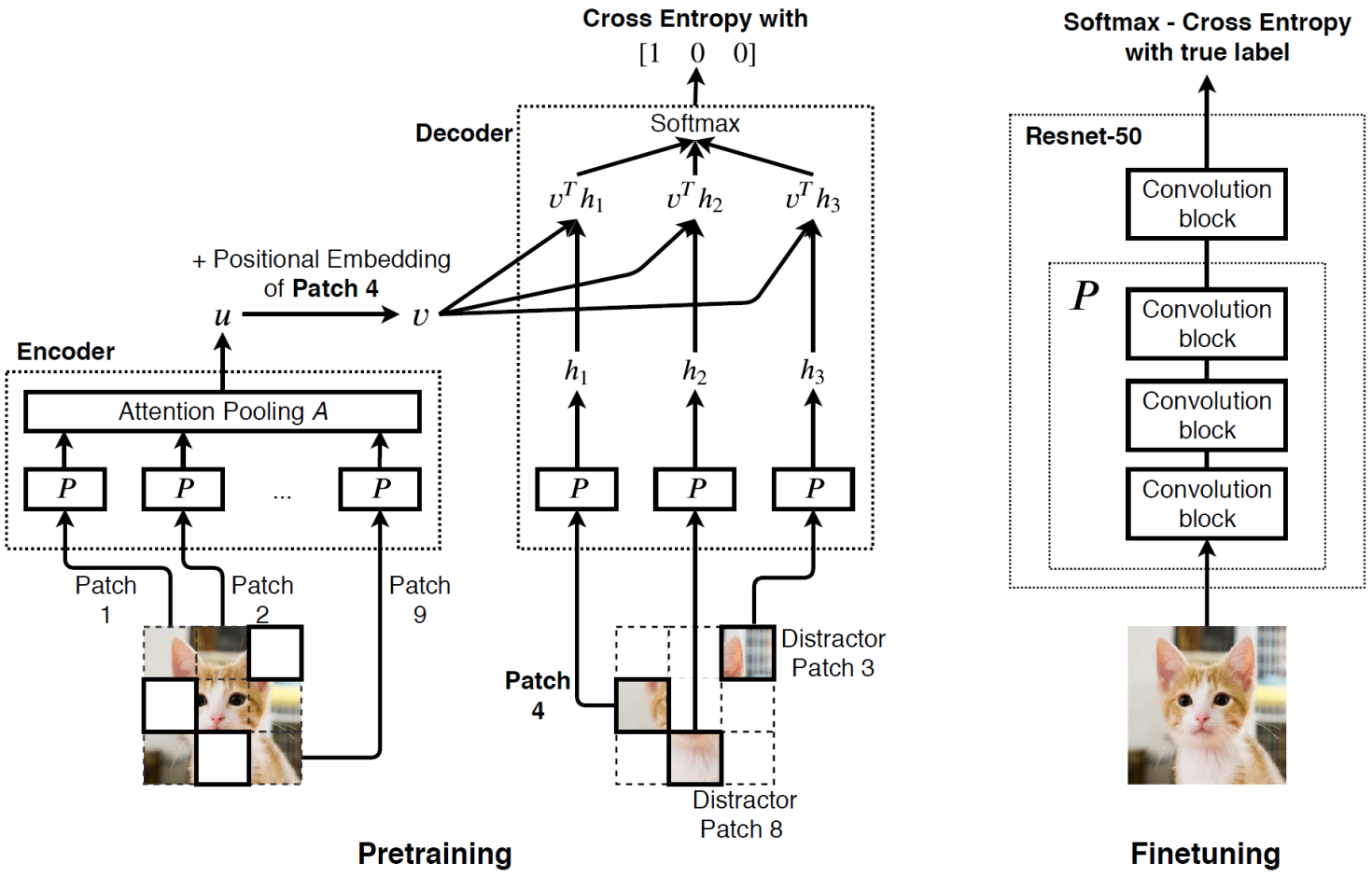




Tomás Mikolov, Wen-tau Yih, Geoffrey Zweig. Linguistic Regularities in Continuous Space Word Representations. HLT-NAACL 2013.

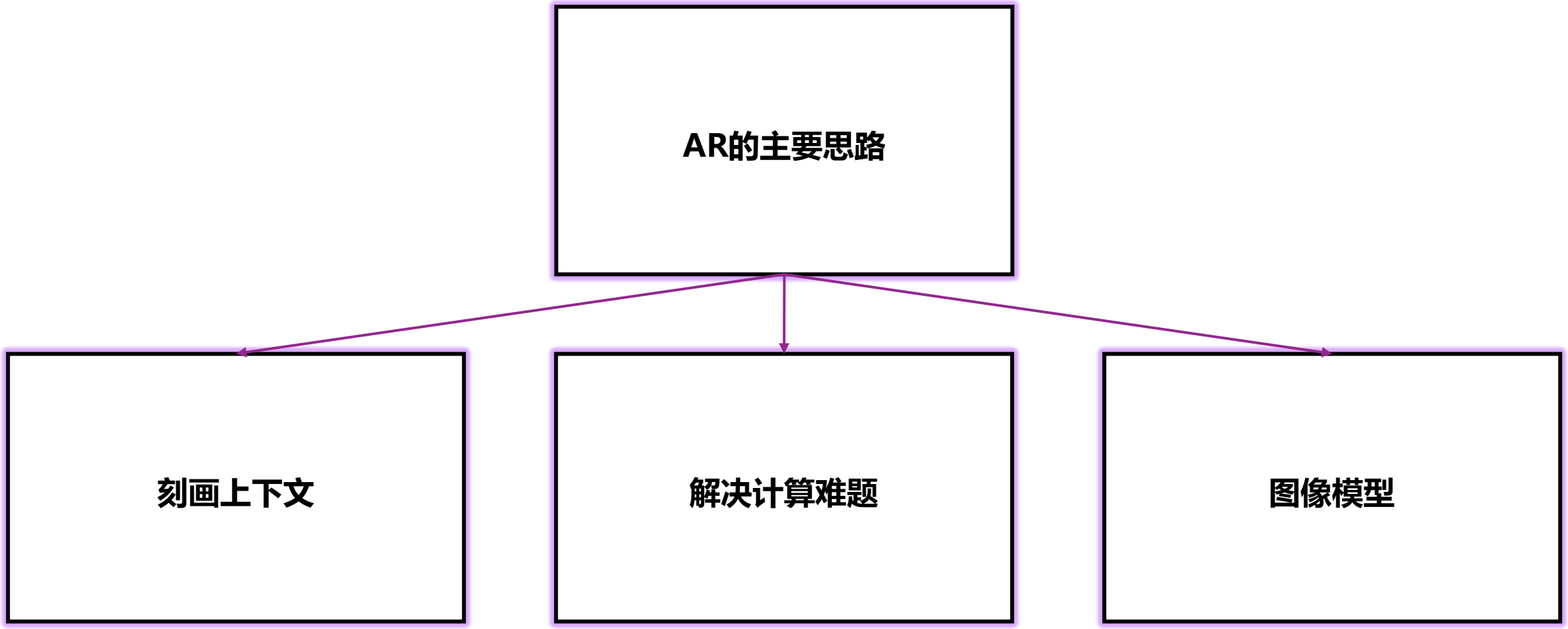
面向图像的自监督学习任务

- 阶段1：针对无标记数据优化特征（使用局部图像预测剩余部分）
  - 针对patch提取特征
  - 融合上下文patch特征得到u
  - 结合位置信息，预测在特定位置上最适合的patch（predict **what patch is being masked out** at a particular location among other distracting patches）
- 阶段2：基于有标记数据微调



Trieu H. Trinh, Minh-Thang Luong, Quoc V. Le. Selfie: Self-supervised Pretraining for Image Embedding. CoRR abs/1906.02940 (2019)

AR思想应用总结



线性过程的平稳与可逆性

- 线性过程：写为以下形式的随机过程，$\{\varepsilon_t\}$为白噪声

$$Y_t = \sum_{i=0}^{\infty} \psi_i \varepsilon_{t-i}$$

$\boldsymbol{\psi}$加权

- 记

$$\psi(\mathbf{B}) = 1 + \sum_{i=1}^{\infty} \psi_i \mathbf{B}^i$$

则

$$Y_t = \psi(\mathbf{B})\varepsilon_t$$

$$\pi(\mathbf{B})\psi(\mathbf{B}) = 1$$

$\boldsymbol{\pi}$加权

- 记

$$\pi(\mathbf{B}) = 1 + \sum_{i=1}^{\infty} \pi_i \mathbf{B}^i$$

则

$$\pi(\mathbf{B})Y_t = \varepsilon_t$$

【平稳性】 $Y_t = \psi(\mathbf{B})\varepsilon_t$平稳

- 若当$|z| > 1$时，$\psi(z) = \sum_{i=0}^{\infty} \psi_i z^{-i}$收敛
- 若当$|\mathbf{B}| < 1$时，$\psi(\mathbf{B}) = \sum_{i=0}^{\infty} \psi_i \mathbf{B}^i$收敛

【可逆性】 $\pi(\mathbf{B})Y_t = \varepsilon_t$可逆:

- 若当$|z| > 1$时，$\pi(z) = \sum_{i=0}^{\infty} \pi_i z^{-i}$收敛
- 若当$|\mathbf{B}| < 1$时，$\pi(\mathbf{B}) = \sum_{i=0}^{\infty} \pi_i \mathbf{B}^i$收敛

特殊的线性过程

- 线性过程：写为以下形式的随机过程

$$Y_t = \sum_{i=0}^{\infty} \psi_i \varepsilon_{t-i}$$

- 设$\boldsymbol{\psi}(\boldsymbol{z}) = \boldsymbol{\theta}(\boldsymbol{z})/\boldsymbol{\phi}(\boldsymbol{z})$, $\theta(z)$, $\phi(z)$为多项式

AR过程

- Y_t表示为

$$Y_t + \phi_1 Y_{t-1} + \cdots + \phi_p Y_{t-p} = \varepsilon_t$$

$\{\varepsilon_t\}$为白噪声

- 从$p+1$开始π权重为0
- 使用延迟算子表示为

$$\left(1 + \phi_1 \mathbf{B} + \cdots + \phi_p \mathbf{B}^p\right) Y_t = \phi(\mathbf{B}) Y_t = \varepsilon_t$$

$$\boldsymbol{\psi}(\boldsymbol{z}) = \boldsymbol{\theta}(\boldsymbol{z})/\boldsymbol{\phi}(\boldsymbol{z})$$

MA过程

- Y_t表示为

$$Y_t = \varepsilon_t + \theta_1 \varepsilon_{t-1} + \cdots + \theta_q \varepsilon_{t-q}$$

$\{\varepsilon_t\}$为白噪声

- 从$q+1$开始ψ权重为0
- 使用延迟算子表示为

$$Y_t = \left(1 + \theta_1 \mathbf{B} + \cdots + \theta_q \boldsymbol{B}^q\right) \varepsilon_t = \theta(\boldsymbol{B}) \varepsilon_t$$

AR过程的平稳性与可逆性

【平稳性】 $Y_t = \psi(\mathbf{B})\varepsilon_t$平稳

- 若当$|z| > 1$时，$\psi(z) = \sum_{i=0}^{\infty} \psi_i z^{-i}$收敛
- 若当$|\mathbf{B}| < 1$时，$\psi(\mathbf{B}) = \sum_{i=0}^{\infty} \psi_i \mathbf{B}^i$收敛

- $AR(1)$过程

$$Y_t - \phi_1 Y_{t-1} = (1 - \phi_1 \mathbf{B}) Y_t = \varepsilon_t$$

- 转化为$MA(\infty)$

$$Y_t = \overbrace{(1 - \phi_1 \mathbf{B})^{-1}}^{\psi(\mathbf{B})} \varepsilon_t = \sum_{i=0}^{\infty} \phi_1^i \, \varepsilon_{t-i}$$

- 判断平稳性，当$|\mathbf{B}| < 1$时，$\psi(\mathbf{B})$收敛，因此有$|\phi_1| < 1$;
- $1 - \phi_1 \mathbf{B} = 0$的根是$\mathbf{B} = \phi_1^{-1}$，因此相当于要求$1 - \phi_1 \mathbf{B} = 0$的根在单位圆**外**;

AR过程的平稳性与可逆性

【平稳性】 $Y_t = \psi(\mathbf{B})\varepsilon_t$平稳

- 若当$|z| > 1$时，$\psi(z) = \sum_{i=0}^{\infty} \psi_i z^{-i}$收敛
- 若当$|\mathbf{B}| < 1$时，$\psi(\mathbf{B}) = \sum_{i=0}^{\infty} \psi_i \mathbf{B}^i$收敛

- 一般$AR(p)$过程

$$\phi(\mathbf{B})\, Y_t = \varepsilon_t$$

- 转化为$MA(\infty)$

$$Y_t = \phi(\mathbf{B})^{-1}\varepsilon_t = \psi(\mathbf{B})\varepsilon_t$$

- 判断平稳性，假设 $G_1^{-1}, \dots, G_p^{-1}$是$\phi(\mathbf{B}) = 0$的根，则有$\phi(\mathbf{B}) = (1 - G_1\mathbf{B}) \cdots (1 - G_p\mathbf{B})$

$$Y_t = \phi(\mathbf{B})^{-1}\varepsilon_t = \sum_{i=1}^{p} \frac{K_i}{1 - G_i\mathbf{B}} \varepsilon_t$$

- 当$|\mathbf{B}| < 1$时，$\psi(\mathbf{B})$收敛，要求$\psi_j = \sum_{i=1}^{p} K_i G_i^j$绝对可和，即$|G_i| < 1$；

- 平稳条件

$$\phi(\mathbf{B}) = 0$$

的根在单位圆**外**；由于

$$\phi(\mathbf{B}) = 1 - \phi_1\mathbf{B} - \cdots - \phi_p\mathbf{B}^p = 0$$

的根和

$$m^p - \phi_1 m^{p-1} - \cdots - \phi_p = 0$$

的根互为**倒数**，因此，相当于上述方程的根在圆内。

- 可逆性

$$\pi(\mathbf{B}) = \phi(\mathbf{B}) = 1 - \phi_1\mathbf{B} - \cdots - \phi_p\mathbf{B}^p$$

为有限项，满足可逆性条件

MA过程的平稳性与可逆性

【可逆性】 $\pi(\mathbf{B})Y_t = \varepsilon_t$可逆:

- 若当$|z| > 1$时，$\pi(z) = \sum_{i=0}^{\infty} \pi_i z^{-i}$收敛
- 若当$|\mathbf{B}| < 1$时，$\pi(\mathbf{B}) = \sum_{i=0}^{\infty} \pi_i \mathbf{B}^i$收敛

- $MA(q)$过程

$$Y_t = \varepsilon_t - \theta_1 \varepsilon_{t-1} - \cdots - \theta_q \varepsilon_{t-q}$$
$$= \left(1 - \theta_1 \mathbf{B} - \cdots - \theta_q \mathbf{B}^q\right)\varepsilon_t = \theta(\mathbf{B})\varepsilon_t$$

- 平稳性

$$\psi(\mathbf{B}) = \theta(\mathbf{B}) = 1 - \theta_1 \mathbf{B} - \cdots - \theta_q \mathbf{B}^q$$

为**有限**项，满足稳定性条件

- 可逆性，可得

$$\varepsilon_t = \theta^{-1}(\mathbf{B})Y_t$$

假设$H_1^{-1}, \ldots, H_q^{-1}$是$\theta(\mathbf{B}) = 0$的根，则有

$$\theta(\mathbf{B}) = \prod_{i=1}^{q}(1 - H_i \mathbf{B})$$

- 基于上述分解，有

$$\pi(\mathbf{B}) = \theta^{-1}(\mathbf{B}) = \sum_{i}^{q}\left(\frac{M_i}{1 - H_i \mathbf{B}}\right)$$

针对可逆性条件，令$\pi_j = \sum_i^q M_i H_i^j$绝对可和，即$|H_i| < 1$，

- 等价于特征方程

$$\theta(\mathbf{B}) = 1 - \theta_1 \mathbf{B} - \cdots - \theta_q \mathbf{B}^q = 0$$

的根$H_i^{-1} > 1$，在单位圆外

AR过程与MA过程

- $AR(p)$过程

$$\phi(\mathbf{B})\,Y_t = \varepsilon_t$$

- 平稳性

$\phi(\mathbf{B}) = 0$的根在单位圆**外**

- 可逆性

直接满足

- ACF

拖尾（指数衰减）

- PACF

截尾（$\phi_{kk} = 0, k > p$）

- $MA(q)$过程

$$Y_t = \theta(\mathbf{B})\varepsilon_t$$

- 平稳性

直接满足

- 可逆性

$\theta(\mathbf{B}) = 0$的根在单位圆**外**

- ACF

截尾（$\rho(k) = 0, k > q$（

- PACF

拖尾（指数衰减）

概要

1. AR模型

2. MA模型

3. AR模型思想的应用

4. ARMA模型：
(平稳时间序列模型)

5. ARIMA模型：
(非平稳时间序列模型)

6. 模型预测

ARMA模型的定义

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

$$\begin{cases} Y_t = \phi_0 + \phi_1 Y_{t-1} + \cdots + \phi_p Y_{t-p} + \varepsilon_t - \theta_1 \varepsilon_{t-1} - \cdots - \theta_q \varepsilon_{t-q} \\ \phi_p \neq 0,\ \theta_q \neq 0 \\ E(\varepsilon_t) = 0,\ Var(\varepsilon_t) = \sigma_\varepsilon^2, E(\varepsilon_t \varepsilon_s) = 0, s \neq t \\ E(Y_s \varepsilon_t) = 0, \forall s < t \end{cases}$$

- 当$\phi_0 = 0$时，称为中心化$ARMA(p,q)$模型
- $ARMA(p,q)$ 模型简记为

$$\phi(\mathbf{B}) Y_t = \theta(\mathbf{B}) \varepsilon_t$$

其中：

$$\phi(\mathbf{B}) = 1 - \phi_1 \mathbf{B} - \phi_2 \mathbf{B}^2 - \cdots - \phi_p \mathbf{B}^p$$
$$\theta(\mathbf{B}) = 1 - \theta_1 \mathbf{B} - \theta_2 \mathbf{B}^2 - \cdots - \theta_q \mathbf{B}^q$$

AR、MA、ARMA模型

$$Y_t = \varepsilon_t + \theta_1 \varepsilon_{t-1} + \cdots + \theta_q \varepsilon_{t-q}$$

$$Y_t + \phi_1 Y_{t-1} + \cdots + \phi_p Y_{t-p} = \varepsilon_t$$

$$Y_t + \phi_1 Y_{t-1} + \cdots + \phi_p Y_{t-p} = \varepsilon_t + \theta_1 \varepsilon_{t-1} + \cdots + \theta_q \varepsilon_{t-q}$$

$$\Longleftrightarrow$$

$$Y_t = \theta(\mathbf{B})\varepsilon_t$$

$$\phi(\mathbf{B})Y_t = \varepsilon_t$$

$$\phi(\mathbf{B})Y_t = \theta(\mathbf{B})\varepsilon_t$$

- $\{\varepsilon_t\}$ 为白噪声，$\theta(\mathbf{B})$ 和$\phi(\mathbf{B})$是关于**B**的多项式

对于$ARMA$过程，

理解为p阶AR，

$$\phi(\mathbf{B})Y_t = e_t,$$

$e_t = \theta(\mathbf{B})\varepsilon_t$为MA过程

理解为q阶MA，

$$Y_t = \theta(\mathbf{B})b_t$$

b_t为p阶AR过程，$\phi(\mathbf{B})b_t = \varepsilon_t$，则

$$\phi(\mathbf{B})Y_t = \theta(\mathbf{B})\phi(\mathbf{B})\, b_t = \theta(\mathbf{B})\varepsilon_t$$

平稳条件与可逆条件

- $ARMA(p,q)$模型的**平稳**条件
  - p阶自回归系数多项式$\phi(\mathbf{B})=0$的根都在单位圆外
  - 即$ARMA(p,q)$模型的平稳性完全由其**自回归部分**的平稳性决定

- $ARMA(p,q)$模型的**可逆**条件
  - q阶移动平均系数多项式$\theta(\mathbf{B})=0$的根都在单位圆外
  - 即$ARMA(p,q)$模型的可逆性完全由其**滑动平均**部分的可逆性决定

ARMA模型的统计性质

- 均值

$$EY_t = \frac{\phi_0}{1 - \phi_1 - \cdots - \phi_p}$$

- 协方差

$$\gamma(k) = \sigma_\varepsilon^2 \sum_{i=0}^{\infty} G_i G_{i+k}$$

- 自相关函数

$$\rho(k) = \frac{\gamma(k)}{\gamma(0)} = \frac{\sum_{j=0}^{\infty} G_j G_{j+k}}{\sum_{j=0}^{\infty} G_j^2}$$

- 自相关函数**拖尾**
- $ARMA(p,q)$模型可以转化为**无穷阶**移动平均模型
- 偏自相关函数**拖尾**
- $ARMA(p,q)$模型可以转化为**无穷阶**自回归模型

ARMA模型的自协方差函数

- 针对ARMA模型

$$Y_t + \phi_1 Y_{t-1} + \cdots + \phi_p Y_{t-p} = \varepsilon_t + \theta_1 \varepsilon_{t-1} + \cdots + \theta_q \varepsilon_{t-q}$$

- 上式两端乘ε_{t-k}求期望

$$\gamma_{\varepsilon Y}(k) + \phi_1 \gamma_{\varepsilon Y}(k-1) + \cdots + \phi_p \gamma_{\varepsilon Y}(k-p) = \theta_k \sigma_\varepsilon^2, \qquad k = 0,1,\ldots$$

- 将ARMA模型转化为$MA(\infty)$模型

$$Y_t = \sum_{i=0}^{\infty} \psi_i \varepsilon_{t-i}$$

ε_t 和当前以及未来的 Y_t 相关，和历史的Y_t无关

- 可得

$$\gamma_{\varepsilon Y}(k) = Cov[\varepsilon_t, Y_{t+k}] = \mathrm{E}[Y_t \varepsilon_{t-k}] \begin{cases} = 0 & \text{for } k < 0 \\ \neq 0 & \text{for } k \geq 0 \end{cases}$$

- 因此

$$\begin{aligned} k = 0: \quad & \gamma_{\varepsilon Y}(0) && = \theta_0 \sigma_\varepsilon^2 \\ k = 1: \quad & \gamma_{\varepsilon Y}(1) + \phi_1 \gamma_{\varepsilon Y}(0) && = \theta_1 \sigma_\varepsilon^2 \end{aligned} \quad \Rightarrow \quad \gamma_{\varepsilon Y}(1) = \overbrace{(\theta_1 - \phi_1 \theta_0)}^{\psi_k} \sigma_\varepsilon^2$$

ARMA模型的自协方差函数

- 针对ARMA模型

$$Y_t + \phi_1 Y_{t-1} + \cdots + \phi_p Y_{t-p} = \varepsilon_t + \theta_1 \varepsilon_{t-1} + \cdots + \theta_q \varepsilon_{t-q}$$

- 上式两端乘Y_{t-k}求期望，得$\forall\ k = 0,1,\ldots$

$$\begin{aligned}\gamma(k) + \phi_1 \gamma(k-1) + \cdots + \phi_p \gamma(k-p) &= \gamma_{\varepsilon Y}(-k) + \theta_1 \gamma_{\varepsilon Y}\big(-(k-1)\big) + \cdots + \theta_k \gamma_{\varepsilon Y}(0) + \cdots + \theta_q \gamma_{\varepsilon Y}(q-k) \\ &= \theta_k \gamma_{\varepsilon Y}(0) + \cdots + \theta_q \gamma_{\varepsilon Y}(q-k)\end{aligned}$$

- 若$k \geq q+1$，则右侧全部为0，得到

$$\gamma(k) + \phi_1 \gamma(k-1) + \cdots + \phi_p \gamma(k-p) = 0$$

- 若$p > q$：自协方差函数和ACF拖尾（包括衰减指数和正弦函数）
- 若$p \leq q$:

$$\begin{gathered}\gamma(p) + \phi_1 \gamma(p-1) + \cdots + \phi_p \gamma(0) = \theta_p \gamma_{\varepsilon Y}(0) + \cdots + \theta_q \gamma_{\varepsilon Y}(q-p) \\ \gamma(p+1) + \phi_1 \gamma(p) + \cdots + \phi_p \gamma(1) = \theta_{p+1} \gamma_{\varepsilon Y}(0) + \cdots + \theta_q \gamma_{\varepsilon Y}(q-p-1) \\ \vdots \\ \gamma(q+1) + \phi_1 \gamma(q) + \cdots + \phi_p \gamma(q+1-p) = 0\end{gathered}$$

$k = q+1-p$后自协方差/ACF拖尾（包括衰减指数和正弦函数）

ARMA模型的自协方差函数

- ARMA(1,1)模型

$$Y_t + \phi_1 Y_{t-1} = \varepsilon_t + \theta_1 \varepsilon_{t-1}$$

- 首先考虑$\gamma_{\varepsilon Y}(k)$

$$\gamma_{\varepsilon Y}(0) = \sigma_\varepsilon^2$$
$$\gamma_{\varepsilon Y}(1) = (\theta_1 - \phi_1)\sigma_\varepsilon^2$$
$$\gamma_{\varepsilon Y}(k) = -\phi_1 \gamma_{\varepsilon Y}(k-1), \text{ for } k \geq 2$$

- 确定$\gamma(k)$

$$\gamma(0) + \phi_1 \gamma(1) = \gamma_{\varepsilon Y}(0) + \theta_1 \gamma_{\varepsilon Y}(1)$$
$$\gamma(1) + \phi_1 \gamma(0) = \theta_1 \gamma_{\varepsilon Y}(0)$$
$$\gamma(k) = -\phi_1 \gamma(k-1), \text{ for } k \geq 2$$

- 因此

$$\gamma(0) + \phi_1 \gamma(1) = \big(1 + \theta_1(\theta_1 - \phi_1)\big)\sigma_\varepsilon^2$$
$$\gamma(1) + \phi_1 \gamma(0) = \theta_1 \sigma_\varepsilon^2$$

- 可得

$$\gamma(0) + \phi_1(\theta_1 \sigma_\varepsilon^2 - \phi_1 \gamma(0)) = \big(1 + \theta_1(\theta_1 - \phi_1)\big)\sigma_\varepsilon^2$$
$$\gamma(1) = -\phi_1 \gamma(0) + \theta_1 \sigma_\varepsilon^2$$

- 则

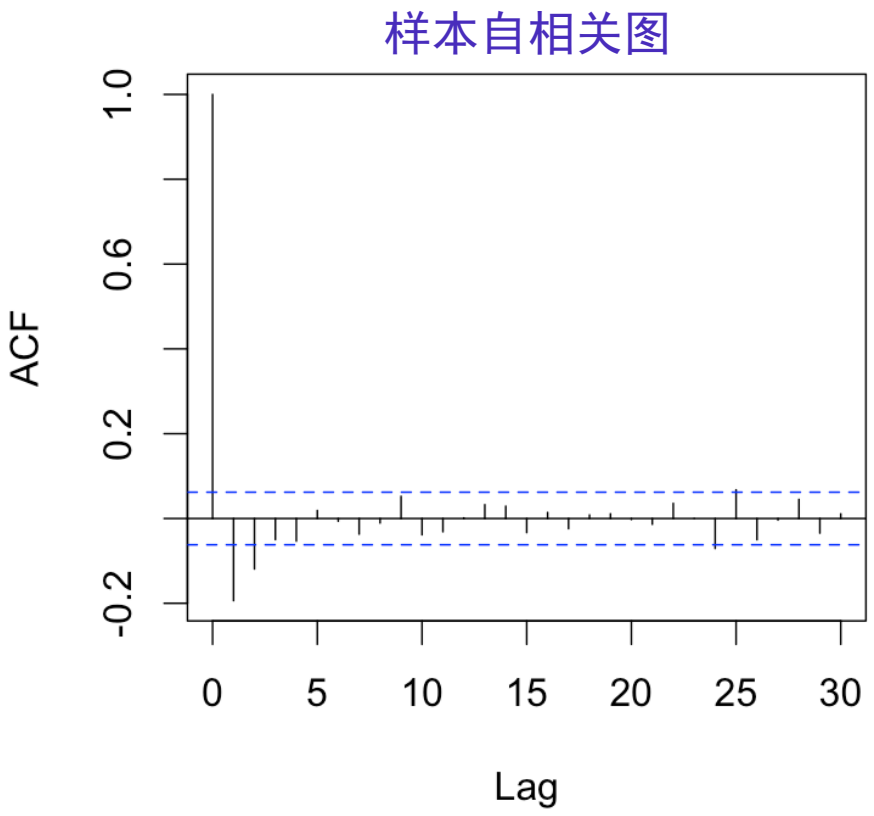
$$\gamma(0) = \frac{1 + \theta_1^2 - 2\phi_1\theta_1}{(1 - \phi_1^2)}\sigma_\varepsilon^2$$

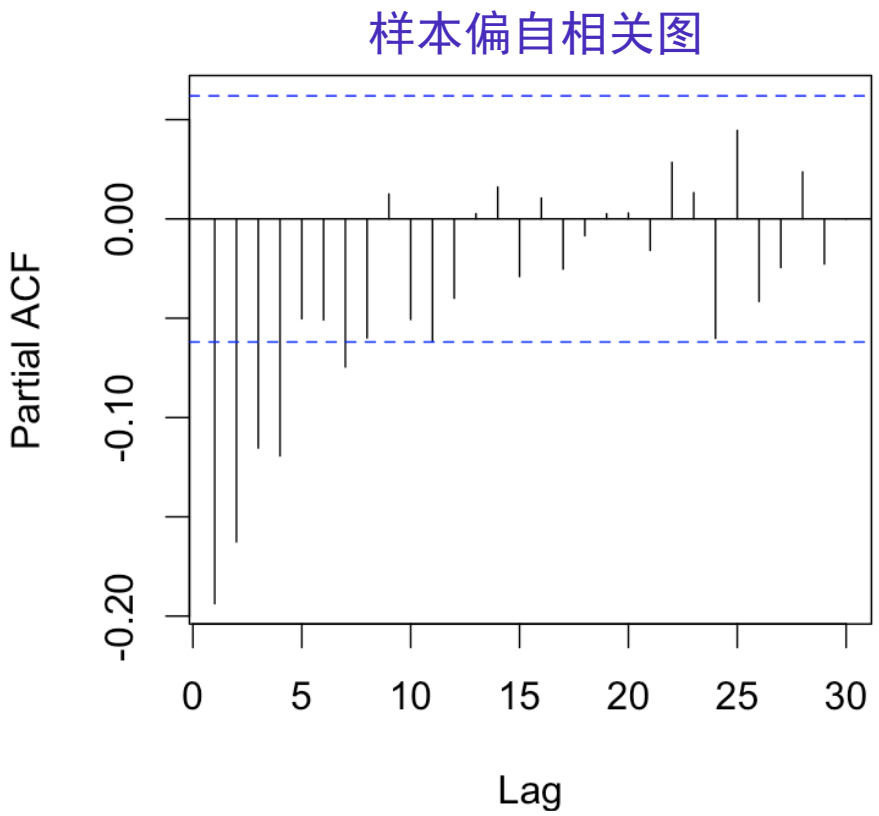
$$\begin{aligned}\gamma(1) &= \frac{\theta_1(1 - \phi_1^2) - \phi_1(1 + \theta_1^2 - 2\phi_1\theta_1)}{(1 - \phi_1^2)}\sigma_\varepsilon^2 \\ &= \frac{(1 - \theta_1\phi_1)(\theta_1 - \phi_1)}{(1 - \phi_1^2)}\sigma_e^2 \\ \gamma(k) &= -\phi_1 \gamma(k-1), k \geq 2\end{aligned}$$

考察ARMA模型的相关性

- 拟合模型$ARMA(1,1)$

$$Y_t - 0.5Y_{t-1} = \varepsilon_t - 0.8\varepsilon_{t-1}$$





ARMA模型相关性特征

模型	自相关函数	偏自相关函数	稳定、可逆性
$AR(p)$	拖尾	p阶截尾	总可逆， 当$\phi(\mathbf{B})=0$的根在单位圆之外时稳定
$MA(q)$	q阶截尾	拖尾	总稳定， 当$\theta(\mathbf{B})=0$的根在单位圆之外时可逆
$ARMA(p,q)$	拖尾	拖尾	当$\phi(\mathbf{B})=0$的根在单位圆之外时稳定； 当$\theta(\mathbf{B})=0$的根在单位圆之外时可逆

自相关函数

- $AR(1)$

$$(1-0.8\mathbf{B})Y_t=\varepsilon_t$$

- 指数衰减

- $MA(2)$

$$Y_t=\left(1+0.9\mathbf{B}+0.8\mathbf{B}^{\mathbf{2}}\right)\varepsilon_t$$

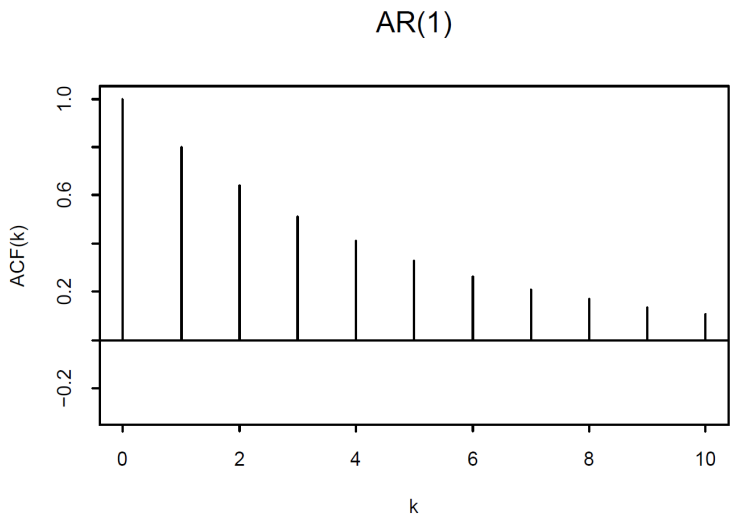
- Lag=2后为0

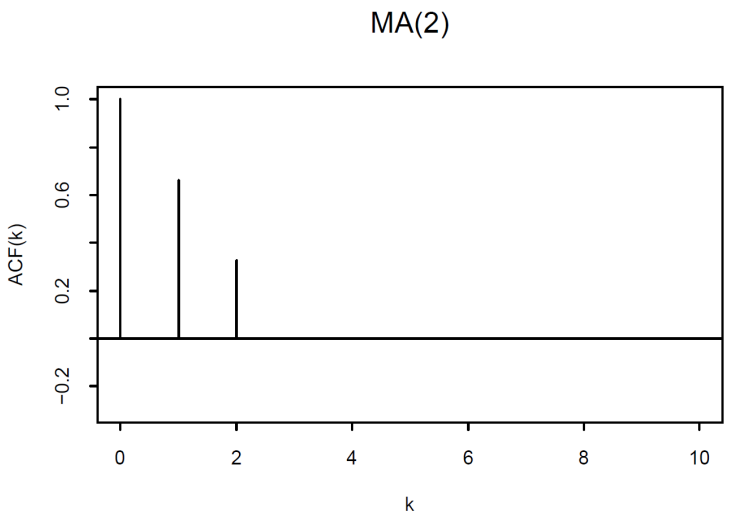
- $ARMA(1,2)$

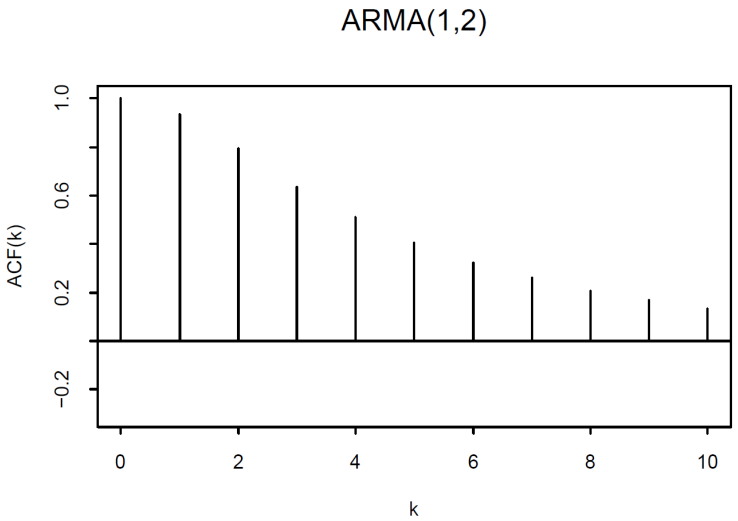
$$(1-0.8\mathbf{B})Y_t$$
$$=\left(1+0.9\mathbf{B}+0.8\mathbf{B}^{\mathbf{2}}\right)\varepsilon_t$$

Lag= $q+1-p=2+1-1$

$=2$后指数衰减







偏自相关函数

- $AR(1)$

$$(1-0.8\mathbf{B})Y_t=\varepsilon_t$$

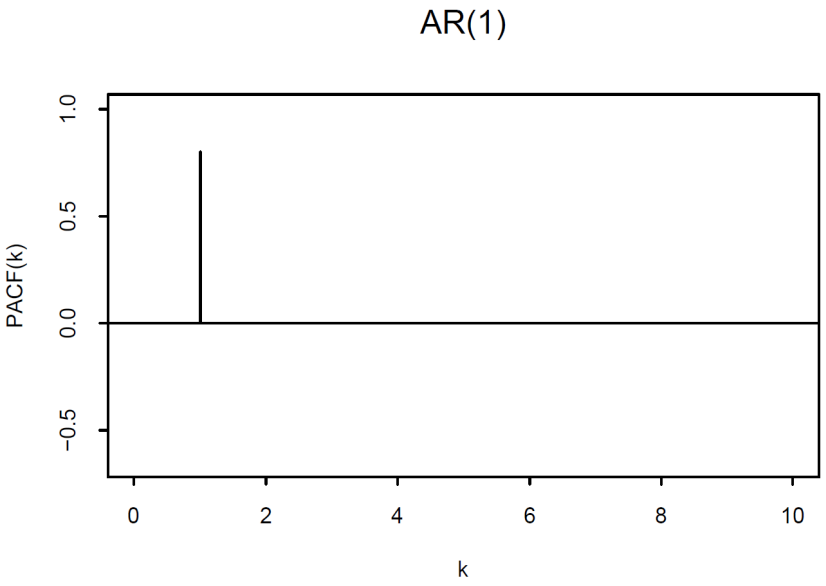
- Lag=1后为0

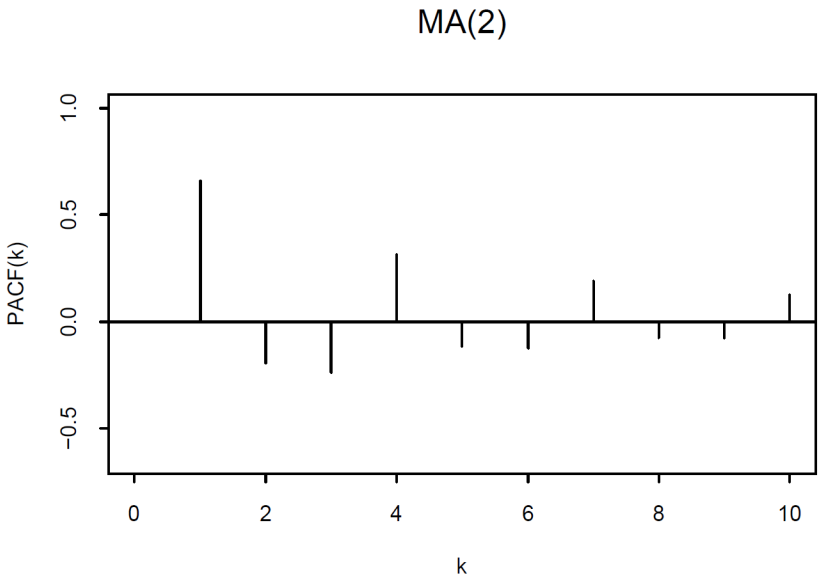
- $MA(2)$

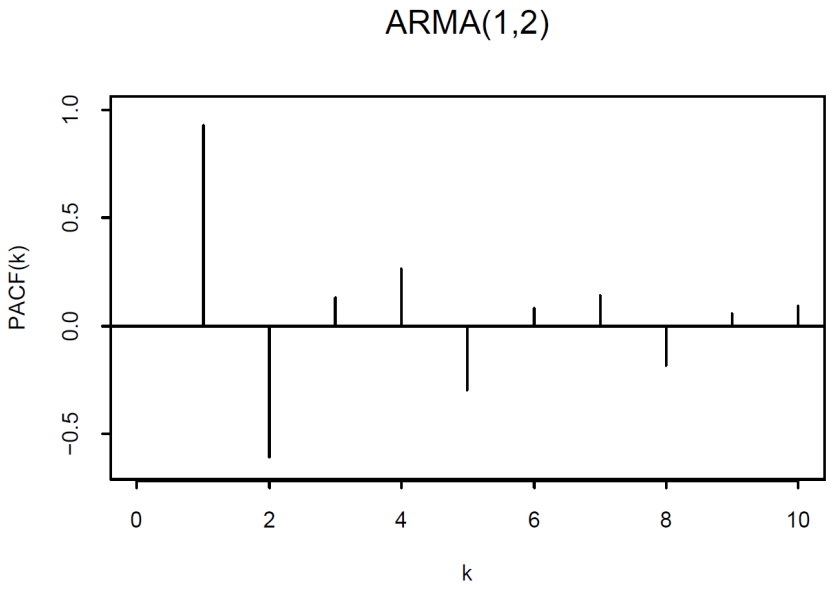
$$Y_t=\left(1+0.9\mathbf{B}+0.8\mathbf{B}^{\mathbf{2}}\right)\varepsilon_t$$

- $ARMA(1,2)$

$$(1-0.8\mathbf{B})Y_t$$
$$=\left(1+0.9\mathbf{B}+0.8\mathbf{B}^{\mathbf{2}}\right)\varepsilon_t$$







概要

1. AR模型

2. MA模型

3. AR模型思想的应用

4. ARMA模型：
(平稳时间序列模型)

5. ARIMA模型：
(非平稳时间序列模型)

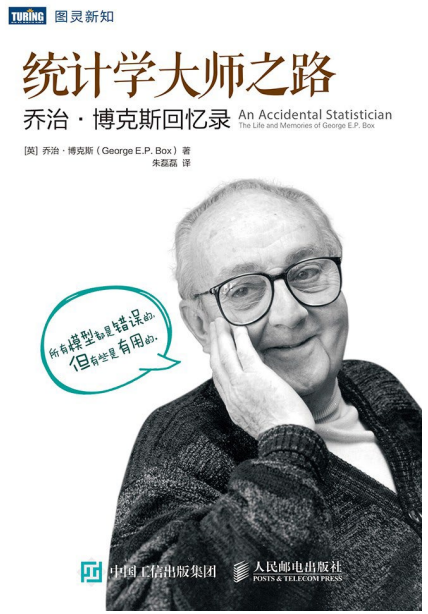
6. 模型预测

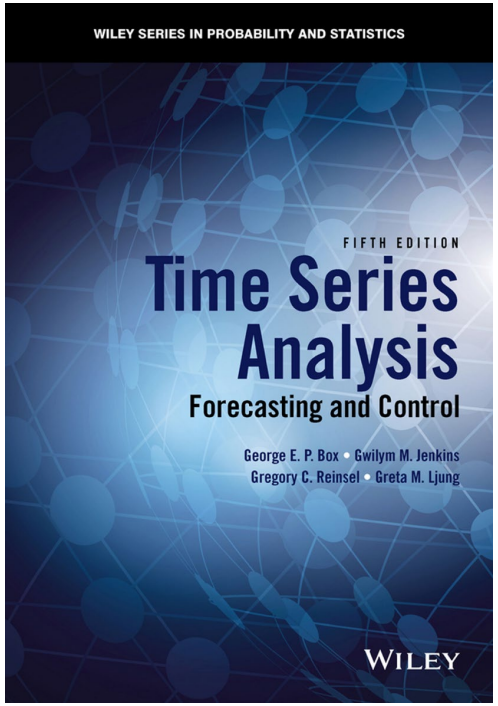
处理不平稳时间序列的ARIMA模型

- 1970年， Box和Jenkins出版了《Time Series Analysis Forecasting and Control》一书，将前人的知识进行了系统的梳理和分析，构造了**ARIMA**模型
- *“All models are wrong, but some are useful”*
- 为了记念Box和Jenkins对时间序列发展的特殊贡献，现在人们也常把ARIMA模型称为Box-Jenkins模型。
- ARIMA模型的实质
  - 单变量、同方差场合的线性模型

George Box
英国统计学家

Gwilym Jenkins
英国统计学家





时间序列中的不平稳性

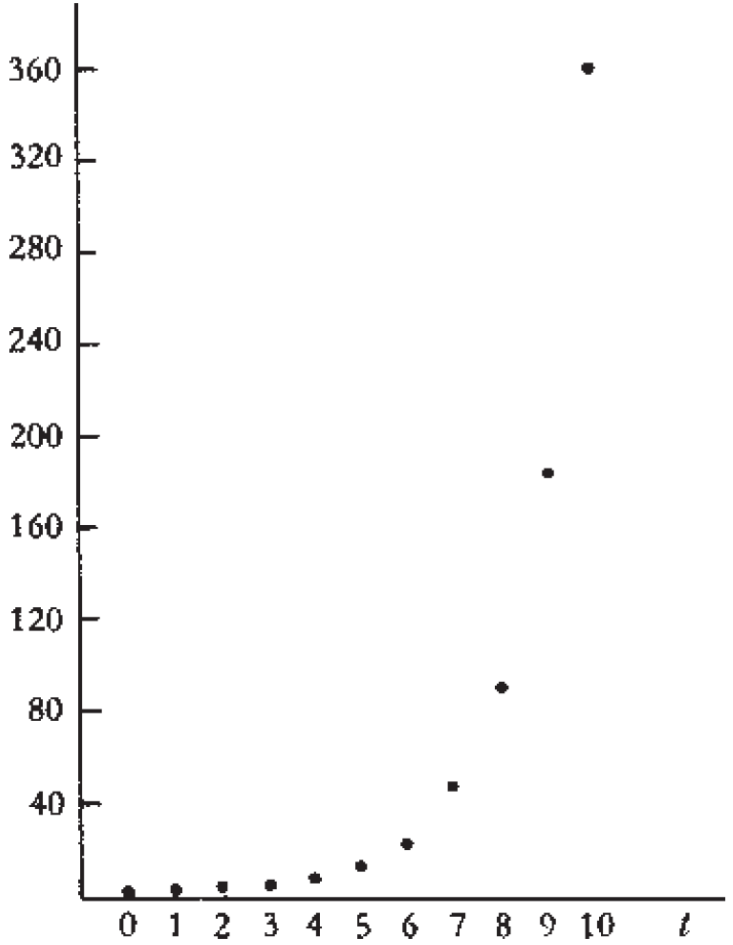
- 对于$AR(1)$模型，

$$(1-\phi\mathbf{B})Y_t=\varepsilon_t$$

- 平稳时有

$$|\phi|<1$$

- 考察“不平稳”的情形，令$\phi=2$
- 平稳：$\phi(\mathbf{B})=0$的根在单位圆**外**
- 不平稳（指数增长）：$\phi(\mathbf{B})=0$的根在单位圆**内**



- $\phi(\mathbf{B})=0$ 的根在单位圆**上**序列性质如何?

时间序列中的不平稳性

- 给定ARMA模型

$$\varphi(\mathbf{B})Y_t = \theta(\mathbf{B})\varepsilon_t$$

- 其中$\varphi(\mathbf{B})$是不平稳AR算子， $\varphi(\mathbf{B}) = 0$有d个根在单位圆**上**，其余根在单位圆外，因此有

$$\varphi(\mathbf{B})Y_t = \phi(\mathbf{B})(1-\mathbf{B})^d Y_t = \theta(\mathbf{B})\varepsilon_t$$

其中$\phi(\mathbf{B})$是平稳AR算子，$(1-\mathbf{B}) = \nabla$为差分算子，有$\phi(\mathbf{B})\nabla^d Y_t = \phi(\mathbf{B})w_t = \theta(\mathbf{B})\varepsilon_t$

- 考虑$w_t = \nabla^d Y_t$的逆过程，$Y_t = S^d w_t$，其中S为求和算子

$$\begin{aligned} Sx_t = \sum_{h=-\infty}^{t} x_h &= (1+\mathbf{B}+\mathbf{B}^2+\cdots)x_t \\ &= (1-\mathbf{B})^{-1}x_t = \nabla^{-1}x_t \end{aligned}$$

有$S = (1-\mathbf{B})^{-1} = \nabla^{-1}$

ARIMA模型结构

- ARIMA（Auto Regressive *Integrated* Moving Average）模型：针对**差分**后的序列使用ARMA模型建模

- $ARIMA(p,d,q)$模型结构

$$\begin{cases} \phi(\mathbf{B})\nabla^d Y_t = \phi(\mathbf{B})(1-\mathbf{B})^d Y_t = W_t = \theta(\mathbf{B})\varepsilon_t \\ E(\varepsilon_t) = 0,\ \ Var(\varepsilon_t) = \sigma_\varepsilon^2, E(\varepsilon_t\varepsilon_s) = 0, s \neq t \\ EY_s\varepsilon_t = 0, \qquad \forall s < t \end{cases}$$

其中， W_t为平稳可逆的$ARMA(p,q)$过程，Y_t通过对平稳可逆的W_t**求和**得到

$$d = 0,$$

$$ARIMA(p,0,q) = ARMA(p,q)$$

$$p = 0$$

$$ARIMA(0,d,q) = IMA(d,q)$$

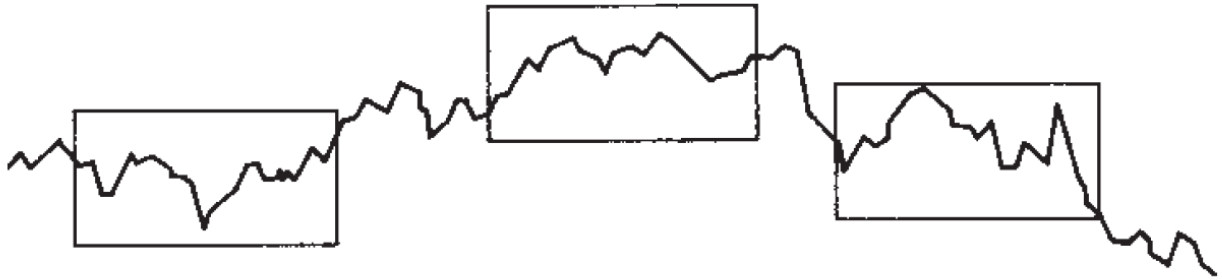
$$q = 0$$

$$ARIMA(p,d,0) = ARI(p,d)$$

$$d = 1, P = q = 0$$

$$ARIMA(p,d,q) \Rightarrow Y_t = Y_{t-1} + \varepsilon_t$$

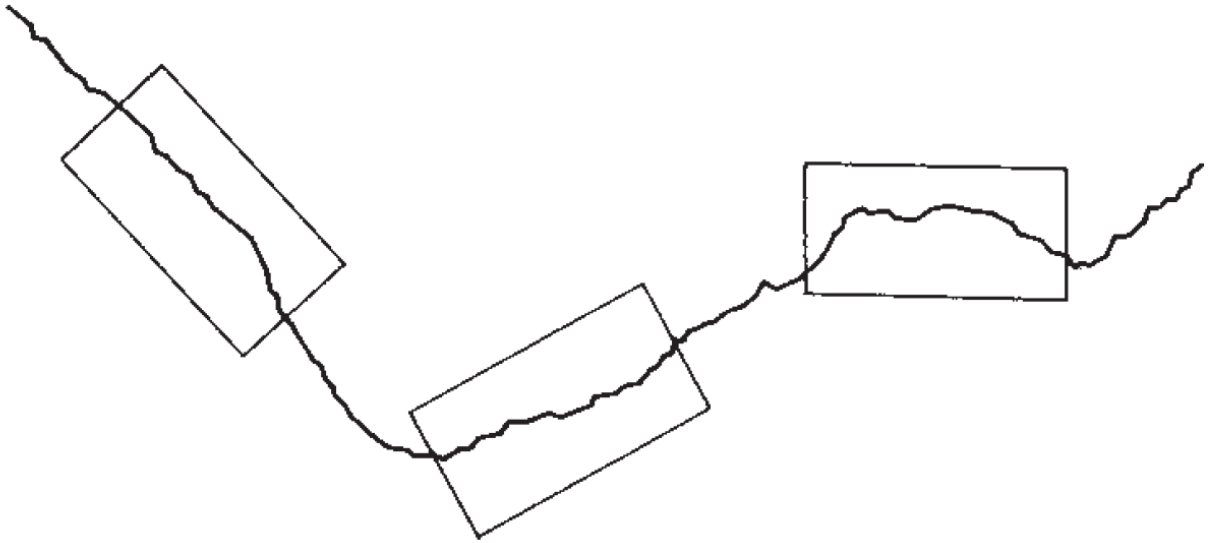
ARIMA模型



(a)

Non-stationarity in level

$$\phi(\mathbf{B})\nabla Y_t = \theta(\mathbf{B})\varepsilon_t$$



(b)

Non-stationarity in level and in slop

$$\phi(\mathbf{B})\nabla^2 Y_t = \theta(\mathbf{B})\varepsilon_t$$

ARIMA模型的平稳性

- $ARIMA(p,d,q)$模型共有$p+d$个特征根，其中p个在单位圆内，d个在单位圆上。所以当

$$d \neq 0$$

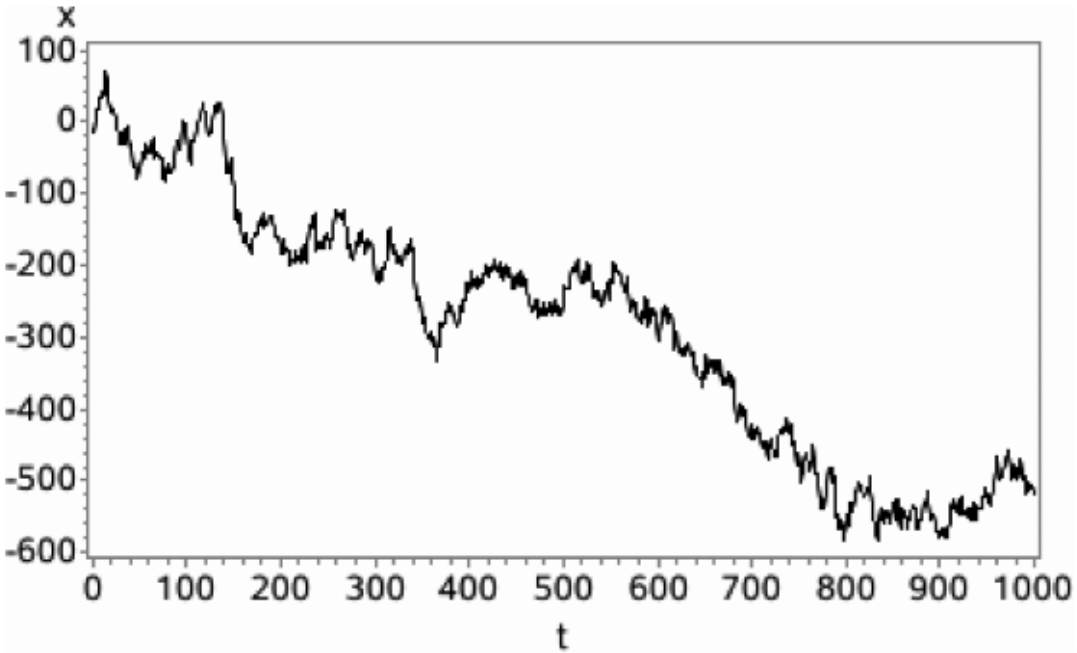
时$ARIMA(p,d,q)$模型非平稳。

- 随机游走（Random Walk）模型

$$\nabla Y_t = \varepsilon_t$$

- 是一个非平稳过程

$ARIMA(0,1,0)$拟合序列时序图



ARIMA模型的方差齐性

- 以$ARIMA(0,1,0)$为例
- $d \neq 0$时，原序列方差**非齐**性

$$
\begin{aligned}
Y_t &= Y_{t-1} + \varepsilon_t \\
&= Y_{t-2} + \varepsilon_t + \varepsilon_{t-1} \\
&\vdots \\
&= Y_0 + \varepsilon_t + \varepsilon_{t-1} + \cdots + \varepsilon_1
\end{aligned}
$$

$$
\begin{aligned}
\mathrm{Var}(Y_t) &= \mathrm{Var}(Y_0 + \varepsilon_t + \varepsilon_{t-1} + \cdots + \varepsilon_1) \\
&= t\sigma_\varepsilon^2
\end{aligned}
$$

- d阶差分后，差分后序列

$$\nabla Y_t = \varepsilon_t$$

- 差分后序列方差齐性

$$Var(\nabla Y_t) = \sigma_\varepsilon^2$$

概要

1. AR模型

2. MA模型

3. AR模型思想的应用

4. ARMA模型：
(平稳时间序列模型)

5. ARIMA模型：
(非平稳时间序列模型)

6. 模型预测

时间序列预测

- 在t时刻有观察序列$Y_t, Y_{t-1}, Y_{t-2}, Y_{t-3}, \dots$
  - 插值（interpolation）
  - 预测（prediction）：针对Y_{t+k}进行预测（$k \geq 1$）
- 预测方法

$$\hat{Y}_{t+k|t} = \mathrm{g}(Y_t, Y_{t-1}, \dots)$$

- 预测准则：最小化均方误差

$$\mathrm{E}\left[\left(Y_{t+k} - \mathrm{g}(Y_t, Y_{t-1}, \dots)\right)^2 \mid Y_t, Y_{t-1}, \dots\right]$$

- 基于预测准则，有

$$\mathrm{g}(Y_t, Y_{t-1}, \dots) = \mathrm{E}[Y_{t+k} \mid Y_t, Y_{t-1}, \dots]$$

- 使用条件期望进行最优预测

$$\hat{Y}_{t+k|t} = \mathrm{E}[Y_{t+k} \mid Y_t, Y_{t-1}, \dots]$$

- 预测误差

$$e_{t+k|t} = Y_{t+k} - \hat{Y}_{t+k|t}$$

最小化预测误差的方差

- 难点：条件期望难以计算

ARIMA模型的预测（ψ加权形式）

- 对于$ARIMA(p,d,q)$过程$\{Y_t\}$，有

$$\phi(B)\nabla^d Y_t = \theta(B)\varepsilon_t$$

$\{\varepsilon_t\}$的方差为σ^2

- 转化为ψ加权形式（MA）

$$Y_t = \varepsilon_t + \psi_1\varepsilon_{t-1} + \psi_2\varepsilon_{t-2} + \cdots$$

- 对于π加权形式（AR），有

$$\varepsilon_t = Y_t + \pi_1 Y_{t-1} + \pi_2 Y_{t-2} + \cdots$$

- 由π加权，当

$$\{Y_s, s = t, t-1, \dots\}$$

已知时

$$\{\varepsilon_s, s = t, t-1, \dots\}$$

也已知。由ψ加权，反之也成立。$\{\varepsilon_t\}$为白噪声，有

$$\mathrm{E}[\varepsilon_{t+k} \mid Y_t, Y_{t-1}, \dots] = \begin{cases} \varepsilon_{t+k} & \text{for } k \le 0 \\ 0 & \text{for } k > 0 \end{cases}$$

历史误差（条件期望）可计算（为预测残差），未来误差期望为0

- 且对于Y_{t+k}有

$$Y_{t+k} = \varepsilon_{t+k} + \psi_1\varepsilon_{t+k-1} + \cdots + \psi_k\varepsilon_t + \psi_{k+1}\varepsilon_{t-1} + \cdots$$

- 计算条件均值

$$\hat{Y}_{t+k|t} = \mathrm{E}[Y_{t+k} \mid Y_t, Y_{t-1}, \dots] = \psi_k\varepsilon_t + \psi_{k+1}\varepsilon_{t-1} + \cdots$$

ARIMA模型的预测（ψ加权形式）

- 且对于Y_{t+k}有

$$Y_{t+k} = \varepsilon_{t+k} + \psi_1 \varepsilon_{t+k-1} + \cdots + \psi_k \varepsilon_t + \psi_{k+1} \varepsilon_{t-1} + \cdots$$

- 计算条件均值

$$\hat{Y}_{t+k|t} = \mathrm{E}[Y_{t+k} \mid Y_t, Y_{t-1}, \ldots] = \psi_k \varepsilon_t + \psi_{k+1} \varepsilon_{t-1} + \cdots$$

- 预测误差

$$e_{t+k|t} = Y_{t+k} - \hat{Y}_{t+k|t} = \varepsilon_{t+k} + \psi_1 \varepsilon_{t+k-1} + \cdots + \psi_{k-1} \varepsilon_{t+1}$$

- 预测误差的方差为

$$\sigma_{k=}^2 = \left(1 + \psi_1^2 + \cdots + \psi_{k-1}^2\right) \sigma_\varepsilon^2$$

- 若$\{\varepsilon_t\}$为正态分布，$(1-\alpha)$置信度的区间估计为

$$\hat{Y}_{t+k|t} \pm u_{\alpha/2} \sigma_k = \hat{Y}_{t+k|t} \pm u_{\alpha/2} \sigma_\varepsilon \sqrt{1 + \psi_1^2 + \cdots + \psi_{k-1}^2}$$

修正预测

- 已知

$$Y_{t+k} = \varepsilon_{t+k} + \psi_1 \varepsilon_{t+k-1} + \cdots + \psi_k \varepsilon_t + \psi_{k+1} \varepsilon_{t-1} + \cdots$$

且

$$\mathrm{E}[\varepsilon_{t+k} \mid Y_t, Y_{t-1}, \ldots] = \begin{cases} \varepsilon_{t+k} & \text{for } k \leq 0 \\ 0 & \text{for } k > 0 \end{cases}$$

- 因此，新获得Y_{t+1}，利用新的信息去获得精度更高的预测值

$$\hat{Y}_{t+k|t+1} = \psi_{k-1} \varepsilon_{t+1} + \psi_k \varepsilon_t + \psi_{k+1} \varepsilon_{t-1} + \cdots = \psi_{k-1} \varepsilon_{t+1} + \hat{Y}_{t+k|t}$$

用于更新预测值（此时ε_{t+1}已知），针对**差分方程**计算条件均值

$$Y_t + \varphi_1 Y_{t-1} + \cdots + \varphi_{p+d} Y_{t-p-d} = \varepsilon_t + \theta_1 \varepsilon_{t-1} + \cdots + \theta_q \varepsilon_{t-q}$$

ARIMA模型的预测（差分方程方法）

- $ARMA(p,q)$模型

$$Y_{t+k}+\phi_1 Y_{t+k-1}+\cdots+\phi_{p+d}Y_{t+k-p}=$$
$$\varepsilon_{t+k}+\theta_1\varepsilon_{t+k-1}+\cdots+\theta_q\varepsilon_{t+k-q}$$

- 两边同时取条件期望

$$\hat{Y}_{t+k|t}=-\phi_1\hat{Y}_{t+k-1|t}-\cdots-\phi_p\hat{Y}_{t+k-p|t}$$
$$+\hat{\varepsilon}_{t+k|t}+\theta_1\hat{\varepsilon}_{t+k-1|t}+\cdots+\theta_q\hat{\varepsilon}_{t+k-q|t}$$

得到递归公式

- 对于ARIMA模型，通过$\varphi(\mathrm{B})=\phi(B)\nabla^d$计算系数$\varphi_1,\ldots,\varphi_{p+d}$，转化为ARMA处理

ARIMA模型的预测（差分方程方法）

- k步预测

$$
\begin{aligned}
&Y_{t+k} + \varphi_1 Y_{t+k-1} + \cdots + \varphi_{p+d} Y_{t+k-p-d} \\
&\quad = \varepsilon_{t+k} + \theta_1 \varepsilon_{t+k-1} + \cdots + \theta_q \varepsilon_{t \div k-4}
\end{aligned}
$$

- 计算条件期望

$$
\begin{aligned}
\hat{Y}_{t+k|t} = &-\varphi_1 \mathrm{E}[Y_{t+k-1} \mid Y_t, Y_{t-1}, \dots] - \cdots \\
&-\varphi_{p+d} \mathrm{E}\left[Y_{t+k-p-d} \mid Y_t, Y_{t-1}, \dots\right] \\
&+\mathrm{E}[\varepsilon_{t+k} \mid Y_t, Y_{t-1}, \dots] + \theta_1 \mathrm{E}[\varepsilon_{t+k-1} \mid Y_t, Y_{t-1}, \dots] + \cdots \\
&+\theta_q \mathrm{E}\left[\varepsilon_{t+k-q} \mid Y_t, Y_{t-1}, \dots\right]
\end{aligned}
$$

其中

$$
\begin{aligned}
&\mathrm{E}\left[Y_{t-j} \mid Y_t, Y_{t-1}, \dots\right] = Y_{t-j} && j = 0,1,2,\dots \\
&\mathrm{E}\left[Y_{t+j} \mid Y_t, Y_{t-1}, \dots\right] = \hat{Y}_{t+j|t} && j = 1,2,\dots \\
&\mathrm{E}\left[\varepsilon_{t-j} \mid Y_t, Y_{t-1}, \dots\right] = \varepsilon_{t-j} && j = 0,1,2,\dots \\
&\mathrm{E}\left[\varepsilon_{t+j} \mid Y_t, Y_{t-1}, \dots\right] = 0 && j = 1,2,\dots
\end{aligned}
$$

$AR(1)$模型的预测

- 模型：

$$Y_{t+1} = \phi Y_t + \varepsilon_{t+1}$$

- 1步预测

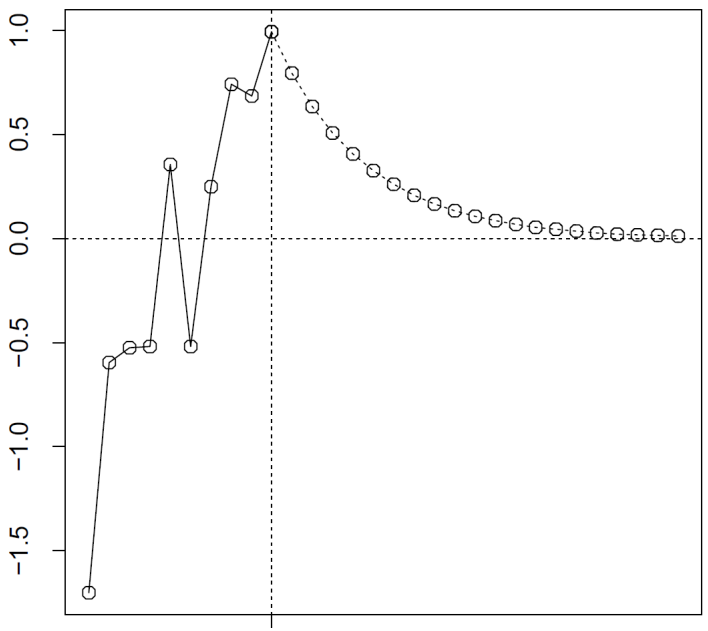
$$\hat{Y}_{t+1|t} = E[Y_{t+1} \mid Y_t, Y_{t-1}, \cdots]$$
$$= E[\phi Y_t + \varepsilon_{t+1} \mid Y_t, Y_{t-1}, \cdots] = \phi Y_t + 0 = \phi Y_t$$

- 2步预测

$$\hat{Y}_{t+2|t} = E[Y_{t+2} \mid Y_t, Y_{t-1}, \cdots]$$
$$= E[\phi Y_{t+1} + \varepsilon_{t+2} \mid Y_t, Y_{t-1}, \cdots]$$
$$= \phi \hat{Y}_{t+1|t} + 0 = \phi^2 Y_t$$

- k步预测

$$\hat{Y}_{t+k|t} = \phi^k Y_t$$



$$Y_t = 0.8Y_{t-1} + \varepsilon_t$$

$AR(1)$模型预测误差的方差

- 预测误差

$$e_{t+k|t} = Y_{t+k} - \hat{Y}_{t+k|t} = Y_{t+k} - \phi^k Y_t$$

则

$$\begin{aligned}
Y_{t+k} &= \phi Y_{t+k-1} + \varepsilon_{t+k} \\
&= \phi(\phi Y_{t+k-2} + \varepsilon_{t+k-1}) + \varepsilon_{t+k} \\
&= \phi^2 Y_{t+k-2} + \phi\varepsilon_{t+k-1} + \varepsilon_{t+k} \\
&= \phi^2(\phi Y_{t+k-3} + \varepsilon_{t+k-2}) + \phi\varepsilon_{t+k-1} + \varepsilon_{t+k} \\
&= \phi^3 Y_{t+k-3} + \phi^2\varepsilon_{t+k-2} + \phi\varepsilon_{t+k-1} + \varepsilon_{t+k} \\
&\vdots \\
&= \phi^k Y_t + \phi^{k-1}\varepsilon_{t+1} + \phi^{k-2}\varepsilon_{t+2} + \cdots + \phi\varepsilon_{t+k-1} + \varepsilon_{t+k}
\end{aligned}$$

- 预测误差的方差

$$\begin{aligned}
V[e_{t+k|t}] &= V[\phi^{k-1}\varepsilon_{t+1} + \phi^{k-2}\varepsilon_{t+2} + \cdots + \phi\varepsilon_{t+k-1} + \varepsilon_{t+k}] \\
&= \left(\phi^{2(k-1)} + \phi^{2(k-2)} + \cdots + \phi^2 + 1\right)\sigma_\varepsilon^2
\end{aligned}$$

$(1-\alpha)\times 100\%$区间预测

$$\hat{Y}_{t+k|t} \pm u_{\alpha/2}\sqrt{V[e_{t+k|t}]}$$

ARIMA(3,1,1)预测

- ARIMA(3,1,1)模型

$$(1-\phi_1\mathbf{B}-\phi_2\mathbf{B}^2-\phi_3\mathbf{B}^3)(1-\mathbf{B})y_t=(1+\theta_1\mathbf{B})\varepsilon_t$$

- 左侧展开

$$[1-(1+\phi_1)\mathbf{B}+(\phi_1-\phi_2)\mathbf{B}^2+(\phi_2-\phi_3)\mathbf{B}^3+\phi_3\mathbf{B}^4]y_t=(1+\theta_1\mathbf{B})\varepsilon_t$$

- 基于**B**的性质

$$y_t-(1+\phi_1)y_{t-1}+(\phi_1-\phi_2)y_{t-2}+(\phi_2-\phi_3)y_{t-3}+\phi_3y_{t-4}=\varepsilon_t+\theta_1\varepsilon_{t-1}$$

- 得到

$$y_t=(1+\phi_1)y_{t-1}-(\phi_1-\phi_2)y_{t-2}-(\phi_2-\phi_3)y_{t-3}-\phi_3y_{t-4}+\varepsilon_t+\theta_1\varepsilon_{t-1}$$

- 则$T+1$步预测

$$y_{T+1}=(1+\phi_1)y_T-(\phi_1-\phi_2)y_{T-1}-(\phi_2-\phi_3)y_{T-2}-\phi_3y_{T-3}+\varepsilon_{T+1}+\theta_1\varepsilon_T$$

ARIMA(3,1,1)预测

- $T+1$步预测为

$$y_{T+1} = (1+\phi_1)y_T - (\phi_1 - \phi_2)y_{T-1} - (\phi_2 - \phi_3)y_{T-2} - \phi_3 y_{T-3} + \varepsilon_{T+1} + \theta_1 \varepsilon_T$$

- 等式两边计算条件期望

$$\mathbb{E}[\varepsilon_{T+1} | Y_t, Y_{t-1}, \dots] = 0$$

$$\mathbb{E}[\varepsilon_T | Y_t, Y_{t-1}, \dots] = \varepsilon_T$$

$$\hat{y}_{T+1|T} = (1+\phi_1)y_T - (\phi_1 - \phi_2)y_{T-1} - (\phi_2 - \phi_3)y_{T-2} - \phi_3 y_{T-3} + \theta_1 \varepsilon_T$$

代入观察值和T步的预测残差

- $T+2$步预测

$$\hat{y}_{T+2|T} = (1+\phi_1)\hat{y}_{T+1|T} - (\phi_1 - \phi_2)y_T - (\phi_2 - \phi_3)y_{T-1} - \phi_3 y_{T-2}$$