应用时间序列分析
课程介绍

主讲老师：席瑞斌

本课件基于李东风老师课件修改

2017年秋季学期

ARMA模型参数估计

- 对时间序列观测数据建立模型可以更深入了解数据的内在规律，也有利于预报。
- 估计模型的一般指导原则是其他条件相同时较简单的模型更有利。
- 建立模型，要兼顾对观测到的数据的拟合情况和模型的可推广性。
- 下面讨论相合性时，由§1.5定理5.2可知，ARMA模型中白噪声项如果是独立同分布的则该平稳列为严平稳遍历。

ARMA模型参数估计

- 对时间序列观测数据建立模型可以更深入了解数据的内在规律，也有利于预报。
- 估计模型的一般指导原则是其他条件相同时较简单的模型更有利。
- 建立模型，要兼顾对观测到的数据的拟合情况和模型的可推广性。
- 下面讨论相合性时，由§1.5定理5.2可知，ARMA模型中白噪声项如果是独立同分布的则该平稳列为严平稳遍历。

ARMA模型参数估计

- 对时间序列观测数据建立模型可以更深入了解数据的内在规律，也有利于预报。
- 估计模型的一般指导原则是其他条件相同时较简单的模型更有利。
- 建立模型，要兼顾对观测到的数据的拟合情况和模型的可推广性。
- 下面讨论相合性时，由§1.5定理5.2可知，ARMA模型中白噪声项如果是独立同分布的则该平稳列为严平稳遍历。

ARMA模型参数估计

- 对时间序列观测数据建立模型可以更深入了解数据的内在规律，也有利于预报。
- 估计模型的一般指导原则是其他条件相同时较简单的模型更有利。
- 建立模型，要兼顾对观测到的数据的拟合情况和模型的可推广性。
- 下面讨论相合性时，由§1.5定理5.2可知，ARMA模型中白噪声项如果是独立同分布的则该平稳列为严平稳遍历。

Yule-Walker估计

- 对AR(p)模型，需要估计系数$a_1, a_2, \ldots, a_p$和白噪声方差σ^2。
- 设$x_1, x_2, \ldots, x_N$为观测数据，建模前一般先中心化:
$$y_t = x_t - \bar{x}_N, \quad t = 1, 2, \ldots, N$$
- 假定$\{y_1, y_2, \ldots, y_N\}$是某个AR($p$)序列的观测样本。先假定$p$已知。
- 如果序列的理论自协方差函数$\{\gamma_k\}$已知则由Yule-Walker方程可以解出$a_1, \ldots, a_p$和σ^2，见P.68定理3.3.
- 从观测样本可以估计样本自协方差$\hat{\gamma}_k, k = 0, 1, \ldots, p$。
- 在Y-W方程中用$\hat{\gamma}_k$代替γ_k可以解出$\hat{a}_1, \ldots, \hat{a}_p, \hat{\sigma}^2$，由§4.2 A的讨论知如果$y_1, \ldots, y_N$不全相同则$\hat{\Gamma}_{p+1}$正定，Y-W方程的解存在唯一。
- 由§2.4定理4.1可知这样解出的$\hat{a}_1, \ldots, \hat{a}_p$满足最小相位条件。

Yule-Walker估计

- 对AR(p)模型，需要估计系数$a_1, a_2, \ldots, a_p$和白噪声方差σ^2。
- 设$x_1, x_2, \ldots, x_N$为观测数据，建模前一般先中心化:

$$y_t = x_t - \bar{x}_N, \quad t = 1, 2, \ldots, N$$

- 假定$\{y_1, y_2, \ldots, y_N\}$是某个AR($p$)序列的观测样本。先假定$p$已知。
- 如果序列的理论自协方差函数$\{\gamma_k\}$已知则由Yule-Walker方程可以解出$a_1, \ldots, a_p$和σ^2，见P.68定理3.3.
- 从观测样本可以估计样本自协方差$\hat{\gamma}_k, k = 0, 1, \ldots, p$。
- 在Y-W方程中用$\hat{\gamma}_k$代替γ_k可以解出$\hat{a}_1, \ldots, \hat{a}_p, \hat{\sigma}^2$，由§4.2 A的讨论知如果$y_1, \ldots, y_N$不全相同则$\hat{\Gamma}_{p+1}$正定，Y-W方程的解存在唯一。
- 由§2.4定理4.1可知这样解出的$\hat{a}_1, \ldots, \hat{a}_p$满足最小相位条件。

Yule-Walker估计

- 对AR(p)模型，需要估计系数$a_1, a_2, \ldots, a_p$和白噪声方差σ^2。
- 设$x_1, x_2, \ldots, x_N$为观测数据，建模前一般先中心化:

$$y_t = x_t - \bar{x}_N, \quad t = 1, 2, \ldots, N$$

- 假定$\{y_1, y_2, \ldots, y_N\}$是某个AR($p$)序列的观测样本。先假定$p$已知。
- 如果序列的理论自协方差函数$\{\gamma_k\}$已知则由Yule-Walker方程可以解出$a_1, \ldots, a_p$和σ^2，见P.68定理3.3.
- 从观测样本可以估计样本自协方差$\hat{\gamma}_k, k = 0, 1, \ldots, p$。
- 在Y-W方程中用$\hat{\gamma}_k$代替γ_k可以解出$\hat{a}_1, \ldots, \hat{a}_p, \hat{\sigma}^2$，由§4.2 A的讨论知如果$y_1, \ldots, y_N$不全相同则$\hat{\Gamma}_{p+1}$正定，Y-W方程的解存在唯一。
- 由§2.4定理4.1可知这样解出的$\hat{a}_1, \ldots, \hat{a}_p$满足最小相位条件。

Yule-Walker估计

- 对AR(p)模型，需要估计系数$a_1, a_2, \ldots, a_p$和白噪声方差σ^2。
- 设$x_1, x_2, \ldots, x_N$为观测数据，建模前一般先中心化:

$$y_t = x_t - \bar{x}_N, \quad t = 1, 2, \ldots, N$$

- 假定$\{y_1, y_2, \ldots, y_N\}$是某个AR($p$)序列的观测样本。先假定$p$已知。
- 如果序列的理论自协方差函数$\{\gamma_k\}$已知则由Yule-Walker方程可以解出$a_1, \ldots, a_p$和σ^2，见P.68定理3.3.
- 从观测样本可以估计样本自协方差$\hat{\gamma}_k, k = 0, 1, \ldots, p$。
- 在Y-W方程中用$\hat{\gamma}_k$代替γ_k可以解出$\hat{a}_1, \ldots, \hat{a}_p, \hat{\sigma}^2$，由§4.2 A的讨论知如果$y_1, \ldots, y_N$不全相同则$\hat{\Gamma}_{p+1}$正定，Y-W方程的解存在唯一。
- 由§2.4定理4.1可知这样解出的$\hat{a}_1, \ldots, \hat{a}_p$满足最小相位条件。

Yule-Walker估计

- 对AR(p)模型，需要估计系数$a_1, a_2, \ldots, a_p$和白噪声方差σ^2。
- 设$x_1, x_2, \ldots, x_N$为观测数据，建模前一般先中心化:

$$y_t = x_t - \bar{x}_N, \quad t = 1, 2, \ldots, N$$

- 假定$\{y_1, y_2, \ldots, y_N\}$是某个AR($p$)序列的观测样本。先假定$p$已知。
- 如果序列的理论自协方差函数$\{\gamma_k\}$已知则由Yule-Walker方程可以解出$a_1, \ldots, a_p$和σ^2，见P.68定理3.3.
- 从观测样本可以估计样本自协方差$\hat{\gamma}_k, k = 0, 1, \ldots, p$。
- 在Y-W方程中用$\hat{\gamma}_k$代替γ_k可以解出$\hat{a}_1, \ldots, \hat{a}_p, \hat{\sigma}^2$，由§4.2 A的讨论知如果$y_1, \ldots, y_N$不全相同则$\hat{\Gamma}_{p+1}$正定，Y-W方程的解存在唯一。
- 由§2.4定理4.1可知这样解出的$\hat{a}_1, \ldots, \hat{a}_p$满足最小相位条件。

Yule-Walker估计

- 对AR(p)模型，需要估计系数$a_1, a_2, \ldots, a_p$和白噪声方差σ^2。
- 设$x_1, x_2, \ldots, x_N$为观测数据，建模前一般先中心化:

$$y_t = x_t - \bar{x}_N, \quad t = 1, 2, \ldots, N$$

- 假定$\{y_1, y_2, \ldots, y_N\}$是某个AR($p$)序列的观测样本。先假定$p$已知。
- 如果序列的理论自协方差函数$\{\gamma_k\}$已知则由Yule-Walker方程可以解出$a_1, \ldots, a_p$和σ^2, 见P.68定理3.3.
- 从观测样本可以估计样本自协方差$\hat{\gamma}_k, k = 0, 1, \ldots, p$。
- 在Y-W方程中用$\hat{\gamma}_k$代替γ_k可以解出$\hat{a}_1, \ldots, \hat{a}_p, \hat{\sigma}^2$，由§4.2 A的讨论知如果$y_1, \ldots, y_N$不全相同则$\hat{\Gamma}_{p+1}$正定，Y-W方程的解存在唯一。
- 由§2.4定理4.1可知这样解出的$\hat{a}_1, \ldots, \hat{a}_p$满足最小相位条件。

Yule-Walker估计

- 对AR(p)模型，需要估计系数$a_1, a_2, \ldots, a_p$和白噪声方差σ^2。
- 设$x_1, x_2, \ldots, x_N$为观测数据，建模前一般先中心化:
$$y_t = x_t - \bar{x}_N, \quad t = 1, 2, \ldots, N$$
- 假定$\{y_1, y_2, \ldots, y_N\}$是某个AR($p$)序列的观测样本。先假定$p$已知。
- 如果序列的理论自协方差函数$\{\gamma_k\}$已知则由Yule-Walker方程可以解出$a_1, \ldots, a_p$和σ^2, 见P.68定理3.3.
- 从观测样本可以估计样本自协方差$\hat{\gamma}_k, k = 0, 1, \ldots, p$。
- 在Y-W方程中用$\hat{\gamma}_k$代替γ_k可以解出$\hat{a}_1, \ldots, \hat{a}_p, \hat{\sigma}^2$，由§4.2 A的讨论知如果$y_1, \ldots, y_N$不全相同则$\hat{\Gamma}_{p+1}$正定，Y-W方程的解存在唯一。
- 由§2.4定理4.1可知这样解出的$\hat{a}_1, \ldots, \hat{a}_p$满足最小相位条件。

Yule-Walker估计(续)

- 为了避免计算$\hat{\Gamma}_p$的逆可以用Levinson递推公式解Y-W方程得到$\hat{a}_1, \ldots, \hat{a}_p$和$\hat{\sigma}^2$。
- 这样得到的AR(p)参数估计叫做Y-W估计或Levinson估计。
- 优点是保证最小相位性，计算简单。

Yule-Walker估计(续)

- 为了避免计算$\hat{\Gamma}_p$的逆可以用Levinson递推公式解Y-W方程得到$\hat{a}_1, \ldots, \hat{a}_p$和$\hat{\sigma}^2$。
- 这样得到的AR(p)参数估计叫做Y-W估计或Levinson估计。
- 优点是保证最小相位性，计算简单。

Yule-Walker估计(续)

- 为了避免计算$\hat{\Gamma}_p$的逆可以用Levinson递推公式解Y-W方程得到$\hat{a}_1, \ldots, \hat{a}_p$和$\hat{\sigma}^2$。
- 这样得到的AR(p)参数估计叫做Y-W估计或Levinson估计。
- 优点是保证最小相位性，计算简单。

Y-W估计的相合性

- 由§2.3定理3.5知AR(p)的Γ_p正定，所以$a_1, \ldots, a_p, \sigma^2$是$\gamma_0, \gamma_1, \ldots, \gamma_p$的连续函数。
- 如果$\hat{\gamma}_k$是γ_k的强相合估计，则$\hat{a}_1, \ldots, \hat{a}_p, \hat{\sigma}^2$也是相应参数的强相合估计。
- 要使$\hat{\gamma}_k$强相合，由§4.2定理2.1, 一个充分条件是白噪声$\{\varepsilon_t\}$独立同分布。这时ARMA模型严平稳遍历。

Y-W估计的相合性

- 由§2.3定理3.5知AR(p)的Γ_p正定，所以$a_1, \ldots, a_p, \sigma^2$是$\gamma_0, \gamma_1, \ldots, \gamma_p$的连续函数。
- 如果$\hat{\gamma}_k$是γ_k的强相合估计，则$\hat{a}_1, \ldots, \hat{a}_p, \hat{\sigma}^2$也是相应参数的强相合估计。
- 要使$\hat{\gamma}_k$强相合，由§4.2定理2.1, 一个充分条件是白噪声$\{\varepsilon_t\}$独立同分布。这时ARMA模型严平稳遍历。

Y-W估计的相合性

- 由§2.3定理3.5知AR(p)的Γ_p正定，所以$a_1, \ldots, a_p, \sigma^2$是$\gamma_0, \gamma_1, \ldots, \gamma_p$的连续函数。
- 如果$\hat{\gamma}_k$是γ_k的强相合估计，则$\hat{a}_1, \ldots, \hat{a}_p, \hat{\sigma}^2$也是相应参数的强相合估计。
- 要使$\hat{\gamma}_k$强相合，由§4.2定理2.1，一个充分条件是白噪声$\{\varepsilon_t\}$独立同分布。这时ARMA模型严平稳遍历。

Y-W估计的相合性(续)

- **定理1.1** 如果AR(p)模型(1.1)中的$\{\varepsilon_t\}$是独立同分布的WN$(0,\sigma^2)$，E$\varepsilon_t^4<\infty$，则当$N\to\infty$时
  1. $\hat\sigma^2\to\sigma^2$, $\hat a_j\to a_j$, a.s., $1\le j\le p$.
  2. $\sqrt{N}(\hat a_1-a_1,\dots,\hat a_p-a_p)^T$ 依分布收敛到p维正态分布$N(0,\sigma^2\Gamma_p^{-1})$。
  3. $\sqrt{N}\sup\limits_{1\le j\le p}|\hat a_j-a_j|=O(\sqrt{\ln\ln N})$, a.s.,
     $\sqrt{N}|\hat\sigma^2-\sigma^2|=O(\sqrt{\ln\ln N})$, a.s.
- 证明:
  1. 由§4.2定理2.1知$\hat\gamma_k$强相合，再根据a_k,σ^2为$\{\gamma_k\}$的连续函数知$\hat a_k,\hat\sigma^2$强相合。
  2. 证明略(详见Brockwell and Davis的教材)。
  3. 证明略。

Y-W估计的相合性(续)

- **定理1.1** 如果AR(p)模型(1.1)中的$\{\varepsilon_t\}$是独立同分布的WN$(0,\sigma^2)$，$\mathrm{E}\varepsilon_t^4<\infty$，则当$N\to\infty$时
  1. $\hat\sigma^2\to\sigma^2$, $\hat a_j\to a_j$, a.s., $1\le j\le p$.
  2. $\sqrt{N}(\hat a_1-a_1,\ldots,\hat a_p-a_p)^T$ 依分布收敛到p维正态分布$N(0,\sigma^2\Gamma_p^{-1})$。
  3. $\sqrt{N}\sup\limits_{1\le j\le p}|\hat a_j-a_j|=O(\sqrt{\ln\ln N})$, a.s.,
     $\sqrt{N}|\hat\sigma^2-\sigma^2|=O(\sqrt{\ln\ln N})$, a.s.
- 证明:
  1. 由§4.2定理2.1知$\hat\gamma_k$强相合，再根据a_k,σ^2为$\{\gamma_k\}$的连续函数知$\hat a_k,\hat\sigma^2$强相合。
  2. 证明略(详见Brockwell and Davis的教材)。
  3. 证明略。

Y-W估计的相合性(续)

- **定理1.1** 如果AR(p)模型(1.1)中的$\{\varepsilon_t\}$是独立同分布的WN$(0,\sigma^2)$，E$\varepsilon_t^4<\infty$，则当$N\to\infty$时
  1. $\hat\sigma^2\to\sigma^2$，$\hat a_j\to a_j$，a.s.，$1\le j\le p$.
  2. $\sqrt{N}(\hat a_1-a_1,\dots,\hat a_p-a_p)^T$ 依分布收敛到p维正态分布$N(0,\sigma^2\Gamma_p^{-1})$。
  3. $\sqrt{N}\sup_{1\le j\le p}|\hat a_j-a_j|=O(\sqrt{\ln\ln N})$, a.s.,
     $\sqrt{N}|\hat\sigma^2-\sigma^2|=O(\sqrt{\ln\ln N})$, a.s.
- 证明:
  1. 由§4.2定理2.1知$\hat\gamma_k$强相合，再根据a_k,σ^2为$\{\gamma_k\}$的连续函数知$\hat a_k,\hat\sigma^2$强相合。
  2. 证明略(详见Brockwell and Davis的教材)。
  3. 证明略。

Y-W估计的相合性(续)

- **定理1.1** 如果AR(p)模型(1.1)中的$\{\varepsilon_t\}$是独立同分布的WN$(0,\sigma^2)$，$\mathrm{E}\varepsilon_t^4<\infty$，则当$N\to\infty$时
  1. $\hat\sigma^2\to\sigma^2$, $\hat a_j\to a_j$, a.s., $1\le j\le p$.
  2. $\sqrt{N}(\hat a_1-a_1,\dots,\hat a_p-a_p)^T$ 依分布收敛到p维正态分布$N(0,\sigma^2\Gamma_p^{-1})$。
  3. $\sqrt{N}\sup\limits_{1\le j\le p}|\hat a_j-a_j|=O(\sqrt{\ln\ln N})$, a.s.,
     $\sqrt{N}|\hat\sigma^2-\sigma^2|=O(\sqrt{\ln\ln N})$, a.s.
- 证明:
  1. 由§4.2定理2.1知$\hat\gamma_k$强相合，再根据a_k,σ^2为$\{\gamma_k\}$的连续函数知$\hat a_k,\hat\sigma^2$强相合。
  2. 证明略(详见Brockwell and Davis的教材)。
  3. 证明略。

Y-W估计的相合性(续)

- **定理1.1** 如果AR(p)模型(1.1)中的$\{\varepsilon_t\}$是独立同分布的WN$(0, \sigma^2)$，E$\varepsilon_t^4 < \infty$，则当$N \to \infty$时
  1. $\hat{\sigma}^2 \to \sigma^2$, $\hat{a}_j \to a_j$, a.s., $1 \leq j \leq p$.
  2. $\sqrt{N}(\hat{a}_1 - a_1, \ldots, \hat{a}_p - a_p)^T$ 依分布收敛到p维正态分布$N(0, \sigma^2 \Gamma_p^{-1})$。
  3. $\sqrt{N} \sup\limits_{1 \leq j \leq p} |\hat{a}_j - a_j| = O(\sqrt{\ln \ln N})$, a.s.,
     $\sqrt{N} |\hat{\sigma}^2 - \sigma^2| = O(\sqrt{\ln \ln N})$, a.s.
- 证明:
  1. 由§4.2定理2.1知$\hat{\gamma}_k$强相合，再根据a_k, σ^2为$\{\gamma_k\}$的连续函数知$\hat{a}_k, \hat{\sigma}^2$强相合。
  2. 证明略(详见Brockwell and Davis的教材)。
  3. 证明略。

Y-W估计的相合性(续)

- **定理1.1** 如果AR(p)模型(1.1)中的$\{\varepsilon_t\}$是独立同分布的WN$(0,\sigma^2)$，$\mathrm{E}\varepsilon_t^4 < \infty$，则当$N \to \infty$时
  1. $\hat{\sigma}^2 \to \sigma^2$, $\hat{a}_j \to a_j$, a.s., $1 \le j \le p$.
  2. $\sqrt{N}(\hat{a}_1 - a_1, \ldots, \hat{a}_p - a_p)^T$ 依分布收敛到p维正态分布$N(0, \sigma^2\Gamma_p^{-1})$。
  3. $\sqrt{N} \sup\limits_{1\le j\le p} |\hat{a}_j - a_j| = O(\sqrt{\ln\ln N})$, a.s.,
     $\sqrt{N}|\hat{\sigma}^2 - \sigma^2| = O(\sqrt{\ln\ln N})$, a.s.
- 证明:
  1. 由§4.2定理2.1知$\hat{\gamma}_k$强相合，再根据a_k, σ^2为$\{\gamma_k\}$的连续函数知$\hat{a}_k, \hat{\sigma}^2$强相合。
  2. 证明略(详见Brockwell and Davis的教材)。
  3. 证明略。

Y-W估计的相合性(续)

- **定理1.1** 如果AR(p)模型(1.1)中的$\{\varepsilon_t\}$是独立同分布的WN$(0,\sigma^2)$，$\mathrm{E}\varepsilon_t^4 < \infty$，则当$N \to \infty$时
  1. $\hat{\sigma}^2 \to \sigma^2$, $\hat{a}_j \to a_j$, a.s., $1 \le j \le p$.
  2. $\sqrt{N}(\hat{a}_1 - a_1, \ldots, \hat{a}_p - a_p)^T$ 依分布收敛到p维正态分布$N(0, \sigma^2 \Gamma_p^{-1})$。
  3. $\sqrt{N} \sup\limits_{1 \le j \le p} |\hat{a}_j - a_j| = O(\sqrt{\ln \ln N})$, a.s.,
     $\sqrt{N}|\hat{\sigma}^2 - \sigma^2| = O(\sqrt{\ln \ln N})$, a.s.
- 证明:
  1. 由§4.2定理2.1知$\hat{\gamma}_k$强相合，再根据a_k, σ^2为$\{\gamma_k\}$的连续函数知$\hat{a}_k, \hat{\sigma}^2$强相合。
  2. 证明略(详见Brockwell and Davis的教材)。
  3. 证明略。

Y-W估计的相合性(续)

- **定理1.1** 如果AR(p)模型(1.1)中的$\{\varepsilon_t\}$是独立同分布的WN$(0,\sigma^2)$，$\mathrm{E}\varepsilon_t^4<\infty$，则当$N\to\infty$时
  1. $\hat{\sigma}^2\to\sigma^2$, $\hat{a}_j\to a_j$, a.s., $1\le j\le p$.
  2. $\sqrt{N}(\hat{a}_1-a_1,\ldots,\hat{a}_p-a_p)^T$ 依分布收敛到p维正态分布$N(0,\sigma^2\Gamma_p^{-1})$。
  3. $\sqrt{N}\sup\limits_{1\le j\le p}|\hat{a}_j-a_j|=O(\sqrt{\ln\ln N})$, a.s.,
     $\sqrt{N}|\hat{\sigma}^2-\sigma^2|=O(\sqrt{\ln\ln N})$, a.s.
- 证明:
  1. 由§4.2定理2.1知$\hat{\gamma}_k$强相合，再根据a_k,σ^2为$\{\gamma_k\}$的连续函数知$\hat{a}_k,\hat{\sigma}^2$强相合。
  2. 证明略(详见Brockwell and Davis的教材)。
  3. 证明略。

Y-W估计的相合性(续2)

- 当定理成立时参数估计有近似正态分布，可以用来作置信区间和假设检验。
- 设σ_{jj}为$\sigma^2\Gamma_p^{-1}$的第$j \times j$元素，则$\hat{a}_j$近似服从$\mathrm{N}(a_j, \sigma_{jj}/N)$，于是$a_j$的近似95% 置信区间为

$$[\hat{a}_j - 1.96\sqrt{\sigma_{jj}}/\sqrt{N},\ \hat{a}_j + 1.96\sqrt{\sigma_{jj}}/\sqrt{N}]$$

其中σ_{jj}可以用$\hat{\sigma}^2\hat{\Gamma}_p^{-1}$的元素$\hat{\sigma}_{jj}$代替。

Y-W估计的相合性(续2)

- 当定理成立时参数估计有近似正态分布，可以用来作置信区间和假设检验。
- 设σ_{jj}为$\sigma^2\Gamma_p^{-1}$的第$j \times j$元素，则$\hat{a}_j$近似服从$\mathrm{N}(a_j, \sigma_{jj}/N)$，于是$a_j$的近似95% 置信区间为

$$[\hat{a}_j - 1.96\sqrt{\sigma_{jj}}/\sqrt{N},\ \hat{a}_j + 1.96\sqrt{\sigma_{jj}}/\sqrt{N}]$$

其中σ_{jj}可以用$\hat{\sigma}^2\hat{\Gamma}_p^{-1}$的元素$\hat{\sigma}_{jj}$代替。

最小二乘估计

- 最小二乘估计计算简单，不需要计算自协方差。
- 最小二乘估计使残差平方和最小:

$$\min S(a_1,\ldots,a_p) \triangleq \sum_{t=p+1}^{N} [y_t - (a_1 y_{t-1} + \cdots + a_p y_{t-p})]^2$$

最小值点$\hat{a}_1,\ldots,\hat{a}_p$称为参数的最小二乘估计。

最小二乘估计

- 最小二乘估计计算简单，不需要计算自协方差。
- 最小二乘估计使残差平方和最小:

$$\min S(a_1, \ldots, a_p) \triangleq \sum_{t=p+1}^{N} [y_t - (a_1 y_{t-1} + \cdots + a_p y_{t-p})]^2$$

最小值点$\hat{a}_1, \ldots, \hat{a}_p$称为参数的最小二乘估计。

- 用线性模型理论，写出最小二乘对应的正规方程为

$$\mathbf{Y} = \begin{pmatrix} y_{p+1} \\ y_{p+2} \\ \cdots \\ y_N \end{pmatrix},$$

$$\mathbf{X} = \begin{pmatrix} y_p & y_{p-1} & \cdots & y_1 \\ y_{p+1} & y_p & \cdots & y_2 \\ \vdots & \vdots & & \vdots \\ y_{N-1} & y_{N-2} & \cdots & y_{N-p} \end{pmatrix} = (y_{p+t-j})_{\substack{t=1,\ldots,N-p; \\ j=1,\ldots,p}}$$

$$\mathbf{a} = \begin{pmatrix} a_1 \\ a_2 \\ \vdots \\ a_p \end{pmatrix}, \quad (\mathbf{X}'\mathbf{X})\mathbf{a} = \mathbf{X}'\mathbf{Y} \tag{1.12}$$

- 最小二乘估计$\tilde{a}_1, \ldots, \tilde{a}_p$是正规方程(1.12)的解。当$p \times p$矩阵$\mathbf{X}^T\mathbf{X}$正定时方程有唯一解

$$\tilde{\mathbf{a}} = (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{Y}$$

- 用线性模型理论，写出最小二乘对应的正规方程为

$$\mathbf{Y} = \begin{pmatrix} y_{p+1} \\ y_{p+2} \\ \cdots \\ y_N \end{pmatrix},$$

$$\mathbf{X} = \begin{pmatrix} y_p & y_{p-1} & \cdots & y_1 \\ y_{p+1} & y_p & \cdots & y_2 \\ \vdots & \vdots & & \vdots \\ y_{N-1} & y_{N-2} & \cdots & y_{N-p} \end{pmatrix} = (y_{p+t-j})_{\substack{t=1,\dots,N-p;\\ j=1,\dots,p}}$$

$$\mathbf{a} = \begin{pmatrix} a_1 \\ a_2 \\ \vdots \\ a_p \end{pmatrix}, \quad (\mathbf{X}'\mathbf{X})\mathbf{a} = \mathbf{X}'\mathbf{Y} \tag{1.12}$$

- 最小二乘估计$\tilde{a}_1, \dots, \tilde{a}_p$是正规方程(1.12)的解。当$p \times p$矩阵$\mathbf{X}^T\mathbf{X}$正定时方程有唯一解

$$\tilde{\mathbf{a}} = (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{Y}$$

依概率有界

- **定义1.1** 设$\{\xi_n\}$是时间序列，$\{c_n\}$是非零常数列，如果对任意$\varepsilon > 0$，存在正数M，使得

$$\sup_n P(|\xi_n| > M) \leq \varepsilon$$

  就称时间序列$\{\xi_n\}$是依概率有界的，记做$\xi_n = O_p(1)$。如果$\{\xi_n/c_n\} = O_p(1)$，就称$\xi_n = O_p(c_n)$。
- 依概率有界就是除去一个概率任意小的集合后序列有界。
- 若$\{\xi_n/c_n\} \overset{\Pr}{\to} 0$则称$\xi_n = o_p(c_n)$。

依概率有界

- **定义1.1** 设$\{\xi_n\}$是时间序列，$\{c_n\}$是非零常数列，如果对任意$\varepsilon > 0$，存在正数M，使得

$$\sup_n P(|\xi_n| > M) \leq \varepsilon$$

就称时间序列$\{\xi_n\}$是依概率有界的，记做$\xi_n = O_p(1)$。如果$\{\xi_n/c_n\} = O_p(1)$，就称$\xi_n = O_p(c_n)$。

- 依概率有界就是除去一个概率任意小的集合后序列有界。
- 若$\{\xi_n/c_n\} \xrightarrow{Pr} 0$则称$\xi_n = o_p(c_n)$。

依概率有界

- **定义1.1** 设$\{\xi_n\}$是时间序列，$\{c_n\}$是非零常数列，如果对任意$\varepsilon > 0$，存在正数M，使得
$$\sup_n P(|\xi_n| > M) \le \varepsilon$$
就称时间序列$\{\xi_n\}$是依概率有界的，记做$\xi_n = O_p(1)$。如果$\{\xi_n / c_n\} = O_p(1)$，就称$\xi_n = O_p(c_n)$。
- 依概率有界就是除去一个概率任意小的集合后序列有界。
- 若$\{\xi_n / c_n\} \xrightarrow{\text{Pr}} 0$则称$\xi_n = o_p(c_n)$。

依概率有界的等价定义

- 称$\{\xi_n\}$依概率有界，若$\forall \varepsilon > 0$, $\exists M > 0$和N使得当$n \geq N$时

$$P(|\xi_n| \leq M) \geq 1 - \varepsilon.$$

- 事实上，当原定义条件成立时显然此等价定义的条件也成立。

依概率有界的等价定义

- 称$\{\xi_n\}$依概率有界，若$\forall \varepsilon > 0$, $\exists M > 0$和N使得当$n \geq N$时

$$P(|\xi_n| \leq M) \geq 1 - \varepsilon.$$

- 事实上，当原定义条件成立时显然此等价定义的条件也成立。

- 若此等价定义条件成立，则对$n \geq N$有

$$P(|\xi_n| \leq M) \geq 1 - \varepsilon, \quad P(|\xi_n| > M) < \varepsilon.$$

- 对$j = 1, 2, \ldots, N$，存在$M_j > 0$使得

$$P(|\xi_j| > M_j) < \varepsilon$$

- 令$M' = \max(M, M_1, M_2, \ldots, M_N)$，则

$$P(|\xi_n| > M') < \varepsilon, \ \forall n \in \mathbb{N}_+,$$
$$\sup_{n \in \mathbb{N}_+} P(|\xi_n| > M') \leq \varepsilon,$$

满足原定义。

- 若此等价定义条件成立，则对$n \geq N$有

$$P(|\xi_n| \leq M) \geq 1 - \varepsilon, \quad P(|\xi_n| > M) < \varepsilon.$$

- 对$j = 1, 2, \ldots, N$，存在$M_j > 0$使得

$$P(|\xi_j| > M_j) < \varepsilon$$

- 令$M' = \max(M, M_1, M_2, \ldots, M_N)$，则

$$P(|\xi_n| > M') < \varepsilon, \ \forall n \in \mathbb{N}_+,$$
$$\sup_{n \in \mathbb{N}_+} P(|\xi_n| > M') \leq \varepsilon,$$

满足原定义。

- 若此等价定义条件成立，则对$n \geq N$有

$$P(|\xi_n| \leq M) \geq 1 - \varepsilon, \quad P(|\xi_n| > M) < \varepsilon.$$

- 对$j = 1, 2, \ldots, N$，存在$M_j > 0$使得

$$P(|\xi_j| > M_j) < \varepsilon$$

- 令$M' = \max(M, M_1, M_2, \ldots, M_N)$，则

$$\begin{aligned} P(|\xi_n| > M') &< \varepsilon, \ \forall n \in \mathbb{N}_+, \\ \sup_{n \in \mathbb{N}_+} P(|\xi_n| > M') &\leq \varepsilon, \end{aligned}$$

满足原定义。

依概率有界(续)

- **引理1**: 若$\xi_n = o_p(c_n)$则$\xi_n = O_p(c_n)$。
- 事实上，按依概率收敛定义，$\forall \delta > 0, \forall \varepsilon > 0$，存在$N$使$n > N$时

$$\Pr(|\xi_n / c_n| > \delta) < \varepsilon$$

取$M \geq \delta$使得

$$\Pr(|\xi_n / c_n| > M) \leq \varepsilon, \quad n = 1, 2, \ldots, N$$

于是

$$\Pr(|\xi_n / c_n| > M) \leq \varepsilon, \quad n = 1, 2, \ldots$$

即$\xi_n / c_n = O_p(1)$.

依概率有界(续)

- **引理1**：若$\xi_n = o_p(c_n)$则$\xi_n = O_p(c_n)$。
- 事实上，按依概率收敛定义，$\forall \delta > 0$, $\forall \varepsilon > 0$，存在N使$n > N$时

$$\Pr(|\xi_n/c_n| > \delta) < \varepsilon$$

取$M \geq \delta$使得

$$\Pr(|\xi_n/c_n| > M) \leq \varepsilon, \quad n = 1, 2, \ldots, N$$

于是

$$\Pr(|\xi_n/c_n| > M) \leq \varepsilon, \quad n = 1, 2, \ldots$$

即$\xi_n/c_n = O_p(1)$.

依概率有界(续2)

- **引理2** 如果$\xi_n = O_p(1)$而$c_n \to \infty$则$\xi_n / c_n = o_p(1)$。
- 见习题1.1。
- 对随机变量ξ，$\xi = O_p(1)$.
- **引理3** 若$\{\xi_n\}$同分布，则$\xi_n = O_p(1)$。
- 事实上，$\forall \varepsilon > 0$, $\exists M > 0$使$\Pr(|\xi_1| > M) < \varepsilon$。由同分布性知$\Pr(|\xi_t| > M) = \Pr(|\xi_1| > M) < \varepsilon$。

依概率有界(续2)

- **引理2** 如果$\xi_n = O_p(1)$而$c_n \to \infty$则$\xi_n / c_n = o_p(1)$。
- 见习题1.1。
- 对随机变量ξ，$\xi = O_p(1)$.
- **引理3** 若$\{\xi_n\}$同分布，则$\xi_n = O_p(1)$。
- 事实上，$\forall \varepsilon > 0$, $\exists M > 0$使$\Pr(|\xi_1| > M) < \varepsilon$。由同分布性知$\Pr(|\xi_t| > M) = \Pr(|\xi_1| > M) < \varepsilon$。

依概率有界(续2)

- **引理2** 如果$\xi_n = O_p(1)$而$c_n \to \infty$则$\xi_n / c_n = o_p(1)$。
- 见习题1.1。
- 对随机变量ξ，$\xi = O_p(1)$.
- **引理3** 若$\{\xi_n\}$同分布，则$\xi_n = O_p(1)$。
- 事实上，$\forall \varepsilon > 0$, $\exists M > 0$使$\Pr(|\xi_1| > M) < \varepsilon$。由同分布性知$\Pr(|\xi_t| > M) = \Pr(|\xi_1| > M) < \varepsilon$。

依概率有界(续2)

- **引理2** 如果$\xi_n = O_p(1)$而$c_n \to \infty$则$\xi_n / c_n = o_p(1)$。
- 见习题1.1。
- 对随机变量ξ，$\xi = O_p(1)$.
- **引理3** 若$\{\xi_n\}$同分布，则$\xi_n = O_p(1)$。
- 事实上，$\forall \varepsilon > 0$, $\exists M > 0$使$\Pr(|\xi_1| > M) < \varepsilon$。由同分布性知$\Pr(|\xi_t| > M) = \Pr(|\xi_1| > M) < \varepsilon$。

依概率有界(续2)

- **引理2** 如果$\xi_n = O_p(1)$而$c_n \to \infty$则$\xi_n / c_n = o_p(1)$。
- 见习题1.1。
- 对随机变量ξ，$\xi = O_p(1)$.
- **引理3** 若$\{\xi_n\}$同分布，则$\xi_n = O_p(1)$。
- 事实上，$\forall \varepsilon > 0$, $\exists M > 0$使$\Pr(|\xi_1| > M) < \varepsilon$。由同分布性知$\Pr(|\xi_t| > M) = \Pr(|\xi_1| > M) < \varepsilon$。

依概率有界(续3)

- **引理4**

$$\begin{aligned} O_p(1) \pm O_p(1) =& O_p(1) \\ O_p(1) \cdot O_p(1) =& O_p(1) \\ O_p(1) \pm o_p(1) =& O_p(1) \\ O_p(1) \cdot o_p(1) =& o_p(1). \end{aligned}$$

依概率有界(续4)

- **引理5** 若ξ_n依概率收敛到ξ则$\{\xi_n\} = O_p(1)$。
- 事实上$\xi_n = \xi + (\xi_n - \xi) = O_p(1) + o_p(1)$。
- 另外，若ξ_n依分布收敛到ξ则$\{\xi_n\} = O_p(1)$。
- **引理6** 若存在非负随机变量ξ使得$|\xi_n| \le \xi$ a.s.则$\xi_n = O_p(1)$。
- 由$P(|\xi_n| \le \xi) = 1$，$\forall \varepsilon > 0, \exists M > 0$使$P(\xi > M) < \varepsilon$，于是$P(|\xi_n| > M) \le P(\xi > M) < \varepsilon, \forall n \in \mathbb{N}$。

依概率有界(续4)

- **引理5**　若ξ_n依概率收敛到ξ则$\{\xi_n\} = O_p(1)$。
- 事实上$\xi_n = \xi + (\xi_n - \xi) = O_p(1) + o_p(1)$。
- 另外，若ξ_n依分布收敛到ξ则$\{\xi_n\} = O_p(1)$。
- **引理6**　若存在非负随机变量ξ使得$|\xi_n| \leq \xi$ a.s.则$\xi_n = O_p(1)$。
- 由$P(|\xi_n| \leq \xi) = 1$，$\forall \varepsilon > 0, \exists M > 0$使$P(\xi > M) < \varepsilon$，于是$P(|\xi_n| > M) \leq P(\xi > M) < \varepsilon, \forall n \in \mathbb{N}$。

依概率有界(续4)

- **引理5**　若ξ_n依概率收敛到ξ则$\{\xi_n\} = O_p(1)$。
- 事实上$\xi_n = \xi + (\xi_n - \xi) = O_p(1) + o_p(1)$。
- 另外，若ξ_n依分布收敛到ξ则$\{\xi_n\} = O_p(1)$。
- **引理6**　若存在非负随机变量ξ使得$|\xi_n| \le \xi$ a.s.则$\xi_n = O_p(1)$。
- 由$P(|\xi_n| \le \xi) = 1$，$\forall \varepsilon > 0, \exists M > 0$使$P(\xi > M) < \varepsilon$，于是$P(|\xi_n| > M) \le P(\xi > M) < \varepsilon, \forall n \in \mathbb{N}$。

依概率有界(续4)

- **引理5** 若ξ_n依概率收敛到ξ则$\{\xi_n\} = O_p(1)$。
- 事实上$\xi_n = \xi + (\xi_n - \xi) = O_p(1) + o_p(1)$。
- 另外，若ξ_n依分布收敛到ξ则$\{\xi_n\} = O_p(1)$。
- **引理6** 若存在非负随机变量ξ使得$|\xi_n| \leq \xi$ a.s.则$\xi_n = O_p(1)$。
- 由$P(|\xi_n| \leq \xi) = 1$，$\forall \varepsilon > 0, \exists M > 0$使$P(\xi > M) < \varepsilon$，于是$P(|\xi_n| > M) \leq P(\xi > M) < \varepsilon, \forall n \in \mathbb{N}$。

依概率有界(续4)

- **引理5**　若ξ_n依概率收敛到ξ则$\{\xi_n\} = O_p(1)$。
- 事实上$\xi_n = \xi + (\xi_n - \xi) = O_p(1) + o_p(1)$。
- 另外，若ξ_n依分布收敛到ξ则$\{\xi_n\} = O_p(1)$。
- **引理6**　若存在非负随机变量ξ使得$|\xi_n| \leq \xi$ a.s.则$\xi_n = O_p(1)$。
- 由$P(|\xi_n| \leq \xi) = 1$，$\forall \varepsilon > 0, \exists M > 0$使$P(\xi > M) < \varepsilon$，于是$P(|\xi_n| > M) \leq P(\xi > M) < \varepsilon, \forall n \in \mathbb{N}$。

最小二乘估计和Y-W估计的比较

- 注意对于$1 \le k \le j \le p$, 矩阵$(1/N)\mathbf{X}^T\mathbf{X}$的第$(j,k)$元素是

$$
\begin{aligned}
\tilde{\gamma}_{jk} =& \frac{1}{N}\sum_{t=1}^{N-p} y_{t+p-j}y_{t+p-k} = \frac{1}{N}\sum_{t=p-j+1}^{N-j} y_t y_{t+j-k} \\
=& \frac{1}{N}\sum_{t=1}^{N-(j-k)} y_t y_{t+j-k} - \frac{1}{N}\sum_{t=1}^{p-j} y_t y_{t+j-k} - \frac{1}{N}\sum_{t=N-j+1}^{N-(j-k)} y_t y_{t+j-k} \\
=& \hat{\gamma}_{j-k} - \frac{1}{N}\sum_{t=1}^{p-j} y_t y_{t+j-k} - \frac{1}{N}\sum_{t=N-j+1}^{N-(j-k)} y_t y_{t+j-k}.
\end{aligned}
$$

这里$\hat{\gamma}_{j-k}$ 由(1.2) 定义.

最小二乘估计和Y-W估计的比较(续)

- 完全类似地得到$(1/N)\mathbf{X}^T\mathbf{Y}$的第$j$个元素

$$
\begin{aligned}
\tilde{\gamma}_j =& \frac{1}{N}\sum_{t=1}^{N-p} y_{t+p-j}y_{t+p} = \frac{1}{N}\sum_{t=p-j+1}^{N-j} y_t y_{t+j} \\
=& \frac{1}{N}\sum_{t=1}^{N-j} y_t y_{t+j} - \frac{1}{N}\sum_{t=1}^{p-j} y_t y_{t+j} \\
=& \hat{\gamma}_j - \frac{1}{N}\sum_{t=1}^{p-j} y_t y_{t+j}.
\end{aligned}
$$

最小二乘估计和Y-W估计的比较(续2)

- $\sum_{t=1}^{p-j} y_t y_{t+j-k}$和$\sum_{t=1}^{p-j} y_t y_{t+j}$是固定的随机变量所以是$O_p(1)$的，$\frac{1}{N}\sum_{t=1}^{p-j} y_t y_{t+j-k} = O_p(\frac{1}{N})$, $\frac{1}{N}\sum_{t=1}^{p-j} y_t y_{t+j} = O_p(\frac{1}{N})$。
- 同时，固定随机变量的$\frac{1}{N}$是a.s.趋于零的。

最小二乘估计和Y-W估计的比较(续2)

- $\sum_{t=1}^{p-j} y_t y_{t+j-k}$和$\sum_{t=1}^{p-j} y_t y_{t+j}$是固定的随机变量所以是$O_p(1)$的，$\frac{1}{N}\sum_{t=1}^{p-j} y_t y_{t+j-k} = O_p(\frac{1}{N})$, $\frac{1}{N}\sum_{t=1}^{p-j} y_t y_{t+j} = O_p(\frac{1}{N})$。
- 同时，固定随机变量的$\frac{1}{N}$是a.s.趋于零的。

最小二乘估计和Y-W估计的比较(续3)

- 由$\{\varepsilon_t\}$独立同分布知AR(p)序列$\{Y_t\}$严平稳遍历，所以$\sum_{t=N-j+1}^{N-(j-k)} y_t y_{t+j-k}$也是严平稳遍历的，且一阶矩存在，由引理3知其为$O_p(1)$的。
- 另一方面，由$\{Y_t\}$的严平稳遍历性则

$$
\begin{aligned}
&\frac{1}{N}\sum_{t=N-j+1}^{N-(j-k)} y_t y_{t+j-k} \\
=&\frac{1}{N}\sum_{j=1}^{N-(j-k)} y_t y_{t+j-k} - \frac{1}{N}\sum_{j=1}^{N-j)} y_t y_{t+j-k} \\
=&\frac{N-(j-k)}{N}\frac{1}{N-(j-k)}\sum_{j=1}^{N-(j-k)} y_t y_{t+j-k} \\
&-\frac{N-j}{N}\frac{1}{N-j}\sum_{j=1}^{N-j)} y_t y_{t+j-k} \\
\to&1\cdot\gamma_{j-k} - 1\cdot\gamma_{j-k} = 0 \quad \text{a.s.}
\end{aligned}
$$

最小二乘估计和Y-W估计的比较(续3)

- 由$\{\varepsilon_t\}$独立同分布知AR(p)序列$\{Y_t\}$严平稳遍历，所以$\sum_{t=N-j+1}^{N-(j-k)} y_t y_{t+j-k}$也是严平稳遍历的，且一阶矩存在，由引理3知其为$O_p(1)$的。
- 另一方面，由$\{Y_t\}$的严平稳遍历性则

$$
\begin{aligned}
&\frac{1}{N}\sum_{t=N-j+1}^{N-(j-k)} y_t y_{t+j-k} \\
=&\frac{1}{N}\sum_{j=1}^{N-(j-k)} y_t y_{t+j-k} - \frac{1}{N}\sum_{j=1}^{N-j)} y_t y_{t+j-k} \\
=&\frac{N-(j-k)}{N}\frac{1}{N-(j-k)}\sum_{j=1}^{N-(j-k)} y_t y_{t+j-k} \\
&-\frac{N-j}{N}\frac{1}{N-j}\sum_{j=1}^{N-j)} y_t y_{t+j-k} \\
\to& 1\cdot\gamma_{j-k} - 1\cdot\gamma_{j-k} = 0 \quad \text{a.s.}
\end{aligned}
$$

最小二乘估计和Y-W估计的比较(续4)

- 于是, 对$1 \le k \le j \le p$有

$$\tilde{\gamma}_{jk} - \hat{\gamma}_{j-k} = O_p(\frac{1}{N}),\ \tilde{\gamma}_j - \hat{\gamma}_j = O_p(\frac{1}{N}),$$

- 且

$$\tilde{\gamma}_{jk} - \hat{\gamma}_{j-k} \to 0,\ \tilde{\gamma}_j - \hat{\gamma}_j \to 0,\ \text{a.s.},$$

- 由$\hat{\gamma}_j$的强相合性知$\tilde{\gamma}_{jk} \to \gamma_{j-k}$, $\tilde{\gamma}_j \to \gamma_j$, a.s.
- 这样,当$N \to \infty$

$$(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{Y} = \left(\frac{1}{N}\mathbf{X}^T\mathbf{X}\right)^{-1}\left(\frac{1}{N}\mathbf{X}^T\mathbf{Y}\right)$$
$$\to \Gamma_p^{-1}\gamma_p = \mathbf{a}, \quad \text{a.s.}$$

即最小二乘估计也是强相合的。

最小二乘估计和Y-W估计的比较(续4)

- 于是, 对$1 \le k \le j \le p$有

$$\tilde{\gamma}_{jk} - \hat{\gamma}_{j-k} = O_p(\frac{1}{N}),\ \tilde{\gamma}_j - \hat{\gamma}_j = O_p(\frac{1}{N}),$$

- 且

$$\tilde{\gamma}_{jk} - \hat{\gamma}_{j-k} \to 0,\ \tilde{\gamma}_j - \hat{\gamma}_j \to 0,\ \text{a.s.},$$

- 由$\hat{\gamma}_j$的强相合性知$\tilde{\gamma}_{jk} \to \gamma_{j-k}$, $\tilde{\gamma}_j \to \gamma_j$, a.s.
- 这样,当$N \to \infty$

$$(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{Y} = \left(\frac{1}{N}\mathbf{X}^T\mathbf{X}\right)^{-1}\left(\frac{1}{N}\mathbf{X}^T\mathbf{Y}\right)$$
$$\to \Gamma_p^{-1}\gamma_p = \mathbf{a}, \quad \text{a.s.}$$

即最小二乘估计也是强相合的。

最小二乘估计和Y-W估计的比较(续4)

- 于是, 对$1 \le k \le j \le p$有

$$\tilde{\gamma}_{jk} - \hat{\gamma}_{j-k} = O_p(\frac{1}{N}),\ \tilde{\gamma}_j - \hat{\gamma}_j = O_p(\frac{1}{N}),$$

- 且

$$\tilde{\gamma}_{jk} - \hat{\gamma}_{j-k} \to 0,\ \tilde{\gamma}_j - \hat{\gamma}_j \to 0,\ \text{a.s.},$$

- 由$\hat{\gamma}_j$的强相合性知$\tilde{\gamma}_{jk} \to \gamma_{j-k}$, $\tilde{\gamma}_j \to \gamma_j$, a.s.
- 这样,当$N \to \infty$

$$(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{Y} = \left(\frac{1}{N}\mathbf{X}^T\mathbf{X}\right)^{-1}\left(\frac{1}{N}\mathbf{X}^T\mathbf{Y}\right) \to \Gamma_p^{-1}\gamma_p = \mathbf{a},\quad \text{a.s.}$$

即最小二乘估计也是强相合的。

最小二乘估计和Y-W估计的比较(续4)

- 于是, 对$1 \le k \le j \le p$有

$$\tilde{\gamma}_{jk} - \hat{\gamma}_{j-k} = O_p(\frac{1}{N}),\ \tilde{\gamma}_j - \hat{\gamma}_j = O_p(\frac{1}{N}),$$

- 且

$$\tilde{\gamma}_{jk} - \hat{\gamma}_{j-k} \to 0,\ \tilde{\gamma}_j - \hat{\gamma}_j \to 0,\ \text{a.s.},$$

- 由$\hat{\gamma}_j$的强相合性知$\tilde{\gamma}_{jk} \to \gamma_{j-k}$, $\tilde{\gamma}_j \to \gamma_j$, a.s.
- 这样,当$N \to \infty$

$$\begin{aligned}(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{Y} &= \left(\frac{1}{N}\mathbf{X}^T\mathbf{X}\right)^{-1}\left(\frac{1}{N}\mathbf{X}^T\mathbf{Y}\right) \\ &\to \Gamma_p^{-1}\gamma_p = \mathbf{a}, \quad \text{a.s.}\end{aligned}$$

即最小二乘估计也是强相合的。

最小二乘估计和Y-W估计的比较(续5)

- 另一方面, 利用$(\tilde{\gamma}_{jk} - \hat{\gamma}_{j-k}) = O_p(1/N)$ 和$(\tilde{\gamma}_j - \hat{\gamma}_j) = O_p(1/N)$,
- 用$\hat{a} = (\hat{a}_1, \hat{a}_2, \cdots, \hat{a}_p)^T$ 和$\tilde{a} = (\tilde{a}_1, \tilde{a}_2, \ldots, \tilde{a}_p)^T$ 分别表示Yule-Walker估计和最小二乘估计.
记$\hat{\gamma}_p = (\hat{\gamma}_1, \hat{\gamma}_2, \ldots, \hat{\gamma}_p)^T$, 就有

$$\begin{aligned}
\tilde{a} - \hat{a} =& (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{Y} - \hat{\Gamma}_p^{-1}\hat{\gamma}_p \\
=& (\frac{1}{N}\mathbf{X}^T\mathbf{X})^{-1}\left(\frac{1}{N}\mathbf{X}^T\mathbf{Y} - \hat{\gamma}_p\right) + \left[(\frac{1}{N}\mathbf{X}^T\mathbf{X})^{-1} - \hat{\Gamma}_p^{-1}\right]\hat{\gamma}_p \\
=& (\frac{1}{N}\mathbf{X}^T\mathbf{X})^{-1}O_p(1/N) + (\frac{1}{N}\mathbf{X}^T\mathbf{X})^{-1}\left[\hat{\Gamma}_p - \frac{1}{N}\mathbf{X}^T\mathbf{X}\right]\hat{\Gamma}_p^{-1}\hat{\gamma}_p \\
=& O_p(1/N), \quad \text{当 } N \to \infty. \qquad (1.14)
\end{aligned}$$

- 上式说明对$1 \le j \le p$, $\hat{a}_j - \tilde{a}_j = O_p(1/N)$. 由于$1/N$收敛到零的速度是很快的, 所以也可以说对较大的$N$, 最小二乘估计和Yule-Walker 估计的差别不大.

最小二乘估计和Y-W估计的比较(续5)

- 另一方面，利用$(\tilde{\gamma}_{jk}-\hat{\gamma}_{j-k})=O_p(1/N)$ 和$(\tilde{\gamma}_j-\hat{\gamma}_j)=O_p(1/N)$,
- 用$\hat{a}=(\hat{a}_1,\hat{a}_2,\cdots,\hat{a}_p)^T$ 和$\tilde{a}=(\tilde{a}_1,\tilde{a}_2,\ldots,\tilde{a}_p)^T$ 分别表示Yule-Walker估计和最小二乘估计.
  记$\hat{\gamma}_p=(\hat{\gamma}_1,\hat{\gamma}_2,\ldots,\hat{\gamma}_p)^T$, 就有

$$\begin{aligned}\tilde{a}-\hat{a}=&(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{Y}-\hat{\Gamma}_p^{-1}\hat{\gamma}_p\\ =&(\frac{1}{N}\mathbf{X}^T\mathbf{X})^{-1}\left(\frac{1}{N}\mathbf{X}^T\mathbf{Y}-\hat{\gamma}_p\right)+\left[(\frac{1}{N}\mathbf{X}^T\mathbf{X})^{-1}-\hat{\Gamma}_p^{-1}\right]\hat{\gamma}_p\\ =&(\frac{1}{N}\mathbf{X}^T\mathbf{X})^{-1}O_p(1/N)+(\frac{1}{N}\mathbf{X}^T\mathbf{X})^{-1}\left[\hat{\Gamma}_p-\frac{1}{N}\mathbf{X}^T\mathbf{X}\right]\hat{\Gamma}_p^{-1}\hat{\gamma}_p\\ =&O_p(1/N),\quad \text{当 } N\to\infty. \end{aligned}\tag{1.14}$$

- 上式说明对$1\le j\le p$, $\hat{a}_j-\tilde{a}_j=O_p(1/N)$. 由于$1/N$收敛到零的速度是很快的，所以也可以说对较大的$N$, 最小二乘估计和Yule-Walker 估计的差别不大.

最小二乘估计和Y-W估计的比较(续5)

- 另一方面, 利用$(\tilde{\gamma}_{jk}-\hat{\gamma}_{j-k})=O_p(1/N)$和$(\tilde{\gamma}_j-\hat{\gamma}_j)=O_p(1/N)$,
- 用$\hat{a}=(\hat{a}_1,\hat{a}_2,\cdots,\hat{a}_p)^T$ 和$\tilde{a}=(\tilde{a}_1,\tilde{a}_2,\ldots,\tilde{a}_p)^T$ 分别表示Yule-Walker估计和最小二乘估计.
  记$\hat{\boldsymbol{\gamma}}_p=(\hat{\gamma}_1,\hat{\gamma}_2,\ldots,\hat{\gamma}_p)^T$, 就有

$$
\begin{aligned}
\tilde{a}-\hat{a} =&(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{Y}-\hat{\Gamma}_p^{-1}\hat{\boldsymbol{\gamma}}_p\\
=&(\frac{1}{N}\mathbf{X}^T\mathbf{X})^{-1}\left(\frac{1}{N}\mathbf{X}^T\mathbf{Y}-\hat{\boldsymbol{\gamma}}_p\right)+\left[(\frac{1}{N}\mathbf{X}^T\mathbf{X})^{-1}-\hat{\Gamma}_p^{-1}\right]\hat{\boldsymbol{\gamma}}_p\\
=&(\frac{1}{N}\mathbf{X}^T\mathbf{X})^{-1}O_p(1/N)+(\frac{1}{N}\mathbf{X}^T\mathbf{X})^{-1}\left[\hat{\Gamma}_p-\frac{1}{N}\mathbf{X}^T\mathbf{X}\right]\hat{\Gamma}_p^{-1}\hat{\boldsymbol{\gamma}}_p\\
=&O_p(1/N),\quad 当\ N\to\infty. \qquad (1.14)
\end{aligned}
$$

- 上式说明对$1\le j\le p$, $\hat{a}_j-\tilde{a}_j=O_p(1/N)$. 由于$1/N$收敛到零的速度是很快的, 所以也可以说对较大的$N$, 最小二乘估计和Yule-Walker 估计的差别不大.

最小二乘估计的极限分布

- **定理1.2** 设AR(p)模型(1.1)中的白噪声$\{\varepsilon_t\}$是独立同分布的，$E\varepsilon_t^4 < \infty$，$(\tilde{a}_1, \tilde{a}_2, \ldots, \tilde{a}_p)$ 是自回归系数$(a_1, a_2, \ldots, a_p)$的最小二乘估计. 则当$N \to \infty$时

$$\sqrt{N}(\tilde{a}_1 - a_1, \tilde{a}_2 - a_2, \ldots, \tilde{a}_p - a_p)$$

依分布收敛到p-维正态分布$N(0, \sigma^2\Gamma_p^{-1})$.

- 即最小二乘估计和Y-W估计有相同的渐近分布。
- 定理1.2 的证明由定理1.1和(1.14)直接得到.

最小二乘估计的极限分布

- **定理1.2**　设AR(p)模型(1.1)中的白噪声$\{\varepsilon_t\}$是独立同分布的, $E\varepsilon_t^4 < \infty$, $(\tilde{a}_1, \tilde{a}_2, \ldots, \tilde{a}_p)$ 是自回归系数$(a_1, a_2, \ldots, a_p)$的最小二乘估计. 则当$N \to \infty$时

$$\sqrt{N}(\tilde{a}_1 - a_1, \tilde{a}_2 - a_2, \ldots, \tilde{a}_p - a_p)$$

依分布收敛到p-维正态分布$N(0, \sigma^2\Gamma_p^{-1})$.

- 即最小二乘估计和Y-W估计有相同的渐近分布。
- 定理1.2 的证明由定理1.1和(1.14)直接得到.

最小二乘估计的极限分布

- **定理1.2** 设AR(p)模型(1.1)中的白噪声$\{\varepsilon_t\}$是独立同分布的, $E\varepsilon_t^4 < \infty$, $(\tilde{a}_1, \tilde{a}_2, \ldots, \tilde{a}_p)$ 是自回归系数$(a_1, a_2, \ldots, a_p)$的最小二乘估计. 则当$N \to \infty$时

$$\sqrt{N}(\tilde{a}_1 - a_1, \tilde{a}_2 - a_2, \ldots, \tilde{a}_p - a_p)$$

依分布收敛到p-维正态分布$N(0, \sigma^2\Gamma_p^{-1})$.

- 即最小二乘估计和Y-W估计有相同的渐近分布。
- 定理1.2 的证明由定理1.1和(1.14)直接得到.

例1.1 Y-W估计和LS估计的模拟比较

- $\{\varepsilon_t\}$为标准正态白噪声，模拟产生AR(4)序列

$$\begin{aligned} X_t =& 1.16X_{t-1} - 0.37X_{t-2} - 0.11X_{t-3} \\ & + 0.18X_{t-4} + \varepsilon_t, t = 1, 2, \ldots, N \end{aligned}$$

- 分别取$N = 25, 100, 400, 1600$，重复$M = 10000$次产生模拟数据并分别用Y-W方法和最小二乘方法估计参数。
- 演示：AR estimation accuracy.
- 结果发现在样本量较小时，系数估计Y-W方法较好，预测方差σ^2估计最小二乘方法较好。样本量大时二者没有明显差别。

例1.1 Y-W估计和LS估计的模拟比较

- $\{\varepsilon_t\}$为标准正态白噪声，模拟产生AR(4)序列

$$\begin{aligned}X_t =& 1.16X_{t-1} - 0.37X_{t-2} - 0.11X_{t-3}\\ &+ 0.18X_{t-4} + \varepsilon_t, t = 1, 2, \ldots, N\end{aligned}$$

- 分别取$N = 25, 100, 400, 1600$，重复$M = 10000$次产生模拟数据并分别用Y-W方法和最小二乘方法估计参数。
- 演示：AR estimation accuracy.
- 结果发现在样本量较小时，系数估计Y-W方法较好，预测方差σ^2估计最小二乘方法较好。样本量大时二者没有明显差别。

例1.1 Y-W估计和LS估计的模拟比较

- $\{\varepsilon_t\}$为标准正态白噪声，模拟产生AR(4)序列

$$\begin{aligned} X_t = & 1.16X_{t-1} - 0.37X_{t-2} - 0.11X_{t-3} \\ & + 0.18X_{t-4} + \varepsilon_t, t = 1, 2, \ldots, N \end{aligned}$$

- 分别取$N = 25, 100, 400, 1600$，重复$M = 10000$次产生模拟数据并分别用Y-W方法和最小二乘方法估计参数。
- 演示：AR estimation accuracy.
- 结果发现在样本量较小时，系数估计Y-W方法较好，预测方差σ^2估计最小二乘方法较好。样本量大时二者没有明显差别。

例1.1 Y-W估计和LS估计的模拟比较

- $\{\varepsilon_t\}$为标准正态白噪声，模拟产生AR(4)序列

$$\begin{aligned} X_t =& 1.16X_{t-1} - 0.37X_{t-2} - 0.11X_{t-3} \\ & + 0.18X_{t-4} + \varepsilon_t, t = 1, 2, \ldots, N \end{aligned}$$

- 分别取$N = 25, 100, 400, 1600$，重复$M = 10000$次产生模拟数据并分别用Y-W方法和最小二乘方法估计参数。
- 演示：AR estimation accuracy.
- 结果发现在样本量较小时，系数估计Y-W方法较好，预测方差σ^2估计最小二乘方法较好。样本量大时二者没有明显差别。

例1.1 Y-W估计和LS估计的模拟比较: $N = 25, 100$

$N = 25$	a_1	a_2	a_3	a_4	σ^2
True pararmeter	1.16	-0.37	-0.11	0.18	1
Y-W avg	0.929	-0.302	-0.052	0.018	1.311
LS avg	1.012	-0.376	-0.056	0.050	0.751
Y-W std	0.208	0.251	0.228	0.158	0.518
LS std	0.252	0.331	0.318	0.226	0.267

$N = 100$	a_1	a_2	a_3	a_4	σ^2
True pararmeter	1.16	-0.37	-0.11	0.18	1
Y-W avg	1.100	-0.343	-0.087	0.133	1.083
LS avg	1.133	-0.373	-0.092	0.147	0.947
Y-W std	0.102	0.144	0.139	0.094	0.178
LS std	0.105	0.156	0.154	0.102	0.141

例1.1 Y-W估计和LS估计的模拟比较: $N = 400, 1600$

$N = 400$	a_1	a_2	a_3	a_4	σ^2
True pararmeter	1.16	-0.37	-0.11	0.18	1
Y-W avg	1.145	-0.362	-0.104	0.169	1.022
LS avg	1.154	-0.370	-0.106	0.173	0.988
Y-W std	0.049	0.074	0.074	0.048	0.075
LS std	0.050	0.076	0.076	0.049	0.071

$N = 1600$	a_1	a_2	a_3	a_4	σ^2
True pararmeter	1.16	-0.37	-0.11	0.18	1
Y-W avg	1.1563	-0.3683	-0.1082	0.1767	1.0060
LS avg	1.1586	-0.3705	-0.1086	0.1777	0.9975
Y-W std	0.0247	0.0377	0.0377	0.0247	0.0355
LS std	0.0247	0.0379	0.0380	0.0249	0.0350

最大似然估计

- Y-W估计是矩估计，最小二乘估计与Y-W估计渐近相同。矩估计容易计算但可能精度不高。
- 最大似然估计一般精度较高。
- 设AR(p)模型(1.1)的白噪声

$$\varepsilon_t = X_t - \sum_{j=1}^{p} a_j X_{t-j}$$

服从正态分布，则$\varepsilon_{p+1}, \varepsilon_{p+2}, \ldots, \varepsilon_N$有联合密度函数

$$\left(\frac{1}{2\pi\sigma^2}\right)^{\frac{N-p}{2}} \exp\left(-\frac{1}{2\sigma^2}\sum_{t=p+1}^{N}\varepsilon_t^2\right). \quad (1.15)$$

最大似然估计

- Y-W估计是矩估计，最小二乘估计与Y-W估计渐近相同。矩估计容易计算但可能精度不高。
- 最大似然估计一般精度较高。
- 设AR(p)模型(1.1)的白噪声

$$\varepsilon_t = X_t - \sum_{j=1}^{p} a_j X_{t-j}$$

服从正态分布，则$\varepsilon_{p+1}, \varepsilon_{p+2}, \ldots, \varepsilon_N$有联合密度函数

$$\left(\frac{1}{2\pi\sigma^2}\right)^{\frac{N-p}{2}} \exp\left(-\frac{1}{2\sigma^2}\sum_{t=p+1}^{N} \varepsilon_t^2\right). \tag{1.15}$$

最大似然估计

- Y-W估计是矩估计，最小二乘估计与Y-W估计渐近相同。矩估计容易计算但可能精度不高。
- 最大似然估计一般精度较高。
- 设AR(p)模型(1.1)的白噪声

$$\varepsilon_t = X_t - \sum_{j=1}^{p} a_j X_{t-j}$$

服从正态分布，则$\varepsilon_{p+1}, \varepsilon_{p+2}, \ldots, \varepsilon_N$有联合密度函数

$$\left(\frac{1}{2\pi\sigma^2}\right)^{\frac{N-p}{2}} \exp\left(-\frac{1}{2\sigma^2}\sum_{t=p+1}^{N}\varepsilon_t^2\right). \tag{1.15}$$

最大似然估计(续)

- 利用(1.15)得到基于$x_1, x_2, \ldots, x_N$的似然函数

$$L(\mathbf{a}, \sigma^2) = \left(\frac{1}{2\pi\sigma^2}\right)^{\frac{N-p}{2}} \exp\left(-\frac{1}{2\sigma^2}\sum_{t=p+1}^{N}(x_t - \sum_{j=1}^{p} a_j x_{t-j})^2\right).$$

- $L(\mathbf{a}, \sigma^2)$的最大值点$(\hat{a}_1, \hat{a}_2, \ldots, \hat{a}_p, \hat{\sigma}^2)$是$(a_1, a_2, \ldots, a_p, \sigma^2)$最大似然估计.

最大似然估计(续)

- 利用(1.15)得到基于$x_1, x_2, \ldots, x_N$的似然函数

$$L(\mathbf{a}, \sigma^2) = \left(\frac{1}{2\pi\sigma^2}\right)^{\frac{N-p}{2}} \exp\left(-\frac{1}{2\sigma^2}\sum_{t=p+1}^{N}(x_t - \sum_{j=1}^{p} a_j x_{t-j})^2\right).$$

- $L(\mathbf{a}, \sigma^2)$的最大值点$(\hat{a}_1, \hat{a}_2, \ldots, \hat{a}_p, \hat{\sigma}^2)$是$(a_1, a_2, \ldots, a_p, \sigma^2)$最大似然估计.

最大似然估计(续II)

- 由于实际计算中首先对数据进行零均值化,所以对数似然函数应当定义为

$$
\begin{aligned}
l(\mathbf{a}, \sigma^2) &= \ln L(\mathbf{a}, \sigma^2) \\
&= -\frac{N-p}{2}\ln(\sigma^2) - \frac{1}{2\sigma^2}\sum_{t=p+1}^{N}\left[y_t - \sum_{j=1}^{p} a_j y_{t-j}\right]^2 + c \\
&= -\frac{N-p}{2}\ln(\sigma^2) - \frac{1}{2\sigma^2}S(\mathbf{a}) + c,
\end{aligned}
$$

- $c = -[(N-p)/2]\ln(2\pi)$是常数.

$$
S(\mathbf{a}) = S(a_1, a_2, \ldots, a_p) = \sum_{t=p+1}^{N}\left[y_t - \sum_{j=1}^{p} a_j y_{t-j}\right]^2
$$

最大似然估计(续II)

- 由于实际计算中首先对数据进行零均值化,所以对数似然函数应当定义为

$$
\begin{aligned}
l(\mathbf{a},\sigma^2) &= \ln L(\mathbf{a},\sigma^2)\\
&= -\frac{N-p}{2}\ln(\sigma^2) - \frac{1}{2\sigma^2}\sum_{t=p+1}^{N}\left[y_t - \sum_{j=1}^{p} a_j y_{t-j}\right]^2 + c\\
&= -\frac{N-p}{2}\ln(\sigma^2) - \frac{1}{2\sigma^2}S(\mathbf{a}) + c,
\end{aligned}
$$

- $c = -[(N-p)/2]\ln(2\pi)$是常数.

$$
S(\mathbf{a}) = S(a_1, a_2, \ldots, a_p) = \sum_{t=p+1}^{N}\left[y_t - \sum_{j=1}^{p} a_j y_{t-j}\right]^2
$$

最大似然估计(续III)

- $l(\mathbf{a}, \sigma^2)$关于σ^2求偏导并令偏导等于零，得

$$\sigma^2 = \frac{1}{N-p} S(\mathbf{a}) \tag{1.16}$$

- 于是最大值点是$S(\mathbf{a})$的最小值点，最大似然估计等同于最小二乘估计。

最大似然估计(续III)

- $l(\mathbf{a}, \sigma^2)$关于σ^2求偏导并令偏导等于零，得

$$\sigma^2 = \frac{1}{N-p} S(\mathbf{a}) \tag{1.16}$$

- 于是最大值点是$S(\mathbf{a})$的最小值点，最大似然估计等同于最小二乘估计。

AR模型定阶

- 对于时间序列观测样本，假设它来自某个AR(p)模型，但是p一般是未知的。
- 要想办法估计p。
- 估计p可以用AR(p)的偏相关p步截尾性质，计算样本偏相关系数$\{\hat{a}_{kk}\}$，看$\{\hat{a}_{kk}\}$在何处截尾。
- 估计p也可以在拟合优度和模型简单之间进行权衡，AIC和BIC准则就是这样的办法。

AR模型定阶

- 对于时间序列观测样本，假设它来自某个AR(p)模型，但是p一般是未知的。
- 要想办法估计p。
- 估计p可以用AR(p)的偏相关p步截尾性质，计算样本偏相关系数$\{\hat{a}_{kk}\}$，看$\{\hat{a}_{kk}\}$在何处截尾。
- 估计p也可以在拟合优度和模型简单之间进行权衡，AIC和BIC准则就是这样的办法。

AR模型定阶

- 对于时间序列观测样本，假设它来自某个AR(p)模型，但是p一般是未知的。
- 要想办法估计p。
- 估计p可以用AR(p)的偏相关p步截尾性质，计算样本偏相关系数$\{\hat{a}_{kk}\}$，看$\{\hat{a}_{kk}\}$在何处截尾。
- 估计p也可以在拟合优度和模型简单之间进行权衡，AIC和BIC准则就是这样的办法。

AR模型定阶

- 对于时间序列观测样本，假设它来自某个AR(p)模型，但是p一般是未知的。
- 要想办法估计p。
- 估计p可以用AR(p)的偏相关p步截尾性质，计算样本偏相关系数$\{\hat{a}_{kk}\}$，看$\{\hat{a}_{kk}\}$在何处截尾。
- 估计p也可以在拟合优度和模型简单之间进行权衡，AIC和BIC准则就是这样的办法。

样本偏相关系数

- 只要观测样本$x_1, \ldots, x_N$不完全相同则样本自协方差阵$\hat{\Gamma}_k$正定($k < N$)。样本偏相关系数$\{\hat{a}_{k,k}\}$可以由样本自协方差唯一决定

$$\begin{pmatrix} \hat{\gamma}_0 & \hat{\gamma}_1 & \cdots & \hat{\gamma}_{k-1} \\ \hat{\gamma}_1 & \hat{\gamma}_0 & \cdots & \hat{\gamma}_{k-2} \\ \vdots & \vdots & & \vdots \\ \hat{\gamma}_{k-1} & \hat{\gamma}_{k-2} & \cdots & \hat{\gamma}_0 \end{pmatrix} \begin{pmatrix} \hat{a}_{k,1} \\ \hat{a}_{k,2} \\ \vdots \\ \hat{a}_{k,k} \end{pmatrix} = \begin{pmatrix} \hat{\gamma}_1 \\ \hat{\gamma}_2 \\ \vdots \\ \hat{\gamma}_k \end{pmatrix}$$

可以用Levinson递推公式计算样本偏相关系数。

- **定理1.3** 如果AR(p)模型(1.1)中的白噪声时独立同分布的，则对任何$k \geq p$，有

$$\lim_{N\to\infty} \hat{a}_{k,j} = \begin{cases} a_j, & 当 j \leq p, \\ 0, & 当 j > p. \end{cases} \quad \text{a.s.} \qquad (1.18)$$

样本偏相关系数

- 只要观测样本$x_1, \ldots, x_N$不完全相同则样本自协方差阵$\hat{\Gamma}_k$正定$(k < N)$。样本偏相关系数$\{\hat{a}_{k,k}\}$可以由样本自协方差唯一决定

$$\begin{pmatrix} \hat{\gamma}_0 & \hat{\gamma}_1 & \cdots & \hat{\gamma}_{k-1} \\ \hat{\gamma}_1 & \hat{\gamma}_0 & \cdots & \hat{\gamma}_{k-2} \\ \vdots & \vdots & & \vdots \\ \hat{\gamma}_{k-1} & \hat{\gamma}_{k-2} & \cdots & \hat{\gamma}_0 \end{pmatrix} \begin{pmatrix} \hat{a}_{k,1} \\ \hat{a}_{k,2} \\ \vdots \\ \hat{a}_{k,k} \end{pmatrix} = \begin{pmatrix} \hat{\gamma}_1 \\ \hat{\gamma}_2 \\ \vdots \\ \hat{\gamma}_k \end{pmatrix}$$

  可以用Levinson递推公式计算样本偏相关系数。
- **定理1.3** 如果AR(p)模型(1.1)中的白噪声时独立同分布的，则对任何$k \geq p$，有

$$\lim_{N \to \infty} \hat{a}_{k,j} = \begin{cases} a_j, & \text{当} j \leq p, \\ 0, & \text{当} j > p. \end{cases} \quad \text{a.s.} \qquad (1.18)$$

定理1.3证明

- 对$k > p$证明，$k = p$情况类似。
- 由§4.2 的定理2.1 知道样本自协方差函数$\hat{\gamma}_k$是γ_k的强相合估计.
- 对任何矩阵$(c_{j,k}(N))$定义极限符号

$$\lim_{N\to\infty}(c_{j,k}(N)) \triangleq \left(\lim_{N\to\infty} c_{j,k}(N)\right).$$

定理1.3证明

- 对$k > p$证明，$k = p$情况类似。
- 由§4.2 的定理2.1 知道样本自协方差函数$\hat{\gamma}_k$是γ_k的强相合估计.
- 对任何矩阵$(c_{j,k}(N))$定义极限符号

$$\lim_{N\to\infty}(c_{j,k}(N)) \triangleq \left(\lim_{N\to\infty} c_{j,k}(N)\right).$$

定理1.3证明

- 对$k > p$证明，$k = p$情况类似。
- 由§4.2 的定理2.1 知道样本自协方差函数$\hat{\gamma}_k$是γ_k的强相合估计.
- 对任何矩阵$(c_{j,k}(N))$定义极限符号

$$\lim_{N\to\infty}\left(c_{j,k}(N)\right) \triangleq \left(\lim_{N\to\infty} c_{j,k}(N)\right).$$

- 对任何$k > p$, 利用Γ_k的正定性得到:

$$
\lim_{N\to\infty}\begin{pmatrix} \hat{a}_{k,1} \\ \hat{a}_{k,2} \\ \vdots \\ \hat{a}_{k,k} \end{pmatrix}
= \lim_{N\to\infty}\begin{pmatrix} \hat{\gamma}_0 & \hat{\gamma}_1 & \cdots & \hat{\gamma}_{k-1} \\ \hat{\gamma}_1 & \hat{\gamma}_0 & \cdots & \hat{\gamma}_{k-2} \\ \vdots & \vdots & \cdots & \vdots \\ \hat{\gamma}_{k-1} & \hat{\gamma}_{k-2} & \cdots & \hat{\gamma}_0 \end{pmatrix}^{-1} \times \lim_{N\to\infty}\begin{pmatrix} \hat{\gamma}_1 \\ \hat{\gamma}_2 \\ \vdots \\ \hat{\gamma}_k \end{pmatrix}
= \begin{pmatrix} \gamma_0 & \gamma_1 & \cdots & \gamma_{k-1} \\ \gamma_1 & \gamma_0 & \cdots & \gamma_{k-2} \\ \vdots & \vdots & & \vdots \\ \gamma_{k-1} & \gamma_{k-2} & \cdots & \gamma_0 \end{pmatrix}^{-1} \times \begin{pmatrix} \gamma_1 \\ \gamma_2 \\ \vdots \\ \gamma_k \end{pmatrix}
= \begin{pmatrix} a_{k,1} \\ a_{k,2} \\ \vdots \\ a_{k,k} \end{pmatrix}.
$$

- 从§2.3 的(3.9)知道

$$(a_{k,1}, a_{k,2}, \ldots, a_{k,k}) = (a_1, a_2, .., a_p, 0, \cdots, 0). \qquad (1.19)$$

所以(1.18)成立.

- 对任何$k > p$, 利用Γ_k的正定性得到:

$$
\begin{aligned}
&\lim_{N\to\infty}\begin{pmatrix}\hat{a}_{k,1}\\ \hat{a}_{k,2}\\ \vdots\\ \hat{a}_{k,k}\end{pmatrix}\\
&=\lim_{N\to\infty}\begin{pmatrix}\hat{\gamma}_0 & \hat{\gamma}_1 & \cdots & \hat{\gamma}_{k-1}\\ \hat{\gamma}_1 & \hat{\gamma}_0 & \cdots & \hat{\gamma}_{k-2}\\ .. & .. & \cdots & ..\\ \hat{\gamma}_{k-1} & \hat{\gamma}_{k-2} & \cdots & \hat{\gamma}_0\end{pmatrix}^{-1}\times\lim_{N\to\infty}\begin{pmatrix}\hat{\gamma}_1\\ \hat{\gamma}_2\\ \vdots\\ \hat{\gamma}_k\end{pmatrix}\\
&=\begin{pmatrix}\gamma_0 & \gamma_1 & \cdots & \gamma_{k-1}\\ \gamma_1 & \gamma_0 & \cdots & \gamma_{k-2}\\ \vdots & \vdots & & \vdots\\ \gamma_{k-1} & \gamma_{k-2} & \cdots & \gamma_0\end{pmatrix}^{-1}\times\begin{pmatrix}\gamma_1\\ \gamma_2\\ \vdots\\ \gamma_k\end{pmatrix}\\
&=\begin{pmatrix}a_{k,1}\\ a_{k,2}\\ ..\\ a_{k,k}\end{pmatrix}.
\end{aligned}
$$

- 从§2.3 的(3.9)知道

$$(a_{k,1}, a_{k,2}, \ldots, a_{k,k}) = (a_1, a_2, .., a_p, 0, \cdots, 0). \qquad (1.19)$$

所以(1.18)成立.

偏相关系数的检验

- 为检验H_0：$a_{k,k}=0$，需要知道$\hat{a}_{k,k}$的分布。但我们只能得到$\hat{a}_{k,k}$的极限分布。
- **定理1.4** 设$(a_{k,1},a_{k,2},\ldots,a_{k,k})$由(1.19)定义，如果AR($p$)模型(1.1)中的白噪声是独立同分布的，$E\varepsilon_t^4<\infty$，则对确定的$k>p$，当$N\to\infty$时，

$$\sqrt{N}(\hat{a}_{k,1}-a_{k,1},\ \hat{a}_{k,2}-a_{k,2},\ \cdots,\ \hat{a}_{k,k}-a_{k,k})$$

依分布收敛到k-维正态分布$N(0,\sigma^2\Gamma_k^{-1})$.
- 详见Brockwell and Davis的教材。
- **推论1.5** 在定理1.4的条件下，对$k>p$，$\sqrt{N}\hat{a}_{k,k}$依分布收敛到标准正态分布$N(0,1)$.

偏相关系数的检验

- 为检验H_0： $a_{k,k}=0$， 需要知道$\hat{a}_{k,k}$的分布。但我们只能得到$\hat{a}_{k,k}$的极限分布。
- **定理1.4** 设$(a_{k,1},a_{k,2},\ldots,a_{k,k})$由(1.19)定义, 如果AR($p$)模型(1.1)中的白噪声是独立同分布的, $E\varepsilon_t^4<\infty$, 则对确定的$k>p$, 当$N\to\infty$时,

$$\sqrt{N}(\hat{a}_{k,1}-a_{k,1},\ \hat{a}_{k,2}-a_{k,2},\ \cdots,\ \hat{a}_{k,k}-a_{k,k})$$

依分布收敛到k-维正态分布$N(0,\sigma^2\Gamma_k^{-1})$.

- 详见Brockwell and Davis的教材。
- **推论1.5** 在定理1.4的条件下, 对$k>p$, $\sqrt{N}\hat{a}_{k,k}$ 依分布收敛到标准正态分布$N(0,1)$.

偏相关系数的检验

- 为检验H_0： $a_{k,k}=0$， 需要知道$\hat{a}_{k,k}$的分布。但我们只能得到$\hat{a}_{k,k}$的极限分布。
- **定理1.4** 设$(a_{k,1}, a_{k,2}, \ldots, a_{k,k})$由(1.19)定义, 如果AR($p$)模型(1.1)中的白噪声是独立同分布的, $E\varepsilon_t^4 < \infty$, 则对确定的$k > p$, 当$N \to \infty$时,
$$\sqrt{N}(\hat{a}_{k,1} - a_{k,1},\ \hat{a}_{k,2} - a_{k,2},\ \cdots,\ \hat{a}_{k,k} - a_{k,k})$$
依分布收敛到k-维正态分布$N(0, \sigma^2\Gamma_k^{-1})$.
- 详见Brockwell and Davis的教材。
- **推论1.5** 在定理1.4的条件下, 对$k > p$, $\sqrt{N}\hat{a}_{k,k}$ 依分布收敛到标准正态分布$N(0,1)$.

偏相关系数的检验

- 为检验H_0： $a_{k,k}=0$， 需要知道$\hat{a}_{k,k}$的分布。但我们只能得到$\hat{a}_{k,k}$的极限分布。
- **定理1.4** 设$(a_{k,1},a_{k,2},\ldots,a_{k,k})$由(1.19)定义，如果AR($p$)模型(1.1)中的白噪声是独立同分布的，$E\varepsilon_t^4<\infty$，则对确定的$k>p$，当$N\to\infty$时，

$$\sqrt{N}(\hat{a}_{k,1}-a_{k,1},\ \hat{a}_{k,2}-a_{k,2},\ \cdots,\ \hat{a}_{k,k}-a_{k,k})$$

依分布收敛到k-维正态分布$N(0,\sigma^2\Gamma_k^{-1})$.
- 详见Brockwell and Davis的教材。
- **推论1.5** 在定理1.4的条件下，对$k>p$，$\sqrt{N}\hat{a}_{k,k}$ 依分布收敛到标准正态分布$N(0,1)$.

推论1.5证明

- 从定理1.4知道只需对$k > p$证明$\sigma^2\Gamma_k^{-1}$的第(k,k)元素是1.
- 用$A_{(j,j)}$表示矩阵A的第(j,j)元素. 利用Γ_k^{-1}的伴随矩阵表示得到

$$\Gamma_k^{-1} = \frac{1}{\det(\Gamma_k)} \begin{pmatrix} \det(\Gamma_{k-1}) & & * \\ & * & \\ * & & \det(\Gamma_{k-1}) \end{pmatrix}.$$

- 于是

$$(\Gamma_k^{-1})_{(k,k)} = (\Gamma_k^{-1})_{(1,1)} = \frac{\det(\Gamma_{k-1})}{\det(\Gamma_k)}, \ \forall k \geq 1.$$

- 对$k \geq p$, 利用Yule-Walker 方程得到

$$\Gamma_{k+1}\begin{pmatrix} 1 \\ -\mathbf{a}_k \end{pmatrix} = \begin{pmatrix} \sigma^2 \\ 0 \\ \vdots \\ 0 \end{pmatrix} \quad \text{从而有} \quad \begin{pmatrix} 1 \\ -\mathbf{a}_k \end{pmatrix} = \Gamma_{k+1}^{-1}\begin{pmatrix} \sigma^2 \\ 0 \\ \vdots \\ 0 \end{pmatrix}$$

- 于是得到

$$(\sigma^2\Gamma_{k+1}^{-1})_{(k+1,k+1)} = (\sigma^2\Gamma_{k+1}^{-1})_{(1,1)} = 1, \ \ k \geq p.$$

推论1.5证明

- 从定理1.4知道只需对$k > p$证明$\sigma^2\Gamma_k^{-1}$的第(k,k)元素是1.
- 用$A_{(j,j)}$表示矩阵A的第(j,j)元素. 利用Γ_k^{-1}的伴随矩阵表示得到

$$\Gamma_k^{-1} = \frac{1}{\det(\Gamma_k)} \begin{pmatrix} \det(\Gamma_{k-1}) & & * \\ & * & \\ * & & \det(\Gamma_{k-1}) \end{pmatrix}.$$

- 于是

$$(\Gamma_k^{-1})_{(k,k)} = (\Gamma_k^{-1})_{(1,1)} = \frac{\det(\Gamma_{k-1})}{\det(\Gamma_k)}, \ \forall k \geq 1.$$

- 对$k \geq p$, 利用Yule-Walker 方程得到

$$\Gamma_{k+1}\begin{pmatrix} 1 \\ -\mathbf{a}_k \end{pmatrix} = \begin{pmatrix} \sigma^2 \\ 0 \\ \vdots \\ 0 \end{pmatrix} \quad \text{从而有} \quad \begin{pmatrix} 1 \\ -\mathbf{a}_k \end{pmatrix} = \Gamma_{k+1}^{-1}\begin{pmatrix} \sigma^2 \\ 0 \\ \vdots \\ 0 \end{pmatrix}$$

- 于是得到

$$(\sigma^2\Gamma_{k+1}^{-1})_{(k+1,k+1)} = (\sigma^2\Gamma_{k+1}^{-1})_{(1,1)} = 1, \ \ k \geq p.$$

推论1.5证明

- 从定理1.4知道只需对$k > p$证明$\sigma^2\Gamma_k^{-1}$的第(k,k)元素是1.
- 用$A_{(i,j)}$表示矩阵A的第(j,j)元素. 利用Γ_k^{-1}的伴随矩阵表示得到

$$\Gamma_k^{-1} = \frac{1}{\det(\Gamma_k)} \begin{pmatrix} \det(\Gamma_{k-1}) & & * \\ & * & \\ * & & \det(\Gamma_{k-1}) \end{pmatrix}.$$

- 于是

$$(\Gamma_k^{-1})_{(k,k)} = (\Gamma_k^{-1})_{(1,1)} = \frac{\det(\Gamma_{k-1})}{\det(\Gamma_k)}, \ \forall k \geq 1.$$

- 对$k \geq p$, 利用Yule-Walker 方程得到

$$\Gamma_{k+1} \begin{pmatrix} 1 \\ -\mathbf{a}_k \end{pmatrix} = \begin{pmatrix} \sigma^2 \\ 0 \\ \vdots \\ 0 \end{pmatrix} \quad \text{从而有} \quad \begin{pmatrix} 1 \\ -\mathbf{a}_k \end{pmatrix} = \Gamma_{k+1}^{-1} \begin{pmatrix} \sigma^2 \\ 0 \\ \vdots \\ 0 \end{pmatrix}$$

- 于是得到

$$(\sigma^2\Gamma_{k+1}^{-1})_{(k+1,k+1)} = (\sigma^2\Gamma_{k+1}^{-1})_{(1,1)} = 1, \ k \geq p.$$

推论1.5证明

- 从定理1.4知道只需对$k > p$证明$\sigma^2\Gamma_k^{-1}$的第(k,k)元素是1.
- 用$A_{(i,j)}$表示矩阵A的第(j,j)元素. 利用Γ_k^{-1}的伴随矩阵表示得到

$$\Gamma_k^{-1} = \frac{1}{\det(\Gamma_k)} \begin{pmatrix} \det(\Gamma_{k-1}) & & * \\ & * & \\ * & & \det(\Gamma_{k-1}) \end{pmatrix}.$$

- 于是

$$(\Gamma_k^{-1})_{(k,k)} = (\Gamma_k^{-1})_{(1,1)} = \frac{\det(\Gamma_{k-1})}{\det(\Gamma_k)}, \quad \forall k \geq 1.$$

- 对$k \geq p$, 利用Yule-Walker 方程得到

$$\Gamma_{k+1} \begin{pmatrix} 1 \\ -\mathbf{a}_k \end{pmatrix} = \begin{pmatrix} \sigma^2 \\ 0 \\ \vdots \\ 0 \end{pmatrix} \quad \text{从而有} \quad \begin{pmatrix} 1 \\ -\mathbf{a}_k \end{pmatrix} = \Gamma_{k+1}^{-1} \begin{pmatrix} \sigma^2 \\ 0 \\ \vdots \\ 0 \end{pmatrix}$$

- 于是得到

$$\left(\sigma^2\Gamma_{k+1}^{-1}\right)_{(k+1,k+1)} = \left(\sigma^2\Gamma_{k+1}^{-1}\right)_{(1,1)} = 1, \ k \geq p.$$

推论1.5证明

- 从定理1.4知道只需对$k > p$证明$\sigma^2\Gamma_k^{-1}$的第(k,k)元素是1.
- 用$A_{(j,j)}$表示矩阵A的第(j,j)元素. 利用Γ_k^{-1}的伴随矩阵表示得到

$$\Gamma_k^{-1} = \frac{1}{\det(\Gamma_k)} \begin{pmatrix} \det(\Gamma_{k-1}) & & * \\ & * & \\ * & & \det(\Gamma_{k-1}) \end{pmatrix}.$$

- 于是

$$(\Gamma_k^{-1})_{(k,k)} = (\Gamma_k^{-1})_{(1,1)} = \frac{\det(\Gamma_{k-1})}{\det(\Gamma_k)}, \ \forall k \ge 1.$$

- 对$k \ge p$, 利用Yule-Walker 方程得到

$$\Gamma_{k+1} \begin{pmatrix} 1 \\ -\mathbf{a}_k \end{pmatrix} = \begin{pmatrix} \sigma^2 \\ 0 \\ \vdots \\ 0 \end{pmatrix} \text{ 从而有 } \begin{pmatrix} 1 \\ -\mathbf{a}_k \end{pmatrix} = \Gamma_{k+1}^{-1} \begin{pmatrix} \sigma^2 \\ 0 \\ \vdots \\ 0 \end{pmatrix}$$

- 于是得到

$$(\sigma^2\Gamma_{k+1}^{-1})_{(k+1,k+1)} = (\sigma^2\Gamma_{k+1}^{-1})_{(1,1)} = 1, \ \ k \ge p.$$

用样本偏相关系数定阶

- N较大时若$k > p$，则$\sqrt{N}\hat{a}_{k,k}$近似服从N(0,1)，$\hat{a}_{k,k}$约以95%概率落入区间

$$(-1.96/\sqrt{N},\ 1.96/\sqrt{N}).$$

- 对某固定的K，以

$$\hat{p} = \max\{j : |\hat{a}_{j,j}| > 1.96/\sqrt{N}, 1 \le j \le K\}$$

  作为p的估计是合理的。
- 实际问题中可以取K为允许的阶数的一个上界。
- 这种办法只要画出样本偏相关系数图就可以定阶。(R中用pacf(x)作样本偏相关系数图)。
- 演示：AR order selection by PACF.

用样本偏相关系数定阶

- N较大时若$k > p$，则$\sqrt{N}\hat{a}_{k,k}$近似服从N(0,1)，$\hat{a}_{k,k}$约以95%概率落入区间

$$(-1.96/\sqrt{N},\ 1.96/\sqrt{N}).$$

- 对某固定的K，以

$$\hat{p} = \max\{j : |\hat{a}_{j,j}| > 1.96/\sqrt{N}, 1 \le j \le K\}$$

作为p的估计是合理的。

- 实际问题中可以取K为允许的阶数的一个上界。
- 这种办法只要画出样本偏相关系数图就可以定阶。(R中用pacf(x)作样本偏相关系数图)。
- 演示：AR order selection by PACF.

用样本偏相关系数定阶

- N较大时若$k > p$，则$\sqrt{N}\hat{a}_{k,k}$近似服从N(0,1)，$\hat{a}_{k,k}$约以95%概率落入区间

$$(-1.96/\sqrt{N},\ 1.96/\sqrt{N}).$$

- 对某固定的K，以

$$\hat{p} = \max\{j : |\hat{a}_{j,j}| > 1.96/\sqrt{N}, 1 \le j \le K\}$$

作为p的估计是合理的。

- 实际问题中可以取K为允许的阶数的一个上界。
- 这种办法只要画出样本偏相关系数图就可以定阶。(R中用pacf(x)作样本偏相关系数图)。
- 演示：AR order selection by PACF.

用样本偏相关系数定阶

- N较大时若$k > p$，则$\sqrt{N}\hat{a}_{k,k}$近似服从N(0,1)，$\hat{a}_{k,k}$约以95%概率落入区间

$$(-1.96/\sqrt{N},\ 1.96/\sqrt{N}).$$

- 对某固定的K，以

$$\hat{p} = \max\{j : |\hat{a}_{j,j}| > 1.96/\sqrt{N}, 1 \le j \le K\}$$

  作为p的估计是合理的。
- 实际问题中可以取K为允许的阶数的一个上界。
- 这种办法只要画出样本偏相关系数图就可以定阶。(R中用pacf(x)作样本偏相关系数图)。
- 演示：AR order selection by PACF.

用样本偏相关系数定阶

- N较大时若$k > p$，则$\sqrt{N}\hat{a}_{k,k}$近似服从N(0,1)，$\hat{a}_{k,k}$约以95%概率落入区间

$$(-1.96/\sqrt{N},\ 1.96/\sqrt{N}).$$

- 对某固定的K，以

$$\hat{p} = \max\{j : |\hat{a}_{j,j}| > 1.96/\sqrt{N}, 1 \le j \le K\}$$

  作为p的估计是合理的。
- 实际问题中可以取K为允许的阶数的一个上界。
- 这种办法只要画出样本偏相关系数图就可以定阶。(R中用pacf(x)作样本偏相关系数图)。
- 演示：AR order selection by PACF.

AIC和BIC定阶

- 可以通过对拟合优度的要求加上对参数个数的惩罚指定一个准则来定阶。
- AIC准则:

$$\text{AIC}(k) = \ln \hat{\sigma}_k^2 + \frac{2k}{N}, \quad k = 0, 1, \dots, P_0$$

(P_0是可取的阶的上限)
- BIC定阶:

$$\text{BIC}(k) = \ln \hat{\sigma}_k^2 + \frac{k \ln N}{N}, \quad k = 0, 1, \dots, P_0$$

- 取AIC或BIC的最小值点作为p的估计。

AIC和BIC定阶

- 可以通过对拟合优度的要求加上对参数个数的惩罚指定一个准则来定阶。
- AIC准则:

$$\mathrm{AIC}(k) = \ln \hat{\sigma}_k^2 + \frac{2k}{N}, \quad k = 0, 1, \ldots, P_0$$

(P_0是可取的阶的上限)

- BIC定阶:

$$\mathrm{BIC}(k) = \ln \hat{\sigma}_k^2 + \frac{k \ln N}{N}, \quad k = 0, 1, \ldots, P_0$$

- 取AIC或BIC的最小值点作为p的估计。

AIC和BIC定阶

- 可以通过对拟合优度的要求加上对参数个数的惩罚指定一个准则来定阶。
- AIC准则:

$$\mathrm{AIC}(k) = \ln \hat{\sigma}_k^2 + \frac{2k}{N}, \quad k = 0, 1, \ldots, P_0$$

(P_0是可取的阶的上限)

- BIC定阶:

$$\mathrm{BIC}(k) = \ln \hat{\sigma}_k^2 + \frac{k \ln N}{N}, \quad k = 0, 1, \ldots, P_0$$

- 取AIC或BIC的最小值点作为p的估计。

AIC和BIC定阶

- 可以通过对拟合优度的要求加上对参数个数的惩罚指定一个准则来定阶。
- AIC准则:

$$\text{AIC}(k) = \ln \hat{\sigma}_k^2 + \frac{2k}{N}, \quad k = 0, 1, \ldots, P_0$$

(P_0是可取的阶的上限)

- BIC定阶:

$$\text{BIC}(k) = \ln \hat{\sigma}_k^2 + \frac{k \ln N}{N}, \quad k = 0, 1, \ldots, P_0$$

- 取AIC或BIC的最小值点作为p的估计。

AIC和BIC定阶效果

- AIC定阶不相合, 倾向于高估；BIC定阶在$N \to \infty$时是强相合的，但对较小的N倾向于低估。
- 演示：AR order selection by AIC and BIC.

AIC和BIC定阶效果

- AIC定阶不相合，倾向于高估；BIC定阶在$N \to \infty$时是强相合的，但对较小的N倾向于低估。
- 演示：AR order selection by AIC and BIC.

模型拟合检验

- 对观测数据$x_1, x_2, \cdots, x_N$，如果得到了AR(p)的阶p和自回归系数$(a_1, a_2, \cdots, a_p)$的估计量$\hat{p}$ 和$(\hat{a}_1, \hat{a}_2, \cdots, \hat{a}_{\hat{p}})$，定义残差

$$\hat{\varepsilon}_t = y_t - \sum_{j=1}^{\hat{p}} \hat{a}_j y_{t-j}, \quad t = \hat{p}+1, \hat{p}+2, \ldots, N. \quad (1.21)$$

- 这里的$y_t = x_t - \bar{x}_N$是x_t零均值化.
- 对上述的残差序列进行白噪声的检验.
- 如果能够判定(1.21)是白噪声，就认为建立的模型是合理的. 否则可以改动$\hat{p}$ 的值后重新计算，或改用MA(q)或ARMA(p, q) 模型.

模型拟合检验

- 对观测数据$x_1, x_2, \cdots, x_N$，如果得到了AR(p)的阶p和自回归系数$(a_1, a_2, \cdots, a_p)$的估计量$\hat{p}$ 和$(\hat{a}_1, \hat{a}_2, \cdots, \hat{a}_{\hat{p}})$，定义残差

$$\hat{\varepsilon}_t = y_t - \sum_{j=1}^{\hat{p}} \hat{a}_j y_{t-j}, \quad t = \hat{p}+1, \hat{p}+2, \ldots, N. \quad (1.21)$$

- 这里的$y_t = x_t - \bar{x}_N$是x_t零均值化.
- 对上述的残差序列进行白噪声的检验.
- 如果能够判定(1.21)是白噪声，就认为建立的模型是合理的. 否则可以改动$\hat{p}$ 的值后重新计算，或改用MA(q)或ARMA(p, q) 模型.

模型拟合检验

- 对观测数据$x_1, x_2, \cdots, x_N$，如果得到了AR(p)的阶p和自回归系数$(a_1, a_2, \cdots, a_p)$的估计量$\hat{p}$ 和$(\hat{a}_1, \hat{a}_2, \cdots, \hat{a}_{\hat{p}})$，定义残差

$$\hat{\varepsilon}_t = y_t - \sum_{j=1}^{\hat{p}} \hat{a}_j y_{t-j}, \quad t = \hat{p}+1, \hat{p}+2, \ldots, N. \quad (1.21)$$

- 这里的$y_t = x_t - \bar{x}_N$是x_t零均值化.
- 对上述的残差序列进行白噪声的检验.
- 如果能够判定(1.21)是白噪声，就认为建立的模型是合理的. 否则可以改动$\hat{p}$ 的值后重新计算，或改用MA(q)或ARMA(p, q) 模型.

模型拟合检验

- 对观测数据$x_1, x_2, \cdots, x_N$，如果得到了AR(p)的阶p和自回归系数$(a_1, a_2, \cdots, a_p)$的估计量$\hat{p}$ 和$(\hat{a}_1, \hat{a}_2, \cdots, \hat{a}_{\hat{p}})$，定义残差

$$\hat{\varepsilon}_t = y_t - \sum_{j=1}^{\hat{p}} \hat{a}_j y_{t-j}, \quad t = \hat{p}+1, \hat{p}+2, \ldots, N. \quad (1.21)$$

- 这里的$y_t = x_t - \bar{x}_N$是x_t零均值化.
- 对上述的残差序列进行白噪声的检验.
- 如果能够判定(1.21)是白噪声，就认为建立的模型是合理的. 否则可以改动$\hat{p}$ 的值后重新计算，或改用MA(q)或ARMA(p, q) 模型.

AR谱密度估计

- 满足AR(p)模型(1.1)的AR(p)序列有谱密度函数

$$f(\lambda) = \frac{\sigma^2}{2\pi}\left|1 - \sum_{j=1}^{p} a_j e^{ij\lambda}\right|^{-2}. \qquad (1.22)$$

- 把σ^2, p 和$(a_1, a_2, \ldots, a_p)$的估计量$\hat{\sigma}^2$, $\hat{p}$ 和$(\hat{a}_1, \hat{a}_2, \ldots, \hat{a}_p)$ 代入(1.22)后，得到$f(\lambda)$ 的估计

$$\hat{f}(\lambda) = \frac{\hat{\sigma}^2}{2\pi}\left|1 - \sum_{j=1}^{\hat{p}} \hat{a}_j e^{ij\lambda}\right|^{-2}. \qquad (1.23)$$

- 通常称$\hat{f}(\lambda)$为AR谱估计或极大熵谱估计.
- 对于AIC或BIC定阶$\hat{p}$, 如果$\{\varepsilon_t\}$是独立同分布的WN$(0, \sigma^2)$, 可以证明$\hat{f}(\lambda)$一致收敛到$f(\lambda)$.

AR谱密度估计

- 满足AR(p)模型(1.1)的AR(p)序列有谱密度函数

$$f(\lambda)=\frac{\sigma^2}{2\pi}\left|1-\sum_{j=1}^{p}a_j e^{ij\lambda}\right|^{-2}. \tag{1.22}$$

- 把σ^2, p 和$(a_1,a_2,\ldots,a_p)$的估计量$\hat\sigma^2$, $\hat p$ 和$(\hat a_1,\hat a_2,\ldots,\hat a_p)$ 代入(1.22)后，得到$f(\lambda)$ 的估计

$$\hat f(\lambda)=\frac{\hat\sigma^2}{2\pi}\left|1-\sum_{j=1}^{\hat p}\hat a_j e^{ij\lambda}\right|^{-2}. \tag{1.23}$$

- 通常称$\hat f(\lambda)$为AR谱估计或极大熵谱估计.
- 对于AIC或BIC定阶$\hat p$, 如果$\{\varepsilon_t\}$是独立同分布的WN$(0,\sigma^2)$, 可以证明$\hat f(\lambda)$一致收敛到$f(\lambda)$.

AR谱密度估计

- 满足AR(p)模型(1.1)的AR(p)序列有谱密度函数

$$f(\lambda)=\frac{\sigma^2}{2\pi}\left|1-\sum_{j=1}^{p}a_j e^{ij\lambda}\right|^{-2}. \tag{1.22}$$

- 把σ^2, p 和$(a_1,a_2,\ldots,a_p)$的估计量$\hat{\sigma}^2$, $\hat{p}$ 和$(\hat{a}_1,\hat{a}_2,\ldots,\hat{a}_p)$ 代入(1.22)后，得到$f(\lambda)$ 的估计

$$\hat{f}(\lambda)=\frac{\hat{\sigma}^2}{2\pi}\left|1-\sum_{j=1}^{\hat{p}}\hat{a}_j e^{ij\lambda}\right|^{-2}. \tag{1.23}$$

- 通常称$\hat{f}(\lambda)$为AR谱估计或极大熵谱估计.
- 对于AIC或BIC定阶$\hat{p}$，如果$\{\varepsilon_t\}$是独立同分布的WN$(0,\sigma^2)$，可以证明$\hat{f}(\lambda)$一致收敛到$f(\lambda)$.

AR谱密度估计

- 满足AR(p)模型(1.1)的AR(p)序列有谱密度函数

$$f(\lambda)=\frac{\sigma^2}{2\pi}\left|1-\sum_{j=1}^{p}a_je^{ij\lambda}\right|^{-2}. \qquad (1.22)$$

- 把σ^2, p 和$(a_1,a_2,\ldots,a_p)$的估计量$\hat{\sigma}^2$, $\hat{p}$ 和$(\hat{a}_1,\hat{a}_2,\ldots,\hat{a}_p)$ 代入(1.22)后，得到$f(\lambda)$ 的估计

$$\hat{f}(\lambda)=\frac{\hat{\sigma}^2}{2\pi}\left|1-\sum_{j=1}^{\hat{p}}\hat{a}_je^{ij\lambda}\right|^{-2}. \qquad (1.23)$$

- 通常称$\hat{f}(\lambda)$为AR谱估计或极大熵谱估计.
- 对于AIC或BIC定阶$\hat{p}$，如果$\{\varepsilon_t\}$是独立同分布的WN$(0,\sigma^2)$，可以证明$\hat{f}(\lambda)$一致收敛到$f(\lambda)$.